2019

Improved methods for statistical inference in the context of various types of parameter variation

https://hdl.handle.net/2144/38750

Boston University
BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

IMPROVED METHODS FOR STATISTICAL INFERENCE
IN THE CONTEXT OF VARIOUS TYPES OF
PARAMETER VARIATION

by

ALESSANDRO CASINI

B.S. and M.Sc., University of Siena, 2011
M.Sc., Universitat Pompeu Fabra, 2012

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2019
Acknowledgments

I wish to express my sincere gratitude to my supervisor Professor Pierre Perron for his invaluable guidance. He taught me Econometrics and how to be a good researcher. He also provided me with all support that I needed. I am fortunate of being his student, research assistant and coauthor. I am also thankful to Professor Zhongjun Qu, Professor Hiroaki Kaido, Professor Iván Fernández-Val and Professor Jean-Jacques Forneron for their support and encouragement. Their comments and suggestions proved to be very useful for the projects I have worked on. I thank the Bank of Italy for the financial support for the first two years in the program, and Undral Byambadalai, Shuowen Chen, Taosong Deng, Junwen Lu, Siyi Luo, Anlong Qin, Guang Zhang and Yi Zhang for their friendship and feedback. I am extremely grateful to my family for their unconditional support and patience over the years. Without their love this thesis would not have been possible.
This dissertation addresses various issues related to statistical inference in the context of parameter time-variation. The problem is considered within general regression models as well as in the context of methods for forecast evaluation.

The first chapter develops a theory of evolutionary spectra for heteroskedasticity and autocorrelation-robust (HAR) inference when the data may not satisfy second-order stationarity. We introduce a class of nonstationary stochastic processes that have a time-varying spectral representation and presents a new positive semidefinite heteroskedasticity- and autocorrelation consistent (HAC) estimator. We obtain an optimal HAC estimator under the mean-squared error (MSE) criterion and show its consistency. We propose a data-dependent procedure based on a “plug-in” approach that determines the bandwidth parameters for given kernels and a given sample size.

The second chapter develops a continuous record asymptotic framework to build inference methods for the date of a structural change in a linear regression model. We impose very mild regularity conditions on an underlying continuous-time model.
assumed to generate the data. We consider the least-squares estimate of the break
date and establish consistency and convergence rate. We provide a limit theory for
shrinking magnitudes of shifts and locally increasing variances.

The third chapter develops a novel continuous-time asymptotic framework for
inference on whether the predictive ability of a given forecast model remains stable
over time. As the sampling interval between observations shrinks to zero the sequence
of forecast losses is approximated by a continuous-time stochastic process possessing
certain pathwise properties. We consider an hypotheses testing problem based on the
local properties of the continuous-time limit counterpart of the sequence of losses.

The fourth chapter develops a class of Generalized Laplace (GL) inference met-
hods for the change-point dates in a linear time series regression model with multi-
ple structural changes. The GL estimator is defined by an integration rather than
optimization-based method and relies on the least-squares criterion function. On the
theoretical side, depending on some smoothing parameter, the class of GL estima-
tors exhibits a dual limiting distribution; namely, the classical shrinkage asymptotic
distribution of the least-squares estimator, or a Bayes-type asymptotic distribution.
Contents

1 Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models 1

1.1 Introduction ................................................................. 1

1.2 The Statistical Environment ............................................. 9
  1.2.1 Segmented Locally Stationary Processes ......................... 12
  1.2.2 HAC Estimation .................................................. 18

1.3 HAC Estimation with Predetermined Bandwidths ..................... 21
  1.3.1 Estimation of Local Covariances ............................... 22
  1.3.2 Results on HAC Estimation with Predetermined Bandwidths ... 23

1.4 Optimal Kernels, Bandwidths and Choice of \( n_T \) ..................... 29
  1.4.1 Optimal \( K_2 (\cdot) \) and \( b_{2,T} \) ......................... 30
  1.4.2 Optimal \( K_1 (\cdot) \) .................................. 32
  1.4.3 Optimal Predetermined Bandwidth Sequence \( b_{1,T} \) ........ 33
  1.4.4 Choice of \( n_T \) ........................................ 35

1.5 Data-Dependent Bandwidths .............................................. 36
  1.5.1 Implementation ............................................... 37
  1.5.2 Theoretical Results ........................................ 40

1.6 Inference about the Break Points ...................................... 42

1.7 HAC Estimation when \( m > 0 \) ...................................... 47

1.8 Small-Sample Evaluations .............................................. 49
3 Tests for Forecast Instability and Forecast Failure under a Continuous Record Asymptotic Framework

3.1 Introduction ..................................................... 143
3.2 The Statistical Environment ................................. 151
  3.2.1 The Forecasting Problem ................................. 152
  3.2.2 The Underlying Continuous-Time Model ............... 154
  3.2.3 The Hypotheses of Interest ............................. 158
    3.2.3.1 The Null and Alternative Hypotheses on Forecast Instability ..................................... 159
  3.2.4 The Test Statistics ........................................ 165
    3.2.4.1 Test Statistics under Quadratic Loss ............. 167
    3.2.4.2 Test Statistics under General Loss Function .... 169
3.3 Continuous Record Distribution Theory for the Test Statistics .... 170
  3.3.1 Asymptotic Distribution under the Null Hypotheses ...... 170
    3.3.1.1 Asymptotic Distribution Under Quadratic Loss Function ..................................... 172
    3.3.1.2 Asymptotic Distribution Under General Loss Function ..................................... 173
3.4 Estimation of the Asymptotic Variance .................. 173
  3.4.1 Estimation of the Asymptotic Variance ................. 173
    3.4.1.1 Stationary Forecast Losses ....................... 174
    3.4.1.2 Heterogeneous Forecast Losses ................... 176
3.5 Continuous Semimartingale Volatility and Asymptotic Local Power . 177
### 3.5.1 Asymptotic Results under Continuous Semimartingale Volatility ................................. 177

### 3.5.2 Asymptotic Local Power ........................................... 179

### 3.6 Extensions ............................................................... 181

### 3.7 Small-Sample Evaluation ........................................... 184

#### 3.7.1 Empirical Size .................................................... 185

#### 3.7.2 Empirical Power .................................................. 187

### 3.8 Conclusions .............................................................. 192

### 3.9 Appendix to Chapter 3 .............................................. 193

#### 3.9.1 Tables ............................................................... 193

#### 3.9.2 Figures .............................................................. 195

##### 3.9.2.1 General Instability ........................................... 195

##### 3.9.2.2 Short-Term Instability ...................................... 200

### 4 Generalized Laplace Inference in Multiple Change-Points Models 204

#### 4.1 Introduction ............................................................ 204

#### 4.2 The Model and its Assumptions ..................................... 209

##### 4.2.1 The Structural Change Model with a Single Change-point 210

##### 4.2.2 The Least-squares Criterion Function ......................... 211

##### 4.2.3 Assumptions ...................................................... 211

#### 4.3 Generalized Laplace Estimation .................................... 213

##### 4.3.1 The Class of Laplace Estimators ............................... 213

##### 4.3.2 Discussion about the GL Approach ............................. 216

##### 4.3.3 Normalized Version of the GL Estimator ..................... 217

##### 4.3.4 Asymptotic Results for Generalized Laplace Estimates .... 222
4.3.4.1 The Asymptotic Distribution of the Quasi-posterior Mean ........................................... 222
4.3.4.2 The Asymptotic Distribution for General Loss Functions ........................................... 226
4.4 Inference based on GL Estimators ................................................................. 227
4.5 Models with Multiple Change-Points .............................................................. 230
4.6 Finite-Sample Evaluation of GL Method .......................................................... 234
   4.6.1 Precision of the Change-point Estimate ...................................................... 235
   4.6.2 Properties of the GL Confidence Sets ....................................................... 236
4.7 Conclusions ........................................................................................................ 239
4.8 Appendix to Chapter 4 ....................................................................................... 240

A Supplement to Chapter 1: Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models ........................................... 245

A.1 Implementation of \( J_T \) HAC in GMM, IV and Structural Change Models ............... 245
   A.1.1 GMM ........................................................................................................... 245
   A.1.2 IV ............................................................................................................... 247
A.2 Mathematical Appendix ...................................................................................... 247
   A.2.1 Proof of the Results of Section 1.2.1 ........................................................... 248
      A.2.1.1 Proof of Theorem 1.2.1 ...................................................................... 248
   A.2.2 Proof of the Results of Section 1.3 ............................................................... 250
      A.2.2.1 Proof of Proposition 1.3.1 ................................................................. 250
      A.2.2.2 Proof of Theorem 1.3.1 ..................................................................... 251
      A.2.2.3 Proof of Theorem 1.3.2 ..................................................................... 262
   A.2.3 Proof of the Results of Section 1.4 ............................................................... 266
      A.2.3.1 Proof of Proposition 1.4.1 ................................................................. 266
      A.2.3.2 Proof of Theorem 1.4.1 ..................................................................... 270
A.2.3.3 Proof of Corollary 1.4.1 . . . . . . . . . . . . . . . . 271
A.2.4 Proofs of the Results of Section 1.5 . . . . . . . . . . . . . . 272
  A.2.4.1 Proof of Theorem 1.5.1 . . . . . . . . . . . . . . . . . 272
A.2.5 Proofs of the Results in Section 1.7 . . . . . . . . . . . . . . 284
  A.2.5.1 Proof of Theorem 1.7.1 . . . . . . . . . . . . . . . . . 284
A.3 Additional Monte Carlo Results . . . . . . . . . . . . . . . . . . . . 293
A.4 Empirical Application . . . . . . . . . . . . . . . . . . . . . . . . . 298
A.5 Tables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 303

B Supplement to Chapter 2: Continuous Record Asymptotics for Structural Change Models 304
  B.1 Mathematical Proofs . . . . . . . . . . . . . . . . . . . . . . . . . . 304
    B.1.1 Additional Notations . . . . . . . . . . . . . . . . . . . . . . . 304
    B.1.2 Preliminary Lemmas . . . . . . . . . . . . . . . . . . . . . . . . 305
    B.1.3 Preliminary Results . . . . . . . . . . . . . . . . . . . . . . . . . 306
      B.1.3.1 Approximate Variation, LLNs and CLTs . . . . . . . 307
    B.1.4 Proofs of Sections 2.3 and 2.4 . . . . . . . . . . . . . . . . . . 310
      B.1.4.1 Additional Notation . . . . . . . . . . . . . . . . . 310
      B.1.4.2 Proof of Proposition 2.3.1 . . . . . . . . . . . . . . 312
      B.1.4.3 Proof of Proposition 2.3.2 . . . . . . . . . . . . . . 315
      B.1.4.4 Proof of Proposition 2.3.3 . . . . . . . . . . . . . . 320
      B.1.4.5 Proof of Proposition 2.4.1 . . . . . . . . . . . . . . 323
      B.1.4.6 Proof of Lemma 2.4.1 . . . . . . . . . . . . . . . . 334
      B.1.4.7 Proof of Proposition 2.5.1 . . . . . . . . . . . . . . 339
    B.1.5 Proofs of Section 2.9.2 . . . . . . . . . . . . . . . . . . . . . 340
      B.1.5.1 Proof of Proposition 2.9.1 . . . . . . . . . . . . . . 343
      B.1.5.2 Proof of Theorem 2.9.1 . . . . . . . . . . . . . . . . 350
B.1.5.3 Proof of Proposition 2.9.2 362
B.1.5.4 Negligibility of the Drift Term 373

B.2 Additional Discussion about the Continuous Record Asymptotic Density Function 374
B.2.1 Further Discussion from Section 2.4.2 374
B.2.2 Further Discussion from Section 2.5 377
B.2.3 Figures 379

C Supplement to Chapter 3: Tests for Forecast Instability and Forecast Failure under a Continuous Record Asymptotic Framework 389
C.1 Mathematical Proofs 389
C.1.1 Additional Notation 389
C.1.2 Localization 390
C.1.3 Preliminary Lemmas 390
C.1.4 Proofs of Section 3.3 390
C.1.4.1 Proof of Theorem 3.3.1 391
C.1.4.1.1 Proof of part (i) of Theorem 3.3.1 391
C.1.4.1.2 Proof of part (ii) of Theorem 3.3.1 421
C.1.4.1.3 Negligibility of the $\mu_{e,t}$ term 424
C.1.4.2 Proof of Corollary 3.3.1 425
C.1.4.3 Proof of Theorem 3.3.2 426
C.1.4.3.1 Proof of part (i) of Theorem 3.3.2 426
C.1.4.3.2 Proof of part (ii) of Theorem 3.3.2 451
C.1.4.3.3 Negligibility of the drift term under general loss functions 457
C.1.4.4 Proof of Corollary 3.3.2 458
C.1.5 Proofs of Section 3.4 458
List of Tables

1.1 Simulation assisted critical values .......................... 64
1.2 Empirical small-sample size of $t_1$-test for model S1 ........ 64
1.3 Empirical small-sample size of $t_1$-test for model S2 ........ 65
1.4 Empirical small-sample size of $t_2$-test for model S3-S4 ....... 65
1.5 Empirical small-sample size of $t$-tests for model S5-S6 ........ 65
1.6 Empirical small-sample size for model S7-S8 .................. 66
1.7 Empirical small-sample rejection rates of $t_1$-test for model P1 . 66
1.8 Empirical small-sample rejection rates of $t_1$-test for model P2 . 66
1.9 Empirical small-sample rejection rates of $t_2$-test for model P3 . 67
1.10 Empirical small-sample rejection rates of $t_2$-test for model P4 . 67
1.11 Empirical small-sample rejection rates of $t_2$-test for model P5 . 67
1.12 Empirical small-sample rejection rates of $t_1$-test for model P6 . 68
1.13 Empirical power of the DM (1995) test ....................... 68

2.1 Coverage rate and length of the confidence set for the example of Section 4.4 ...................................................... 137
2.2 Small-sample coverage rate and length of the confidence set for model M1 .............................................................. 137
2.3 Small-sample coverage rate and length of the confidence set for model M2 .............................................................. 138
2.4 Small-sample coverage rate and length of the confidence set for model M3 ................................................................. 138
2.5 Small-sample coverage rate and length of the confidence set for model M4 ................................................................. 139
2.6 Small-sample coverage rate and length of the confidence set for model M5 ................................................................. 139
2.7 Small-sample coverage rate and length of the confidence set for model M6 ................................................................. 140
2.8 Small-sample coverage rate and length of the confidence set for model M7 ................................................................. 140
2.9 Small-sample coverage rate and length of the confidence sets for model M8 ................................................................. 141
2.10 Small-sample coverage rate and length of the confidence sets for model M9 ................................................................. 141
2.11 Small-sample coverage rate and length of the confidence set for model M10 ................................................................. 142
3.1 Empirical small sample size of forecast instability tests based on model S1 ................................................................. 193
3.2 Empirical small sample size of forecast instability tests based on model S2 ................................................................. 194
4.1 Small-sample accuracy of the estimates of the break point $T^0_b$ for model M1 ................................................................. 240
4.2 Small-sample accuracy of the estimates of the break point $T^0_b$ for model M2 ................................................................. 241
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>Small-sample accuracy of the estimates of the break point $T^0_b$ for model M3</td>
<td>242</td>
</tr>
<tr>
<td>4.4</td>
<td>Small-sample coverage rates and lengths of the confidence sets for model M1</td>
<td>243</td>
</tr>
<tr>
<td>4.5</td>
<td>Small-sample coverage rates and lengths of the confidence sets for model M2</td>
<td>243</td>
</tr>
<tr>
<td>4.6</td>
<td>Small-sample coverage rates and lengths of the confidence sets for model M3</td>
<td>244</td>
</tr>
<tr>
<td>A.1</td>
<td>Empirical small-sample size of the $t_1$-test for model M1-M2</td>
<td>295</td>
</tr>
<tr>
<td>A.2</td>
<td>Empirical small-sample size of $t$-test for model M3-M4</td>
<td>295</td>
</tr>
<tr>
<td>A.3</td>
<td>Empirical small-sample size of $F$-test for model M5-M6</td>
<td>296</td>
</tr>
<tr>
<td>A.4</td>
<td>Empirical small-sample rejection rates of $t_1$-test for model M1</td>
<td>296</td>
</tr>
<tr>
<td>A.5</td>
<td>Empirical small-sample rejection rates of $t_2$-test for model M3</td>
<td>296</td>
</tr>
<tr>
<td>A.6</td>
<td>Empirical small-sample rejection rates of $F$-test for model M9</td>
<td>297</td>
</tr>
<tr>
<td>A.7</td>
<td>Giacomini and Rossi (2009) $t$-test</td>
<td>303</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Power functions of forecast breakdown t-test for Model M8 with $T = 200$. 69
1.2 Power functions of forecast breakdown t-test for Model M8 with $T = 400$. 69
1.3 Power functions of forecast breakdown t-test for Model M8 with $T = 800$. 70

2.1 Distributions with $\rho^2 = 0.2$. . . . . . . . . . . . . . . . . . . . . . 130
2.2 Distributions with $\rho^2 = 0.3$. . . . . . . . . . . . . . . . . . . . . . 130
2.3 Distributions with $\rho^2 = 0.5$. . . . . . . . . . . . . . . . . . . . . . 131
2.4 Distributions with $\rho^2 = 0.8$. . . . . . . . . . . . . . . . . . . . . . 131
2.5 Distributions for model (2.5.1) with $\delta^0 = 0.2$. . . . . . . . . 132
2.6 Distributions for model (2.5.1) with $\delta^0 = 0.3$. . . . . . . . . 133
2.7 Distributions for model (2.5.1) with $\delta^0 = 0.5$. . . . . . . . . 134
2.8 Distributions for model (2.5.1) with $\delta^0 = 1$. . . . . . . . . . 135
2.9 Highest Density Regions . . . . . . . . . . . . . . . . . . . . . . . . . 136

3.1 Power functions for model P1a with $T = 100$ and $T = 150$. . . . 195
3.2 Power functions for model P1a with $T = 200$ and $T = 300$. . . . 196
3.3 Power functions for model P4 with $T = 200$ and $T = 300$. . . . 197
3.4 Power functions for model P4 with $T = 400$ and $T = 500$. . . . 198
3.5 Power functions for model P7 with $T = 200$ and $T = 300$. . . . 199
3.6 Power functions for model P1a with short-term instability and with
   $T = 100$ and 150 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 200
3.7 Power functions for model P1a with $T = 200$ and $T = 300$. . . . 201
C.2 Power functions for model P1b with $T = 200$ and $T = 300$ . . . . . . 476
C.3 Power functions for model P2 with $T = 100$ and $T = 150$ . . . . . . . 477
C.4 Power functions for model P4 with $T = 200$ and $T = 300$ . . . . . . . 478
C.5 Power functions for model P3 with $T = 400$ and $T = 500$ . . . . . . . 479
C.6 Power functions for model P6 with $T = 200$ and $T = 300$ . . . . . . . 480
C.7 Power functions for model P4 with $T = 400$ and $T = 500$ . . . . . . . 481
C.8 Power functions for model P5 with $T = 200$ and $T = 300$ . . . . . . . 482
C.9 Power functions for model P5 with $T = 300$ and $T = 400$ . . . . . . . 483
C.10 Power functions for model P6 with $T = 200$ and $T = 300$ . . . . . . . 484
C.11 Power functions for model P6 with $T = 200$ and $T = 300$ . . . . . . . 485
C.12 Power functions for model P1b with $T = 100$ and $T = 150$ . . . . . . . 486
C.13 Power functions for model P1b with $T = 100$ and $T = 200$ . . . . . . . 487
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Autoregressive Process</td>
</tr>
<tr>
<td>CR</td>
<td>Continuous Record</td>
</tr>
<tr>
<td>DM</td>
<td>Diebold and Mariano</td>
</tr>
<tr>
<td>GL</td>
<td>Generalized Laplace</td>
</tr>
<tr>
<td>GR</td>
<td>Giacomini and Rossi</td>
</tr>
<tr>
<td>HAC</td>
<td>Heteroskedasticity and Autocorrelation Consistent</td>
</tr>
<tr>
<td>HAR</td>
<td>Heteroskedasticity and Autocorrelation Robust</td>
</tr>
<tr>
<td>LS</td>
<td>Least Squares</td>
</tr>
<tr>
<td>MAE</td>
<td>mean absolute error</td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary Least Squares</td>
</tr>
<tr>
<td>RMSE</td>
<td>root-mean-squared error</td>
</tr>
<tr>
<td>Std</td>
<td>standard deviation</td>
</tr>
</tbody>
</table>
Chapter 1

Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models

1.1 Introduction

This chapter develops a theory of evolutionary spectra for heteroskedasticity- and autocorrelation-robust (HAR) inference when the data may not satisfy second-order stationarity. This situation occurs regularly in empirical economics because economic models are, in general, misspecified and/or because economic data is nonstationary, i.e., the model parameters that govern the data change over time [cf. Perron (1989), Stock and Watson (1996) and the surveys of Ng and Wright (2013) and Giacomini and Rossi (2015)].\footnote{In this chapter, by nonstationary we consider processes that exhibit nonstationarity but whose sum of absolute autocovariances is finite. That is, we rule out processes with unbounded second moments (e.g., unit root). For unit root or trending time series, one has to first apply some differencing or detrending technique. Then, our discussion here applies to such transformed time series.} HAR inference builds upon the estimation of covariance matrices of parameter estimators in linear and nonlinear models. Such heteroskedasticity and autocorrelation consistent (HAC) estimators are then used to construct confidence intervals for model parameters and to conduct hypotheses testing about them. For example, for the linear model HAR inference requires estimation of the limit of the covariance matrix (often referred to as the long-run variance matrix) of $V_t \triangleq x_t e_t$ where...
$x_t$ is a vector of regressors and $e_t$ is an unobservable disturbance. The fundamental works in econometrics of Newey and West (1987; 1994) and Andrews (1991) have provided HAC estimators of the covariance matrix of $V_t$ essentially under the assumption that $\{V_t\}$ is fourth-order stationary. This leads to an elegant theory which relates the limit of the covariance matrix of $V_t$ to the spectral density matrix at frequency $\omega = 0$ of $V_t$. The latter matrix is usually denoted as $f(\omega)$ and when evaluated at $\omega = 0$ it is equal to the sum of all autocovariances $\mathbb{E}(V_tV'_{t-k})$ of $\{V_t\}$. Under the assumption of second-order stationarity, $\mathbb{E}(V_tV'_{t-k})$ depends on $k$ but not on $t$. This is not an innocuous assumption—it is usually made for mathematical convenience because it provides the basis for general asymptotic considerations. In practice, economic variables exhibit a nonstationary behavior. To elaborate on this, the early study of Stock and Watson (1996) documented that 70-80% of the macroeconomic variables were found to be unstable in that models describing them exhibited some kind of parameter instability. In addition, even if the data are stationary, model misspecification may induce a nonstationary pattern for the unexplained part of economic models (i.e., for the disturbance sequence $\{e_t\}$). When this occurs, $\mathbb{E}(V_tV'_{t-k})$ depends also on $t$ in addition to $k$. The spectral density matrix $f(\omega)$ of $\{V_t\}$ is no longer well-defined and the main theory behind the classical HAC estimators of Newey and West and Andrews (1991)—based on stationarity—is not immediately applicable. Our simulation study shows that for standard data-generating mechanisms and mild form of misspecification in linear regression models, the usual $t$-tests for the significance of a regression coefficient might have little or no power. This problem becomes even more serious for economic forecasting. Popular tests for forecast evaluation [e.g., Die-

---

2However, Newey and West (1987) established consistency under nonstationarity and Andrews (1991) in his Section 8 considered the case of unconditionally heteroskedastic random variables and derived some asymptotic properties of the classical HAC estimators in that context. Unfortunately, it seems that the quality of the approximations in small-samples is poor as we show in this chapter (see also the references provided below).
bold and Mariano (1995)] and tests for forecast instability [cf. Casini (2018b) and Giacomini and Rossi (2009)] when standardized by the classical HAC estimators may suffer from issues such as non-monotonic power and little or no power. Classical HAC estimators estimate an average of a time-varying spectrum. Misspecification and/or nonstationarity make the series appear much more persistent. As a consequence, HAC standard errors are too large and when used as normalization of test statistics, the tests have little or no power.

We introduce a new class of nonstationary stochastic processes which possess a spectrum which varies both over frequencies and time. We work in an infill asymptotic setting akin to the one used in nonparameteric regression [cf. Robinson (1989)] whereby we rescale the original time scale $[1, T]$ by dividing each $t$ by $T$. This generates a new time scale with index $u \triangleq t/T$ where $u \in [0, 1]$. For a process $\{V_t\}_{t=1}^T$, its spectrum at frequency $\omega$ and time $u$ is denoted as $f(u, \omega)$. We allow $f(u, \omega)$ to change slowly yet continuously as well as to change abruptly in $u$ at a finite number of time points—the latter feature is important because it allows for structural breaks in the spectrum of $V_t$ which is a prominent feature of economic data. We name this class Segmented Locally Stationary. It is related to the locally stationary processes introduced by Dahlhaus (1997)—who formalized the ideas on time-varying spectrum of Priestley (1965)—that have the characterizing property of behaving as a stationary process in a small neighborhood $[u - \varepsilon, u + \varepsilon]$ of $u$, for each $u \in [0, 1]$. This is achieved via smoothness assumptions on $f(u, \omega)$ in $u$. This class rules out important models used in applied economics such as structural change and regime switching-type models [cf. Bai and Perron (1998), Casini and Perron (2017a; 2017b; 2018a; 2018b), Elliott and Müller (2007) and Hamilton (1989)]. The class of Segmented Locally Stationary processes extends some of the analysis of Dahlhaus (1997).
to processes that can have a time-varying spectrum with a finite number of discontinuity in $u \in [0, 1]$. We use inference methods from Casini (2018c) for testing and estimating the dates at which such discontinuities occur.

Next, we discuss an alternative HAC estimation method based on the assumption that $\{V_t\}$ is Segmented Locally Stationary. The estimator is defined on the time domain and is positive semidefinite. In addition to the usual smoothing procedure over the autocovariance lag orders akin to the classical kernel HAC estimators, it is essential to apply a second smoothing procedure over time. The classical HAC estimator would estimate $E(V_t V_t')$ by a sample average over $t = k + 1, \ldots, T$. Our proposed HAC estimator applies a kernel smoothing over $t$ thereby using realizations of $V_t$ close to time $t$. Since $\{V_t\}$ is Segmented Locally Stationary, $E(V_t V_t')$ changes smoothly in $t$—as long as $t$ is away from the change-points in the spectrum $f(t/T, \omega)$. This yields good estimates for the time path of $E(V_t V_t')$ for all $k$. Introducing a double smoothing procedure requires us to control the relative degree of smoothing of the two procedures. More specifically, we need to study the rate at which the bandwidth sequences each associated to a given kernel converge to zero. In order to make our HAC estimator operational, we determine optimal values for the bandwidth sequences used to define the estimators.

We obtain an optimal HAC estimator under a mean-squared error (MSE) optimality criterion. The optimal HAC estimator uses the Quadratic Spectral (QS) kernel for smoothing over $k$ and a quadratic-type kernel for smoothing over $t$. The latter is a transformation of the Epanechnikov’s kernel [cf. Epanechnikov (1969)]. The consistency of the HAC estimator is established under stronger conditions on the growth rate of the bandwidth sequence corresponding to the kernel smoothing over $k$ than what is required for the classical HAC estimators under stationary data as
established in Andrews (1991). We require $b_{1,T}b_{2,T} = o\left(T^{-1/2}\right)$ where $b_{1,T} \to 0$ and $b_{2,T} \to 0$ are the bandwidth sequences associated to the two smoothing procedures, respectively. Recall that Andrews (1991) only required $b_{1,T} = o\left(T^{-1/2}\right)$. However, this difference is of little practical importance because optimal bandwidth sequences are typically much slower than $o\left(T^{1/2}\right)$—for the QS kernel HAC estimator of Andrews (1991), $b_{1,T}^{\text{opt}} = O\left(T^{-1/5}\right)$. This implies that when the optimal bandwidths are employed, the rate of convergence of the HAC estimators proposed in this chapter is only an order $O\left(T^{1/10}\right)$ slower than the rate of convergence of the QS kernel HAC estimator of Andrews (1991); the latter being, however, only applicable to stationary data.

It is crucial in our analysis to deal with breaks in the spectrum $f(u, \omega)$. At the break points, $f(u, \omega)$ is no longer continuous in $u$. However, left-continuity is preserved and thus a simple modification to the original HAC estimator allows us to extend the analysis to more general form of nonstationary random variables $\{V_t\}$. This modification relies on a pre-test for breaks in $f(u, \omega)$. When the test detects the breaks, the dates at which the breaks occur are estimated. This information is used in the construction of the HAC estimator. These inference methods about the break points are developed in Casini (2018c). In particular, the test is a two-sample $t$-test over asymptotically small blocks of data and follows an extreme value limiting distribution.

We propose a data-dependent procedure that determines the bandwidth parameter for given kernels and a given sample size. The procedure is based on the so-called “plug-in” approach used previously by both Newey and West (1994) and Andrews (1991). This approach is characterized by plugging-in estimates into the analytical formula for the optimal bandwidth. These estimates are based on an approximating
parametric model. Unlike Andrews (1991), assuming a standard AR(1) or VAR(1) as the approximating model is no longer a useful assumption because candidate parametric models should lead to processes with a time-varying spectrum. Our candidate parametric models do not have constant parameters; rather, they are described by parameter curves which can be estimated by applying ordinary methods to local data.


The econometric works on HAC estimation builds upon the theory behind spectral density estimation for stationary processes developed in statistics and time series analyses. Fundamental papers are Bartlett (1950), Berk (1974), Grenander and Rosenblatt (1953), Parzen (1957) and Priestley (1962; 1981). For nonstationary processes, HAC estimation has not been discussed so far. Our framework builds upon the work of Priestley (1965) and Dahlhaus (1997). The former introduces the concept of oscillatory processes while the latter develops theoretical results for locally stationary processes [see Koo and Linton (2012) and Vogt (2012) for alternative definitions of
local stationarity]. We contribute to the literature on nonstationary processes by introducing a framework for the study of processes whose spectra evolve continuously over time except at a finite number of time points where the spectrum can change abruptly.

The main contribution of the chapter is the development of a HAC estimation method that can be used even when the underlying time series is not second-order stationary. Our simulation study shows that usual \( F \)- or \( t \)-tests in the linear regression models standardized by classical HAC estimators suffer not only from size distortions—as it was already documented in the literature—but also significant power losses when either the errors or regressors are not second-order stationary and/or when models are subject to misspecification or parameter instability (even when that is accounted for). Note that the latter result is stronger than the one emerging from Deng and Perron (2008), Kim and Perron (2009) and Chang and Perron (2018) who analyzed the effects of unaccounted structural breaks on HAC estimates. When classical HAC estimates are applied to commonly used \( t \)-tests for forecast evaluation [e.g., the tests of Diebold and Mariano (1995)] and the tests for forecast instability [cf. Casini (2018b) and Giacomini and Rossi (2009)], these tests have little or no power.

As explained above, this occurs because nonstationary in \( \{V_t\} \) induces classical HAC methods to overestimate the extent of the dependence or variation in \( V_t \). Then HAC estimates are too large and robust test statistics have in turn little or no power in addition to substantial size distortions. In contrast, the theory of evolutionary spectra presented in this chapter and the associated HAC estimation method, does not present such issues and HAR inference can be safely conducted in the usual way. The issues with classical HAC estimators in the context of economic forecasting have been recently documented by Casini (2018b), Crainiceanu and Vogelsang (2007), Fossati
More recently, Lazarus et al. (2017) used higher-order expansions to provide a size-power frontier for kernel and orthogonal series tests using nonstandard fixed-\(b\) critical values in the context of the Gaussian multivariate location model and showed that the QS spectral kernel and the Empirical Weighted Periodogram (EWP) allow achieving the frontier. Based on their results, Lazarus et al. (2018) recommended the use of either the Newey-West (1987) HAC estimator or a version of the EWP estimator for constructing HAR inference tests, each implemented with a long bandwidth. Their theoretical and many of their simulation results, which are tied to stationary data and correctly specified models\(^3\), do not necessarily extrapolate to contexts where the data are nonstationary and/or the models are misspecified. First, such HAR inference tests can be overly undersized. Second, as explained above classical HAC estimators already face power losses in such contexts because they do not account of time variation in covariance structure and a long bandwidth makes these issues ever more severe. Intuitively, implementation of classical HAC with long bandwidths does account of the high serial dependence better but inevitably implies that more data with greater differences in covariance structure is combined together to estimate autocovariances of large lag order. Such covariances are then overestimated and HAR inference tests have in turn no power in such contexts.\(^4\) The new HAC estimator uses bandwidths that are smaller than the ones recommended by Lazarus et al. (2018) but similar to the ones derived under MSE criterion by Andrews (1991) and Newey and

\(^3\)In Lazarus et al. (2018) the authors actually also use designs in which the data are generated from an estimated dynamic factor model under stationarity. In this chapter, we use standard linear regression models as in Lazarus et al. (2017).

\(^4\)This suggests that also the near-optimality properties established for the EWP (or some version thereof) for the location model with stationary Gaussian errors [cf. Lazarus et al. (2017) and Duo (2018)] may not extend to more general situations.
West (1994). The associated HAR inference tests then control the size well and have always good power.

The remainder of the chapter is organized as follows. Section 3.2 introduces the statistical setting, a new class of nonstationary processes and the new HAC estimation method. We first deal with the case when there are no breaks in the spectrum. Section 1.3 presents consistency, rates of convergence and the asymptotic MSE results for the HAC estimators. Asymptotically optimal kernels and bandwidth parameters are derived in Section 1.4. A data-dependent method for choosing the bandwidth parameters and its asymptotic properties are introduced in Section 1.5. We then consider processes that can have discontinuity in the spectrum. Section 1.6 discusses inference methods about the break points. Section 1.7 extends the results of Section 1.3-1.5 to such processes. Section 1.8 presents Monte Carlo results regarding the small-sample behavior of HAR inference tests based on the HAC estimators. Section 4.7 concludes the chapter. An appendix contains all mathematical proofs, additional simulations and an empirical application.

1.2 The Statistical Environment

To motivate our approach, consider the linear regression model estimated by a least-squares (LS) method: \( y_t = x_t' \beta_0 + e_t \) \((t = 1, \ldots, T)\), where \( \beta_0 \in \Theta \subset \mathbb{R}^p \), \( y_t \) is an observation on the dependent variable, \( x_t \) is a \( p \)-vector of regressors and \( e_t \) is an unobservable disturbance. The LS estimator is given by \( \hat{\beta} = (X'X)^{-1} X'Y \), where \( Y = (y_1, \ldots, y_T)' \) and \( X = (x_1, \ldots, x_T)' \). Inference about \( \beta_0 \) requires estimation of
Var \left( \sqrt{T} \left( \hat{\beta} - \beta_0 \right) \right) defined as

\[
\text{Var} \left( \sqrt{T} \left( \hat{\beta} - \beta_0 \right) \right) \triangleq \left( T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E} \left( e_s x_s (e_t x_t)' \right) \left( T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1}.
\]

Consistent estimation of \( \text{Var} \left( \sqrt{T} \left( \hat{\beta} - \beta_0 \right) \right) \) relies on consistent estimation of

\[
\lim_{T \to \infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} \left( e_s x_s (e_t x_t)' \right).
\]

More generally, one needs a consistent estimate of \( J \triangleq \lim_{T \to \infty} J_T \) where \( J_T = T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} \left( V_s (\beta_0) V_t (\beta_0)' \right) \) with \( V_t (\beta) \) being a random \( p \)-vector for each \( \beta \in \Theta \).

For the linear regression model, \( V_t (\beta) = (y_t - x_t' \beta) x_t \). Having \( y_t = x_t' \beta_0 + e_t \) implicitly implies that the model is correctly specified. However, the literature on HAC estimation also covers some form of misspecification; for example, in Andrews (1991) one can interpret \( \beta_0 = \beta^* \) as the pseudo-true parameter and his analysis remains valid as long as \( \sqrt{T} \left( \hat{\beta} - \beta^* \right) = O_p(1) \). We adopt the same approach here so that \( \beta_0 = \beta^* \) when the model is misspecified. We then assume below the usual condition \( \sqrt{T} \left( \hat{\beta} - \beta_0 \right) = O_p(1) \) no matter whether the model is correctly specified or not.

Our statistical problem is to estimate \( J = \lim_{T \to \infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} \left( V_s V_t' \right) \) when \( \{V_t\} \) is a Segmented Locally Stationary process. This latter class of processes is defined in Section 1.2.1. Such an estimate of \( J_T \) can then be used to conduct HAR inference in the usual way. By a change of variables, \( J_T \) can be rewritten as

\[
J_T = \sum_{k=-T+1}^{T-1} \Gamma_{T,k}, \quad \text{where} \quad \Gamma_{T,k} = \begin{cases} T^{-1} \sum_{l=k+1}^{T} \mathbb{E} \left( V_l V_{l-k}' \right) & \text{for } k \geq 0, \\ T^{-1} \sum_{l=-k+1}^{T} \mathbb{E} \left( V_{l+k} V_l' \right) & \text{for } k < 0. \end{cases}
\]

\(^5\)In Section A.1 of the Supplement we review other examples such as GMM [cf. Hansen (1982)], and IV.
and \( V_t = V_t(\beta_0), \ t = 1, \ldots, T \). The rest of this section is structured as follows. In Section 1.2.1 we introduce a class of time series that have an evolutionary spectrum. Section 1.2.2 presents a new HAC estimation method. Throughout we adopt the following notational conventions. All limits are taken as \( T \to \infty \). \( \text{vec}(\cdot) \) is the vectorization operator. The \( j \)th element of a vector \( x \) is indicated by \( x^{(j)} \) while the \( (j, l) \)th element of a matrix \( X \) is indicated as \( X^{(j,l)} \). \( \text{tr}(\cdot) \) denotes the trace function and \( \otimes \) denotes the tensor (or Kronecker) product operator. The \( p^2 \times p^2 \) matrix \( C_{pp} \) is a commutation matrix that transforms \( \text{vec}(A) \) into \( \text{vec}(A') \), i.e., \( C_{pp} = \sum_{j=1}^p \sum_{l=1}^p t_j t'_l \otimes t_l t'_j \), where \( t_i \) is the \( i \)th elementary \( p \)-vector. \( \lambda_{\text{max}}(A) \) denotes the largest eigenvalue of the matrix \( A \). The notation \( W \) is used for a \( p^2 \times p^2 \) weight matrix. For a sequence of matrices \( \{A_T\} \), we write \( A_T = o_p(1) \) if each of its elements is \( o_p(1) \) and likewise for \( O_p(1) \). Let \( 0 < \lambda_0 < \ldots < \lambda_{m+1} < 1 \). A function \( G(\cdot) : [0, 1] \times \mathbb{R} \to \mathbb{C} \) is said to be piecewise (Lipschitz) continuous with \( m + 1 \) segments if it is (Lipschitz) continuous within each segment. For example, it is piecewise Lipschitz continuous if for each segment \( j = 1, \ldots, m + 1 \) it satisfies \( \sup_{u \neq v} |G(u) - G(v)| \leq K |u - v| \) with \( \lambda_{j-1} < u, v \leq \lambda_j \) for some \( K < \infty \). If we say piecewise Lipschitz continuous with index \( \vartheta > 0 \), then the above inequality is replaced by \( \sup_{u \neq v} |G(u) - G(v)| \leq K |u - v|^\vartheta \). Let \( \|\cdot\| \) denote the Euclidean norm of a vector or matrix. If \( x \) is a stochastic vector, the same notation is used for the \( L^2 \) norm. We use \( \lfloor \cdot \rfloor \) to denote the largest smaller integer function. We use \( \overset{\text{P}}{\to} \) to denote convergence in probability. \( \mathbb{R} \) is used for the set of real numbers while \( \mathbb{C} \) is used for the set of complex numbers. The notation \( \bar{A} \) is used for the complex conjugate of \( A \in \mathbb{C} \). The symbol \( \overset{\Delta}{=} \) is for definitional equivalence.
1.2.1 Segmented Locally Stationary Processes

Suppose we are given a stochastic process \( \{V_t\}_{t=1}^T \) defined on an abstract probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where \(\Omega\) is the sample space, \(\mathcal{F}\) is the \(\sigma\)-algebra and \(\mathbb{P}\) is a probability measure. In order to introduce a framework for analyzing time series models that have a time-varying spectrum it is necessary to introduce an infill asymptotic setting whereby we rescale the original discrete time horizon \([1, T]\) by dividing each \(t\) by \(T\). Letting \(u = t/T\) and \(T \to \infty\), this defines a new time scale \(u \in [0, 1]\) which we interpret as saying that as \(T \to \infty\) we observe more and more realizations of \(V_t\) close to time \(t\) (i.e., we observe the rescaled process \(V_{Tu}\) on the interval \([u - \varepsilon, u + \varepsilon]\), where \(\varepsilon > 0\) is a small number).

In order to define a general class of nonstationary processes we shall start from processes that have a time-varying spectral representation as follows:

\[
V_{t,T} = \mu(t/T) + \int_{-\pi}^{\pi} \exp(i\omega t) A(t/T, \omega) d\xi(\omega),
\]

where \(i \triangleq \sqrt{-1}\), \(\mu(t/T)\) is the trend function, \(A(t/T, \omega)\) is the transfer function and \(\xi(\omega)\) is some stochastic process whose properties are specified below. Observe that this representation is similar to the spectral representation of stationary processes [see Anderson (1971), Brillinger (1975), Hannan (1970) and Priestley (1981) for an accessible treatise of some of the introductory concepts of this section]. We shall see that the main difference is that \(A(t/T, \omega)\) and \(\mu(t/T)\) are not constant in \(t\).\(^6\) Dahlhaus (1997) used the time-varying spectral representation to define the so-called locally stationary processes which are characterized, broadly speaking, by smoothness conditions on \(\mu(\cdot)\) and \(A(\cdot, \cdot)\). Locally stationary processes have been

\(^6\)It should be noted that in the context of HAR inference, a minimal assumption on \(V_t\) is that it has zero mean. This would imply that \(\mu(t/T) = 0\) for all \(t\). However, in the definitions of this subsection we allow for arbitrary \(\mu(t/T)\) so as to introduce a general framework.
used widely in both statistics and economics, although in the latter field they are best known as time-varying parameter processes [see e.g. Cai (2007) and Chen and Hong (2012)]. Due to the smoothness restrictions, these processes exclude many prominent econometric models that account for time variation in the parameters.\footnote{It should be mentioned that the estimation method of Hamilton (1989) assumes nonlinear stationarity to model nonstationary (or regime switching) time series.} For example, structural change and regime switching-type models do not belong to the class of locally stationary processes because parameter changes occur suddenly at a particular point in time and not smoothly over short periods of time. Due to this important limitation, we propose a class of nonstationarity processes which allow both continuous and discontinuous changes in the parameters. Stationarity and local stationarity are recovered as special cases.

Likewise to locally stationary processes, we begin with the representation (1.2.1) and note that it will hold only approximately as we explain in some examples below.

**Definition 1.2.1.** (Segmented Locally Stationary Processes) A sequence of stochastic processes $V_{t,T}$ $(t = 1, \ldots, T)$ is called \textbf{Segmented Locally Stationary} with $m + 1$ regimes, transfer function $A^0$ and trend $\mu$, if there exists a representation

$$V_{t,T} = \mu_j(t/T) + \int_{-\pi}^{\pi} \exp(i\omega t) A^0_{j,t,T}(\omega) d\xi(\omega), \quad (t = T^0_{j-1} + 1, \ldots, T^0_j),$$

for $j = 1, \ldots, m + 1$, where by convention $T^0_0 = 0$ and $T^0_{m+1} = T$ and the following holds:

(i) $\xi(\lambda)$ is a stochastic process on $[-\pi, \pi]$ with $\xi(-\omega) = \xi(-\omega)$ and

$$\text{cum}\{d\xi(\omega_1), \ldots, d\xi(\omega_r)\} = \varphi \left( \sum_{j=1}^{r} \omega_j \right) g_r(\omega_1, \ldots, \omega_{r-1}) d\omega_1 \ldots d\omega_r,$$
where \( \text{cum} \{ \cdots \} \) denotes the cumulant spectra of \( r \)th order, \( g_1 = 0, g_2 (\omega) = 1, \)

\[ |g_r (\omega_1, \ldots, \omega_{r-1})| \leq M_r \]

for all \( r \) with \( M_r \) being a constant that may depend on \( r \),

and \( \varphi (\omega) = \sum_{j=-\infty}^{\infty} \delta (\omega + 2\pi j) \) is the period \( 2\pi \) extension of the Dirac delta function \( \delta (\cdot) \).

(ii) There exists a constant \( K \) (which depends on \( j \)) and a piecewise continuous function \( A : [0, 1] \times \mathbb{R} \to \mathbb{C} \) such that, for each \( j \), there exists a \( 2\pi \)-periodic function \( A_j : (\lambda_{j-1}, \lambda_j] \times \mathbb{R} \to \mathbb{C} \) with \( A_j (u, -\omega) = \overline{A_j (u, \omega)} \), \( \lambda_0^j \triangleq T_0^j / T \) and

\[
A (u, \omega) = A_j (u, \omega) \quad \text{for} \quad \lambda_{j-1}^0 < u \leq \lambda_j^0,
\]

\[
\sup_{1 \leq j \leq m+1} \sup_{\lambda_{j-1}^0 < t \leq \lambda_j^0} \left| A_{j,t,T}^0 (\omega) - A_j (t/T, \omega) \right| \leq KT^{-1},
\]

for all \( T \), where \( j = 1, \ldots, m+1 \).

(iii) \( \mu_j (t/T) \) is piecewise continuous.

The smoothness properties of \( A \) in \( u \) guarantees that \( V_{t,T} \) has a piecewise locally stationary behavior. Later we will require additional smoothness properties for \( A \), namely left-differentiability in \( u \) and differentiability in \( \omega \). For the rest of this section, \( t \) and \( s \) always denote time points in the original time interval \([1, T]\) while \( u \) and \( v \) are time points in the rescaled time interval \([0, 1]\). We collect the break dates in \( \mathcal{T} \triangleq \{ T_1^0, \ldots, T_m^0 \} \).

**Example 1.2.1.** (i) Suppose \( X_t \) is a stationary process with spectral representation

\[
X_t = \int_{-\pi}^{\pi} \exp (i\omega t) A (\omega) d\xi (\omega),
\]

and \( \mu, \sigma : [0, 1] \to \infty \) are piecewise continuous. Then,

\[
V_{t,T} = \mu_j (t/T) + \sigma_j (t/T) X_t, \quad T_{j-1}^0 < t \leq T_j^0, 1 \leq j \leq m+1,
\]
is a Segmented Locally Stationary process with \(m + 1\) regimes where \(A_{j,t,T}^0(\omega) = A_j(t/T, \omega) = \sigma_j(t/T)A(\omega)\). Within each segment, \(V_{t,T}\) is locally stationary. When \(t = Tu\) is away from the change-points, as \(T \to \infty\) more and more realizations of \(V_{[Tu],T}\) with \(u \in [u - \varepsilon, u + \varepsilon]\) are observed, that is, realizations with amplitude close to \(\sigma_j(u)\) for the appropriate \(j\).

(ii) Suppose \(e_t\) is an i.i.d. sequence and \(V_{t,T} = \sum_{k=0}^{\infty} a_{j,k}(t/T)e_{t-k}, T_{j-1} \leq t < T_j, 1 \leq j \leq m + 1\). Then, \(V_{t,T}\) is Segmented Locally Stationary with \(A_{j,t,T}^0(\omega) = A_j(t/T, \omega) = \sum_{k=0}^{\infty} a_{j,k}(t/T)\exp(-i\omega k)\).

(iii) Autoregressive processes with time-varying coefficients—known as TVAR—augmented with structural breaks are Segmented Locally Stationary. In this case, we do not have the exact relationship \(A_{j,t,T}^0(\omega) = A_j(t/T, \omega)\) but only the approximate relationship (1.2.4).

If \(\mu\) and \(A^0\) do not depend on \(t\) nor \(T\), then \(V_{t,T}\) is stationary and the ordinary spectral representation of stationary processes applies. Thus, the classical asymptotic theory for stationary processes is a special case of our approach. On the other hand, if there is only a single regime (i.e., \(m = 0\)) then \(V_{t,T}\) is locally stationary and the asymptotic theory of Dahlhaus (1997) applies. However, \(m = 0\) rules out structural change and regime switching models—only continuously time-varying parameter models are allowed. By allowing \(m > 0\) we essentially propose a framework whereby parameter variation can occur either smoothly or abruptly—both are relevant features for economic time series.

Early work on nonstationary processes with a time-varying spectral representation was first carried out by Priestley (1965, 1981). However, his definition of oscillatory processes does not lend itself to a framework for asymptotic considerations since he was mainly concerned with the development of a stochastic representation.
Dahlhaus (1997) introduced locally stationary processes and his framework actually allowed for a rigorous asymptotic treatment of statistical inference problems involving nonstationary processes. However, his analysis does not allow for discontinuous changes in the spectrum.

We define the spectrum of $V_{t,T}$ in (1.2.1) (for fixed $T$) as

$$f_{j,T}(u, \omega) \triangleq \begin{cases} (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \text{Cov}(V_{uT-3|s|/2,T}, V_{uT-|s|/2,T}) \exp(-i\omega s), & Tu \in \mathcal{T}, \ u = \lambda^0_j \\ (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \text{Cov}(V_{uT-s/2,T}, V_{uT+s/2,T}) \exp(-i\omega s), & Tu \notin \mathcal{T}, \ \lambda^0_{j-1} < u < \lambda^0_j \end{cases}$$

with $A_{0,t,T}^0(\omega) = A_1(0, \omega)$ for $t < 1$ and $A_{m+1,t,T}^0(\omega) = A_{m+1}(1, \omega)$ for $t > T$. Our definition coincides with the Wigner-Ville spectrum [cf. Martin and Flandrin (1985)] only when there are no change-points (i.e., $m = 0$). Below we show that $f_{j,T}(u, \omega)$ tends in mean-squared to $f_j(u, \omega) \triangleq |A_j(u, \omega)|^2$ for $T^0_{j-1}/T < u = t/T \leq T^0_j/T$ which is the spectrum that corresponds to the spectral representation. Therefore, we call $f_j(u, \omega)$ the time-varying spectral density matrix of the process. Note that for the boundary point $u = 0$ (resp., $u = 1$) the definition of $f_{j,T}(u, \omega)$ is modified in such a way that only observations on the right (resp., on the left) of the boundary point are used.

**Assumption 1.1.** $A(u, \omega)$ is piecewise Lipschitz continuous in the first component and uniformly Lipschitz continuous in the second component, with index $\vartheta > 1/2$ for both.

**Theorem 1.2.1.** Assume $V_{t,T}$ is Segmented Locally Stationary with $m + 1$ regimes
and Assumption (1.2.1) holds. Then, for all $u \in (0, 1)$,

$$
\int_{-\pi}^{\pi} \sum_{j=1}^{m+1} |f_{j,T}(u, \omega) - f_j(u, \omega)|^2 d\omega = o(1).
$$

Let $f(u, \omega) = f_j(u, \omega)$ if $Tu \in (T_{j-1}^0, T_j^0]$ so as to suppress the subscript $j$ from $f$. It is well-known that, even when $m = 0$, the spectral representation (1.2.2) is not unique [cf. Priestley (1981), Chapter 11.1]. A consequence of Theorem 1.2.1 is that $\{f_j(u, \omega) = |A_j(u, \omega)|^2, j = 1, \ldots, m + 1\}$ is uniquely determined from the whole triangular array. However, there may exist other non-smooth representations. Furthermore, it should be also noted that $f_j(u, \omega)$ is not the limit of the Wigner-Ville spectrum, the difference being that the latter spectrum is required to be smooth. The equivalence holds only point-wise for each $u$ away from the break points, i.e., $Tu \notin T$.

For $Tu \notin T$ with $T_{j-1}^0/T < u = t/T < T_j^0/T$, only the realizations of $V_{t,T}$ in the time interval $u \in [u - n/T, u + n/T]$ with $n \to \infty$ contribute to $f_j(u, \omega)$. Since this interval is fully contained in a segment $j$ where $A_j(u, \omega)$ is smooth and given that the length of this interval tends to zero, $V_{t,T}$ becomes “asymptotically stationary” on this interval. The length of the interval in which $V_{t,T}$ can be considered stationary is given by $n \ln n/T^0 \to 0$. For $Tu \in T$, the arguments are different. Suppose $Tu = T_j^0$. The spectrum $f_{j,T}(u, \omega)$ is defined in such a way that only observations prior to $T_j^0$ are used in order to construct an approximation to $f_j(u, \omega)$. Since the length of this interval tends to zero and $A_j(u, \omega)$ is left-Lipschitz continuous, then those observations become “asymptotically stationary” and thus provide the same information about $f_j(u, \omega)$. Given $f(u, \omega)$, we can define the local covariance of $V_{t,T}$ at rescaled time $u$ with $Tu \notin T$ and lag $k \in \mathbb{Z}$ as $c(u, k) \triangleq \int_{-\pi}^{\pi} e^{i\omega k} f(u, \omega) d\omega$. The same definition is also used when $Tu \in T$ and $k \geq 0$. For $Tu \in T$ and $k < 0$ it is defined as $c(u, k) \triangleq \int_{-\pi}^{\pi} e^{i\omega k} A(u, \omega) A(u - k/T, -\omega) d\omega$. 


In Section 1.6 we present methods for testing for breaks in the spectrum \( f(u, \omega) \) and methods for estimating the break dates \( T_j^0 \) once the null hypothesis is rejected.

1.2.2 HAC Estimation

In model (1.2.2) if \( m = 0 \) and \( A^0 \) is constant in its first argument, then \( \{V_{t,T}\} \) is second-order stationary. Its spectral density matrix is then well-defined and equal to

\[
\begin{align*}
    f(\omega) &\triangleq (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \Gamma(k) e^{-i\omega k} \\
    \Gamma(k) &\triangleq \mathbb{E}\left( V_{t,T} V'_{t-k,T} \right) 
\end{align*}
\]

The spectral density matrix at frequency \( \omega = 0 \) plays a prominent role because the limit as \( T \to \infty \) of the estimand \( J_T \) equals \( 2\pi f(0) \). Nonstationarity implies that the spectral density \( f(\omega) \) is not well-defined, or more precisely, it is time-varying since \( \mathbb{E}\left( V_{t,T} V'_{t-k} \right) \) depends on \( k \) as well as on \( t \). The class of Segmented Locally Stationary processes introduced above accommodates this property because they have a time-varying spectrum \( f(u, \omega) \).

Accordingly, we introduce the notation \( \Gamma_u(k) \triangleq \mathbb{E}\left( V_{T_u,T} V'_{T_u-k,T} \right) \) where \( u = t/T \). Adapting the arguments from Dahlhaus (1997, 2012), we have \( \Gamma_u(k) = c(u, k) + O(T^{-1}) \) uniformly in \( 1 \leq j \leq m + 1, \ T u \leq T_j^0 \) and \( k \in \mathbb{Z} \). Under the rescaling \( u = t/T, \ u \in [0, 1] \), the limit of the estimand \( J_T \) is given by,

\[
J \triangleq \lim_{T \to \infty} J_T = \int_0^1 c(u, 0) \, du + \sum_{k=1}^{\infty} \int_0^1 \left( c(u, k) + c(u, k)' \right) \, du.
\]

It can be shown that \( J = 2\pi \int_0^1 f(u, 0) \, du \). Dahlhaus (2012) discussed how to estimate \( f(u, \omega) \) for the scalar case under smoothness in both arguments using the smoothed local periodogram. Our goals are to estimate \( J \) using a time-domain method and to relax the smoothness assumption in \( u \). Also, our estimator can be shown to be consistent for \( J_T \) even when \( V_t \) is not segmented locally stationary. The class
of estimators of \( J \) relies on kernel estimation of the local covariances \( c ( \cdot, \cdot ) \),

\[
\hat{J}_T = \hat{J}_T ( b_{1,T}, b_{2,T} ) \triangleq \frac{T}{T - p} \sum_{k=-T+1}^{T-1} K_1 ( b_{1,T} k ) \hat{\Gamma} ( k ),
\]

with

\[
\hat{\Gamma} ( k ) \triangleq \frac{n_T}{T - n_T} \sum_{r=0}^{\lfloor (T - n_T)/n_T \rfloor} \hat{c}_T ( rn_T/T, k ),
\]

where \( K_1 ( \cdot ) \) is a real-valued kernel in the class \( K_1 \) defined below, \( b_{1,T} \) is a bandwidth sequence, \( n_T \to \infty \) satisfying the conditions given below, and

\[
\hat{c}_T ( rn_T/T, k ) \triangleq \begin{cases} 
(Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2 \left( \frac{( (r+1)n_T - (s+k/2) )/T}{b_{2,T}} \right) \tilde{V}_{s-k}^t \tilde{V}_{s}^t, & k \geq 0 \\
(Tb_{2,T})^{-1} \sum_{s=-k+1}^{T} K_2 \left( \frac{( (r+1)n_T - (s-k/2) )/T}{b_{2,T}} \right) \tilde{V}_{s+k}^t \tilde{V}_{s}^t, & k < 0
\end{cases}
\]

(1.2.5)

with \( K_2 \) being a real-valued kernel in the class \( K_2 \) and \( b_{2,T} \) is a bandwidth sequence. \( \hat{c}_T ( u, k ) \) is an estimate of the local covariance \( c ( u, k ) \) of lag \( k \) at time \( u = rn_T/T \). Estimation of \( c ( u, k ) \) for locally stationary processes was considered by Dahlhaus (2012). Here we deal with the multivariate case and allow for segmented locally stationary.

We note that in order to guarantee positive semi-definiteness in finite-sample of \( J_T \) for all kernels in \( K_2 \) in practice one has to replace \( K_2 \left( ((r+1)n_T - (s+k/2)) / Tb_{2,T} \right) \) above by

\[
(K_2 \left( ((r+1)n_T - s) / Tb_{2,T} \right))^{1/2} (K_2 \left( ((r+1)n_T - (s-k)) / Tb_{2,T} \right))^{1/2} \text{ for } k \geq 0,
\]

and replace \( K_2 \left( ((r+1)n_T - (s-k/2)) / Tb_{2,T} \right) \) by

\[
(K_2 \left( ((r+1)n_T - s) / Tb_{2,T} \right))^{1/2} (K_2 \left( ((r+1)n_T - (s+k)) / Tb_{2,T} \right))^{1/2} \text{ for } k < 0.
\]
To see why, note that we need each $V_t (t = 1, \ldots, T)$ to be assigned the same weight across different $k$ for any given $r$. Then, letting

$$\hat{V}_t = (K_2 (((r + 1) n_T - t) / Tb_{2,T}))^{1/2} \hat{V}_t,$$

we can use the same arguments as in Andrews (1991) applied now to $\hat{V}_t$ to show that $J_T$ is positive semidefinite.

The factor $T / (T - p)$ is an optional small-sample degrees of freedom adjustment introduced to offset the effect of estimation of the $p$-vector $\beta$. In Section 1.3-1.4, we consider estimators $\hat{J}_T$ for which $b_{1,T}$ and $b_{2,T}$ are given sequences. In Section 1.5, we consider adaptive estimators $\hat{J}_T$ for which $b_{1,T}$ and $b_{2,T}$ are data-dependent. Observe that the optimal $b_{2,T}$ actually depends on the properties of $\{V_{t,T}\}$ in any given block [i.e., $b_{2,T} = b_{2,T} (t/T)$]. Since the order of $b_{2,T} (\cdot)$ is the same across blocks, we omit this notation for the developments of the asymptotic results. However, when we determine the data-dependent estimate of $b_{2,T} (\cdot)$, we will estimate $b_{2,T} (rn_T/T)$ for each block so as to reflect the time-varying properties of the optimal $b_{2,T} (\cdot)$. The estimator $\hat{J}_T$ involves two kernels. One kernel smooths the lags of the autocovariance to be estimated—akin to the classical HAC estimators—while the other kernel applies smoothing over time on the observations to be used for estimating a given lag of the autocovariance at a given point in time.

The class of kernels $K_1$ is the same as the one considered by Andrews (1991):

$$K_1 = \{ K_1 (\cdot) : \mathbb{R} \to [-1, 1] : K_1 (0) = 1, K_1 (x) = K_1 (-x), \forall x \in \mathbb{R}, \int_{-\infty}^{\infty} K_1^2 (x) dx < \infty, \text{ and } K_1 (\cdot) \text{ is continuous at 0 and at all but finite numbers of points} \}.$$

For $k$ small relative to $T$, the conditions $K_1 (0) = 1$ and $K_1 (\cdot)$ is continuous at zero
implies that the weight given to $\hat{\Gamma}(0)$ is close to one. We shall show below that the Quadratic Spectral (QS) kernel has certain optimality properties. It takes the following form:

$$K_{1}^{QS}(x) = \frac{25}{12\pi^{2}x^{2}} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right).$$

Other examples of kernels in $K_1$ include the Truncated, Bartlett, Parzen and Tukey-Hanning kernel. Their respective expressions can be found in e.g., Priestley (1981). Classical HAC estimators $\hat{J}_T$ corresponding to the Truncated, Bartlett, and Parzen kernels are the estimators proposed by White (1984), Newey and West (1987), and Gallant (1987), respectively. The Tukey-Hanning and QS kernels were analyzed by Andrews (1991). In the spectral and probability density estimation literature, investigation of the kernels in $K_1$ had been conducted by, among others, Priestley (1962) and Epanechnikov (1969).

The quantity $1/b_{1,T}$ can also act as a lag truncation sequence for which lags of order $k > b_{1,T}^{-1}$ receive zero weight. This requires $K_1(x) = 0$ for $|x| > 1$ (and $K_1(x) \neq 0$ for some $|x|$ arbitrarily close to 1). This applies to all kernels discussed above except the QS kernel.

### 1.3 HAC Estimation with Predetermined Bandwidths

In Section 1.3.1 we analyze the asymptotic properties of the local covariance estimate $\tilde{c}(\cdot, \cdot)$. We then use these results in Section 1.3.2 in order to establish consistency, rate of convergence and MSE properties of predetermined bandwidths HAC estimator. Throughout this section, we present results for the case $m = 0$ (i.e., no breaks in the time-varying spectrum). We extend the results to the case $m > 0$ in Section 1.7. Let $\tilde{J}_T$ denote the pseudo-estimator identical to $\hat{J}_T$ but based on the unobserved sequence
\{V_{t,T}\} = \{V_{t,T} (\beta_0)\}$ rather than on $\{\hat{V}_{t,T}\} = \{V_{t,T} (\hat{\beta})\}$.

### 1.3.1 Estimation of Local Covariances

We consider the following class of kernels:

\[
K_2 = \{K_2 (\cdot) : \mathbb{R} \to [0, \infty] : K_2 (x) = K_2 (1 - x), \quad (1.3.1)
\]

\[
\int K_2 (x) \, dx = 1, \quad K_2 (x) = 0, \text{ for } x \notin [0, 1].
\]

The kernels in $K_2$ give zero weight for $x \notin [0, 1]$ and are symmetric around $x = 1/2$. This class was also considered by Dahlhaus (2012) and Dahlhaus and Giraitis (1998) in the context of the estimation of the local covariance of locally stationary processes.

Let $\tilde{c}_T (u, k)$ denote the estimator that uses $\{V_{t,T}\}$. The following proposition provides expressions for the bias, variance and MSE of $\tilde{c}_T (u_0, k)$ for some $u_0 \in (0, 1)$. Let

\[
\text{MSE} (\tilde{c}_T (u_0, k)) = Tb_{2,T} \mathbb{E} \left[ \text{vec} (\tilde{c}_T (u_0, k) - c(u_0, k))' W \text{vec} (\tilde{c}_T (u_0, k) - c(u_0, k)) \right],
\]

where $W$ is some $p^2 \times p^2$ weight matrix.

**Proposition 1.3.1.** Suppose $V_{t,T}$ is Locally Stationary with zero mean and Assumption 1.1 holds with $\vartheta = 1$. Suppose $b_{2,T} \to 0$ as $T \to \infty$. Then, for all $u_0 \in (0, 1),

\[
\mathbb{E} [\tilde{c}_T (u_0, k)] = c(u_0, k) + \frac{1}{2} b_{2,T}^2 \int_0^1 x^2 K_2 (x) \, dx \left[ \frac{\partial^2}{\partial u^2} c(u_0, k) \right] + o \left( b_{2,T}^2 \right) + O \left( 1 / (b_{2,T} T) \right) \quad \text{and}
\]
Var [vec (\( \tilde{c}_T (u_0, k) \))] 

\[
= \frac{1}{Tb_{2,T}} \int_0^1 K^2_2 (x) \, dx \sum_{l=-\infty}^{\infty} \text{vec} (c (u_0, l)) \\
\times \left[ \text{vec} (c (u_0, l))' + \text{vec} (c (u_0, l + 2k))' \right] \\
+ o(1/ (b_{2,T} T)).
\]

If \( Tb_{2,T}^5 \to \eta \in (0, \infty) \), then

\[
\lim_{T \to \infty} \text{MSE} (\tilde{c}_T (u_0, k)) \\
= \frac{\eta}{4} \left( \int_0^1 x^2 K_2 (x) \, dx \right)^2 \left[ \frac{\partial^2}{\partial u^2} \text{vec} (c (u_0, k)) \right]' W \left[ \frac{\partial^2}{\partial u^2} \text{vec} (c (u_0, k)) \right] \\
+ \int_0^1 K^2_2 (x) \, dx \text{tr} W \sum_{l=-\infty}^{\infty} \text{vec} (c (u_0, l)) \\
\times \left[ \text{vec} (c (u_0, l))' + \text{vec} (c (u_0, l + 2k))' \right],
\]

and \( \tilde{c}_T (u_0, k) - c (u_0, k) = O_P \left( \sqrt{Tb_{2,T}} \right) \) for all \( u_0 \in (0, 1) \).

The bias and variance expressions in Proposition 1.3.1 extend the results for the univariate case from Dahlhaus (2012). The bias of order \( b_{2,T}^2 \) is due to nonstationarity which is measured by \( \partial^2 c (u_0, k) / \partial u^2 \). The latter derivative is null for a stationary process. The bias of order \( (Tb_{2,T})^{-1} \) is due to the smoothing procedure as it involves an effective number of observations equals to \( Tb_{2,T} \). In Section 1.7, we show that the same results hold when \( m > 0 \) as long as \( u_0 \) is away from the change-points. The rate of convergence of \( \tilde{c}_T (u_0, k) \) is given by \( O_P \left( \sqrt{Tb_{2,T}} \right) \).

1.3.2 Results on HAC Estimation with Predetermined Bandwidths

Since the estimator \( \hat{J}_T \) involves two kernels, it is important to control their relative degree of smoothing. In particular, the bandwidth sequence \( b_{2,T} \) restricts the rate
at which $b_{1,T}$ goes to zero. Under certain conditions, the asymptotic bias of $\hat{J}_T$ depends on the smoothness of the kernel $K_1 (\cdot)$ at zero and on the smoothness of the (integrated) spectral density matrix $f_1^1 f (u, \omega) du$ at $\omega = 0$. Following Parzen (1957), we define $K_{1,q} \triangleq \lim_{x \to 0} (1 - K_1 (x)) / |x|^q$ for $q \in [0, \infty)$. The value of $q$ increases with the smoothness of $K_1 (\cdot)$ with the largest value being such that $K_{1,q}$ is finite. Further, it is well-known that when $q$ is an even integer, then $K_{1,q} = - (d^n K_1 (x) / dx^n)_{|x=0}/q!$ and $K_{1,q} < \infty$ if and only if $K_1 (x)$ is $q$ times differentiable at zero. For the QS kernel, $K_{1,q} = 0$ for $q < 2$, $K_{1,2} = 1.421223$ and $K_{1,q} = \infty$ for $q > 2$.

We define the index of smoothness of $f (u, \omega)$ at $\omega = 0$ by

$$f^{(q)} (u, 0) \triangleq (2\pi)^{-1} \sum_{k=\infty}^{\infty} |k|^q c (u, k),$$

for $q \in [0, \infty)$. If $q$ is even, then $f^{(q)} (u, 0) = (-1)^{q/2} (d^n f (u, \omega) / d\omega^n)_{|\omega=0}$. Further, $\|f^{(q)} (u, 0)\| < \infty$ if and only if $f (u, \omega)$ is $q$ times differentiable at $\omega = 0$. We define

$$\text{MSE} (Tb_{1,T} b_{2,T}, \tilde{J}_T, W) = Tb_{1,T} b_{2,T} \mathbb{E} \left[ \text{vec} (\tilde{J}_T - J_T)' W \text{vec} (\tilde{J}_T - J_T) \right]. \tag{1.3.2}$$

We also need to impose conditions on the temporal dependence of $\{V_t\}$ (we omit the second subscript $T$ when it is clear from the context). Let

$$\kappa_{V,t}^{(a,b,c,d)} (u, v, w)$$

$$\triangleq \kappa_{(a,b,c,d)}^{(a,b,c,d)} (t, t + u, t + v, t + w) - \kappa_{(a,b,c,d)}^{(a,b,c,d)} (t, t + u, t + v, t + w)$$

$$\triangleq \mathbb{E} \left( V_{t}^{(a)} - \mathbb{E} V_{t}^{(a)} \right) \left( V_{t+u}^{(b)} - \mathbb{E} V_{t+u}^{(b)} \right) \left( V_{t+v}^{(c)} - \mathbb{E} V_{t+v}^{(c)} \right) \left( V_{t+w}^{(d)} - \mathbb{E} V_{t+w}^{(d)} \right)$$

$$- \mathbb{E} \left( V_{\mathcal{N},t}^{(a)} - \mathbb{E} V_{\mathcal{N},t}^{(a)} \right) \left( V_{\mathcal{N},t+u}^{(b)} - \mathbb{E} V_{\mathcal{N},t+u}^{(b)} \right)$$

$$\times \left( V_{\mathcal{N},t+v}^{(c)} - \mathbb{E} V_{\mathcal{N},t+v}^{(c)} \right) \left( V_{\mathcal{N},t+w}^{(d)} - \mathbb{E} V_{\mathcal{N},t+w}^{(d)} \right),$$

where $\{V_{\mathcal{N},t}\}$ is a Gaussian sequence with the same mean and covariance structure as
\{V_t\}. \kappa^{(a,b,c,d)}_{V_t}(u, v, w) is the time-\(t\) fourth-order cumulant of \((V^{(a)}_t, V^{(b)}_{t+u}, V^{(c)}_{t+v}, V^{(d)}_{t+w})\) while \(\kappa^{(a,b,c,d)}_0(t, t+u, t+v, t+w)\) is the time-\(t\) centered fourth moment of \(V_t\) if \(V_t\) were Gaussian.

**Assumption 1.2.** (i) \(\{V_{t,T}\}\) is a mean zero Locally Stationary process, \(A(u, \omega)\) is twice continuously differentiable in \(u\) for all \(u \in (0, 1)\) and uniformly Lipschitz continuous with index \(\vartheta = 1\) in \(\omega\), the first and second derivatives of \(A(u, \omega)\) with respect to \(u\) are uniformly bounded, \(\sum_k^\infty \sup_{u \in [0,1]} \|c(u, k)\| < \infty\) and

\[
\sum_k^\infty \sum_j^\infty \sum_l^\infty \sup_{u \in [0,1]} \kappa^{(a,b,c,d)}_{V_t[Tu]}(k, j, l) < \infty,
\]

for all \(a, b, c, d \leq p\). (ii) For all \(a, b, c, d \leq p\) there exists a constant \(K\) and a function \(\tilde{\kappa}_{a,b,c,d} : [0, 1] \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}\) such that

\[
\sup_{u \in [0,1]} \left| \kappa^{(a,b,c,d)}_{V_t[Tu]}(k, s, l) - \tilde{\kappa}_{a,b,c,d}(u, k, s, l) \right| \leq KT^{-1};
\]

the function \(\tilde{\kappa}_{a,b,c,d}(u, k, s, l)\) is uniformly Lipschitz continuous in \(u\) for all \(a, b, c, d \leq p\).

If \(\{V_{t,T}\}\) is stationary then the cumulant condition of Assumption 1.2-(i) reduces to the standard one used in the time series literature [see also Assumption A in Andrews (1991)]. We do not require fourth-order stationarity but only that the time-\(t = Tu\) fourth order cumulant is locally constant in a neighborhood of \(u\). As in Andrews (1991) we can show that \(\alpha\)-mixing and moment conditions imply that the cumulant condition of Assumption 1.2 holds [see Lemma 1 Andrews (1991)].

**Theorem 1.3.1.** Suppose \(K_1(\cdot) \in K_1\), Assumption 1.2 holds with \(\vartheta = 1, b_{1,T}, b_{2,T} \rightarrow \)


0, \ n_T \to \infty, \ n_T/T \to 0 \ and \ 1/Tb_{1,T}b_{2,T} \to 0. \ We \ have: \ (i) \\

\lim_{T \to \infty} Tb_{1,T}b_{2,T} \Var [\text{vec}(\tilde{J}_T)] \\
= 4\pi^2 \left[ \gamma K^2_{1,q} \text{vec} \left( \int_0^1 f^{(q)}(u, 0) \, du \right) \right] \text{vec} \left( \int_0^1 f^{(q)}(u, 0) \, du \right) \\
+ \int K^2_1(y) \, dy \int_0^1 K^2_2(x) \, dx \, (I - C_{pp}) \\
\times \left( \int_0^1 f (u, 0) \, du \right) \otimes \left( \int_0^1 f (v, 0) \, dv \right). \\

(ii) \ If 

1/Tb^q_{1,T}b_{2,T} \to 0, \ n_T/Tb^q_{1,T} \to 0 \ and \ b^q_{2,T}/b^q_{1,T} \to 0 \ for \ some \ q \in [0, \infty) \ for \ which \ K_{1,q}, \ \left\| \int_0^1 f^{(q)}(u, 0) \, du \right\| \in [0, \infty), \ then \\

\lim_{T \to \infty} b^{-q}_{1,T} \text{E} (\tilde{J}_T - J_T) = -2\pi K_{1,q} \int_0^1 f^{(q)}(u, 0) \, du. \\

(iii) \ If \ n_T/Tb^q_{1,T} \to 0, \ b^q_{2,T}/b^q_{1,T} \to 0 \ and \ T \gamma_{1,T}b^q_{2,T} \to \gamma \in (0, \infty) \ for \ some \ q \in [0, \infty) \ for \ which \ K_{1,q}, \ \left\| \int_0^1 f^{(q)}(u, 0) \, du \right\| \in [0, \infty), \ then \\

\lim_{T \to \infty} \text{MSE} (Tb_{1,T}b_{2,T}, \tilde{J}_T, W) \\
= 4\pi^2 \left[ \gamma K^2_{1,q} \text{vec} \left( \int_0^1 f^{(q)}(u, 0) \, du \right) \right] \text{vec} \left( \int_0^1 f^{(q)}(u, 0) \, du \right) \\
+ \int K^2_1(y) \, dy \int_0^1 K^2_2(x) \, dx \, \text{tr}W (I - C_{pp}) \\
\times \left( \int_0^1 f (u, 0) \, du \right) \otimes \left( \int_0^1 f (v, 0) \, dv \right). \\

If \ b^q_{2,T}/b^q_{1,T} \to \nu < \infty \ replaces \ b^q_{2,T}/b^q_{1,T} \to 0 \ in \ part \ (ii), \ then \ the \ asymptotic \ bias \ becomes \\

\lim_{T \to \infty} b^{-q}_{1,T} \text{E} (\tilde{J}_T - J_T) = -2\pi K_{1,q} \int_0^1 f^{(q)}(u, 0) \, du \\
+ \frac{\nu}{2} \int_0^1 x^2 K_2(x) \sum_{k=-\infty}^{\infty} \int_0^1 \frac{\partial^2}{\partial u^2} c(u, k) \, du. \quad (1.3.3)
The second summand on the right-hand side of (1.3.3) cancels when

$$\int_0^1 \int_0^1 \left( \frac{\partial^2}{\partial^2 u} c (u, k) \right) du = 0.$$ 

The latter occurs when the process is stationary. Dahlhaus (2012) derived MSE results for a pointwise estimate of $f(u, \omega)$ under continuity in both components by applying smoothing over $u$ and $\omega$. Its results depends on the local behavior of $f(u, \omega)$ at time $u$ and frequency $\omega$ whereas in our problem the MSE results depend on properties of all time path of $f(u, 0)$. In order to obtain asymptotic results for the estimator $\hat{J}_T$ we state the following assumption.

**Assumption 1.3.** (i) $\sqrt{T} (\beta - \beta_0) = O_P(1)$; (ii) $\sup_{u \in [0, 1]} \mathbb{E} \left\| V_{[T u]} \right\|^2 < \infty$; (iii) $\sup_{u \in [0, 1]} \mathbb{E} \sup_{\beta \in \Theta} \left\| (\partial/\partial \beta') V_{[T u]} (\beta) \right\|^2 < \infty$; (iv) $\int_{-\infty}^{\infty} |K_1 (x)| dx < \infty$; (v) $\int_0^1 |K_2 (x)| dx < \infty$; (vi) $\int_0^1 K_2^2 (x) dx < \infty$.

Assumption 1.3 is easy to verify and is the same as Assumption B in Andrews (1991). Part (i)-(iii) were also used by Newey and West (1987). As remarked above, we interpret $\beta_0$ as the pseudo-true parameter $\beta^*$ when the model is misspecified. Part (iv)-(v) of the assumption are satisfied by most commonly used kernels. Theorem 1.3.2 below shows that the effect of using $\hat{\beta}$ rather than $\beta_0$ when constructing $\hat{J}_T$ is at most $o_P(1)$.

In order to obtain rate of convergence results we replace Assumption 1.2 with the following assumptions. Let $\kappa_{r, t}^{(a_1, \ldots, a_8)} (j_1, \ldots, j_7)$ denote the cumulant of $(V_{t}^{(a_1)}, V_{t+j_1}^{(a_2)}, \ldots, V_{t+j_7}^{(a_8)})$ where $a_1, \ldots, a_8$ range are positive integers less than $p+1$ and

---

8There are a few situations where $\hat{\beta}$ has infinite second moments (e.g., the two-stage LS estimator in some cases). This would imply that $\hat{\beta}$ may dominate the MSE criterion (1.3.2). Andrews (1991) suggested to use a truncated MSE criterion as opposed to the standard MSE criterion (1.3.2). This would avoid that undue influence of $\hat{\beta}$ can affect the criterion of performance. This approach can, of course, be extended to our context as well. However, to avoid further notational burden we present results assuming that $\beta$ is well-behaved and thus rule out the aforementioned pathological cases.
Assumption 1.4. (i) Assumption 1.2 holds with $V_{i,T}$ replaced by

$$
\left( V'_{Tu}, \text{vec} \left( \left( \frac{\partial}{\partial \beta'} V_{Tu} (\beta_0) \right) - \mathbb{E} \left( \frac{\partial}{\partial \beta'} V_{Tu} (\beta_0) \right) \right) \right)'.
$$

(ii) $\sup_{u \in [0,1]} \mathbb{E} \left( \sup_{\beta \in \Theta} \left\| \frac{\partial^2}{\partial \beta \partial \beta'} V_{Tu} (\beta_0) \right\|^2 \right) < \infty$ for all $a = 1, \ldots, p$.

Assumption 1.5. (i) Let $W_T$ denote a $p^2 \times p^2$ weight matrix such that $W_T \overset{p}{\to} W$.

Theorem 1.3.2. Suppose $K_1(\cdot) \in K_1$, $b_{1,T}, b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$ and $1/Tb_{1,T}b_{2,T} \to 0$. We have

(i) If Assumption 1.2-1.3 hold, $\sqrt{Tb_{1,T}b_{2,T}} \to \infty$, then $\hat{J}_T - J_T \overset{p}{\to} 0$ and $\tilde{J}_T - \tilde{J}_T \overset{p}{\to} 0$.

(ii) If Assumption 1.3-1.4 hold, $n_T/Tb_{1,T} \to 0$, $n_T/Tb_{1,T}^q \to 0$ and $Tb_{1,T}^{q+1}b_{2,T} \to \gamma \in (0, \infty)$ for some $q \in [0, \infty)$ for which $K_{1,q} \left\| \int_0^1 f(q)(u,0) du \right\| \in [0, \infty)$, then $\sqrt{Tb_{1,T}b_{2,T}} (\hat{J}_T - J_T) = O_p(1)$ and $\sqrt{Tb_{1,T}} (\tilde{J}_T - \tilde{J}_T) = o_p(1)$.

(iii) Under the conditions of part (ii) and Assumption 1.5,

$$
\lim_{T \to \infty} \text{MSE} \left( Tb_{1,T}b_{2,T}, \hat{J}_T, W_T \right) = \lim_{T \to \infty} \text{MSE} \left( Tb_{1,T}b_{2,T}, \tilde{J}_T, W \right).
$$

The consistency result of $\hat{J}_T$ in part (i) applies to kernels $K_1(\cdot)$ with unbounded support and to bandwidths $b_{1,T}$ and $b_{2,T}$ such that $1/b_{1,T}b_{2,T}$ grows at rate $o \left( \sqrt{T/b_{2,T}} \right)$. Part (ii) yields the consistency of $\tilde{J}_T$ with $b_{1,T}$ only required to be $o(Tb_{2,T})$. This rate is slower than the corresponding rate $o(T)$ of the classical kernel HAC estimators as shown by Andrews (1991) in his Theorem 1-(b). However, this property is of little practical import because optimal growth rates typically are less than $T^{1/2}$—for the QS kernel HAC estimator the optimal growth rate is $T^{1/5}$ while
it is $T^{1/3}$ for the Newey-West HAC estimator. Part (ii) of the theorem presents the rate of convergence of $\hat{J}_T$ which is $\sqrt{Tb_{2,T}b_{1,T}}$. In Section 1.4, we compare the rate of convergence of $\hat{J}_T$ with that of the classical HAC estimators when the respective optimal bandwidths are used.

### 1.4 Optimal Kernels, Bandwidths and Choice of $n_T$

In this section, we show the optimality of quadratic-type kernels under a mean-squared criterion. For the kernel $K_1$, the result states that the QS kernel minimizes the asymptotic MSE for any $K_2(\cdot)$ and any given positive semidefinite weight matrix $W$. Let

$$\text{MSE} \left( b_{2,T}^{-4}, \tilde{c}_T (u_0, k), \tilde{W}_T \right) = b_{2,T}^{-4} \mathbb{E} \left[ \text{vec} \left( \tilde{c}_T (u_0, k) - c(u_0, k) \right) \right] \tilde{W}_T \left[ \text{vec} \left( \tilde{c}_T (u_0, k) - c(u_0, k) \right) \right],$$

where $\tilde{W}_T$ is some $p \times p$ positive semidefinite matrix. The optimal bandwidths $b_{1,T}^{\text{opt}}$ and $b_{2,T}^{\text{opt}}$ satisfy the following MSE criterion:

$$\text{MSE} \left( Tb_{1,T}^{\text{opt}}b_{2,T}^{\text{opt}}, \hat{J}_T \left( b_{1,T}^{\text{opt}}, b_{2,T}^{\text{opt}} \right), W_T \right) \leq \text{MSE} \left( Tb_{1,T}^{\text{opt}} \bar{b}_{2,T}^{\text{opt}}, \hat{J}_T \left( b_{1,T}, \bar{b}_{2,T}^{\text{opt}} \right), W_T \right)$$

where $\bar{b}_{2,T}^{\text{opt}} = \int_0^1 b_{2,T}^{\text{opt}}(u) \, du$ and

$$b_{2,T}^{\text{opt}}(u) = \arg\min_{b_{2,T}} b_{2,T}^{-4} \mathbb{E} \left( \tilde{c}_T (u_0, k) - c(u_0, k), \tilde{W}_T \right).$$

The notation $\hat{J}_T \left( b_{1,T}, b_{2,T}^{\text{opt}} \right)$ indicates the estimator $\hat{J}_T$ that uses $b_{1,T}$ and $b_{2,T}^{\text{opt}}$. The first inequality above has to hold as $T \rightarrow \infty$. The above criterion determines the globally optimal $b_{1,T}^{\text{opt}}$ given the integrated locally optimal $b_{2,T}^{\text{opt}}(u)$. Thus, $b_{1,T}^{\text{opt}}$ and $b_{2,T}^{\text{opt}}$ need not be the same to the bandwidths $(\tilde{b}_{1,T}^{\text{opt}}, \tilde{b}_{2,T}^{\text{opt}})$ that jointly minimize the global
asymptotic MSE,
\[
\lim_{T \to \infty} \text{MSE} \left( Tb_{1,T}b_{2,T}, \hat{J}_T (b_{1,T}, b_{2,T}), W_T \right).
\]

Unfortunately, the solution of the latter problem is a high-degree polynomial in \( b_1 \) and \( b_2 \) which is challenging to solve analytically. The difficulty arises from the bias expression (1.3.3) under \( b_2^2 / b_1^2 \to \nu < \infty \). With this bias expression, the form of the asymptotic MSE is complex and its minimization with respect to \( b_1 \) and \( b_2 \) leads to multiple solutions. Also, Theorem 1.3.1-(ii) states that, under the condition \( b_2^2 / b_1^2 \to 0 \), the bias only depends on the smoothing over autocovariance lag orders but not on \( b_{2,T} \). Then, the solution \( \tilde{b}_{2,T}^{\text{opt}} \) becomes trivial: \( b_{2,T} \) affects the MSE only through the variance term and optimality requires to set the bandwidth as large as possible. In contrast, the MSE criterion (1.4.1) where the MSE is given by Theorem 1.3.1-(iii) leads to a unique solution which can be obtained analytically. It gives the locally optimal \( b_{2,T} \) and determine the globally optimal \( b_{1,T} \) given the integrated optimal \( b_{2,T} \).

We begin with deriving the optimal bandwidth \( b_{2,T} \) and the optimal kernel \( K_2 (\cdot) \) that minimize the asymptotic MSE of \( \hat{c}_T (u, k) \). In particular the optimal bandwidth is such that \( 1 / b_{2,T} \) grows at rate \( T^{1/5} \) and \( K_2 (\cdot) \) has a quadratic form.

### 1.4.1 Optimal \( K_2 (\cdot) \) and \( b_{2,T} \)

Let \( D_1 (u_0) \triangleq \text{vec} (\partial^2 c (u_0, k) / \partial u^2)' \bar{W} \text{vec} (\partial^2 c (u_0, k) / \partial u^2) \),

\[
D_2 (u_0) \triangleq \text{tr} \bar{W} (I + C_{pp}) \sum_{l=-\infty}^{\infty} c(u_0, l) \otimes [c(u_0, l) + c(u_0, l + 2k)],
\]

\(^{9}\)An interesting extension would be to let also \( b_{1,T} \) depend on \( u \). This may even make the determination of the optimal \( b_{1,T} \) and \( b_{2,T} \) easier under an appropriate MSE criterion. However, in this chapter we consider \( b_{1,T} \) to be independent of \( u \)—leading to a more direct comparison to Newey and West (1987) and Andrews (1991).
\[ F (K_2) \triangleq \int_0^1 K_2^2 (x) \, dx, \text{ and } H (K_2) = \left( \int_0^1 x^2 K_2 (x) \, dx \right)^2. \]

**Proposition 1.4.1.** Suppose Assumption 1.3-1.5 hold and \( \tilde{W}_T \to \tilde{W} \). We have,

\[
\text{MSE} \left( 1, \hat{c}_T (u_0, k), \tilde{W}_T \right) = \frac{1}{4} b_{2,T}^4 \left( \int_0^1 x K_2 (x) \, dx \right)^2 \text{vec} \left( \frac{\partial^2}{\partial^2 u} c (u_0, k) \right) \tilde{W}_T \text{vec} \left( \frac{\partial^2}{\partial^2 u} c (u_0, k) \right) + \frac{1}{T b_{2,T}} \int_{-1}^0 K_2^2 (x) \, dx \text{tr} \tilde{W}_T (I + C) \\
\sum_{l=-\infty}^{\infty} c (u_0, l) \otimes [c (u_0, l) + c (u_0, l + 2k)] + o \left( b_{2,T}^4 \right) + O \left( 1/(b_{2,T}T)^2 \right).
\]

MSE \( \left( b_{2,T}^{-1}, \hat{c}_T (u_0, k) - c (u_0, k), \tilde{W}_T \right) \) is minimized for

\[ b_{2,T}^{\text{opt}} (u_0) = [(3/4) H \left( K_2^{\text{opt}} \right) D_1 (u_0)]^{-1/5} \left( F \left( K_2^{\text{opt}} \right) D_2 (u) \right)^{1/5} T^{-1/5}, \]

and \( K_2^{\text{opt}} (x) = 6x (1 - x) \), \( 0 \leq x \leq 1 \).

The optimal kernel \( K_2^{\text{opt}} (x) \) is a transformation of the Epanechnikov kernel. Optimality of quadratic kernels under a MSE criterion has been shown in many contexts, including estimation of the spectral density function [cf. Priestley (1962; 1981)] and of probability densities [cf. Epanechnikov (1969)]. The optimal bandwidth sequence decreases at rate \( T^{-1/5} \) which is the same optimal rate derived in the context of Yule-Walker estimates of parameters of locally stationary processes by Dahlhaus and Giraitis (1998). Overall, our results are similar to kernel estimation in nonparametric regression. The term \( D_1 (u_0) \) is due to nonstationary, while the term \( D_2 (u_0) \) measures the variability of \( \hat{c}_T (u_0, k) \) at time \( u_0 \). The bandwidth \( b_{2,T}^{\text{opt}} \) converges to zero at a slower rate as the process becomes closer to stationary (i.e., as the square
root of $D_1(u_0)$ decreases).

### 1.4.2 Optimal $K_1(\cdot)$

We next determine the optimal kernel $K_1$ and optimal bandwidth sequence $b_{1,T}$ given any $K_2$ and any $b_{2,T}$ of order $O\left(T^{-1/5}\right)$, i.e., the same order of $b_{2,T}^{\text{opt}}(u)$ for any $u \in [0, 1]$. Let $\tilde{J}_T^{\text{QS}}$ denote $\tilde{J}_T$ when the latter is based on the QS kernel. For some results below, we consider a subset of $K_1$. Let $\tilde{K}_1 = \{K_1(\cdot) \in K_1 \mid \tilde{K}(\omega) \geq 0 \forall \omega \in \mathbb{R}\}$,

where $\tilde{K}(\omega) = (2\pi)^{-1} \int_{-\infty}^{\infty} K_1(x) e^{-ix\omega} dx$. The function $\tilde{K}(\omega)$ is referred to as the spectral window generator corresponding to the kernel $K_1(\cdot)$. The set $\tilde{K}_1$ contains all kernels $K_1$ that necessarily generate positive semidefinite estimators in finite samples. $K_1$ contains the Bartlett, Parzen, and QS kernels, but not the truncated or Tukey-Hanning kernels.

We adopt the notation $\tilde{J}_T(b_{1,T}) = \tilde{J}_T(b_{1,T}, b_{2,T}, K_2)$ to denote the estimator $\tilde{J}_T$ that uses $b_{1,T}$, $b_{2,T} = b_{2,T}^{\text{opt}} + o\left(T^{-1/5}\right)$, and $K_2(\cdot)$. We then compare two kernels $K_1$ using comparable bandwidths $b_{1,T}$ which are defined as follows. Given $K_1(\cdot) \in \tilde{K}_1$, the QS kernel $K_1^{\text{QS}}(\cdot)$, and a bandwidth sequence $\{b_{1,T}\}$ to be used with the QS kernel, define a comparable bandwidth sequence $\{b_{1,T,K_1}\}$ for use with $K_1(\cdot)$ such that both kernel/bandwidth combinations have the same asymptotic variance when scaled by the same factor $Tb_{1,T}b_{2,T}$. This means that $b_{1,T,K_1}$ is such that

$$\lim_{T \to \infty} \text{MSE} \left( T b_{1,T} b_{2,T}, \tilde{J}_T^{\text{QS}}(b_{1,T}) - \mathbb{E} \left( \tilde{J}_T^{\text{QS}}(b_{1,T}) \right) + J_T, W_T \right)$$

$$= \lim_{T \to \infty} \text{MSE} \left( T b_{1,T} b_{2,T}, \tilde{J}_T(b_{1,T,K_1}) - \mathbb{E} \left( \tilde{J}_T(b_{1,T,K_1}) \right) + J_T, W_T \right).$$

This definition yields $b_{1,T,K_1} = b_{1,T}/\left(\int K_1^2(x) dx\right)$. Note that for the QS kernel
Theorem 1.4.1. Suppose Assumption 1.3-1.5 hold, \( \int_{0}^{1} \left\| f^{(2)}(u, 0) \right\| du < \infty, b_{2,T} \to 0, b_{2,T}^5 T \to \eta \in (0, \infty) \), \( \left( \text{vec} \left( \int_{0}^{1} f^{(q)}(u, 0) du \right) \right)^{\prime} W \text{vec} \left( \int_{0}^{1} f^{(q)}(u, 0) du \right) > 0 \) and \( W \) is positive semidefinite. For any bandwidth sequence \( \{b_{1,T}\} \) such that \( b_{2,T}/b_{1,T} \to 0 \), \( T b_{1,T}^5 b_{2,T} \to \gamma \in (0, \infty) \) and for any kernel \( K_{1}(\cdot) \in \tilde{K}_1 \) used to construct \( \hat{J}_{T} \), the QS kernel is preferred to \( K_{1}(\cdot) \) in the sense that

\[
\lim_{T \to \infty} \left( \text{MSE} \left( Tb_{1,T}b_{2,T}, \hat{J}_{T}(b_{1,T},K_{1}) \right) - \text{MSE} \left( Tb_{1,T}b_{2,T}, \hat{J}_{T}^{QS}(b_{1,T}) \right) \right)
= 4\gamma \pi^2 \left( \text{vec} \left( \int_{0}^{1} f^{(q)}(u, 0) du \right) \right)^{\prime} W \text{vec} \left( \int_{0}^{1} f^{(q)}(u, 0) du \right)
\times \int_{0}^{1} \left( K_{2,\text{opt}}(x) \right)^2 dx \left[ K_{1,2}^2 \left( \int K_{1}^2(y) dy \right)^4 - \left( K_{1,2}^{QS} \right)^2 \right] \geq 0.
\]

The inequality is strict if \( K_{1}(x) \neq K_{1,\text{QS}}(x) \) with positive Lebesgue measure.

The requirement \( \int_{0}^{1} \left\| f^{(2)}(u, 0) \right\| du < \infty \) is not stringent and it reduces to the one used by Andrews (1991) when \( \{V_{t,T}\} \) is second-order stationary. As in Andrews (1991), if \( \int_{0}^{1} \left\| f^{(q)}(u, 0) \right\| du < \infty \) only for some \( 1 \leq q < 2 \), one can show that any kernel with \( K_{1,q} = 0 \) has smaller asymptotic MSE than a kernel with \( K_{1,q} > 0 \). In particular, the QS, Parzen, and Tukey-Hanning kernels have \( K_{1,q} = 0 \) for \( 1 \leq q < 2 \), whereas the Bartlett kernel has \( K_{1,q} > 0 \) for \( 1 \leq q < 2 \). Thus, the asymptotic superiority of the former kernels over the Bartlett kernel holds even if \( \int_{0}^{1} \left\| f^{(q)}(u, 0) \right\| du < \infty \) only for \( 1 \leq q < 2 \).

1.4.3 Optimal Predetermined Bandwidth Sequence \( b_{1,T} \)

We now present the predetermined bandwidth sequence that minimizes the asymptotic MSE. This optimality result applies to each kernel \( K_{1}(\cdot) \in K_1 \) for which
$K_{1,q} \in (0, \infty)$ for some $q \in (0, \infty)$. Thus, most commonly used kernels are allowed with the exception of the truncated kernel. Let

$$\phi(q) = \frac{\text{vec} \left( \int_0^1 f^{(q)}(u, 0) \, du \right)'}{\text{tr} \left( W (I + C_{pp}) \left( \int_0^1 f(u, 0) \, du \right) \otimes \left( \int_0^1 f(v, 0) \, dv \right) \right)}.$$

The optimal bandwidth is

$$b_{1,T}^{\text{opt}} = \left( 2q K_{1,q}^2 \phi(q) T b_{2,T}^{\text{opt}} / \left( \int K_1^2(y) \, dy \int_0^1 K_2^2(x) \, dx \right) \right)^{-1/(2q+1)},$$

where $\phi(q)$ is a function of the unknown integrated spectral density matrix $f(\cdot, \cdot)$. Hence, the optimal bandwidth $b_{1,T}^{\text{opt}}$ is unknown in practice, and we consider data-dependent estimates of $\phi(q)$ in Section 1.5. In the statement of the following corollary, $\hat{J}_T$ denotes the estimator that uses any $b_{1,T}$ while the notation $\hat{J}_T(b_{1,T}^{\text{opt}})$ is reserved for the estimator that uses $b_{1,T}^{\text{opt}}$.

**Condition 1.** $b_{1,T}, b_{2,T} \to 0$ with $b_{2,T}/b_{1,T} \to 0$, and $T b_{1,T}^{2q+1} b_{2,T} \to \gamma \in (0, \infty)$ for some $q \in [0, \infty)$ for which $K_{1,q}, \left\| \int_0^1 f^{(q)}(u, 0) \, du \right\| \in [0, \infty)$.

**Corollary 1.4.1.** Suppose Assumption 1.3-1.5 hold, $\left\| \int_0^1 f^{(q)}(u, \omega) \, du \right\| < \infty$, $\phi(q) \in (0, \infty)$, and $W$ is positive semidefinite. Consider a kernel $K_1(\cdot) \in K_1$ for which $K_{1,q} \in (0, \infty)$ for some $q \in (0, \infty)$. The bandwidth sequence $\{b_{1,T}^{\text{opt}}\}$ is optimal among the sequences $\{b_{1,T}\}$ that satisfy Condition 1 in the sense that

$$\lim_{T \to \infty} \left( \text{MSE} \left( \left( T b_{2,T} \right)^{2q/(2q+1)} , \hat{J}_T(b_{1,T}) , W_T \right) - \text{MSE} \left( \left( T b_{2,T} \right)^{2q/(2q+1)} , \hat{J}_T(b_{1,T}^{\text{opt}}) , W_T \right) \right) \geq 0,$$

where $b_{2,T} = O \left( T^{-1/5} \right)$. The inequality is strict unless

$$b_{1,T} = b_{1,T}^{\text{opt}} + o \left( (T b_{2,T})^{-1/(2q+1)} \right).$$
In Corollary 1.4.1, \( q = 2 \) for the QS kernel and so

\[
b_{1,T}^{\text{opt}} = 0.6584 \left( \phi (2) T b_{2,T}^{\text{opt}} \right)^{-1/5} \left( \int_0^1 K_2^2 (y) \, dy \right)^{1/5}.
\]

For \( K_2 (y) = K_2^{\text{opt}} (y) \), the latter reduces to,

\[
b_{1,T}^{\text{opt}} = 0.6828 \left( \phi (2) T b_{2,T}^{\text{opt}} \right)^{-1/5}.
\] (1.4.2)

The optimal bandwidth is of order \( T^{-4/25} \). Thus, the optimal bandwidth sequence decreases to zero at a slower rate than the optimal bandwidth sequence for the QS kernel-based HAC estimator of Andrews (1991), for which the rate is of order \( T^{-1/5} \). The slower rate is due to the fact that our estimator smooths the spectrum over time—through \( K_2 (\cdot) \)—and this restricts the smoothing of \( K_1 (\cdot) \) over autocovariance lag orders. When \( b_{1,T} \) and \( b_{2,T} \) are chosen optimally, the convergence rate from Theorem 1.3.2 reduces to \( T^{8/25} \). Thus, the rate is slower than the corresponding one for the kernel-based HAC estimators considered in Andrews (1991). However, it is misleading to compare our HAC estimator with the classical HAC estimators only on the basis of the rate of convergence. In fact, the limit of our HAC estimator is different. The limit is the same only when the \( \{X_t e_t\} \) is second-order stationary (i.e., it has constant spectrum). When the spectrum is time-varying—a relevant case in practice—it is hard to make such comparison as the the main theory behind the kernel-based HAC estimators is not applicable, as our procedure is robust to model misspecification and general nonstationarity, a beneficial feature.

### 1.4.4 Choice of \( n_T \)

Our MSE analysis does not indicate an optimal value for \( n_T \). It only suggests growth rate bounds. They are \( n_T \to \infty, n_T / T \to 0, n_T / T h_{1,T} \to 0 \) and \( n_T / T b_{1,T}^q \to 0 \); when
$K_1^{QS}$ is used the latter restriction reduces to $n_T \approx T^{2/3}$. This turns out to be very similar to the condition on $n_T$ required for the testing procedure for detection of the break points in the spectrum (see Section 1.6). Throughout, in the simulations and applications, we set $n_T = T^{0.6} \approx T^{2/3-\epsilon}$ for a small $\epsilon > 0$. That is, we choose the number of observations $n_T$ to be the largest possible as allowed by the condition. Choosing a smaller $n_T$ as the condition potentially allows would lead to blocks of observations that are too small which in turn may result into size distortions for the HAR inference tests. This was investigated in our sensitivity analyses (not reported).

Finally, in this chapter we have considered the case $n_T \to \infty$. It is possible, and indeed the derivations are even easier, to let $n_T$ be fixed. However, this approach can be problematic when $m > 0$ (see Section 1.7).

1.5 Data-Dependent Bandwidths

In this section we consider estimators $\hat{J}_T$ that use bandwidths $b_{1,T}$ and $b_{2,T}$ whose values are determined via data-dependent (automatic) methods. We interpret $b_{1,T}$ and $b_{2,T}$ as bandwidth parameters—rather than sequences—and estimate them from the data. Following Andrews (1991) we use the so-called “plug-in” method which is characterized by plugging-in estimates of unknown quantities into an asymptotic formula for an optimal bandwidth parameter (i.e., the expressions for $b_{1,T}^{\text{opt}}$ and $b_{2,T}^{\text{opt}}$ from Section 1.4). In principle, one can use either parametric or nonparametric methods to obtain such estimates. Here, as in Andrews (1991), we consider parametric methods for pragmatic reasons. Section 1.5.1 explains how to construct the automatic bandwidths while Section 1.5.2 presents the corresponding theoretical results.
1.5.1 Implementation

The optimal bandwidths account for nonstationarity and misspecification as can be easily seen from their definitions. This implies that our estimates of \( b_{1,T}^{\text{opt}} \) and \( b_{2,T}^{\text{opt}} \) are actually functions of local data rather than being functions of the whole sample data. Let us begin with \( b_{1,T}^{\text{opt}} \) and then move to \( b_{2,T}^{\text{opt}} \), though the procedures are similar. The first step for the construction of data-dependent bandwidth parameters is to specify \( p \) univariate parametric models for the elements of \( V_t = (V_t^{(1)}, \ldots, V_t^{(p)}) \). An alternative to this first step is to specify a single multivariate parametric model for \( \{V_t\} \). The second step involves the estimation of the parameters of the parametric models. In our context, standard estimation methods are local (weighted) least-squares (i.e., least-squares method applied to rolling windows), nonparametric kernel methods and generalized Whittle’s (1953) method [cf. Dahlhaus (1997)]. In a third step, we replace the unknown parameters in \( \phi(q) \) with corresponding estimates. Such estimate \( \hat{\phi}(q) \) of \( \phi(q) \) is then substituted into the expression for the optimal bandwidth parameter \( b_{1,T}^{\text{opt}} \) to yield the data-dependent bandwidth parameter \( \hat{b}_{1,T} \):

\[
\hat{b}_{1,T} = \left( 2qK_{1,q}^2 \hat{\phi}(q) T \hat{b}_{2,T} / \left( \int K_1^2(y) dy \int K_2^2(x) dx \right) \right)^{-1/(2q+1)},
\]

where \( \hat{b}_{2,T} = (n_T/T) \sum_{r=1}^{[T/n_T]} -1 \hat{b}_{2,T} (rn_T/T). \) \( \hat{b}_{2,T} \) is an average of the estimated bandwidths \( \hat{b}_{2,T} (\cdot) \). Since \( b_{2,T} \) is applied to each block of data, \( b_{2,T} \) depends on \( u \). It is then more efficient to estimate it for each block of data as its optimal value can change over the sample. In practice, a reasonable candidate to be used as an approximating parametric model is the class of first order autoregressive [AR(1)] models for \( \{V_t^{(r)}\}, r = 1, \ldots, p \) (with different parameters for each \( r \)) or a first order vector autoregressive [VAR(1)] model for \( \{V_t\} \). This class was also used by Andrews (1991); it is parsimonious and it has been commonly adopted in the HAC estimation literature.
However, in our context it is reasonable to augment the canonical AR(1) model by allowing the parameters to be time-varying. For parsimony, we consider time-varying autoregression of order one with no break points in the spectrum.

The use of $p$ univariate parametric models requires a simple form for the weight matrix $W$ that appears in (1.5.1). In particular, $W$ has to be a diagonal matrix which in turn implies that $\phi(q)$ reduces to

$$\phi(q) = 2^{-1} \sum_{r=1}^{p} W^{(r,r)} \left( \int_{0}^{1} f^{(q)(r,r)}(u, 0) \, du \right)^2 / \sum_{r=1}^{p} W^{(r,r)} \left( \int_{0}^{1} f^{(r,r)}(u, 0) \, du \right)^2.$$  

The usual choice is $W^{(r,r)} = 1$ for $r = 1, \ldots, p$ or for all $r$ except that which corresponds to an intercept for which it is set to zero. An estimate of $f^{(r,r)}(u, 0)$ ($r = 1, \ldots, p$) is $\hat{f}^{(r,r)}(u, 0) = (2\pi)^{-1} \left( \hat{\sigma}^{(r)}(u) \right)^2 |1 - \hat{a}_1^{(r)}(u)|^{-2}$ while $f^{(2)(r,r)}(u, 0)$ can be estimated by $\hat{f}^{(2)(r,r)}(u, 0) = -3\pi^{-1} \left( \hat{\sigma}^{(r)}(u) \hat{a}_1(u) \right)^2 |1 + \hat{a}_1^{(r)}(u)|^{-4}$ where $\hat{a}_1^{(r)}(u)$ and $\hat{\sigma}^{(r)}(u)$ are the least-squares estimates computed using data close to time $u = t/T$:

$$\hat{a}_1^{(r)}(u) = \frac{\sum_{j=t-n_T+1}^{t} \hat{V}_{j}^{(r)} \hat{V}_{j-1}^{(r)}}{\sum_{j=t-n_T+1}^{t} \left( \hat{V}_{j-1}^{(r)} \right)^2}, \quad \hat{\sigma}^{(r)}(u) = \left( \frac{\sum_{j=t-n_T+1}^{t} \left( \hat{V}_{j}^{(r)} - \hat{a}_1^{(r)}(u) \hat{V}_{j-1}^{(r)} \right)^2}{\sum_{j=t-n_T+1}^{t} \left( \hat{V}_{j-1}^{(r)} \right)^2} \right)^{1/2},$$  

(1.5.2)

where $n_T$ is the same as the one used in the definition of $\hat{J}_T$, though in general it does
not need to be the same. Then, for the QS kernel \( K_1 \)—for which \( q = 2 \)—we have,

\[
\hat{\phi}(2) = \frac{\sum_{r=1}^{p} W^{(r,r)} \left( 36 \left( \frac{nT}{T} \sum_{j=0}^{[T/nT]-1} \left( \hat{\sigma}^{(r)} ((jn_T + 1) / T) \right)^2 \hat{a}_1 ((jn_T + 1) / T) \right)^2 \right) / \left( \frac{nT}{T} \sum_{j=0}^{[T/nT]-1} \left| 1 + \hat{a}_1^{(r)} ((jn_T + 1) / T) \right|^2 \right)^2}{\sum_{r=1}^{p} W^{(r,r)} \left( \frac{nT}{T} \sum_{j=0}^{[T/nT]-1} \left( \hat{\sigma}^{(r)} ((jn_T + 1) / T) \right)^2 \right) / \left( \frac{nT}{T} \sum_{j=0}^{[T/nT]-1} \left| 1 + \hat{a}_1^{(r)} ((jn_T + 1) / T) \right|^2 \right)^2}.
\]

After plugging-in \( \hat{\phi}(2) \) into the formula (1.4.2), we have

\[
\hat{b}_{1,T} = 0.6828 \left( \hat{\phi}(2) T \hat{b}_{2,T} \right)^{-1/5}.
\]

We now propose a data-dependent procedure for the bandwidth \( b_{2,T} (u_r) \), where \( u_r = rn_T / T \) for \( r = 1, \ldots, \lfloor (T - n_T) / n_T \rfloor \). We assume that the parameters of the approximating time-varying AR(1) models change slowly such that the smoothness of \( f(\cdot, \omega) \) and thus of \( c(\cdot, \cdot) \) is the same to the one that would arise if \( a_1(u) = 0.8 (\cos 1.5 + \cos 4\pi u) \) and \( \sigma(u) = 1 \) for all \( u \in [0, 1] \) [cf. Dahlhaus (2012)]. The reason for imposing this condition is that it is otherwise difficult to estimate the second derivative of \( c(u, \omega) \), which enters \( D_1(u) \), from the data. Under the above specification, the exact expression of \( D_1(u) \) can be computed analytically:

\[
D_{1,\theta} (u) \triangleq \frac{1}{\pi} \left( 1 + (0.8 (-4\pi \sin (4\pi u))) 0.8 (-4\pi \sin (4\pi u)) \right) + \frac{1}{\pi} \left( 1 + (0.8 (-4\pi \sin (4\pi u))) 0.8 (-16\pi^2 \cos (4\pi u)) \right).
\]

It remains to derive an estimate of \( D_2(u) \) since \( F(K_2) \) and \( H(K_2) \) can be computed for a given \( K_2(\cdot) \). Since \( c(u, k) \) can be consistently estimated by Proposition 1.3.1,
an estimate of
\[ \hat{D}_2(u_0) \triangleq \sum_{l=-\lfloor T^{4/25} \rfloor}^{\lfloor T^{4/25} \rfloor} \hat{c}_T(u_0, l) \left[ \hat{c}_T(u_0, l) + \hat{c}_T(u_0, l + 2k) \right], \]
where the number of summands (or autocovariance lags) grows at the same rate as the inverse of the optimal bandwidth \( b_{1,T}^{\text{opt}} \); a different choice is allowed as long as it grows at a slower rate than \( T^{4/25} \). Hence, the estimate of the optimal bandwidth \( b_{2,T} \) is given by
\[ \hat{b}_{2,T}(u_r) = 1.7781 \left( D_{1,\theta}(u_r) \right)^{-1/5} \left( \hat{D}_2(u_r) \right)^{1/5} T^{-1/5}, \quad \text{where} \quad u_r = rn_T/T. \]

\section*{1.5.2 Theoretical Results}

Next, we establish consistency, rate of convergence and asymptotic MSE results for the estimator \( \hat{J}_T \left( \hat{b}_{1,T}, \hat{b}_{2,T} \right) \) that uses the data-dependent bandwidths \( \hat{b}_{1,T} \) and \( \hat{b}_{2,T} \). As in Andrews (1991), we need to restrict the class of admissible kernels to the following class:

\[ K_3 = \left\{ K_3(\cdot) \in K_1 : (i) \left| K_1(x) \right| \leq C_1 \left| x \right|^{-b} \text{ for some } b > 1 + 1/q \text{ and some } C_1 < \infty, \text{ where } q \in (0, \infty) \text{ is such that } K_{1,q} \in (0, \infty), \text{ and } (ii) \right. \]
\[ \left. \left| K_1(x) - K_1(y) \right| \leq C_2 \left| x - y \right| \forall x, y \in \mathbb{R} \text{ for some constant } C_2 < \infty \right\}. \]

Let \( \hat{\theta} \) denote the estimator of the parameter of the approximate (time-varying) parametric model(s) introduced above. For example, with univariate AR(1) approximating parametric models,
\[ \hat{\theta} = \left( \int_0^1 \hat{\alpha}_1(u) \, du, \int_0^1 \hat{\sigma}_1^2(u) \, du, \ldots, \int_0^1 \hat{\alpha}_p^2(u) \, du, \int_0^1 \hat{\sigma}_p^2(u) \, du \right)'. \]
Let \( \theta^* \) denote the probability limit of \( \hat{\theta} \). \( \hat{\phi} (q) \) is the value of \( \phi (q) \) with \( \hat{\theta} \) instead of \( \theta \). The probability limit of \( \hat{\phi} (q) \) is denoted by \( \phi_{\theta^*} \).

**Assumption 1.6.** (i) \( \hat{\phi} (q) = O_p (1) \) and \( 1/\hat{\phi} (q) = O_p (1) \); (ii) \( \sqrt{n_T} (\hat{\phi} (q) - \phi_{\theta^*}) = O_p (1) \) for some \( \phi_{\theta^*} \in (0, \infty) \); (iii) \( \sup_{u \in [0, 1]} \lambda_{\max} (\Gamma_u (k)) \leq C_3 k^{-l} \) for all \( k \geq 0 \) for some \( C_3 < \infty \) and some \( l > \max \{2, 1 + 2q/(q + 2)\} \), where \( q \) is as in \( K_3 \); (iv) uniformly in \( u \in [0, 1] \), \( D_2 (u) = O_p (1) \) and \( 1/D_2 (u) = O_p (1) \); (v) \( \sqrt{T_b_{2,T} (u) (D_2 (u) - D_2 (u)) = O_p (1) \) for all \( u \in [0, 1] \); (vi) \( K_2 \) includes kernels that satisfy \( |K_2 (x) - K_2 (y) | \leq C_4 |x - y| \) for all \( x, y \in \mathbb{R} \) and some constant \( C_4 < \infty \).

Parts (i), (iv) and (vi) are sufficient for the consistency of \( \hat{J}_T (b_{1,T}, b_{2,T}) \). Parts (ii), (iii) and (v) are required for the rate of convergence and asymptotic MSE results. Note that \( \phi_{\theta^*} \) coincides with the optimal value \( \phi (q) \) only when the approximate parametric model indexed by \( \theta^* \) corresponds to the true data-generating mechanism.

Part (v), (iv) and (vi) correspond to the kernel \( K_2 \) and associated bandwidth \( b_{2,T} \). Part (v) follows from the asymptotic results about \( \hat{c}_T (u, k) \).

Let \( b_{\theta^* T} = \left( 2q K_{1, q}^2 \phi_{\theta^*} T b_{2,T} \right) / \int K_{1, q}^2 (y) dy \int_0^1 K_{2, q}^2 (x) dx \right)^{-1/(2q+1)} \), where

\[
\tilde{b}_{\theta^* T} \triangleq \int_0^1 [(3/4) H (K_2) D_{1, \theta} (u)]^{-1/5} (F (K_2) D_2 (u))^{1/5} T^{-1/5} du.
\]

The asymptotic properties of \( \hat{J}_T (b_{1,T}, b_{2,T}) \) are shown to be equivalent to those of \( \hat{J}_T (b_{\theta^* T}, b_{2,T}) \).

**Theorem 1.5.1.** Suppose \( K_1 (\cdot) \in \mathcal{K}_3 \), \( q \) is as in \( K_3 \) and \( \| \int_0^1 f^{(q)} (u, 0) \| < \infty \). Then, we have:

(i) If Assumption 1.2-1.3 and 1.6-(i, iv, vi) hold, and \( q > 1/2 \), then

\[
\hat{J}_T (b_{1,T}, b_{2,T}) - J_T \xrightarrow{p} 0.
\]
(ii) If Assumption 1.3-1.4 and 1.6-(ii,iii,v,vi) hold and $n_T / T b_{1,T} \rightarrow 0$, $n_T / T b_{1,T}^2 \rightarrow 0$, $n_T / T q \rightarrow 0$, $n_T / T Q \rightarrow \infty$, then

$$
\sum T b_{1,T} \theta_{1,T} b_{2,T} \left( J_T \left( b_{1,T}, b_{2,T} \right) - J_T \right) = O_p(1).
$$

In addition, if $q \leq 2$, then

$$
\sum T b_{1,T} \theta_{1,T} b_{2,T} \left( J_T \left( b_{1,T}, b_{2,T} \right) - J_T \left( b_{1,T}, b_{2,T} \right) \right) = o_p(1).
$$

(iii) Let $\gamma_\theta = 2q K^2_1 q \varphi_\theta / \left( \int K^2_1 (y) dy \int_0^1 K^2_2 (x) dx \right)$. If Assumption 1.3-1.5 and 1.6-(ii,iii) hold, then

$$
\lim_{T \rightarrow \infty} \text{MSE} \left( T b_{\theta_1,T} b_{\theta_2,T}, J_T \left( b_{1,T}, b_{2,T} \right), W_T \right) = \lim_{T \rightarrow \infty} \text{MSE} \left( T b_{\theta_1,T} b_{\theta_2,T}, J_T \left( b_{\theta_1,T}, b_{\theta_2,T} \right), W_T \right)
$$

$$
= 4\pi^2 \left[ \gamma_\theta K^2_1 q \text{vec} \left( \int_0^1 f^{(q)} (u, 0) du \right) \text{vec} \left( \int_0^1 f^{(q)} (u, 0) du \right) \right]
$$

$$
+ \int K^2_1 (y) dy \int K^2_2 (x) dx \text{tr} (I - C_{pp})
$$

$$
\times \left( \int_0^1 f (u, 0) du \right) \otimes \left( \int_0^1 f (v, 0) dv \right)
$$

The condition $q \leq 2$ is only needed for the second result of part (ii); the most commonly used kernels satisfy this restriction. When the chosen parametric model indexed by $\theta$ is correct, it follows that $\varphi_{\theta^*} = \varphi (q)$ and $\hat{\varphi} (q) \xrightarrow{p} \varphi (q)$. The theorem then implies that $\hat{J}_T \left( b_{1,T}, b_{2,T} \right)$ exhibits the same optimality properties presented in Theorem 1.4.1 and Corollary 1.4.1. We omit the details.

### 1.6 Inference about the Break Points

A more efficient HAC estimator uses information about the location of the (possible) break points in the spectrum $f (u, \omega)$. Casini (2018c) proposed frequency-domain methods to test for breaks in the time-varying spectrum of a time series along with methods for the estimation of the break points. Here we briefly explain how those methods can be applied to the current context.
The methods in Casini (2018c) are based on the local smoothed periodogram. The procedure is applied to univariate time series. Thus, here we apply the procedure to each series \( \{V_t^{(k)}\} \), \( k = 1, \ldots, p \), separately. One could apply the procedure to a single series (i.e., any \( k \)) only if a common break assumption is imposed, i.e., for every \( k = 1, \ldots, p \) the breaks in the spectrum of \( \{V_t^{(k)}\} \) occur at the same time. Here we present the discussion for a given \( k \). Denote by \( f_{V^{(k)}}(u, \omega) \) the spectrum of \( V^{(k)} \). Let \( I_{r,j,T}(\omega) \triangleq (2n\pi)^{-1}\left| \sum_{s=1}^{n} \exp \left( -i(\rho n + j + s) \omega \right) V_{rn+j+s,T}^{(k)} \right|^2 \), where \( n \) is some sequence such that \( n \to \infty \) as \( T \to \infty \). \( I_{r,j,T}(\omega) \) is a rescaled local version of the periodogram over blocks of the partition \([rn + j, (r + 1)n + j] \). Under the null hypothesis of no break in the spectrum, \( V_t,T \) is a locally stationary process (i.e., \( m = 0 \) in Definition 1.2.1). Let \( \chi^2_k \) denote a chi-squared random variable with \( k \) degrees of freedom. Then, since for large \( T \), \( I_{r,j,T}(\omega) \approx f_{V^{(k)}}((rn T + j) / T, \omega) \chi^2_2 \) for \( \omega \in (-\pi, \pi) \) (with \( \chi^2_1 \) in place of \( \chi^2_2 \) when \( \omega = -\pi \) or \( \pi \)), we have \( 2^{-1}E(I_{r,j,T}(\omega)) \approx f_{V^{(k)}}((rn T + j) / T, \omega) \) for \( \omega \in (-\pi, \pi) \). However, the local periodogram \( I_{r,j,T}(\omega) \) is an inconsistent estimate of \( f_{V^{(k)}}((rn T + j) / T, \omega) \). Therefore, a better proxy for \( f_{V^{(k)}}(u, \omega) \) is the smoothed periodogram. We smooth \( I_{r,j,T}(\omega) \) over an odd number \( N_T \) of adjacent frequencies:

\[
\text{SI}_{r,j,T}(\omega) \triangleq \frac{1}{N_T} \sum_{l=-(N_T-1)/2}^{(N_T-1)/2} I_{r,j,T}(\omega + l), \quad r = 0, \ldots, [T/n_T] - 1, \ j = 1, \ldots, n.
\]

(1.6.1)

\( \text{SI}_{r,j,T}(\omega) \) is asymptotically distributed as \( f_{V^{(k)}}(t/T, \omega) \chi^2_{2N_T}/2N_T \) with \( t = rn T + j \), with one fewer degree of freedom if \( \omega + l = -\pi \) or \( \omega + l = \pi \). If \( N_T \to \infty \) and \( N_T/n \to 0 \) as \( n \to \infty \), \( \text{SI}_{r,j,T}(\omega) \) is a consistent estimator of the power spectrum at time \( t \) and frequency \( \omega \), \( f_{V^{(k)}}(t/T, \omega) \). Therefore, we may try to identify breaks where by looking for too large deviations between any two successive local estimators
of the spectrum. Let

$$\widehat{SB}_{r,T}(\omega) \triangleq n^{-1} \sum_{j=1}^{n} SI_{r,j,T}(\omega),$$

which estimates a block-wise proxy of the spectrum $f((r-1)n/T, \omega)$ on the respective blocks. As mentioned above, a large distance between $\widehat{SI}_{r,T}(\omega)$ and $\widehat{SI}_{r-1,T}(\omega)$ suggests the presence of a jump or unsmooth break in the spectrum close to time $rn$. Casini (2018c) proposed a local two-sample $t$-test over asymptotically vanishing time blocks:

$$S_{\text{max},T}(\omega) \triangleq \max_{r=0, \ldots, [T/n]-2} \left| \frac{\widehat{SB}_{r+1,T}(\omega)}{\widehat{SB}_{r,T}(\omega)} - 1 \right|, \quad \omega \in [-\pi, \pi]. \quad (1.6.2)$$

Here we focus on the version of the statistic that uses all overlapping blocks of $n$ increments:

$$MS_{\text{max},T}(\omega) \triangleq \max_{r=n, \ldots, T-n} \left| \frac{n^{-1} \sum_{s=r+1}^{r+n} \sum_{l=-(N_T-1)/2}^{(N_T-1)/2} \widehat{MSI}_{s,T}(\omega + l)}{n^{-1} \sum_{s=r-n+1}^{r-n} \sum_{l=-(N_T-1)/2}^{(N_T-1)/2} \widehat{MSI}_{s,T}(\omega + l)} - 1 \right|, \quad (1.6.3)$$

where

$$\widehat{MSI}_{s,T}(\omega + l) \triangleq \frac{1}{2n\pi} \sum_{j=1}^{n} \exp \left( -i(s - 1 + j)(\omega + l) \right) X_{s-1+j}. \quad$$

The statistics $\widehat{MSI}_{r-1,T}(\omega)$ are dependent as the smoothed periodogram introduces short-range dependence over $\omega$. Thus, in order to apply limit theorems toward extreme value distribution, which essentially require independence, we cannot take the maximum over all possible frequencies but only over a subset thereof. The test statistic is a double-sup test which seeks for maximal sample evidence for breaks in the spectrum. Let $\Pi' \triangleq \{\omega_1, \omega_{2+N_T}, \ldots, \omega_{n_{\omega}-N_T-1}, \omega_{n_{\omega}}\}$, $n'_{\omega} \triangleq \lfloor 2\pi/(N_T + 1) \rfloor$ and $\mathcal{V}$ be a random variable defined by $\mathbb{P}(\mathcal{V} \leq v) = \exp \left( -\pi^{-1/2} \exp \left( -v \right) \right)$. The null
hypotheses in Casini (2018c) specifies that \( f(u, \omega) \) is Lipschitz continuous:

\[
F(a_T) = \left\{ \{ f(u, \omega) \}_{u \in [0, 1], \omega \in [-\pi, \pi]} \mid \sup_{\omega \in [-\pi, \pi], u, v \in [0, 1], |v-u|<h} |f(u, \omega) - f(v, \omega)| \leq a_T h \right\},
\]

for an appropriate sequence \( a_T \to \infty \). The formulation of the hypothesis testing problem in Casini (2018c) is actually formulated in a more technical manner; more specifically, in terms of the so-called minimax detection boundary which is defined as the minimum break magnitude such that we are still able to uniformly control the type I and type II errors. For our purposes, here we test the hypotheses \( H_0 \) that \( f_{V(k)}(u, \omega) \in F(a) \) where \( a < \infty \). Then, under \( H_0 \) the results in Casini (2018c) imply that

\[
\max_{\omega_k \in \Omega} \sqrt{\log(m_T) n_T^{1/2} N_T^{1/2} \text{MS}_{\max,T}(\omega_k)} - (2 \log(m_T) + (1/2) \log(\log m_T)) - \log 3 - \log(n_\omega) \Rightarrow \mathcal{V},
\]

where \( m_T \triangleq \lfloor T/n \rfloor \). The null limiting distribution of the test is an Extreme Value distribution. The proof of this result relies on Gaussian approximation to partial sums. These are known as strong invariance principles [see Wu (2007), Wu and Zhou (2011) and reference therein]. Casini (2018c) extended such Gaussian approximations to partial sums of frequency-domain estimates allowing for nonstationarity.

Once the null hypothesis is rejected, the next step is to estimate the locations of the breaks in the spectrum of \( V_{t,T}^{(k)} \). Let us assume that there are \( m^{(k)} \) break points in \( f_{V(k)}(u, \omega) \). This means that \( f_{V(k)}(T_{t,+}^0/T, \omega) \) satisfies the following hypotheses:

\[
H_1 : \left\{ f\left(T_{t,+}^0/T, \omega\right) - f\left(T_t^0/T, \omega\right) = \delta_l \neq 0, \quad \text{for} \ 1 \leq l \leq m^{(k)} \text{and} \ \omega \in [-\pi, \pi] \right\},
\]
where $T_{t,+}^0 = \lim_{s \downarrow T_{t}^0} s$. Collect the break points in the set $\Lambda \triangleq \{ \lambda_1^0, \ldots, \lambda_m^0 \}$. Consider the following statistic:

$$D_{r,T}(\omega) = \frac{1}{\sqrt{n}} \left| \sum_{s=r+1}^{r+n} \widehat{MSI}_{s,T}(\omega) - \sum_{s=r-n+1}^{r} \widehat{MSI}_{s,T}(\omega) \right|,$$

where $r = n_T, \ldots, T - n_T$. Unlike $\text{MS}_{\text{max,T}}$, $D_{r,T}(\omega)$ does not involve ratios but only differences over adjacent (overlapping) blocks. Let $\mathcal{I} \subseteq \{n, \ldots, T - n\}$ denote a generic index set. Let $h_T \to \infty$ with $h_T/T \to 0$. Based on the following estimator, Casini (2018c) proposed Algorithm 1 below for the detection of the number of breaks $m_k$ in the spectrum of $\{V_{t,T}^{(k)}\}$, where use is made of the generic estimate,

$$T \hat{\lambda}_T(\mathcal{I}) = \arg \max_{r \in \mathcal{I}} \max_{\omega \in [-\pi,\pi]} D_{r,T}(\omega), \quad (1.6.5)$$

for some set $\mathcal{I}$.

**Algorithm 1.** Set $\hat{\mathcal{I}} = \{n, \ldots, T - n\}$ and $\hat{T} = \emptyset$. (1) If $H_0$ is not rejected, return $\hat{T} = \emptyset$. Otherwise proceed with step 2; (2) Estimate the change-point via (1.6.5) by using $\hat{\mathcal{I}}$. Call it $\hat{\lambda}_T(\hat{\mathcal{I}})$; (3) Set

$$\hat{\mathcal{I}} = \mathcal{I} \backslash \{ [T \hat{\lambda}_T(\hat{\mathcal{I}})] - h_T, \ldots, [T \hat{\lambda}_T(\hat{\mathcal{I}})] + h_T \}.$$

Repeat step 1.

Algorithm 1 has to be repeated for each $k$. This returns an estimated number of breaks $\hat{m} = \sum_{k=1}^{p} \hat{m}_k$ along with $\hat{m}$ break dates. $^{10}$ The latter should be placed in chronological order. That is, if $\hat{m} \geq 1$ we will have $\hat{T}_j, j = 1, \ldots, \hat{m}$. The estimated break dates $\hat{T}_j$ are used in the construction of the HAC estimator as described in the next section.

$^{10}$Here the notation assumes that there are non-repetitive breaks across each series $k = 1, \ldots, p$. If there are commons breaks then evidently we include them once.
1.7 HAC Estimation when \( m > 0 \)

When there are break points in the spectrum of \( \{V_{t,T}\} \) the estimator \( \hat{J}_T \) takes a different form. More specifically, the summation in \( \hat{\Gamma} (k) \) should not be over arbitrary blocks of length \( n_T \). Efficiency requires that information from the estimation of the break dates is used in constructing \( \hat{\Gamma} (k) \). Let

\[
\mathcal{T}_J \triangleq \{0, n_T, \ldots, \hat{T}_1 - n_T, \hat{T}_1, \hat{T}_1 + n_T, \ldots, \hat{T}_m - n_T, \hat{T}_m, \hat{T}_m + n_T, \ldots, T - n_T, T\},
\]

where we have assumed that \( \hat{T}_1 - n_T > n_T \) and \( \hat{T}_m < T - n_T \) for notational convenience (i.e., if \( \hat{T}_1 - n_T \leq n_T \) then the smallest two indexes in \( \mathcal{T}_J \) are 0 and \( \hat{T}_1 - n_T \)). Let \( |\mathcal{A}| \) denote the cardinality of a set \( \mathcal{A} \). \( \hat{J}_T \) takes the following form: \( \hat{J}_T = \hat{J}_T (b_1, T) = \sum_{k=-T+1}^{T-1} K_1 (b_1, T k) \hat{\Gamma} (k) \) where \( \hat{\Gamma} (k) = |\mathcal{T}_J|^{-1} \sum_{r \in \mathcal{T}_J} \hat{c} (r/T, k) \) and

\[
\hat{c} (r n_T, k) \triangleq \begin{cases} 
(T b_2, T)^{-1} \sum_{s=k+1}^{T} K_2 \left( \frac{((r+1)n_T-(s-k/2))/T}{b_2, T} \right) \hat{V}_{s-k} \hat{V}'_{s} & , k \geq 0 \\
(T b_2, T)^{-1} \sum_{s=-k+1}^{T} K_2 \left( \frac{((r+1)n_T-(s+k/2))/T}{b_2, T} \right) \hat{V}_{s+k} \hat{V}'_{s} & , k < 0
\end{cases}
\]

The following theorem presents the asymptotic results corresponding to Theorem 1.3.1-1.3.2. As for the latter theorems, the proof of Theorem 1.7.1 relies on MSE and consistency results concerning \( \hat{c}_T (\cdot, \cdot) \). Those results—which are the counterpart of Proposition 1.3.1—are proved as part of the proof of Theorem 1.7.1.

**Assumption 1.7.** (i) \( \{V_{t,T}\} \) is a mean zero Segmented Locally Stationary process, \( A (u, \omega) \) is twice continuously left-differentiable in \( u \) and uniformly Lipschitz continuous with index \( \vartheta = 1 \) in \( \omega \), \( \sum_{k=-\infty}^{\infty} \sup_{u \in [0,1]} \| c (u, k) \| < \infty \) and

\[
\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sup_{u \in [0,1]} \kappa_{V_{1,T}}^{(a,b,c,d)} (k, j, l) < \infty,
\]
for all $a, b, c, d \leq p$. 

(ii) For all $a, b, c, d \leq p$ there exists a constant $K$ and a function $\tilde{\kappa}_{a,b,c,d} : [0, 1] \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ such that

$$\sup_{1 \leq j \leq m+1} \sup_{\lambda_0^{j-1} < u \leq \lambda_0^j} |\kappa^{(a,b,c,d)}_{\nu,[Tu]}(k, s, l) - \tilde{\kappa}_{a,b,c,d}(u, k, s, l)| \leq KT^{-1};$$

The function $\tilde{\kappa}_{a,b,c,d}(u, k, s, l)$ is uniformly piecewise Lipschitz continuous in $u$ for all $a, b, c, d \leq p$.

When $t = T_j^0$ for some $j$, the requirement reduces to left-Lipschitz continuity in $u$ of $\tilde{\kappa}_{a,b,c,d}(u, k, s, l)$ with $t = Tu$.

**Theorem 1.7.1.** The results of Theorem 1.3.1 and 1.3.2-(i,iii) continue to hold when $m > 0$. With the additional condition $\sqrt{b_{1,T}b_{2,T}n_T}/T \to 0$, 1.3.2-(ii) continue to hold when $m > 0$.

The results about optimality and data-dependent bandwidths can also be shown to hold using Theorem 1.7.1. The optimal bandwidth $b_{2,T}$ is slightly different. The expression for $\partial^2 c(u_0, k)/\partial u^2$ in $D_1(u_0)$ depends on whether $Tu_0 \in \mathcal{T}$ or $Tu_0 \notin \mathcal{T}$.

For the former case, $b_{2,T}^{opt}$ remains as before. When $Tu_0 \in \mathcal{T}$ and $k < 0 \partial^2 c(u_0, k)/\partial u^2$ is replaced by

$$\int_{-\pi}^{\pi} \exp(i\omega k)(C_1(u_0, \omega) + C_2(u_0, \omega) + C_3(u_0, \omega)) d\omega,$$

where, assuming $Tu_0 = T_j^0$,

$$C_1(u_0, \omega) = 2\frac{\partial A_j(u_0, -\omega)}{\partial u} \frac{\partial A_{j+1}(v, \omega)}{\partial v}, \quad C_2(u_0, \omega) = \frac{\partial^2 A_{j+1}(v, \omega)}{\partial^2 v^2} A_j(u_0, -\omega),$$

$$C_3(u_0, \omega) = \frac{\partial^2 A_j(u_0, \omega)}{\partial^2 u^2} A_{j+1}(v, \omega),$$

with $\partial^2 A_j(u_0, \omega)/\partial u^2$ ($\partial^2 A_j(v, \omega)/\partial v^2$) being the second left- (right-) derivative.
at $u_0$ ($v = u_0 - k/T$). When $Tu_0 \in \mathcal{T}$ and $k > 0$, $\partial^2 c(u_0, k) / \partial u^2$ is replaced by $\partial^2 c(u_0, k) / \partial u^2$.

For the data-dependent method, the parametric time-varying AR(1) assumption about the approximating model has to be replaced by a time-varying AR(1) model with a certain number of breaks in the spectrum of the series. Thus, one has first to test and estimate the breaks (cf. Section 1.6). This information is then used in the construction of the data-dependent method. For example, if there is a single break at date $T_2^0$ then (1.5.2) would be replaced by

\[
\hat{\alpha}_1^{(r)}(u) = \sum_{j=T_2-nT+1}^{T_2} \frac{\hat{V}_{j-1}^{(r)}\hat{V}_j^{(r)}}{\sum_{j=T_2-nT+1}^{T_2}(\hat{V}_{j-1}^{(r)})^2}, \quad \hat{\sigma}^{(r)}(u) = \left( \sum_{j=T_2-nT+1}^{T_2} (\hat{V}_j^{(r)} - \hat{\alpha}_1^{(r)}(u)\hat{V}_{j-1}^{(r)})^2 \right)^{1/2},
\]

where $u = \hat{T}_2/T$ with $\hat{T}_2$ being the estimate of $T_2^0$. The rest of the data-dependent procedure remains unchanged even thought the optimal $b_{2,T}$ is different. The reason is that we can extend the same parametric assumption about the smoothness of the transfer function $A_j(u, \omega)$ to all regimes $1 \leq j \leq m + 1$. Thus, the form of $D_1(u_0)$ remains the same as in Section 1.5.

An additional interesting issue is how to efficiently determine the time points which separate the regimes. This issue arises when the testing procedure for the detection of breaks in the spectrum finds a certain number of break points. On this purpose, we are currently working on an algorithm based on the principle of dynamic programming—akin to Bai and Perron (2003) for estimating models with multiple structural changes.

1.8 Small-Sample Evaluations

In this section, we conduct a Monte Carlo analysis to evaluate the properties of HAR inference based on the HAC estimator $\hat{J}_T$ relative to the traditional HAR inference
based on the classical kernel-based HAC estimators. We consider tests in the linear regression models as well as \( t \)-tests employed in the forecast evaluation literature, namely the Diebold-Mariano test [cf. Diebold and Mariano (1995)] and the forecast breakdown test of Giacomini and Rossi (2009). The linear regression models have an intercept and a stochastic regressors. Recall the notations for the linear model introduced in Section 1.2.2. We focus on the \( t \)-statistics

\[
t_r = \sqrt{T} \left( \hat{\beta}^{(r)} - \beta^{(r)}_0 \right) / \sqrt{\hat{J}_T^{(r,r)}}
\]

where \( \hat{J}_T \) is a consistent estimate of the asymptotic variance of \( \text{Var} \left( \sqrt{T} (\hat{\beta} - \beta_0) \right) \) and \( r = 1, 2 \). \( t_1 \) is the \( t \)-statistic for the parameter associated to the intercept while \( t_2 \) is associated to the stochastic regressor \( x_t \). Results for the \( F \)-test are reported in Section C.2 of the Supplement. Six basic regression models are considered. We run a \( t \)-test on the intercept in model S1-S2 and S6 whereas a \( t \)-test on the coefficient of the stochastic regressor is run in model S3-S5. Model S1 is a location model

\[
y_t = \beta^{(1)}_0 + e_t \quad \text{for} \quad t = 1, \ldots, T,
\]

where \( e_t = 0.6e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \). All the rest of the models are based on,

\[
y_t = \beta^{(1)}_0 + \delta + \beta^{(2)}_0 x_t + e_t, \quad t = 1, \ldots, T, \quad (1.8.1)
\]

for the \( t \)-test on the intercept (i.e., \( t_1 \)) and

\[
y_t = \beta^{(1)}_0 + \left( \beta^{(2)}_0 + \delta \right) x_t + e_t, \quad t = 1, \ldots, T, \quad (1.8.2)
\]

for the \( t \)-test on \( \beta^{(2)}_0 \) (i.e., \( t_2 \)) where \( \delta = 0 \) under the null hypotheses. Model S2 involves segmented locally stationary errors

\[
e_t = \rho_t e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1), \quad \rho_t = -0.8 (\cos (1.5 - \cos (10t/T))) \quad \text{for} \quad t < 4T/5, \quad e_t = 0.8e_{t-1} + 2u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \]

for \( t \geq 4T/5 \) and \( x_t \sim \text{i.i.d.} \mathcal{N}(1, 1) \). Model S3-S4 involve some misspecification that
induces nonstationarity in the errors. Model S3 is given by

\[ y_t = \beta_0^{(1)} + \left( \beta_0^{(2)} + \delta \right) x_t + w_t 1 \{ t \geq 4T/5 \} + e_t, \quad t = 1, \ldots, T, \]

where \( e_t = 0.4e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 1), \ x_t \sim \text{i.i.d. } \mathcal{N}(1, 1), \) and \( w_t \) has the same distribution as \( x_t \) but it is independent from \( x_t \). Model S4 is the same as S3 but the autoregressive coefficient of \( e_t \) is 0.2. In model S5 we have locally stationary errors \( e_t = \rho_t e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 1), \rho_t = -0.4 (\cos(1.5 - \cos(12t/T))) \). Model S6 involves segmented locally stationary errors \( e_t = \rho_t e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 1), \rho_t = - (\cos(1.5 - \cos(10t/T)))) \) for

\[ t \in \{ 1, \ldots, T/5 - 1 \} \cup \{ T/2 + 1, \ldots, 4T/5 \}, \]

\( e_t = 0.6e_{t-1} + 2v_t, \ v_t \sim \text{i.i.d. } \mathcal{N}(0, 1) \) for \( T/5 \leq t \leq 2T/5, \ e_t = 0.8e_{t-1} + 2v_t, \ v_t \sim \text{i.i.d. } \mathcal{N}(0, 1) \) for \( t \geq 4T/5, \) and \( x_t \sim \text{i.i.d. } \mathcal{N}(1, 1) \) for \( 1 \leq t \leq T/8 - 1 \) and \( x_t = 1 + 0.7x_{t-1} + 2u_{X,t}, \ u_{X,t} \sim \text{i.i.d. } \mathcal{N}(0, 1) \) for \( t \geq T/8. \)

Next, we move to the forecast evaluation tests. The Diebold-Mariano test statistic is defined as

\[ t_{DM} \triangleq \sqrt{T_n \bar{d}_L} / \sqrt{\hat{J}_{dl,T}}, \]

where \( \bar{d}_L \) is the average of the loss differentials between two competing forecast models, \( \hat{J}_{dl,T} \) is an estimate of asymptotic variance of the loss differential series and \( T_n \) is the number of observations in the out-of-sample. Throughout our study we use the quadratic loss. In model S7 we consider an out-of-sample forecasting exercise with a fixed forecasting scheme where, given a sample of \( T \) observations, \( 0.5T \) observations are used for the in-sample and the remaining half is used for prediction. The true model for the target variable is given by \( y_t = \beta_0^{(1)} + \beta_0 x_{t-1}^0 + e_t \) where \( x_{t-1}^0 \sim \text{i.i.d. } \mathcal{N}(1, 1), \ e_t = 0.3e_{t-1} + u_t \) with \( u_t \sim \text{i.i.d. } \mathcal{N}(0, 1) \) and we set \( \beta_0^{(1)} = \beta_0^{(2)} = 1. \) The two competing models both

\[ ^{11}\text{Some of the locations of the changes in the parameters are toward the end of the sample. The results are equivalent when the changes in the parameters occur in other parts of the sample.} \]
involve an intercept but differ on the predictor used in place of $x_0^t$. The first forecast model uses $x_1^t$ while the second uses $x_2^t$ where $x_1^t$ and $x_2^t$ are independent i.i.d. $\mathcal{N}(1, 1)$ sequences, both independent from $x_0^t$. Each forecast model generates a sequence of $\tau (= 1)$-step ahead out-of-sample losses $L_i^t$ ($i = 1, 2$) for $t = T/2 + 1, \ldots, T - \tau$. Then $d_t \triangleq L_2^t - L_1^t$ denotes the loss differential at time $t$. The Diebold-Mariano test rejects the null of equal predictive ability when (after normalization) $d_t$ is sufficiently far from zero.

Finally, we consider model S8 which we use for investigating the performance of a $t$-test for forecast breakdown [cf. Giacomini and Rossi (2009)]. Suppose we want to forecast a variable $y_t$ which follows the following equation: $y_t = \beta_0^{(1)} + \beta_0^{(2)} x_{t-1} + e_t$ where $x_t \sim$ i.i.d. $\mathcal{N}(1, 1.2)$ and $e_t = 0.3e_{t-1} + u_t$ with $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. For a given forecast model and forecasting scheme, the test of Giacomini and Rossi (2009) detects a forecast breakdown when the average of the out-of-sample losses differs significantly from the average of the in-sample losses. The in-sample is used to obtain estimates of $\beta_0^{(1)}$ and $\beta_0^{(2)}$ which are in turn used to construct out-of-sample forecasts $\hat{y}_t = \hat{\beta}_0^{(1)} + \hat{\beta}_0^{(2)} x_{t-1}$. We set $\beta_0^{(1)} = \beta_0^{(2)} = 1$. We consider a fixed forecasting scheme and one-step ahead forecasts. GR’s (2009) test statistic is defined as $t_{GR} \triangleq \sqrt{T_n SL} / \sqrt{J_{SL}}$ where $\overline{SL} \triangleq T_n^{-1} \sum_{t=T_m}^{T-t} SL_{t+\tau}$, $SL_{t+\tau}$ is the surprise loss at time $t + \tau$ (i.e., the difference between the time $t + \tau$ out-of-sample loss and in-sample loss, $SL_{t+\tau} = L_{t+\tau} - \overline{L}_{t+\tau}$), $T_n$ is the sample size in the out-of-sample, $T_m$ is the sample size in the in-sample and $J_{SL}$ is an HAC estimator. We restrict attention to $\tau = 1$.

Throughout our study we consider the following HAC estimators: $\hat{J}_T$ with a Bartlett kernel $K_1$ and predetermined bandwidth, $\hat{J}_T$ with a QS kernel $K_1$ and predetermined bandwidth; $\hat{J}_T$ with a QS kernel $K_1$, automatic bandwidths and no pre-test for breaks; $\hat{J}_T$ with a QS kernel $K_1$, automatic bandwidths and pre-test for breaks;
Andrews’s (1991) HAC estimator with automatic bandwidth; Andrews’s (1991) HAC estimator with automatic bandwidth and the prewhitening procedure of Andrews and Monahan (1992); Newey and West’s (1987) HAC estimator with predetermined bandwidth set equal to the so-called “rule” \( b_{1,T} = (4T/100)^{2/9} \); Newey and West’s (1987) HAC estimator with the automatic bandwidth as proposed in Newey and West (1994); Newey and West’s (1987) HAC estimator with the automatic bandwidth as proposed in Newey and West (1994) and prewhitening procedure; Newey and West’s (1987) HAC estimator with predetermined bandwidth \( b_{1,T} = 1.5T/8 \) and fixed-\( b \) critical values; the Empirical Weighted Periodogram (EWP) of Lazarus et al. (2017) with eight degrees of freedom.

For all estimators \( \hat{J}_T \), \( K_2 \) is chosen to be the optimal kernel as suggested by Proposition 1.4.1 and \( n_T = T^{0.6} \) as explained in Section 1.4.4. For the estimators \( \hat{J}_T \) that are implemented with predetermined bandwidths we set \( b_{1,T} = T^{-4/25} \) and \( b_{2,T} = T^{-1/5} \) which correspond to the order of their optimal asymptotic values, respectively. We employ the data-dependent procedures described in Section 1.5 for constructing the automatic bandwidths. We set \( \beta_0^{(1)} = 0 \) in S1-S6 and \( \beta_0^{(2)} = 1 \) in all models.

We consider the following sample sizes: \( T = 125, 200, 400 \). Simulation results for additional data-generating processes involving ARMA, ARCH and heteroskedastic errors are not discussed here because the results are qualitatively equivalent. The supplement contains additional results about size and power of \( t_1 \) and \( t_2 \) for several other models. The significance level is \( \alpha = 0.05 \) throughout the study.

### 1.8.1 Empirical Sizes of HAR Inference Tests

Table 3.1-1.6 report the rejection rates for model S1-S8. We begin with the \( t \)-test in the linear regression models. As a general pattern, we confirm previous evidence that the Newey-West (1987) HAC estimator using the “rule” to determine the bandwidth
leads to $t$-tests that are largely oversized. On average, the empirical rejection rates are 5-10% above the desired nominal level. Increasing the sample size brings only a small improvement. For models S1 and S6 (cf. Table 3.1 and 1.5), the rejection rates are more than 10% higher than the nominal level even for $T = 200$. For the same models, when $T = 400$ the rejection rates are still more than eight percentage points above the nominal level. The same problem arises for the Newey-West (1987) HAC estimator with automatic bandwidth which turns out to lead to $t$-tests that are systematically oversized—in general, 10-15% above the exact size. The latter method used together with the prewhitening procedure helps to reduce the oversize problem but it often remains deficient (cf. model S6 and model M2 in the supplement). $t$-tests that use Andrews’s (1991) HAC estimator with automatic bandwidths also have an empirical size beyond the nominal level. For some data-generating mechanisms the oversize problem can be severe. The prewhitening procedure helps reducing the oversize problem only marginally and sometimes can also lead to even worse rejection rates (cf. M2 in the supplement). Our simulations also confirm that the Newey-West’s (1987) HAC estimator implemented with a large bandwidth and fixed-$b$ critical values reduces the oversize problem relative to using the same HAC estimator with small bandwidths and with asymptotic critical values. However, the opposite issue arises. In fact, the rejection rates tend to be systematically below the nominal level by a substantial amount. In model S2-S5 the $t$-tests that use Newey-West’s (1987) HAC estimator with large bandwidths and fixed-$b$ critical values or the EWP of Lazarus et al. (2017) (which also uses a large bandwidth), show evident size distortions as they are considerably undersized. For example, in model S2 with $T = 400$ the latter two tests display rejection rates equal to 0.005 and 0.006, respectively. In model S3, the rejections rates are 0.015 and 0.009 when $T = 200$. Moreover, increasing the sample
size does not necessarily help in moving the empirical size close to the nominal level. This property can be costly in terms of power losses under the alternative hypotheses, as we will show below. The same distortions also affect the \( F \)-tests in a similar manner (cf. Table A.3 in the supplement).

Finally, when there is high serial dependence in the errors as in model S6, all the classical HAC estimators (including the ones that use fixed-\( b \) asymptotics and EWP) lead to \( t \)-tests that are significantly oversized. In model S6 with \( T = 200 \), the rejection rate that is closest to the nominal level is 0.108 which for the Newey-West (1987) with automatic bandwidth and prewhitening. The Newey-West’s (1987) HAC with fixed-\( b \) and EWP have rejection rates equal to 0.110 and 0.111, respectively. The largest distortion is achieved by the Newey-West’s (1987) HAC estimator with automatic bandwidth (i.e., 0.158).

Turning to the tests that use the HAC estimators proposed in this chapter, the ones that use automatic bandwidths control the size well in general. For the methods that do not employ automatic bandwidths, we note that, when there is high serial dependence such as in model S1, the \( t_1 \)-tests on the intercept can be slightly liberal (cf. Table S1, label “no simulation-assisted cv”). Thus, we propose a finite-sample simple refinement [cf. Zhang and Wu (2012)] which lead to better size and make the power comparisons more indicative. This refinement is described in Section 1.8.2 and it involves obtaining the critical value via a simulation-assisted method. We only apply it to \( t_1 \) but not to the tests on the stochastic regressors because they do not suffer from the same issue. Alternatively, we have developed independently a data-dependent method for obtaining the critical value which can also be useful outside the framework of this chapter. The method and its theoretical results will be presented in separate work. In addition, one could propose more complex size-refinement procedures (e.g.,
prewhitening, etc.). However, these procedures, as currently developed, are not valid under our context. Hence, the development of such procedures for our setting will be considered in the future as they are beyond the scope of this chapter. Note that the simulation-assisted method should not be confused with a size-adjustment procedure.

We observe from the tables the tests implemented with the HAC estimators proposed in this chapter are well-sized in that rejection rates are close to the nominal level in general. We do not observe a general tendency for their rejection rates to be below or above the nominal level. In any circumstance, the deviations are very small. Differences across the kernel used and/or with predetermined or automatic bandwidth are minor, the most notable being that $\hat{J}_T$ implemented with the Bartlett kernel and predetermined bandwidth is associated to lower rejection rates than $\hat{J}_T$ implemented with the QS kernel. $\hat{J}_T$ that uses the data-dependent bandwidths performs better than the one that uses predetermined bandwidths.

Moving to the HAR inference tests in the forecasting context, Table 1.6 shows that in model S7 the $\hat{J}_T$ estimators with predetermined bandwidth leads to slightly liberal tests. This does not occur when the data-dependent bandwidths are employed. Andrews’s (1991) HAC estimator (both with and without prewhitening) tends to be slightly oversized. Newey-West HAC estimators display a very unstable performance: when the “rule” is used to select the bandwidth or when the prewhitening is used, the tests are oversized, while without prewhitening the rejection rates are close to zero. EWP may suffer from somewhat large size distortions even when $T = 400$. Similar features regarding the $\hat{J}_T$ estimators, EWP and Newey-West HAC estimator with predetermined bandwidth remain valid for model S8. In contrast, Andrews’s (1991) and Newey and West’s (1987) HAC estimators with automatic bandwidths both have rejection rates close to zero.
In summary, the $\hat{J}_T$ HAC estimators yield $t$-test in regression models with rejection rates that are relatively close to the exact size. Thus, overall they perform better than the classical HAC methods. Our results confirm the oversize problem of the classical HAC estimators documented in the literature. In addition, our Monte Carlo study suggests that size distortions also arise for tests that use HAC estimators with fixed-$b$ critical values and long bandwidths. For example, Newey-West HAC with fixed-$b$ critical values and EWP can be significantly undersized. This may hold even when the data are stationary. Prewhitening applied to classical HAC estimators can be helpful in some cases and detrimental in others.

In our power function comparisons below, we consider all HAC estimators with the exception of the Newey-West estimator with either predetermined bandwidth or automatic bandwidth since they are excessively oversized.

### 1.8.2 Simulation-Assisted Critical Values for $t_1$-Test without Automatic Bandwidths

This approach is often used in statistics [cf. Zhang and Wu (2012)] in order to provide a refined approximation to the null distribution of test statistics when convergence to the asymptotic distribution is slow. Consider the $t$ statistic on the intercept:

$$t_1^0 = \sqrt{T} \left( \hat{\beta}^{(1),0} - \beta_0^{(1)} \right) / \sqrt{\hat{J}_T^{(1,1),0}}.$$  

A two-sided $t$-tests rejects $H_0$ at level $\alpha \in (0, 1)$ if $t_1^0 < cv_{\alpha/2}$ or $t_1^0 > cv_{1-\alpha/2}$, where $cv_{\alpha}$ is a $\alpha$ level critical value. We generate data from the following DGP: $y_t = \varepsilon_t$, where $\varepsilon_t = 0.5\varepsilon_{t-1} + u_t$ with $u_t \sim \mathcal{N}(0, 1)$. We compute the $t$ statistic $t_1^0$ where $\hat{\beta}^{(1),0} = T^{-1} \sum_{t=1}^T y_t, \beta_0^{(1)} = 0$ and $\hat{J}_T^{(1,1),0}$ is an HAC estimator of the long-run variance of $\{\varepsilon_t\}$. We repeat this for 100,000 times and obtain the empirical quantiles $cv_{\alpha/2}$ and $cv_{1-\alpha/2}$ of $t_1^0$. Different $\hat{J}_T$ (e.g., using different kernels $K_1$) give rise to different critical values. The critical values for $t$-tests are tabulated in the Appendix. The advantage of using this procedure as opposed to just using the
asymptotic critical values (e.g., $[-1.96, 1.96]$ for a two-sided tests with $\alpha = 0.05$) is that it can help to make the null rejection rates closer to the nominal level in small-samples. More complex procedures will be considered in future work. We remark that the simulation-assisted method is applied only to the $t$-tests on the intercept in the linear model for the HAC estimator that use predetermined bandwidths. The other HAR inference tests considered in this chapter (e.g., $t_2$-test $F$-test, Diebold-Mariano test, etc.) do not need any adjustment as they control the size well with the asymptotic critical values.

1.8.3 Empirical Power of HAR Inference Tests

Model P1-P4 correspond to model S1-S4 (i.e., $t_1$ for S1-S2 and $t_2$ for S3-S4), respectively. Model P5 (for $t_2$-test) corresponds to model S5 and involves misspecification via a smooth change in the coefficient $\beta_2$ toward the end of the sample. This situation is very common in practice and it is motivated by the model for the variable “cay” from Bianchi et al. (2018) (cf. figure 4 in their paper).$^{12}$ The model is given by

$$y_t = \beta_0^{(1)} + \delta + \left(\beta_0^{(2)} + \rho_t \mathbf{1} \{t \geq 4.5T/5\}\right)x_t + e_t, \quad t = 1, \ldots, T,$$

where $\rho_t = 4\delta (t - 4.5T/5) / T$, $e_t = 0.2e_{t-1} + u_t$, $u_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$, and $x_t = 1 + 0.2x_{t-1} + u_{X,t}$, $u_{X,t} \sim \text{i.i.d.} \mathcal{N}(0, 1)$. For $t > 4.5T/5$ the coefficient on $x_t$ increases slowly in small increments of magnitude $\rho_t$. We do not report the power results for model S6 because all the classical HAC estimators were associated with oversized tests and thus it is difficult to compare the power. Model P6 (for $t_1$) involves an

$^{12}$See also model M6 in the supplement where the misspecification is in the intercept as in their application. It should be noted that the authors do not essentially conduct HAR inference in their paper and so the issues shown here do not have any consequence for their analysis.
excluded relevant regressor \( w_t \),

\[
y_t = \beta^{(1)}_0 + \delta + \beta^{(2)}_0 x_t + \beta^{(3)}_0 w_t \mathbb{1} \{ t \geq 4T/5 \} + e_t, \quad t = 1, \ldots, T,
\]

where \( e_t = 0.6e_{t-1} + u_t, \ u_t \sim \text{i.i.d. } \mathcal{N}(0, 1) \), \( x_t = 0.2x_{t-1} + u_{X,t}, \ u_{X,t} \sim \text{i.i.d. } \mathcal{N}(0, 1) \), \( w_t = 3 + 0.2w_{t-1} + u_{W,t}, \ u_{W,t} \sim \text{i.i.d. } \mathcal{N}(0, 1) \), and \( \beta^{(3)}_0 = 1.13 \). For model P1-P6 we report the values of the power in Table 1.7-1.12. The sample size is \( T = 200 \). We set \( \beta^{(1)}_0 = 0 \) and \( \beta^{(2)}_0 = 1 \) in model P1-P6. Power functions for the Diebold-Maraino and for the forecast breakdown test are presented next.

In Table 1.7-1.8, Andrews’s (1991) and Newey and West’s (1987) HAC estimators with asymptotic critical values lead to \( t \)-tests that have good power and similar to the power of the \( \tilde{J}_T \) HAC-based \( t \)-tests. However, the power is substantially lower for the HAC estimators that use the fixed-\( b \) critical values (cf. the last two rows in each table). Newey-West (1987) HAC with fixed-\( b \) and EWP display evident power losses presumably due to the fact that they are often undersized as we showed above. Their power can be about one half lower than the power corresponding to the \( \tilde{J}_T \) HAC estimators. For example, in model P2 with \( \delta = 0.2 \), the power of Newey-West (1987) with fixed-\( b \) critical value is 0.099 while it is 0.201 for the \( \tilde{J}_T \) HAC with automatic bandwidth. Also Andrews’s (1991) HAC estimators suffer of large power losses (see e.g. model M1 in the supplement). The power losses are not a special feature of the \( t \)-test on the intercept but they are also present for \( t_2 \) (cf. Table 1.9-1.11). In particular, the power is considerably lower when \( \delta = 0.2 \) for the HAC estimators that use the prewhitening in model P4.

The results become most striking when there is some misspecification in the linear model. In Table 1.11-1.12, the power of all traditional HAC-based \( t \)-tests (both

\[\text{[13] The size for this model is reported in the first column of the table.}\]
for \( t_1 \) and \( t_2 \) is much lower than the power of the \( \hat{J}_T \) HAC-based tests. For example, in Table 1.12 with \( \delta = 0.2 \), the power of the test associated to the \( \hat{J}_T \) HAC with automatic bandwidth is 0.795 which is about 40 percentage points higher than the power of Andrews’s (1991) HAC estimator with automatic bandwidth and about 60-70 percentage points larger than the traditional HAC estimators with fixed-\( b \) critical values (cf. the last two rows of the table). The power losses persist as we raise \( \delta \). Interestingly, the loss in power becomes more severe as the mean of the excluded regressor increases (not reported). This feature highlights a severe issue with the traditional HAC estimator which we discuss more in detail below. The same problem extends to the \( F \)-test as it can be easily seen from Table A.6 in the supplement; see also Figure A.1-A.2 in the supplement which plot the power functions for \( t_1 \) and \( F \)-test, respectively, for models similar to P5.

Next, let us move to the evaluation of the power properties of the \( t \)-tests used in the forecasting literature. We begin with the Diebold-Mariano test. For this test, the separation between the null and alternative hypothesis does not depend on the value of a single parameter. Thus, the data-generating mechanism is different from the one under the null. The two competing forecast models are as follows: the first model uses the actual true data-generating process while the second model differs in that in place of \( x^0_{t-1} \) it uses \( x^2_{t-1} = x^0_{t-1} + u_{X_2,t} \) for \( t \leq 3T/4 \) and \( x^2_{t-1} = \delta + x^0_{t-1} + u_{X_2,t} \) for \( t > 3T/4 \), with \( u_{X_2,t} \sim \text{i.i.d. } N(0, 1) \). Evidently, the null hypotheses of equal predictive ability should be rejected by the Diebold-Mariano test whenever \( \delta > 0 \). Table 1.13 reports the power for several values of \( \delta \). When \( \delta = 1.5 \), no matter whether one uses the Bartlett or QS spectral kernel, the classical HAC estimators have very low power relative to the \( \hat{J}_T \) HAC estimators. The largest departure is about two times lower power. When \( \delta = 3 \) the same pattern arises though the power differences
are smaller. It appears that the $\hat{J}_T$ HAC estimator with QS kernel and automatic bandwidth performs slightly worse than with predetermined bandwidth. As we raise $\delta$ to 6, all tests, with the exception of Andrews’s (1991) HAC estimator with automatic bandwidth, show very high power, above 90%. As we increase $\delta$ further, the $\hat{J}_T$ HAC estimators lead to $t$-tests that first attain and then maintain unit power. In contrast, the tests standardized by the HAC estimators of Andrews (1991) and Newey and West (1987) display non-monotonic power gradually converging to zero. Newey-West (1997) HAC with predetermined bandwidth and EWP are associated to $t$-tests that have monotonic power, though the power is lower than that with $\hat{J}_T$ HAC estimators even if we do not take into account that the former can be oversized as discussed above.

Finally, we move to the $t$-test of Giacomini and Rossi (2009). The data-generating process under $H_1 : \mathbb{E} \left( S \right) \neq 0$ is given by

$$y_t = 1 + x_{t-1} + \delta x_{t-1} \mathbb{1} \left\{ t > T_1^0 \right\} + e_t,$$

where $x_{t-1} \sim \text{i.i.d. } \mathcal{N} (0, 1.4)$, $e_t = 0.4 e_{t-1} + u_t$, $u_t \sim \text{i.i.d. } \mathcal{N} (0, 1)$ and $T_1^0 = T \lambda_1^0$ with $\lambda_1^0 = 0.9$. Under this specification there is a break in the coefficient associated to the predictor $x_{t-1}$. Thus, there is a forecast instability or failure as defined in Casini (2018b) and the test of Giacomini and Rossi (2009) should reject $H_0$. Figure 1.1-1.3 plot the power functions for $T = 200, 400$ and $800$, respectively. From the plots it appears that all versions of the classical HAC estimators of Andrews (1991) and Newey and West (1987) lead to $t$-tests that have, essentially, zero power for all $\delta$. In contrast, the $t$-test standardized by the $\hat{J}_T$ HAC estimators proposed in this chapter have good power. Among the latter HAC estimators, no one prevails on the others. The failure of the classical HAC estimators cannot be attributed to the sample size.
because as we raise $T$ to 400 or 800, the tests still display no power with exception of the one with Andrews’s (1991) HAC estimator which enjoys some increase in power. However, even the latter $t$-test cannot be said to perform well since when the tests using the $\hat{J}_T$ HAC estimators achieve unit power the test using Andrews’s (1991) HAC estimator with prewhitening has only about 10% power when $T = 400$.

The failure of the classical HAC estimators when used as standardizations of $t$-tests or $F$-tests occurring in some of the data-generating mechanisms reported here can be simply reconciled with the fact that in such data-generating mechanisms the spectrum of $V_t$ is not constant. In other words, the covariance structure of $V_t$ depends not only on the lag order but also on $t$. The main theory of Andrews (1991) and Newey and West (1987) does not allow for such a feature. Classical HAC estimators estimate an average of a time-varying spectrum. Because of this instability in the spectrum, classical HAC estimators overestimate the extent of the dependence or variation in $V_t$. This reconciles with a well-known result in the unit root literature where tests for unit root struggle to reject the unit root hypotheses if a process is second-order stationary (i.e., no unit root) but it is contaminated by breaks in the mean or trend [cf. Perron (1989) and Perron (1990)]. Similarly, theoretical results in the long-memory literature documents that a short-memory sequence contaminated by structural breaks can approximate a long-memory series in the sense that the autocorrelation function has the same properties [cf. Diebold and Inoue (2001), Granger and Hyung (2004), Hillebrand (2005) and Mikosch and Stărica (2004)]. That is, parameter variation makes the series appear much more persistent. As a consequence, HAC standard errors are too large and when used as normalization of test statistics, the tests have little or no power. In contrast, the theory of evolutionary spectra and the associated HAC estimation method, does not face this difficulty and HAR
inference can be safely conducted in the usual way.

1.9 Conclusions

Economic time series are highly nonstationary. Methods constructed under the assumption of stationarity might then have undesirable properties. Both applied and theoretical works involving economic time series should make more effort to account for nonstationarity. This chapter has developed a theoretical framework for inference in settings where the data may be nonstationary. A class of nonstationary processes that have a time-varying spectral representation is introduced. This class is then used as a building block for a new theory of heteroskedasticity- and autocorrelation-robust (HAR) inference valid in nonstationary environments and/or when models are misspecified. A new heteroskedasticity and autocorrelation consistent (HAC) estimator is presented. In addition to the usual smoothing procedure over autocovariance lag order—akin to the classical HAC estimators—the estimator applies a smoothing procedure over time. This is crucial in order to account properly for the variation over time of the structural properties of the economic time series and the noise associated with the model. Optimality results under MSE criterion concerning bandwidths and kernels have been established. A data-dependent method based on the “plug-in” approach has been proposed. A Monte Carlo study has showed the benefits of the proposed approach. In particular, there are empirical relevant circumstances where usual $t$-tests, either in linear regression models or in other econometric contexts, standardized by classical HAC estimators perform poorly. These may result in size distortions as well as significant power losses, even when the sample size is large. In contrast, when the proposed HAC estimator is used the same $t$-tests do not suffer from those issues and inference is then more reliable. An empirical application on
the detection of changes in the predictive ability of the Phillips curve for inflation, which is included in the supplement, shows that HAR inference based on classical HAC estimates leads to misleading conclusions which are difficult to reconcile with the findings documented in the empirical literature.

1.10 Appendix to Chapter 1

1.10.1 Simulation-Assisted Critical Values

Table 1.1: Simulation assisted critical values

<table>
<thead>
<tr>
<th></th>
<th>( \alpha = 0.05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{J}_T ), Bartlett kernel</td>
<td>[-2.9122, 2.9164]</td>
</tr>
<tr>
<td>( \hat{J}_T ), QS kernel</td>
<td>[-2.7693, 2.7649]</td>
</tr>
</tbody>
</table>

1.10.2 Tables

Table 1.2: Empirical small-sample size of \( t_1 \)-test for model S1

<table>
<thead>
<tr>
<th></th>
<th>( T = 125 )</th>
<th>( T = 200 )</th>
<th>( T = 400 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J_T ), Bartlett kernel, no simulation-assisted cv</td>
<td>0.119</td>
<td>0.091</td>
<td>0.077</td>
</tr>
<tr>
<td>( \hat{J}_T ), QS kernel, no simulation-assisted cv</td>
<td>0.125</td>
<td>0.107</td>
<td>0.099</td>
</tr>
<tr>
<td>( J_T ), Bartlett kernel</td>
<td>0.065</td>
<td>0.054</td>
<td>0.053</td>
</tr>
<tr>
<td>( \hat{J}_T ), QS kernel</td>
<td>0.082</td>
<td>0.077</td>
<td>0.083</td>
</tr>
<tr>
<td>( \hat{J}_T ), QS kernel, auto, no breaks</td>
<td>0.065</td>
<td>0.064</td>
<td>0.055</td>
</tr>
<tr>
<td>( \hat{J}_T ), QS kernel, auto</td>
<td>0.042</td>
<td>0.073</td>
<td>0.052</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.112</td>
<td>0.089</td>
<td>0.075</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.075</td>
<td>0.063</td>
<td>0.060</td>
</tr>
<tr>
<td>Newey-West (1987), “rule”</td>
<td>0.158</td>
<td>0.118</td>
<td>0.110</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.201</td>
<td>0.158</td>
<td>0.148</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.096</td>
<td>0.079</td>
<td>0.060</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-b</td>
<td>0.061</td>
<td>0.064</td>
<td>0.049</td>
</tr>
<tr>
<td>EWP</td>
<td>0.051</td>
<td>0.055</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Table 1.3: Empirical small-sample size of \( t_1 \)-test for model S2

<table>
<thead>
<tr>
<th>( \alpha = 0.05 )</th>
<th>( T = 125 )</th>
<th>( T = 200 )</th>
<th>( T = 400 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J_T ), Bartlett kernel</td>
<td>0.039</td>
<td>0.053</td>
<td>0.055</td>
</tr>
<tr>
<td>( J_T ), QS kernel</td>
<td>0.047</td>
<td>0.094</td>
<td>0.090</td>
</tr>
<tr>
<td>( J_T ), QS kernel, auto, no breaks</td>
<td>0.052</td>
<td>0.088</td>
<td>0.083</td>
</tr>
<tr>
<td>( J_T ), QS kernel, auto</td>
<td>0.068</td>
<td>0.065</td>
<td>0.075</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.083</td>
<td>0.078</td>
<td>0.092</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.101</td>
<td>0.121</td>
<td>0.169</td>
</tr>
<tr>
<td>Newey-West (1987), “rule”</td>
<td>0.085</td>
<td>0.094</td>
<td>0.128</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.102</td>
<td>0.094</td>
<td>0.114</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.085</td>
<td>0.067</td>
<td>0.089</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-( b )</td>
<td>0.020</td>
<td>0.014</td>
<td>0.005</td>
</tr>
<tr>
<td>EWP</td>
<td>0.031</td>
<td>0.022</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Table 1.4: Empirical small-sample size of \( t_2 \)-test for model S3-S4

<table>
<thead>
<tr>
<th>( \alpha = 0.05 )</th>
<th>( T = 125 )</th>
<th>( T = 200 )</th>
<th>( T = 400 )</th>
<th>( T = 125 )</th>
<th>( T = 200 )</th>
<th>( T = 400 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J_T ), Bartlett kernel</td>
<td>0.060</td>
<td>0.051</td>
<td>0.063</td>
<td>0.047</td>
<td>0.081</td>
<td>0.055</td>
</tr>
<tr>
<td>( J_T ), QS kernel</td>
<td>0.054</td>
<td>0.060</td>
<td>0.066</td>
<td>0.052</td>
<td>0.087</td>
<td>0.072</td>
</tr>
<tr>
<td>( J_T ), QS kernel, auto, no break</td>
<td>0.048</td>
<td>0.078</td>
<td>0.067</td>
<td>0.059</td>
<td>0.061</td>
<td>0.071</td>
</tr>
<tr>
<td>( J_T ), QS kernel, auto</td>
<td>0.064</td>
<td>0.062</td>
<td>0.052</td>
<td>0.046</td>
<td>0.056</td>
<td>0.061</td>
</tr>
<tr>
<td>Andrews (1991), auto,</td>
<td>0.040</td>
<td>0.043</td>
<td>0.024</td>
<td>0.037</td>
<td>0.032</td>
<td>0.011</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.022</td>
<td>0.000</td>
<td>0.005</td>
<td>0.003</td>
<td>0.002</td>
<td>0.006</td>
</tr>
<tr>
<td>Newey-West (1987), “rule”</td>
<td>0.042</td>
<td>0.029</td>
<td>0.037</td>
<td>0.007</td>
<td>0.016</td>
<td>0.042</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.038</td>
<td>0.032</td>
<td>0.038</td>
<td>0.012</td>
<td>0.016</td>
<td>0.041</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.030</td>
<td>0.003</td>
<td>0.012</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-( b )</td>
<td>0.016</td>
<td>0.015</td>
<td>0.007</td>
<td>0.005</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>EWP</td>
<td>0.022</td>
<td>0.009</td>
<td>0.016</td>
<td>0.007</td>
<td>0.015</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 1.5: Empirical small-sample size of \( t \)-tests for model S5-S6

<table>
<thead>
<tr>
<th>( \alpha = 0.05 )</th>
<th>( T = 125 )</th>
<th>( T = 200 )</th>
<th>( T = 400 )</th>
<th>( T = 125 )</th>
<th>( T = 200 )</th>
<th>( T = 400 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J_T ), Bartlett kernel</td>
<td>0.062</td>
<td>0.064</td>
<td>0.062</td>
<td>0.061</td>
<td>0.071</td>
<td>0.056</td>
</tr>
<tr>
<td>( J_T ), QS kernel</td>
<td>0.064</td>
<td>0.074</td>
<td>0.066</td>
<td>0.073</td>
<td>0.086</td>
<td>0.071</td>
</tr>
<tr>
<td>( J_T ), QS kernel, auto, no break</td>
<td>0.068</td>
<td>0.075</td>
<td>0.071</td>
<td>0.064</td>
<td>0.084</td>
<td>0.060</td>
</tr>
<tr>
<td>( J_T ), QS kernel, auto</td>
<td>0.055</td>
<td>0.045</td>
<td>0.045</td>
<td>0.074</td>
<td>0.068</td>
<td>0.051</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.052</td>
<td>0.042</td>
<td>0.051</td>
<td>0.164</td>
<td>0.143</td>
<td>0.112</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.054</td>
<td>0.056</td>
<td>0.056</td>
<td>0.124</td>
<td>0.114</td>
<td>0.097</td>
</tr>
<tr>
<td>Newey-West (1987), “rule”</td>
<td>0.052</td>
<td>0.046</td>
<td>0.052</td>
<td>0.160</td>
<td>0.142</td>
<td>0.137</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.044</td>
<td>0.040</td>
<td>0.044</td>
<td>0.178</td>
<td>0.158</td>
<td>0.140</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.054</td>
<td>0.050</td>
<td>0.054</td>
<td>0.124</td>
<td>0.108</td>
<td>0.089</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-( b )</td>
<td>0.022</td>
<td>0.026</td>
<td>0.022</td>
<td>0.110</td>
<td>0.110</td>
<td>0.078</td>
</tr>
<tr>
<td>EWP</td>
<td>0.038</td>
<td>0.034</td>
<td>0.038</td>
<td>0.108</td>
<td>0.111</td>
<td>0.098</td>
</tr>
</tbody>
</table>
Table 1.6: Empirical small-sample size for model S7-S8

<table>
<thead>
<tr>
<th>$\alpha = 0.05$</th>
<th>Model S7</th>
<th>Model S8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 125$</td>
<td>$T = 200$</td>
<td>$T = 400$</td>
</tr>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.101</td>
<td>0.088</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel</td>
<td>0.109</td>
<td>0.090</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.073</td>
<td>0.055</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.074</td>
<td>0.087</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.071</td>
<td>0.084</td>
</tr>
<tr>
<td>Newey-West (1987), “rule”</td>
<td>0.098</td>
<td>0.146</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.075</td>
<td>0.073</td>
</tr>
<tr>
<td>EWP</td>
<td>0.119</td>
<td>0.083</td>
</tr>
</tbody>
</table>

Table 1.7: Empirical small-sample rejection rates of $t_1$-test for model P1

<table>
<thead>
<tr>
<th>$\alpha = 0.05$, $T = 200$</th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
<th>$\delta = 1.6$</th>
<th>$\delta = 2.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.192</td>
<td>0.614</td>
<td>0.994</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel</td>
<td>0.212</td>
<td>0.675</td>
<td>0.997</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no break</td>
<td>0.223</td>
<td>0.648</td>
<td>0.994</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.253</td>
<td>0.625</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.246</td>
<td>0.652</td>
<td>0.992</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.242</td>
<td>0.680</td>
<td>0.993</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.194</td>
<td>0.582</td>
<td>0.978</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>EWP</td>
<td>0.174</td>
<td>0.530</td>
<td>0.975</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 1.8: Empirical small-sample rejection rates of $t_1$-test for model P2

<table>
<thead>
<tr>
<th>$\alpha = 0.05$, $T = 200$</th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
<th>$\delta = 1.6$</th>
<th>$\delta = 2.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.158</td>
<td>0.410</td>
<td>0.830</td>
<td>0.990</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel</td>
<td>0.214</td>
<td>0.484</td>
<td>0.861</td>
<td>0.994</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no break</td>
<td>0.208</td>
<td>0.480</td>
<td>0.857</td>
<td>0.990</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.201</td>
<td>0.495</td>
<td>0.888</td>
<td>0.985</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.238</td>
<td>0.473</td>
<td>0.859</td>
<td>0.982</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.284</td>
<td>0.572</td>
<td>0.870</td>
<td>0.988</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.235</td>
<td>0.524</td>
<td>0.853</td>
<td>0.985</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.099</td>
<td>0.373</td>
<td>0.782</td>
<td>0.965</td>
<td>0.995</td>
</tr>
<tr>
<td>EWP</td>
<td>0.107</td>
<td>0.370</td>
<td>0.796</td>
<td>0.966</td>
<td>0.998</td>
</tr>
</tbody>
</table>
Table 1.9: Empirical small-sample rejection rates of $t_2$-test for model P3

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
<th>$\delta = 1.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.290</td>
<td>0.648</td>
<td>0.968</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel</td>
<td>0.278</td>
<td>0.628</td>
<td>0.970</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no breaks</td>
<td>0.281</td>
<td>0.756</td>
<td>0.963</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.245</td>
<td>0.755</td>
<td>0.973</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.384</td>
<td>0.766</td>
<td>0.932</td>
<td>0.999</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.120</td>
<td>0.502</td>
<td>0.888</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.220</td>
<td>0.536</td>
<td>0.884</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.212</td>
<td>0.650</td>
<td>0.908</td>
<td>0.998</td>
</tr>
<tr>
<td>EWP</td>
<td>0.230</td>
<td>0.664</td>
<td>0.904</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Table 1.10: Empirical small-sample rejection rates of $t_2$-test for model P4

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
<th>$\delta = 1.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.312</td>
<td>0.705</td>
<td>0.961</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel</td>
<td>0.332</td>
<td>0.689</td>
<td>0.964</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no breaks</td>
<td>0.422</td>
<td>0.734</td>
<td>0.972</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.456</td>
<td>0.755</td>
<td>0.975</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.445</td>
<td>0.751</td>
<td>0.935</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.140</td>
<td>0.554</td>
<td>0.846</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.175</td>
<td>0.577</td>
<td>0.895</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.255</td>
<td>0.632</td>
<td>0.891</td>
<td>0.999</td>
</tr>
<tr>
<td>EWP</td>
<td>0.295</td>
<td>0.657</td>
<td>0.895</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Table 1.11: Empirical small-sample rejection rates of $t_2$-test for model P5

<table>
<thead>
<tr>
<th></th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.6$</th>
<th>$\delta = 0.8$</th>
<th>$\delta = 1.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.582</td>
<td>0.852</td>
<td>0.975</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td>$J_T$, QS kernel</td>
<td>0.561</td>
<td>0.842</td>
<td>0.981</td>
<td>0.997</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no break</td>
<td>0.564</td>
<td>0.855</td>
<td>0.975</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.552</td>
<td>0.856</td>
<td>0.976</td>
<td>0.997</td>
<td>0.999</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.501</td>
<td>0.825</td>
<td>0.973</td>
<td>0.996</td>
<td>0.999</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.531</td>
<td>0.831</td>
<td>0.972</td>
<td>0.994</td>
<td>0.996</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.510</td>
<td>0.832</td>
<td>0.972</td>
<td>0.983</td>
<td>0.994</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.285</td>
<td>0.557</td>
<td>0.556</td>
<td>0.614</td>
<td>0.741</td>
</tr>
<tr>
<td>EWP</td>
<td>0.360</td>
<td>0.655</td>
<td>0.692</td>
<td>0.774</td>
<td>0.884</td>
</tr>
</tbody>
</table>
Table 1.12: Empirical small-sample rejection rates of $t_1$-test for model P6

<table>
<thead>
<tr>
<th>$\alpha = 0.05$, $T = 200$</th>
<th>$\delta = 0$</th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
<th>$\delta = 1.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.047</td>
<td>0.800</td>
<td>0.970</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel</td>
<td>0.079</td>
<td>0.870</td>
<td>0.975</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel, auto, no break</td>
<td>0.067</td>
<td>0.645</td>
<td>0.975</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel, auto</td>
<td>0.055</td>
<td>0.685</td>
<td>0.945</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.071</td>
<td>0.360</td>
<td>0.775</td>
<td>0.995</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.049</td>
<td>0.435</td>
<td>0.770</td>
<td>0.995</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.067</td>
<td>0.414</td>
<td>0.735</td>
<td>0.985</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-b</td>
<td>0.053</td>
<td>0.085</td>
<td>0.320</td>
<td>0.915</td>
<td>1.000</td>
</tr>
<tr>
<td>EWP</td>
<td>0.054</td>
<td>0.095</td>
<td>0.370</td>
<td>0.925</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 1.13: Empirical power of the DM (1995) test

<table>
<thead>
<tr>
<th>$\alpha = 0.05$, $T = 400$</th>
<th>$\delta = 1.5$</th>
<th>$\delta = 3$</th>
<th>$\delta = 6$</th>
<th>$\delta = 10$</th>
<th>$\delta = 15$</th>
<th>$\delta = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.682</td>
<td>0.805</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel</td>
<td>0.652</td>
<td>0.793</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel, auto</td>
<td>0.525</td>
<td>0.745</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.495</td>
<td>0.630</td>
<td>0.396</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Andrews (1991), auto, pre</td>
<td>0.520</td>
<td>0.725</td>
<td>0.977</td>
<td>0.404</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>Newey-West (1987), “rule”</td>
<td>0.620</td>
<td>0.750</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.355</td>
<td>0.572</td>
<td>0.924</td>
<td>0.670</td>
<td>0.623</td>
<td>0.596</td>
</tr>
<tr>
<td>Newey-West (1987), auto, pre</td>
<td>0.490</td>
<td>0.685</td>
<td>0.943</td>
<td>0.313</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>EWP</td>
<td>0.345</td>
<td>0.610</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
1.10.3 Figures

**Figure 1.1:** Power functions of forecast breakdown $t$-test for Model M8 with $T = 200$.

**Figure 1.2:** Power functions of forecast breakdown $t$-test for Model M8 with $T = 400$. 
Figure 1.3: Power functions of forecast breakdown $t$-test for Model M8 with $T = 800$. 
Chapter 2

Continuous Record Asymptotics for Structural Change Models$^1$

2.1 Introduction

Parameter instability in linear regression models is a common problem and more so when the span of the data is large. In the context of a partial structural change in a linear regression model with a single break point, we develop a continuous record asymptotic framework and inference methods for the break date. Our model is specified in continuous time but estimated with discrete-time observations using a least-squares method. We have $T$ observations with a sampling frequency $h$ over a fixed time horizon $[0, N]$, where $N = Th$ denotes the time span of the data. We consider a continuous record asymptotic framework whereby $T$ increases by shrinking the time interval $h$ to zero while keeping time span $N$ fixed. We impose very mild conditions on an underlying continuous-time model assumed to generate the data, basically continuous Itô semimartingales. Using an infill asymptotic setting, the uncertainty about the unknown parameters is assessed from the sample paths of the processes, which differs from the standard large-$N$ asymptotics whereby it is assessed from features of the distributions or moments of the processes. This allows us to impose mild pathwise regularity conditions and to avoid any ergodic or weak-dependence assumption. Our setting includes most linear models considered in the

\footnote{$^1$This chapter is based on joint work with Pierre Perron at Boston University}
structural change literature based on large-$N$ asymptotics, which essentially involve processes satisfying some form of mixing conditions.

An extensive amount of research addressed structural change problems under the classical large-$N$ asymptotics. Early contributions are Hinkley (1971), Bhattacharya (1987), and Yao (1987), who adopted a Maximum Likelihood (ML) approach, and for linear regression models, Bai (1997), Bai and Perron (1998) and Perron and Qu (2006). Qu and Perron (2007) generalized this work by considering multivariate regressions. Extensions to models with endogenous regressors were considered by Perron and Yamamoto (2014) [see also Hall et al. (2010)], though Perron and Yamamoto (2015) argue that standard least-squares methods are still applicable, and indeed preferable, in such cases. Notable also are the contributions on testing for structural changes by Hawkins (1977), Picard (1985), Kim and Siegmund (1989), Andrews (1993), Horváth (1993), Andrews and Ploberger (1994) and Bai and Perron (1998), among others. See the reviews of Csörgő and Horváth (1997), Perron (2006) and references therein. In this literature, the resulting large-$N$ limit theory for the estimate of the break date depends on the exact distribution of the regressors and disturbances. Therefore, a so-called shrinkage asymptotic theory was adopted whereby the magnitude of the shift converges to zero as $T$ increases, which leads to a limit distribution invariant to the distributions of the regressors and errors.

We study a general change-point problem under a continuous record asymptotic framework and develop inference procedures based on the derived asymptotic distribution. As $h \downarrow 0$, identification of the break point translates to the detection of a change in the slope coefficients for the continuous local martingale part of locally square-integrable semimartingales. We establish consistency at rate-$T$ convergence for the least-squares estimate of the break date, assumed to occur at time $N_h^0$. Given
the fast rate of convergence, we introduce a limit theory with shrinking magnitudes of shifts and increasing variance of the residual process local to the change-point. The asymptotic distribution corresponds to the location of the extremum of a function of the (quadratic) variation of the regressors and of a Gaussian centered martingale process over some time interval. The properties of this limit theory, in particular how the magnitude of the shift and how the span versus the sample size affect the precision of the break date estimate are then discussed. The knowledge of such features of the distribution of the estimator is important from a theoretical perspective and cannot be gained from the classical large-$N$ asymptotics. It is also very useful to provide guidelines as to the proper method to use to construct confidence sets.

Our continuous record limit distribution is characterized by some notable aspects. With the time horizon $[0, N]$ fixed, we can account for the asymmetric informational content provided by the pre- and post-break sample observations, i.e., the time span and the position of the break date $N^0_b$ convey useful information about the finite-sample distribution. In contrast, this is not achievable under the large-$N$ shrinkage asymptotic framework because both pre- and post-break segments expand proportionately as $T$ increases and, given the mixing assumptions imposed, only the neighborhood around the break date remains relevant. Furthermore, the domain of the extremum depends on the position of the break point $N^0_b$ and therefore the distribution is asymmetric, in general. The degree of asymmetry increases as the true break point moves away from mid-sample. This holds unless the magnitude of the break is large, in which case the density is symmetric irrespective of the location of the break. This accords with simulation evidence which documents that in small samples, the break point estimate is less precise and the coverage rates of the confidence intervals less reliable when the break is not at mid-sample. These results
are natural consequences of our continuous record asymptotic theory, which indicate that the time span, location and magnitude of the break and statistical properties of the errors and regressors all jointly play a primary role in shaping the limit distribution of the break date estimator. For example, when the shift magnitude is small, the probability density displays three modes. As the shift magnitude increases, this tri-modality vanishes.\(^2\)

Furthermore, unless the magnitude of the break is large, the asymptotic distribution is symmetric only if both: (i) the break date is located at mid-sample, (ii) the distribution of the errors and regressors do not differ “too much” across regimes. Given the fixed-span setting, our limit theory treats the volatility of the regressors and errors as random quantities. We thus use the concept of stable convergence in distribution. As for the impact of the sample size relative to the span of the data on the precision of the estimate, we find that the span plays a more pronounced role. We also show, via simulations, that our continuous record asymptotics provides good approximations to the finite-sample distributions of the estimate of the break date.

Our continuous record asymptotic theory is not limited to providing a better approximation to the finite-sample distribution. It can also be exploited to address the problem of conducting inference about the break date. This issue has received considerable attention. Besides the original asymptotic arguments used by Bai (1997) and Bai and Perron (1998), Elliott and Müller (2007) proposed to invert Nyblom’s (1990) statistic, while Eo and Morley (2015) introduced a procedure based on the likelihood-ratio statistic of Qu and Perron (2007). The latter methods were mainly motivated by the fact that the empirical coverage rates of the confidence intervals

---

\(^2\)In work that we became aware of after the first draft of this chapter, Jiang et al. (2018) studied the finite-sample bias of a break point estimator based on maximum likelihood for a simple univariate diffusion with constant volatility and a change-point in the drift. They also find that the span can be important and the distribution can be asymmetric. We comment on the differences in the Appendix.
obtained from Bai’s (1997) method are below the nominal level with small breaks. The method of Elliott and Müller (2007) delivers the most accurate coverage rates, though at the expense of increased average lengths of the confidence sets especially with large breaks [cf. Chang and Perron (2018)]. What is still missing is a method that, uniformly over break magnitudes, achieves both accurate coverage rates and satisfactory average lengths of the confidence sets for a wide range of data-generating processes. Given the peculiar properties of the continuous record asymptotic distribution, we propose an inference method which is rather non-standard and relates to Bayesian analyses. We use the concept of Highest Density Region to construct confidence sets for the break date. Our method is simple to implement and has a frequentist interpretation.

The simulation analysis conducted indicates that our approach has two notable properties. First, it provides adequate empirical coverage rates over all data-generating mechanisms considered and, importantly, for any size and/or location of the break, a notoriously difficult problem. Second, the lengths of the confidence sets are always shorter than those obtained using Elliott and Müller’s (2007) approach. Often, the reduction in length is substantial and increases with the size of the break. Also, our method performs markedly better when lagged dependent variables are present in the model. Compared to Bai’s (1997) method, our approach yields better coverage rates, especially when the magnitude of the break is small. With large breaks, the two methods are basically equivalent. Of particular interest is the fact that our confidence set can be the union of disjoint intervals. This is illustrated in Section 4.4.

The chapter is organized as follows. Section 2.2 introduces the model, the estimation method and extensions to predictable processes. Section 2.3 contains results
about the consistency and rate of convergence for fixed shifts. Section 2.4 develops the asymptotic theory. We compare our limit theory with the finite-sample distribution in Section 2.5. Section 4.4 describes how to construct the confidence sets. Simulation results about its adequacy are reported in Section 2.7. Section 4.7 provides brief concluding remarks. Additional details and some proofs for the main results are included in an appendix. The Supplement contains most of the proofs as well as additional materials.

2.2 Model and Assumptions

Section 2.2.1 introduces the benchmark model of interest, the main assumptions, the estimation method and the relation of our setup with the traditional large-$N$ asymptotic framework. In Section 2.2.2 we extend the benchmark model to include predictable processes. The following notations are used throughout. Recall the relation $N = Th$. We shall use $T \to \infty$ and $h \downarrow 0$ interchangeably. All vectors are column vectors. For two vectors $a$ and $b$, we write $a \leq b$ if the inequality holds component-wise. We denote the transpose of a matrix $A$ by $A'$ and the $(i, j)$ elements of $A$ by $A^{(i,j)}$. For a sequence of matrices $\{A_T\}$, we write $A_T = o_p(1)$ if each of its elements is $o_p(1)$ and likewise for $O_p(1)$. $\mathbb{R}$ denotes the set of real numbers. We use $\|\cdot\|$ to denote the Euclidean norm of a linear space, i.e., $\|x\| = \left(\sum_{i=1}^{p} x_i^2\right)^{1/2}$ for $x \in \mathbb{R}^p$. We use $\lfloor \cdot \rfloor$ to denote the largest smaller integer function and for a set $A$, the indicator function of $A$ is denoted by $1_A$. The symbol $\otimes$ denotes the product of $\sigma$-fields. A sequence $\{u_{kh}\}_{k=1}^{T}$ is i.i.d. (resp., i.n.d) if the $u_{kh}$ are independent and identically (resp., non-identically) distributed. We use $P \Rightarrow$, $\Rightarrow$, and $\Rightarrow^s$ to denote convergence in probability, weak convergence and stable convergence in law, respectively. For semimartingales $\{S_t\}_{t \geq 0}$ and $\{R_t\}_{t \geq 0}$, we denote their covariation process by $[S, R]_t$. 


and their predictable counterpart by \( (S, R)_t \). The symbol "\( \triangleq \)" denotes definitional equivalence. Finally, note that in general \( N \) is not identified and could be normalized to one. However, we keep a generic \( N \) throughout to allow a better intuitive understanding of the results.

### 2.2.1 The Benchmark Model

We consider the following partial structural change model with a single break point:

\[
Y_t = D_t' \pi^0 + Z_t' \delta^0_1 + e_t, \quad (t = 0, 1, \ldots, T_0^b) \tag{2.2.1}
\]

\[
Y_t = D_t' \pi^0 + Z_t' \delta^0_2 + e_t, \quad (t = T_0^b + 1, \ldots, T),
\]

where \( Y_t \) is the dependent variable, \( D_t \) and \( Z_t \) are, respectively, \( q \times 1 \) and \( p \times 1 \) vectors of regressors and \( e_t \) is an unobservable disturbance. The vector-valued parameters \( \pi^0, \delta^0_1 \) and \( \delta^0_2 \) are unknown with \( \delta^0_1 \neq \delta^0_2 \). Our main purpose is to develop inference methods for the unknown break date \( T_0^b \) when \( T + 1 \) observations on \((Y_t, D_t, Z_t)\) are available. Before moving to the re-parametrization of the model, we discuss the underlying continuous-time model assumed to generate the data. The processes \( \{D_s, Z_s, e_s\}_{s \geq 0} \) are continuous-time processes, defined on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \geq 0}, P)\), where \( s \) can be interpreted as the continuous-time index. We observe realizations of \((Y_s, D_s, Z_s)\) at discrete points of time. Below, we impose very minimal "pathwise" assumptions on these continuous-time stochastic processes which imply mild restrictions on the observed discrete-time counterparts. We discuss what these assumptions imply for our model and the distributional properties of the errors and regressors.

The sampling occurs at regularly spaced time intervals of length \( h \) within a fixed
time horizon \([0, N]\) where \(N\) denotes the span of the data. We observe

\[
\{hY_k, hD_k, hZ_k; k = 0, 1, \ldots, T = N/h\}.
\]

\(hD_k \in \mathbb{R}^q\) and \(hZ_k \in \mathbb{R}^p\) are random vector step functions which jump only at times 0, \(h\), \(2h\), \(Th\). We shall allow \(hD_k\) and \(hZ_k\) to include both predictable processes and locally-integrable semimartingales, though the case with predictable regressors is more delicate and discussed in Section 2.2.2. Recall the Doob-Meyer decomposition [cf. Doob (1953) and Meyer (1967)]\(^3\) from which it follows that any locally-integrable semimartingale process can be decomposed into a “predictable” and a “martingale” part. The discretized processes \(hD_k\) and \(hZ_k\) are assumed to be adapted to the increasing and right-continuous filtration \(\mathcal{F}_t\) \(t \geq 0\). For any process \(X\) we denote its “increments” by \(\Delta hX_k = X_k - X_{k-1}\). For \(k = 1, \ldots, T\), let \(\Delta hD_k \triangleq \mu_{D,k} h + \Delta hM_{D,k}\) and \(\Delta hZ_k \triangleq \mu_{Z,k} h + \Delta hM_{Z,k}\) where the “drifts” \(\mu_{D,t} \in \mathbb{R}^q, \mu_{Z,t} \in \mathbb{R}^p\) are \(\mathcal{F}_t\)-measurable (exact assumptions will be given below), and \(M_{D,k} \in \mathbb{R}^q, M_{Z,k} \in \mathbb{R}^p\) are continuous local martingales with finite conditional covariance matrix \(P\)-a.s.,

\[
\mathbb{E}\left(\Delta hM_{D,t}\Delta hM'_{D,t} | \mathcal{F}_{t-h}\right) = \Sigma_{D,t-h} \Delta t \quad \text{and} \quad \mathbb{E}\left(\Delta hM_{Z,t}\Delta hM'_{Z,t} | \mathcal{F}_{t-h}\right) = \Sigma_{Z,t-h} \Delta t \quad (\Delta t \quad \text{and} \quad h \quad \text{are used interchangeably}).
\]

Let \(\lambda_0 \in (0, 1)\) denote the fractional break date (i.e., \(T^0_b = \lfloor T \lambda_0 \rfloor\)). Via the Doob-Meyer Decomposition, model (2.2.1) can be expressed as

\[
\Delta hY_k \triangleq \begin{cases} 
(\Delta hD_k)' \pi^0 + (\Delta hZ_k)' \delta^0_{Z,1} + \Delta h\varepsilon_k^*, & (k = 1, \ldots, \lfloor T \lambda_0 \rfloor) \\
(\Delta hD_k)' \pi^0 + (\Delta hZ_k)' \delta^0_{Z,2} + \Delta h\varepsilon_k^*, & (k = \lfloor T \lambda_0 \rfloor + 1, \ldots, T) 
\end{cases}, \quad (2.2.2)
\]

where the error process \(\{\Delta h\varepsilon_t^*, \mathcal{F}_t\}\) is a continuous local martingale difference sequence with conditional variance \(\mathbb{E}\left[(\Delta h\varepsilon_t^*)^2 | \mathcal{F}_{t-h}\right] = \sigma_{\varepsilon,t-h}^2 \Delta t \quad P\)-a.s. finite. The

---

\(^3\)A treatment of the probabilistic material can be found in Aït-Sahalia and Jacod (2014), Karatzas and Shreve (1996), Protter (2005), Jacod and Shiryaev (2003) and Jacod and Protter (2012). For measure theoretical aspects we refer to Billingsley (1995).
underlying continuous-time data-generating process can thus be represented (up to $P$-null sets) in integral equation form as

$$
D_t = D_0 + \int_0^t \mu_{D,s} ds + \int_0^t \sigma_{D,s} dW_{D,s}, \quad Z_t = Z_0 + \int_0^t \mu_{Z,s} ds + \int_0^t \sigma_{Z,s} dW_{Z,s},
$$

(2.2.3)

where $\sigma_{D,t}$ and $\sigma_{Z,t}$ are the instantaneous covariance processes taking values in $\mathcal{M}_q^{\text{càdlàg}}$ and $\mathcal{M}_p^{\text{càdlàg}}$ [the space of $p \times p$ positive definite real-valued matrices whose elements are càdlàg]; $W_D$ (resp., $W_Z$) is a $q$ (resp., $p$)-dimensional standard Wiener process; $e^* = \{e^*_t\}_{t \geq 0}$ is a continuous local martingale which is orthogonal (in a martingale sense) to $\{D_t\}_{t \geq 0}$ and $\{Z_t\}_{t \geq 0}$; and $D_0$ and $Z_0$ are $\mathcal{F}_0$-measurable random vectors.

In (2.2.3), $\int_0^t \mu_{D,s} ds$ is a continuous adapted process with finite variation paths and $\int_0^t \sigma_{D,s} dW_{D,s}$ corresponds to a continuous local martingale.

**Assumption 2.1.** (i) $\mu_{D,t}$, $\mu_{Z,t}$, $\sigma_{D,t}$ and $\sigma_{Z,t}$ satisfy $P$-a.s., $\sup_{\omega \in \Omega, 0 < t \leq \tau_T} \|\mu_{D,t}(\omega)\| < \infty$, $\sup_{\omega \in \Omega, 0 < t \leq \tau_T} \|\mu_{Z,t}(\omega)\| < \infty$, $\sup_{\omega \in \Omega, 0 < t \leq \tau_T} \|\sigma_{D,t}(\omega)\| < \infty$ and $\sup_{\omega \in \Omega, 0 < t \leq \tau_T} \|\sigma_{Z,t}(\omega)\| < \infty$ for some localizing sequence $\{\tau_T\}$ of stopping times. Also, $\sigma_{D,s}$ and $\sigma_{Z,s}$ are càdlàg; (ii) $\int_0^t \mu_{D,s} ds$ and $\int_0^t \mu_{Z,s} ds$ belong to the class of continuous adapted finite variation processes; (iii) $\int_0^t \sigma_{D,s} dW_{D,s}$ and $\int_0^t \sigma_{Z,s} dW_{Z,s}$ are continuous local martingales with $P$-a.s. finite positive definite conditional variances (or spot covariances) defined by $\Sigma_{D,t} = \sigma_{D,t} \sigma_{D,t}'$ and $\Sigma_{Z,t} = \sigma_{Z,t} \sigma_{Z,t}'$, which for all $t < \infty$ satisfy $\int_0^t \Sigma_{D,s}^{(j,j)} ds < \infty$ ($j = 1, \ldots, q$) and $\int_0^t \Sigma_{Z,s}^{(j,j)} ds < \infty$ ($j = 1, \ldots, p$). Furthermore, for every $j = 1, \ldots, q$, $r = 1, \ldots, p$, and $k = 1, \ldots, T$, $h^{-1} \int_{(k-1)h}^{kh} \Sigma_{D,s}^{(j,r)} ds$ and $h^{-1} \int_{(k-1)h}^{kh} \Sigma_{Z,s}^{(r,r)} ds$ are bounded away from zero and infinity, uniformly in $k$ and $h$; (iv) $e^*_t$ is such that $e^*_t \triangleq \int_0^t \sigma_{e,s} dW_{e,s}$ with $0 < \sigma_{e,t}^2 < \infty$, where $W_e$ is a one-dimensional standard Wiener process. Furthermore, $\langle e, D \rangle_t = \langle e, Z \rangle_t = 0$ identically for all $t \geq 0$.

Part (i) restricts the processes to be locally bounded and part (ii) requires the
80

drifts to be adapted finite variation processes. These are standard regularity conditions in the high-frequency statistics literature [cf. Barndorff-Nielsen and Shephard (2004), Li et al. (2017) and Li and Xiu (2016)]. Part (iii) imposes restrictions on the regressors which require them to have finite integrated covariance. The second part of condition (iii) means that the process $\Sigma_{t,t}$ is bounded away from zero and infinity on any bounded time interval. Part (iv) specifies the error term to be contemporaneously uncorrelated with the regressors. We also rule out jump processes. This is a natural restriction to impose since it essentially implies that the structural change in our model arises from the shift in the parameter $\delta_{Z,t}$ after $T_b$ only. Hence, our results are not expected to provide good approximations for applications involving high-frequency data for which jumps are likely to be important. Our intended scope is for models involving data sampled at, say, the daily or lower frequencies. Since this is an important point we restate it as a separate assumption:

**Assumption 2.2.** $D, Z, e$ and $\Sigma^0 \triangleq \{\Sigma_{t}, \sigma_{e,t}\}_{t \geq 0}$ have $P$-a.s. continuous sample paths.

The assumption above implies that the variables in our model are diffusion processes if one further assumes that the volatilities are deterministic. We shall not impose the latter condition. As a consequence, the processes $D_t$ and $Z_t$ belong to the class of continuous Itô semimartingales with stochastic volatility. Our choice of modeling volatility as a latent factor is justified on multiple grounds. First, a setting in which the variance process is stochastic seems to be more appropriate for the development of a fixed-span asymptotic experiment since sampling uncertainty cannot be averaged out with a limited span of data. Second, some estimates will follow a mixed Gaussian distribution asymptotically, which may lead to better approximations. Third, it does not impose any substantial impediment for the development of
our theoretical results. Fourth, such results will be valid under general conditions on
the variance processes, e.g., nonstationarity and long-memory.

An interesting issue is whether the theoretical results to be derived for model
(2.2.2) are applicable to classical structural change models for which an increasing
span of data is assumed. This requires establishing a connection between the assump-
tions imposed on the stochastic processes in both settings. Roughly, the classical
long-span setting uses approximation results valid for weakly dependent data; e.g.,
ergodic and mixing processes. Such assumptions are not needed under our fixed-span
asymptotics. Nonetheless, we can impose restrictions on the probabilistic properties
of the latent volatility processes in our model and thereby guarantee that ergodic
and mixing properties are inherited by the corresponding observed processes. This
follows from Theorem 3.1 in Genon-Catalot et al. (2000) together with Proposition 4
in Carrasco and Chen (2002). For example, these results imply that the observations
\( \{Z_{kh}\}_{k \geq 1} \) (with fixed \( h \)) can be viewed (under certain conditions) as a hidden Mar-
kov model which inherits the ergodic and mixing properties of \( \{\sigma_{Z,t}\}_{t \geq 0} \). Hence, our
model encompasses those considered in the structural change literature that uses a
long-span asymptotic setting. We shall extend model (2.2.2) to allow for predictable
processes (e.g., a constant and/or lagged dependent variable) in a separate section.

**Assumption 2.3.** \( N_0^b = \lfloor N \lambda_0 \rfloor \) for some \( \lambda_0 \in (0, 1) \).

Assumption 2.3 dictates the asymptotic framework adopted and implies that the
change-point occurs at the observation-index \( T_0^b = \lfloor T \lambda_0 \rfloor \), where \( T_0^b = \lfloor N_0^b/h \rfloor \). Our
framework requires us to distinguish between the actual break date \( N_0^b \) and the index
of the observation associated with the break point, \( T_0^b \). From a practical perspective,
the assumption states that the change-point is bounded away from the starting and
end points. It implies that the pre- and post-break segments of the sample remain
fixed whereas the usual assumption under the large-N asymptotics implies that the
time horizons before and after the break date grow proportionately. This, along
with the usual mixing assumptions imply that only a small neighborhood around
the true break date is relevant asymptotically, thereby ruling out the possibility for
the long-span asymptotics to discern features simply caused by the location of the
break. As opposed to the large-N asymptotics, the continuous record asymptotic
framework preserves information about the data span and the location of the break.
This feature is empirically relevant; simulations reported in Elliott and Müller (2007)
suggests that the location of the break affects the properties of its estimate in small
samples. We show below that our theory reproduces these small-sample features and
provide accurate approximations to the finite-sample distributions.

It is useful to re-parametrize (2.2.2). Let $y_kh = \Delta_h Y_k, x_kh = (\Delta_h D'_h, \Delta_h Z'_h)'$, $z_kh = \Delta_h Z_k, e_kh = \Delta_h e'_k, \beta^0 = \left( (\pi^0), \left( \delta^0_{Z,1} \right) \right)'$ and $\delta^0 = \delta^0_{Z,2} - \delta^0_{Z,1}$. (2.2.2) can be
expressed as:

$$
y_kh = x'_{kh} \beta^0 + e_kh, \quad \left( k = 1, \ldots, T_0 \right) \quad (2.2.4)
$$

$$
y_kh = x'_{kh} \beta^0 + z'_{kh} \delta^0 + e_kh, \quad \left( k = T_0 + 1, \ldots, T \right),
$$

where the true parameter $\theta^0 = \left( (\beta^0)', (\delta^0)'^{'} \right)$ takes value in a compact space $\Theta \subset \mathbb{R}^{\dim(\theta)}$. Also, define $z_kh = R' x_kh$, where $R$ is a $(q+p) \times p$ known matrix with full
column rank. We consider a partial structural change model for which $R = (0, I)'$ with $I$ an identity matrix.

The final step is to write the model in matrix format which will be useful for
the derivations. Let $Y = (y_h, \ldots, y_{Th})'$, $X = (x_h, \ldots, x_{Th})'$, $e = (e_h, \ldots, e_{Th})'$,
\[ X_1 = (x_h, \ldots, x_{T_b} h, 0, \ldots, 0)', \ X_2 = (0, \ldots, 0, x_{(T_b+1)h}, \ldots, x_{T_b})' \]

and

\[ X_0 = (0, \ldots, 0, x_{(T_0_b+1)h}, \ldots, x_{T_b})'. \]

Note that the difference between \( X_0 \) and \( X_2 \) is that the latter uses \( T_b \) rather than \( T_0^b \). Define \( Z_1 = X_1 R, Z_2 = X_2 R \) and \( Z_0 = X R \). (2.2.4) in matrix format is: \( Y = X \beta^0 + Z_0 \delta^0 + e \). We consider the least-squares estimator of \( T_b \), i.e., the minimizer of \( S_T (T_b) \), the sum of squared residuals when regressing \( Y \) on \( X \) and \( Z_2 \) over all possible partitions, namely: \( \hat{T}^{LS}_b = \arg\min_{p+q \leq T_b \leq T} S_T (T_b) \). It is straightforward to show that \( \hat{T}^{LS}_b = \arg\min_{p+q \leq T_b \leq T} QT (T_b) \) where \( QT (T_b) \triangleq \hat{\delta}_T (Z_2' M Z_2) \hat{\delta}_T \), \( \hat{\delta}_T \) is the least-squares estimator of \( \delta^0 \) when regressing \( Y \) on \( X \) and \( Z_2 \), and \( M = I - X (X' X)^{-1} X' \).

For brevity, we will write \( \hat{T}_b \) for \( \hat{T}^{LS}_b \) with the understanding that \( \hat{T}_b \) is a sequence indexed by \( T \) or \( h \). The estimate of the break fraction is then \( \hat{\lambda}_b = \hat{T}_b / T \).

### 2.2.2 The Extended Model with Predictable Processes

The assumptions on \( D_t \) and \( Z_t \) specify that they are continuous semimartingale of the form (2.2.3). This precludes predictable processes, which are often of interest in applications; e.g., a constant and/or a lagged dependent variable. Technically, these require a separate treatment since the coefficients associated with predictable processes are not identified under a fixed-span asymptotic setting. We consider the following extended model:

\[ \Delta_h Y_k = \] (2.2.5)

\[
\begin{cases}
\mu_{1,h} h + \alpha_{1,h} Y_{(k-1)h} + (\Delta_h D_k)' \pi^0 + (\Delta_h Z_k)' \delta_{Z,1}^0 + \Delta_h e_k^*, \quad (k \leq \lfloor T^0_b \rfloor) \\
\mu_{2,h} h + \alpha_{2,h} Y_{(k-1)h} + (\Delta_h D_k)' \pi^0 + (\Delta_h Z_k)' \delta_{Z,2}^0 + \Delta_h e_k^*, \quad (k > \lfloor T^0_b \rfloor, \ldots, T)
\end{cases}
\]
for some given initial value $Y_0$. We specify the parameters associated with the constant and the lagged dependent variable as being of higher order in $h$, or lower in $T$, as $h \downarrow 0$ so that some fixed true parameter values can be identified, i.e., $\mu_{1,h} \triangleq \mu_1^0 h^{-1/2}$, $\mu_{2,h} \triangleq \mu_2^0 h^{-1/2}$, $\mu_{\delta,h} \triangleq \mu_{2,h} - \mu_{1,h}$, $\alpha_{1,h} \triangleq \alpha_{1}^0 h^{-1/2}$, $\alpha_{2,h} \triangleq \alpha_{2}^0 h^{-1/2}$ and $\alpha_{\delta,h} \triangleq \alpha_{2,h} - \alpha_{1,h}$. Our framework is then similar to the small-diffusion setting studied previously [cf. Ibragimov and Has’minskii (1981), Galtchouk and Konev (2001), Laredo (1990) and Sørensen and Uchida (2003)]. With $\mu_{.,h}$ and $\alpha_{.,h}$ independent of $h$ and fixed, respectively, at the true values $\mu_0^0$ and $\alpha_0^0$, the continuous-time model is then equivalent to

$$Y_t = Y_0 + \int_0^t \left( \mu_1^0 + \mu_2^0 1_{\{s > N_b^0\}} \right) ds + \int_0^t \left( \alpha_{1}^0 + \alpha_{2}^0 1_{\{s > N_b^0\}} \right) Y_s ds + \int_0^t \left( \delta_{Z,1}^0 + \delta_{1}^0 1_{\{s > N_b^0\}} \right)' dZ_s + e^*_t,$$

for $t \in [0, N]$, where $Y_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h Y_k$, $D_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h D_k$, $Z_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h Z_k$ and $e^*_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h e^*_k$. The results to be discussed below go through in this extended framework. However, some additional technical details are needed. Hence, we treat both cases with and without predictable components separately. Note that the model and results can be trivially extended to allow for more general forms of predictable processes, at the expense of additional technical details of no substance.

2.3 Consistency and Convergence Rate under Fixed Shifts

We now establish the consistency and convergence rate of the least-squares estimator under fixed shifts. Under the classical large-$N$ asymptotics, related results have been established by Bai (1997), Bai and Perron (1998) and also Perron and Qu (2006) who relaxed the conditions used. Early important results for a mean-shift appeared in Yao (1987) and Bhattacharya (1987) for an i.i.d. series, Bai (1994) for linear processes and
Picard (1985) for a Gaussian autoregressive model. In order to proceed, we impose the following identification conditions.

**Assumption 2.4.** There exists an $l_0$ such that for all $l > l_0$, the matrices $(lh)^{-1} \sum_{k=l}^{T} x_{kh}x'_{kh}$, $(lh)^{-1} \sum_{k=T-l+1}^{T} x_{kh}x'_{kh}$, and $(lh)^{-1} \sum_{k=T_0-l+1}^{T_0} x_{kh}x'_{kh}$, have minimum eigenvalues bounded away from zero in probability.

**Assumption 2.5.** Let $Q_0(T_0, \theta^0) \triangleq E [Q_T(T_0, \theta^0) - Q_T(T_0^0, \theta^0)]$. There exists a $T_0$ such that $Q_0(T_0, \theta^0) > \sup_{(T_0, \theta^0) \in B} Q_0(T_0, \theta^0)$, for every open set $B$ that contains $(T_0, \theta^0)$.

Assumption 2.4 is similar to A2 in Bai and Perron (1998) and requires enough variation around the break point and at the beginning and end of the sample. The factor $h^{-1}$ normalizes the observations so that the assumption is implied by a weak law of large numbers. Assumption 2.5 is a standard uniqueness identification condition.

We then have the following results.

**Proposition 2.3.1.** Under Assumption 3.1-2.3 and 2.4-2.5, for any $\varepsilon > 0$ and $K > 0$, and all large $T$, $P \left( \left| \hat{\lambda}_b - \lambda_0 \right| > K \right) < \varepsilon$.

**Proposition 2.3.2.** Under Assumption 3.1-2.3 and 2.4-2.5 for any $\varepsilon > 0$, there exists a $K > 0$ such that for all large $T$, $P \left( T \left| \hat{\lambda}_b - \lambda_0 \right| > K \right) = P \left( \left| \hat{T}_b - T_0 \right| > K \right) < \varepsilon$.

We have the same $T$-convergence rate as under large-$N$ asymptotics. Let $\theta^0 = \left( (\beta^0)', (\delta_1^0)', (\delta_2^0)' \right)'$. The fast $T$-rate of convergence implies that the least-squares estimate of $\theta^0$ is the same as when $\lambda_0$ is known. A natural estimator for $\theta^0$ is $\arg\min_{\beta \in \mathbb{R}^{p+q}, \delta \in \mathbb{R}^p} \| Y - X \beta - \hat{Z}_2 \delta \|^2$, where we use $T_b = \hat{T}_b$ in the construction of $\hat{Z}_2$. Then we have the following result, akin to an extension of corresponding results in Section 3 of Barndorff-Nielsen and Shephard (2004). As a matter of notation, let $\Sigma^* \triangleq \{ \mu, t, \Sigma, \sigma \}$ and denote expectation taken with respect to $\Sigma^*$ by $E^*$. 
Proposition 2.3.3. Under Assumption 3.1-2.3 and 2.4-2.5, we have as \( T \to \infty \) (\( N \) fixed), conditionally on \( \Sigma^* \), \( \left( \sqrt{T/N} \left( \hat{\beta} - \beta^0 \right), \sqrt{T/N} \left( \hat{\delta} - \delta^0 \right) \right)' \xrightarrow{d} \mathcal{MN}(0, V) \) where \( \mathcal{MN} \) denotes a mixed Gaussian distribution, with

\[
V \triangleq \nabla^{-1} \lim_{T \to \infty} \frac{T}{\sum_{k=1}^{T} \mathbb{E}^* (x_{kh}x'_{kh}e^2_{kh}) + \sum_{k=T_0}^{T} \mathbb{E}^* (x_{kh}z'_{kh}e^2_{kh}) + \sum_{k=T_0}^{T} \mathbb{E}^* (z_{kh}z'_{kh}e^2_{kh})}{\sum_{k=T_0}^{T} \mathbb{E}^* (x_{kh}x'_{kh}) + \sum_{k=T_0}^{T} \mathbb{E}^* (x_{kh}z'_{kh}) + \sum_{k=T_0}^{T} \mathbb{E}^* (z_{kh}z'_{kh})},
\]

and

\[
\nabla \triangleq \lim_{T \to \infty} \left[ \frac{\sum_{k=1}^{T} \mathbb{E}^* (x_{kh}x'_{kh}) + \sum_{k=T_0}^{T} \mathbb{E}^* (x_{kh}z'_{kh}) + \sum_{k=T_0}^{T} \mathbb{E}^* (z_{kh}z'_{kh})}{\sum_{k=1}^{T} \mathbb{E}^* (x_{kh}x'_{kh}) + \sum_{k=1}^{T} \mathbb{E}^* (x_{kh}z'_{kh}) + \sum_{k=1}^{T} \mathbb{E}^* (z_{kh}z'_{kh})} \right].
\]

The limit law of the regression parameters is mixed Gaussian, where the variance matrix \( V \) is stochastic. Hence, the theorem is also useful because it approximates a setting where the uncertainty about the break date transmits to a limit law for the regression parameters that has heavier tails than the Gaussian law; this turns out to be often the case in practice. Under the assumption of deterministic variances, the limit law would be a normal variate.

2.4 Asymptotic Distribution under a Continuous Record

We now present results about the limiting distribution of the least-squares estimate of the break date under a continuous record framework. As in the classical large-\( N \) asymptotics, it depends on the exact distribution of the data and the errors for fixed break sizes [c.f., Hinkley (1971)]. This has forced researchers to consider a shrinkage asymptotic theory where the size of the shift is made local to zero as \( T \) increases, an approach developed by Picard (1985) and Yao (1987). We continue with this avenue. Section 2.4.1 presents the main theoretical results. The features of the asymptotic distribution obtained are discussed in Section 2.4.2.
2.4.1 Main Theoretical Results

We first discuss the main arguments of our derivation. Given the consistency result, we know that there exists some \( h^* \) such that for all \( h < h^* \) with high probability

\[
\eta Th \leq \hat{N}_b - N^b_0 = O_p(T^{-1}),
\]

i.e., \( \hat{N}_b \) is in a shrinking neighborhood of \( N^b_0 \), which, however, shrinks too fast and impedes the development of a feasible limit theory. Hence, we rescale time and work with the objective function in a small neighborhood of the true break date under this “new” time scale. We begin with the following assumption which specifies that i) we use a shrinking condition on \( \delta^0 \); ii) we introduce a locally increasing variance condition on the residual process. The first is similarly used under classical large-\( N \) asymptotics, while the second is new and necessary in our context in order to accurately approximate the change-point problem. In addition, it also leads to a limit distribution that is influenced by parameters that can be consistently estimated, so that a feasible method of inference can be conducted.

**Assumption 2.6.** Let \( \delta_h = \delta^0 h^{1/4} \) and assume that for all \( t \in (N^b_0 - \epsilon, N^b_0 + \epsilon) \), with \( \epsilon \downarrow 0 \) and \( T^{1-\kappa} \epsilon \to B < \infty \), \( 0 < \kappa < 1/2 \),

\[
\mathbb{E} \left[ \left( \Delta_h e_t^* \right)^2 | \mathcal{F}_{t-h} \right] = \sigma^2_{h,t-h} \Delta t \text{ P-a.s.,}
\]

where \( \sigma_{h,t} \triangleq \sigma_h \sigma_{e,t} \), \( \sigma_h \triangleq \sigma h^{-1/4} \) and \( \sigma \triangleq \int_0^N \sigma_{e,s}^2 ds \).

The vector of scaled true parameters is \( \theta_h \triangleq \left( (\beta^0)' , \delta^0_h \right)' \). Define

\[
\Delta_h \bar{e}_t \triangleq \begin{cases} 
\Delta_h e_t^* , & t \notin (N^b_0 - \epsilon, N^b_0 + \epsilon) \\
 h^{1/4} \Delta_h e_t^* , & t \in (N^b_0 - \epsilon, N^b_0 + \epsilon)
\end{cases}.
\]

We shall refer to \( \{\Delta_h \bar{e}_t, \mathcal{F}_t\} \) as the normalized residual process. Under this framework, the rate of convergence is now \( T^{1-\kappa} \) with \( 0 < \kappa < 1/2 \). Due to the fast rate of convergence of the change-point estimator, the objective function oscillates too rapidly as \( h \downarrow 0 \). By scaling up the volatility of the errors around the change-point,
we make the objective function behave as if it were a function of a standard diffusion process. The neighborhood in which the errors have relatively higher variance is shrinking at a rate $1/T^{1-\kappa}$, the rate of convergence of $\hat{N}_b$. Hence, in a neighborhood of $N^0_b$ in which we study the limiting behavior of the break point estimator, the rescaled criterion function is regular enough so that a feasible limit theory can be developed. The rate of convergence $T^{1-\kappa}$ is still sufficiently fast to guarantee a $\sqrt{T}$-consistent estimation of the slope parameters, as stated in the following proposition.

**Proposition 2.4.1.** Under Assumption 3.1-2.3, 2.4-2.5 and 2.6, (i) $\hat{\lambda}_b \overset{p}{\to} \lambda_0$; (ii) for every $\varepsilon > 0$ there exists a $K > 0$ such that for large $T$, $P(T^{1-\kappa} | \hat{\lambda}_b - \lambda_0 | > K \| \delta^0 \|^2 \sigma^2) < \varepsilon$; and (iii) for $\kappa \in (0, 1/4]$, $\left( \sqrt{T/N} (\hat{\beta} - \beta^0), \sqrt{T/N} (\hat{\delta} - \delta^0) \right)' \overset{d}{\to} \mathcal{M} \mathcal{N} (0, V)$ as $T \to \infty$, with $V$ given in Proposition 2.3.3.

Consider the set $D(C) \triangleq \{ N_b : N_b \in \{ N^0_b + Ch^{1-\kappa} \} , |C| < \infty \}$, on the original time scale. Let $Z_\Delta \triangleq (0, \ldots, 0, z(T_{b+1})_h, \ldots, z_{T_b}^h, 0, \ldots, 0)$ if $T_b < T^0_b$ and $Z_\Delta \triangleq (0, \ldots, 0, z(T^0_{b+1})_h, \ldots, z_{T_b}^h, 0, \ldots, 0)$ if $T_b > T^0_b$; also set $\psi_h \triangleq h^{1-\kappa}$. The following lemma will be needed in the derivations.

**Lemma 2.4.1.** Under Assumption 3.1-2.3, 2.4-2.5 and 2.6, uniformly in $T_b$,

$$
\left( Q_T (T_b) - Q_T (T^0_b) \right) / \psi_h = -\delta_h (Z'_\Delta Z_\Delta / \psi_h) \delta_h
$$

$$
+ 2\delta'_h (Z'_\Delta e / \psi_h) \text{sgn} (T^0_b - T_b) + o_p \left( h^{1/2} \right).
$$

Lemma 2.4.1 shows that only the terms involving the regressors whose parameters are allowed to shift have a first-order effect on the asymptotic analysis. For brevity, we use the notation $\pm$ in place of $\text{sgn} (T^0_b - T_b)$ hereafter.

The conditional first moment of the centered criterion function $Q_T (T_b) - Q_T (T^0_b)$ is of order $O (h^{1-\kappa})$, i.e., “oscillates” rapidly as $h \downarrow 0$. Hence, in order to approximate
the behavior of \( \{ T_b - T_0 \} \) we rescale “time”. For any \( C > 0 \), let \( L_C \triangleq N_0^b - Ch^{1-\kappa} \) and \( R_C \triangleq N_0^b + Ch^{1-\kappa} \), where \( L_C \) and \( R_C \) are the left and right boundary points of \( D(C) \), respectively. We then have \( |R_C - L_C| = O(Ch^{1-\kappa}) \). Now, take the vanishingly small interval \([L_C, R_C]\) on the original time scale, and stretch it into a time interval \([T^{1-\kappa}L_C, T^{1-\kappa}R_C]\) on a new “fast time scale”. Since the criterion function is scaled by \( \psi^{-1}_h \), all scaled processes are \( O_p(1) \). Now, let \( N_b(v) = N_0^b - vh^{1-\kappa}, \ v \in [-C, C] \).

Using Lemma 2.4.1 and Assumption 2.6 (see the appendix),

\[
\psi^{-1}_h \left( Q_T(T_b(v)) - Q_T(T_0^b) \right) = - \delta_h \left( \sum_{k=T_b(v)+1}^{T_0^b} \frac{z_{kh}'}{\sqrt{\psi h}} \frac{z'_{kh}}{\sqrt{\psi h}} \right) \delta_h + 2 \left( \delta^0 \right)' \sum_{k=T_b(v)+1}^{T_0^b} \frac{z_{kh}}{\sqrt{\psi h}} \frac{\bar{e}_{kh}}{\sqrt{\psi h}} + o_p \left( h^{1/2} \right).
\]

In addition, in view of (2.2.3), we let \( dZ_{\psi, s} = \psi^{-1/2}_h \sigma_{Z,s} dW_{Z,s} \) for

\[
s \in \left[ N_b^0 - vh^{1-\kappa}, N_b^0 + vh^{1-\kappa} \right].
\]

Applying the time scale change \( s \to t \triangleq \psi^{-1}_h s \) to all processes including \( \Sigma^0 \), we have \( dZ_{\psi, t} = \sigma_{Z,t} dW_{Z,t} \) with \( t \in D^*(C) \), where

\[
D^*(C) \triangleq \left\{ t : t \in \left[ N_b^0 + v \left\| \delta^0 \right\|^2 / \sigma^2 \right], \ |v| \leq C \right\}.
\]

Therefore,

\[
\psi^{-1}_h \left( Q_T(T_b(v)) - Q_T(T_0^b) \right) = - \delta_h \left( \sum_{k=T_b(v)+1}^{T_0^b} z_{\psi, kh} z'_{\psi, kh} \right) \delta_h + 2 \left( \delta^0 \right)' \sum_{k=T_b(v)+1}^{T_0^b} z_{\psi, kh} \bar{e}_{\psi, kh} + o_p \left( h^{1/2} \right),
\]

with \( NT_b(v)/T = N_b(v) = N_b^0 + v \), where \( z_{\psi, kh} \triangleq z_{kh}/\sqrt{\psi_h} \) and \( \bar{e}_{\psi, kh} \triangleq \bar{e}_{kh}/\sqrt{\psi_h} \).

That is, because of the change of time scale all processes in the last display are scaled
up to be $O_p(1)$ and thus behave as diffusion-like processes. On this new “fast time scale”, we have $T^{1-\kappa}R_C - T^{1-\kappa}L_C = O(1)$ and $Q_T(T_b(v)) - Q_T(T_b^0)$ is restored to be $O_p(1)$. Observe that changing the time scale does not affect any statistic which depends on observations from $k = 1$ to $k = \lfloor L_C/h \rfloor$. By symmetry, it does not affect any statistic which involves observations from $k = \lfloor R_C/h \rfloor$ to $k = T$ (since these involve a positive fraction of data). However, it does affect quantities which include observations that fall in $[T_b h, T_b^0 h]$ (assuming $T_b < T_b^0$). In particular, on the original time scale, the processes $\{D_t\}$, $\{Z_t\}$ and $\{e_t\}$ are well-defined and scaled to be $O_p(1)$ while $Q_T(T_b) - Q_T(T_b^0)$ (asymptotically) oscillates more rapidly than a simple diffusion-type process. On the new “fast time scale”, $\{D_t\}$, $\{Z_t\}$ and $\{e_t\}$ are not affected since they have the same order in $[T^{1-\kappa}L_C, T^{1-\kappa}R_C]$ as $h \downarrow 0$. That is, the first conditional moments are $O(h)$ while the corresponding moments for $Q_T(T_b) - Q_T(T_b^0)$ on $D^*(C)$ are restored to be $O(h)$. As the continuous-time limit is approached, the rescaled criterion function $(Q_T(T_b(v)) - Q_T(T_b^0))/h^{1/2}$ operates on a “fast time scale” on $D^*(C)$.

Our analysis is local; we examine the limiting behavior of the centered and rescaled criterion function process in a neighborhood $D^*(C)$ of the true break date $N^0_b$ defined on a new time scale. We first obtain the weak convergence results for the statistic $(Q_T(T_b(v)) - Q_T(T_b^0))/h^{1/2}$ and then apply a continuous mapping theorem for the argmax functional. However, it is convenient to work with a reparametrized objective function. Proposition 2.4.1 allows us to use

$$Q_T(\theta^*) = \left( Q_T(\theta^*_b, T_b(v)) - Q_T(\theta^0, T_b^0) \right)/h^{1/2},$$

where $\theta^* \triangleq (\theta^*_b, v)'$ with $T_b(v) \triangleq T_b^0 + [v/h]$ and $T_b(v)$ is the time index on the “fast time scale”. When $v$ varies, $T_b(v)$ visits all integers between 1 and $T$, with
the normalizations $T_b(v) = 1$ if $T_b(v) \leq 1$ and $T_b(v) = T$ if $T_b(v) \geq T$. On the old time scale $N_b(u) = N_b^0 + u$ with $u \to \psi_h^{-1}u$, so that $N_b(u)$ is in a vanishing neighborhood of $N_b^0$. On $D^*(C)$, we index the process $Q_T(\theta_h, T_b(v)) - Q_T(\theta^0, T_b^0)$ by two time subscripts: one referring to the time $T_b$ on the original time scale and one referring to the time elapsed since $T_bh$ on the “fast time scale”. For simplicity, we omit the former; the optimization problem is not affected by the change of time scale. In fact, by Proposition 2.4.1, $u = Th(\lambda - \lambda_0) = KO_p(h^{1-\kappa})$ on the old time scale; whereas on the new “fast time scale”, $v = Th(\lambda - \lambda_0) = O_p(1)$. The maximization problem is not changed because $v/h$ can take any value in $\mathbb{R}$. The process $Q_T(\theta_h, T_b(v)) - Q_T(\theta^0, T_b^0)$ is thus analyzed on a fixed horizon since $v$ now varies over $\left[-N_b^0/\left(\|\delta^0\|^{-2}\sigma^2\right), (N - N_b^0)/\left(\|\delta^0\|^{-2}\sigma^2\right)\right]$. Hence, redefine

$$D^*(C) = \{(\beta^0, \delta_h, v) : \|\theta^0\| \leq C; T_b(v) = T_b^0 + vN^{-1}\|\delta^0\|^{-2}\sigma^2; \ - \frac{N_b^0}{\|\delta^0\|^{-2}\sigma^2} \leq v \leq \frac{N - N_b^0}{\|\delta^0\|^{-2}\sigma^2}\}.$$  

Note that $D^*(C)$ is compact. Let $D(D^*(C), \mathbb{R})$ denote the space of all càdlàg functions from $D^*(C)$ into $\mathbb{R}$. Endow this space with the Skorokhod topology and note that $D(D^*(C), \mathbb{R})$ is a Polish space. The faster rate of convergence of $\hat{\lambda}_b$ established in Proposition 2.4.1-(ii) combined with the $\sqrt{T}$-rate for the regression parameters allow us to apply the continuous mapping theorem for the argmax functional [cf. Kim and Pollard (1990)]. Under a continuous record, we can apply limit theorems for statistics involving (co)variation between regressors and errors. This enables us to deduce the limiting process for $\overline{Q}_T(\theta^*)$. These asymptotic results mainly rely upon the work of Jacod (1994; 1997) and Jacod and Protter (1998).

To guide intuition, note that under the new re-parametrization, the limit law of
$\overline{Q}_T(\theta^*)$ is, according to Lemma 2.4.1, the same as the limit law of

$$-h^{-1/2}\delta_h' (Z'_\Delta Z_\Delta) \delta_h \pm 2h^{-1/2}\delta_h' (Z'_\Delta \bar{e})
\equiv - \left(\delta^0\right)' (Z'_\Delta Z_\Delta) \delta^0 \pm 2h^{-1/2} \left(\delta^0\right)' h^{1/4} \left(Z'_\Delta h^{-1/4} \bar{e}\right),$$

where $\equiv$ denotes (first order) equivalence in law, $\bar{e}_{kh} \equiv h^{1/4} e_{kh}$ and since (approximately) $e_{kh} \sim i.n.d. \mathcal{N} \left(0, \sigma^2_{h,k-1} h\right)$, $\sigma_{h,k} = \sigma_h \sigma_{e,k}$ then $\bar{e}_{kh} \sim i.n.d. \mathcal{N} \left(0, \sigma^2_{e,k-1} h\right)$. Hence, the limit law of $\overline{Q}_T(\theta^*)$ is, to first-order, equivalent to the law of

$$- \left(\delta^0\right)' (Z'_\Delta Z_\Delta) \delta^0 \pm 2 \left(\delta^0\right)' \left(h^{-1/2} Z'_\Delta \bar{e}\right). \tag{2.4.3}$$

We apply a law of large numbers to the first term and a stable convergence in law under the Skorokhod topology to the second. Assumption 2.6 combined with the normalizing factor $h^{-1/2}$ in $\overline{Q}_T(\theta^*)$ account for the discrepancy between the deterministic and stochastic component in (2.4.3).

Having outlined the main steps in the arguments used to derive the continuous records limit distribution of the break date estimate, we now state the main result of this section. The full details are relegated to the Appendix. Part of the proof involves showing the stable convergence in distribution [cf. Rényi (1963) and Aldous and Eagleson (1978)] toward an $\mathcal{F}$-conditionally two-sided Gaussian process. The limiting process is realized on an extension of the original probability space and we relegate this description to Section 2.9.1 in the appendix.

**Theorem 2.4.1.** Under Assumption 3.1-2.3, 2.4-2.5 and 2.6, and under the “fast
time scale”,

\[
N \left( \hat{\lambda}_b - \lambda_0 \right) \xrightarrow{L^2} \argmax_{v \in \left[ -N_0^b/\left(\|\delta^0\|^2\right), (N-N_0^b)/\left(\|\delta^0\|^2\right) \right]} \left\{ -\left( \delta^0 \right)' \langle Z_\Delta, Z_\Delta \rangle (v) \delta^0 + 2 \left( \delta^0 \right)' \mathcal{W} (v) \right\},
\]

where \( \langle Z_\Delta, Z_\Delta \rangle (v) \) is the predictable quadratic variation process of \( Z_\Delta \). The process \( \mathcal{W} (v) \) is, conditionally on the \( \sigma \)-field \( \mathcal{F} \), a two-sided centered Gaussian martingale with independent increments and variances given in Section 2.9.1 in the Appendix.

Note that the theorem is in accordance with Proposition 2.4.1 because it holds under the new “fast time scale”. The theoretical results from Section 2.3 and Theorem 2.4.1 allow one to draw the following features. Under fixed shifts, the fractional break date is super-consistent—i.e., \( N \left( \hat{\lambda}_b - \lambda_0 \right) = O_p \left( h/\|\delta^0\|^2 \right) \). In contrast, the estimate of the break point \( \hat{T}_b \) is not even consistent. Simply letting the magnitude of the shift shrink to zero does not result in a useful approximation when the shifts are small. When one augments the shrinking shifts assumption with locally increasing variances (cf. Assumption 2.6), the rate of convergence of \( \hat{\lambda}_b \) becomes slower. Further, through a change of time scale, Theorem 2.4.1 suggests that the span of the data is more important than the sample size when shifts are small. Further work will report on formalizing the complex relationships between the sampling frequency, sample size, span of the data and shift magnitude. For example, we can show that \( \hat{T}_b \) is itself consistent if the shift is large, i.e., \( \delta \to \infty \), even if the sample size and the sampling frequency are fixed.

For comparison purposes, recall that the classical large-\( N \) limiting distribution is related to the location of the maximum of a two-sided Wiener process over the interval \((-\infty, \infty)\). Its probability density is symmetric for the case of stationary regimes and
has thicker tails and a higher peak than the density of a Gaussian variate. In contrast, the limiting distribution in Theorem 2.4.1 involves the location of the maximum of a function of the (quadratic) variation of the regressors and of a two-sided centered Gaussian martingale process over the interval

\[-N_b^0/\left(\|\delta^0\|^{-2}\sigma^2\right), (N - N_b^0) / \left(\|\delta^0\|^{-2}\sigma^2\right)\]

Notably, this domain depends on the true value of the break point $N_b^0$ and therefore the limit distribution is asymmetric, in general. The degree of asymmetry increases as the true break point moves away from mid-sample. This holds even when the distributions of the errors and regressors are the same in the pre- and post-break regimes.

Additional relevant remarks follow; more details are provided in Section 2.4.2. The size of the shift plays a key role in determining the density of the asymptotic distribution. More precisely, for an appropriately defined “signal-to-noise” ratio, the density displays interesting properties which change when this quantity as well as other parameters of the model change. Moreover, the distribution in Theorem 2.4.1 is able to reproduce important features of the small-sample results obtained via simulations [e.g., Bai and Perron (2006)]. First, the second moments of the regressors impact the asymptotic mean as well as the second-order behavior of the break point estimator. This complies with simulation evidence pointing out that, for instance, the persistence of the regressors influences the finite-sample performance of the estimator. Second, the continuous record setting manages to preserve information about the time span $N$ of the data and this is clearly an advantage since the location of the true break point matters for the small-sample distribution of the estimator. It has been shown via simulations that in small-samples the break point estimator tends to
be imprecise if the break size is small, and some bias arises if the break point is not at mid-sample. In our framework, the time horizon \([0, N]\) is fixed and thus we can distinguish between the statistical content of the segments \([0, N^0_b]\) and \([N^0_b, N]\). In contrast, this is not feasible under the classical shrinkage large-\(N\) asymptotics because both the pre- and post-break segments increase to infinity proportionately and mixing conditions are imposed so that the only relevant information is a neighborhood around the true break date. As for the relative impact of time span \(N\) versus sample size \(T\) on the precision of the estimator, the time span plays a key role and has a more pronounced impact relative to the sample size. We shall see in the next section that the asymptotic distribution derived under a continuous record provides an accurate approximation to the finite-sample distribution and the approximation is remarkably better than that resulting from the classical shrinkage large-\(N\) asymptotics [cf. Bai (1997) and Yao (1987)]. Details on how to simulate the limiting distribution in Theorem 2.4.1 are given in Section 2.9.4.

We further characterize the asymptotic distribution by exploiting the (\(\mathcal{F}\) - conditionally) Gaussian property of the limit process. The analysis also holds unconditionally if we assume that the volatility processes are non-stochastic. Thus, as in the classical setting, we begin with a second-order stationarity assumption within each regime. The following assumption guarantees that the results below remain valid without the need to condition on \(\mathcal{F}\).

**Assumption 2.7.** \(\Sigma^0\) is (possibly time-varying) deterministic; \(\{z_{kh}, e_{kh}\}\) is second-order stationary within each regime. For \(k = 1, \ldots, T_b^0\), \(\mathbb{E}(z_{kh}z'_{kh} | \mathcal{F}_{(k-1)h}) = \Sigma_{Z,1} h\), \(\mathbb{E}(e^2_{kh} | \mathcal{F}_{(k-1)h}) = \sigma^2_{e,1} h\) and \(\mathbb{E}(z_{kh}z'_{kh}e^2_{kh} | \mathcal{F}_{(k-1)h}) = \Omega_{W,1} h^2\) while for \(k = T_b^0 + 1, \ldots, T\), \(\mathbb{E}(z_{kh}z'_{kh} | \mathcal{F}_{(k-1)h}) = \Sigma_{Z,2} h\), \(\mathbb{E}(e^2_{kh} | \mathcal{F}_{(k-1)h}) = \sigma^2_{e,2} h\) and \(\mathbb{E}(z_{kh}z'_{kh}e^2_{kh} | \mathcal{F}_{(k-1)h}) = \Omega_{W,2} h^2\).
Let $W^*_i$, $i = 1, 2$, be two independent standard Wiener processes defined on $[0, \infty)$, starting at the origin when $s = 0$. Let

$$
\mathcal{V}(s) = \begin{cases} 
-\frac{|s|}{2} + W_1^*(s), & \text{if } s < 0 \\
-\frac{(\delta^0)'\Sigma Z, Z, \delta^0}{(\delta^0)'\Sigma Z, \delta^0} \frac{|s|}{2} + \left(\frac{(\delta^0)'\Omega_{\mathcal{W}, 1}(\delta^0)}{(\delta^0)'\Omega_{\mathcal{W}, 1}(\delta^0)}\right)^{1/2} W_2^*(s), & \text{if } s \geq 0.
\end{cases}
$$

**Theorem 2.4.2.** Under Assumption 3.1-2.3, 2.4-2.5 and 2.6-2.7, and under the “fast time scale”,

$$
\frac{((\delta^0)'(Z, Z)_{11}\delta^0)^2}{((\delta^0)'\Omega_{\mathcal{W}, 1}\delta^0)} N\left(\tilde{\lambda}_b - \lambda_0\right) \Rightarrow \argmax_{s \in \left[-\frac{N_0}{\|\delta^0\|^2 - \pi^2}, \frac{N - N_0}{\|\delta^0\|^2 - \pi^2}\right]} \mathcal{V}(s).
$$

Unlike the asymptotic distribution derived under classical long-span asymptotics, the probability density function of the argmax process in (2.4.5) is not available in closed form. Furthermore, the limiting distribution depends on unknown quantities. We first discuss the probabilistic properties of the infeasible density and then explain how we can derive a feasible counterpart. This will be useful to characterize the main features of interest that will guide us in devising methods to construct confidence sets for $T_b^0$.

**2.4.2 Infeasible Density of the Asymptotic Distribution**

An important parameter is $\rho \triangleq \left((\delta^0)'(Z, Z)_{11}\delta^0\right)^2 / \left((\delta^0)'\Omega_{\mathcal{W}, 1}\delta^0\right)$. We plot the probability density functions of the infeasible distribution of $\rho N\left(\tilde{\lambda}_b - \lambda_0\right)$ for $N = 100$, as given in Theorem 2.4.2, and compare it with the corresponding distribution in Bai (1997). We first consider cases for which the first and second moments of regressors...
and errors do not vary “too much” across regimes; cases that satisfy:

$$\frac{1}{\varpi} \leq \frac{\rho}{\xi_1}, \frac{\rho}{\xi_2} \leq \varpi,$$

$$\xi_1 = \frac{(\delta^0)' \langle Z, Z \rangle_2 \delta^0}{(\delta^0)' \langle Z, Z \rangle_1 \delta^0}, \quad \xi_2 = \left(\frac{(\delta^0)' \Omega W \delta^0}{(\delta^0)' \Omega W \delta^0}\right),$$

(2.4.6)

for some number \(\varpi\) (see below). Then \(\mathcal{Y}(s)\) in (2.4.5) becomes

$$\mathcal{Y}(s) = \begin{cases} 
-\frac{|s|}{2} + W_1^* (-s), & \text{if } s < 0 \\
-\frac{|s|}{2} \xi_1 + \sqrt{\xi_2} W_2^* (s), & \text{if } s \geq 0.
\end{cases}$$

We consider the case of “nearly stationary regimes” where we set \(\varpi = 1.5\) so that the degree of heterogeneity across regimes is restricted. Figure 4.1 displays the density of \(\rho \left(\hat{T}_b - T_0\right)\) for \(\lambda_0 = 0.3, 0.5, 0.7\) and for a low signal-to-noise ratio \(\rho^2 = 0.2\). In addition, we set \(\xi_1 = \xi_2 = 1\) so that each regime has the same distribution. We also plot the density of Bai’s (1997) large-\(N\) shrinkage asymptotic distribution [see also Yao (1987)]. The corresponding plots when \(\rho^2 = 0.3, 0.5, 0.8\) are reported in Figure 4.2-2.4. Note that the restrictions in (2.4.6) imply that a high value of \(\rho\) corresponds to a large shift size \(\delta^0\).

Several interesting observations appear at the outset. First, the density of the large-\(N\) shrinkage asymptotic distribution does not depend on the location of the break, and thus it is always unimodal and symmetric about the origin. Second, it has thicker tails and a much higher peak than the density of a standard normal variate. None of these features are shared by the density derived under a continuous record. When the true break date is at mid-sample \((\lambda_0 = 0.5)\), the density function is symmetric and centered at zero. However, when the signal-to-noise ratio is low \((\rho^2 = 0.2, 0.3)\), the density features three modes. The highest mode is not at the true break date \(\lambda_0 = 0.5\) when the signal is very low \((\rho^2 = 0.2)\). When the break date
is not at mid-sample, the density is asymmetric despite having homogenous regimes. This tri-modality vanishes as the signal-to-noise ratio increases ($\rho^2 = 0.8$) (Figure 2.4, middle panel). When $\rho^2 = 0.3$ and the true break date is not at mid sample ($\lambda_0 = 0.3$ and $\lambda_0 = 0.7$; left and right panel, respectively, Figure 4.2) the density is asymmetric; for values of $\lambda_0$ less (larger) than 0.5, the probability density is right (left) skewed. Such feature is more apparent when the signal-to-noise ratio is low ($\rho^2 = 0.2, 0.3$ and 0.5; Figure 4.1-2.3, side panels). When the signal is low and $\lambda_0$ is less (larger) than 0.5, the probability density has highest mode at values that correspond to $\lambda_b$ being close to the starting (end) sample point than centered at $\lambda_0$. However, as in the case of $\lambda_0 = 0.5$, when the signal-to-noise ratio increases ($\rho^2 = 0.5, 0.8, 1.5$) the highest mode is centered at a value which corresponds to $\lambda_b$ being close to $\lambda_0$ (cf. Figure 2.3-2.4, side panels). Indeed, the density is still asymmetric when $\lambda_0 = 0.3$ and $\lambda_0 = 0.7$ if $\rho^2 = 0.5$.

The interpretation of these features are straightforward. For example, asymmetry reflects the fact that the span of the data and the actual location of the break play a crucial role on the behavior of the estimator. If the break occurs early in the sample there is a tendency to overestimate the break date and vice-versa if the break occurs late in the sample. The marked changes in the shape of the density as we raise $\rho$ confirms that the magnitude of the shift matters a great deal as well. The tri-modality of the density when the shift size is small reflects the uncertainty in the data as to whether a structural change is present at all; i.e., the least-squares estimator finds it easier to locate the break at either the beginning or the end of the sample.

The supplement in Chapter 6 contains an extended description of the features

---

\(^4\)Asymmetry and multi-modality of the finite-sample distribution of the break point estimator were also found by Perron and Zhu (2005) and Deng and Perron (2006) in models with a trend.
of the asymptotic distribution where we consider many other cases for \( \lambda_0 \) and \( \rho^2 \) as well as an analysis of cases allowing differences between the distribution of errors and regressors in the pre- and post-break regimes (referred to as “nonstationary regimes”). In the latter case, we show that even if the signal-to-noise ratio is moderately high the continuous record asymptotic distribution is asymmetric even when the break occurs at mid-sample. This is in stark contrasts to the “nearly stationary” scenario since the density was shown to be always symmetric no matter the value taken by \( \rho \) if \( \lambda_0 = 0.5 \). This means that the asymptotic distribution attributes different weights to the informational content of the two regimes since they possess heterogeneous characteristics.

The results for the densities under the two different scenarios, “nearly stationary” versus “nonstationary regimes”, allows one to deduce the following feature. The asymptotic distribution of \( \hat{T}_b \) is symmetric if both (i) the break date is located at exactly mid-sample, and (ii) the distributions of the errors and regressors do not differ “too much” across regimes. This holds unless the break magnitude is very large in which case the density is symmetric.

2.5 Approximation to the Finite-Sample Distribution

In order to use the continuous record asymptotic distribution in practice one needs consistent estimates of the unknown quantities. In this section, we compare the finite-sample distribution of the least-squares estimator of the break date with a feasible version of the continuous record asymptotic distribution obtained with plug-in estimates. We obtain the finite-sample distribution of \( \rho \left( \hat{T}_b - T^y_b \right) \) based on 100,000
simulations from the following model:

\[ Y_t = D_t \pi^0 + Z_t \beta^0 + Z_t \delta^0 1_{\{t > T_0\}} + \epsilon_t, \quad t = 1, \ldots, T, \tag{2.5.1} \]

where \( Z_t = 0.5 Z_{t-1} + u_t \) with \( u_t \sim \text{i.i.d.} \mathcal{N} \left( 0, 1 \right) \) independent of \( \epsilon_t \sim \text{i.i.d.} \mathcal{N} \left( 0, \sigma^2_\epsilon \right) \), \( \sigma^2_\epsilon = 1 \), \( \pi^0 = 1 \), \( Z_0 = 0 \), \( D_t = 1 \) for all \( t \), and \( T = 100 \). We set \( T_0 = \lceil T \lambda_0 \rceil \) with \( \lambda_0 = 0.3, 0.5, 0.7 \) and consider different break sizes \( \delta^0 = 0.2, 0.3, 0.5, 1 \). The infeasible continuous record asymptotic distribution is computed assuming knowledge of the data generating process (DGP) as well as of the model parameters, i.e., using Theorem 2.4.2 where we set \( N^0_b, \| \delta^0 \|^2 \sigma^2, \xi_1, \xi_2 \) and \( \rho \) at their true values. The feasible counterparts are constructed with plug-in estimates of \( \xi_1, \xi_2, \rho \) and \( \left( N^0_b \| \delta^0 \|^2 / \sigma^2 \right) \rho \).

In practice we need to use a normalization for \( N \). A common choice is \( N = 1 \). Then \( \hat{\lambda}_b = T_b / T \) is a natural estimate of \( \lambda_0 \). The estimates \( \hat{\xi}_1 \) and \( \hat{\xi}_2 \) are given, respectively, by

\[
\hat{\xi}_1 = \frac{\tilde{\delta}' \left( T - T_b \right)^{-1} \sum_{k=1}^{T} z_{kh} z_{kh}' \delta}{\tilde{\delta}' \left( T_b \right)^{-1} \sum_{k=1}^{T} z_{kh} z_{kh}' \delta}, \quad \hat{\xi}_2 = \frac{\tilde{\delta}' \left( T - T_b \right)^{-1} \sum_{k=1}^{T} \hat{c}_{kh} z_{kh} z_{kh}' \delta}{\tilde{\delta}' \left( T_b \right)^{-1} \sum_{k=1}^{T} \hat{c}_{kh} z_{kh} z_{kh}' \delta},
\]

where \( \tilde{\delta} \) is the least-squares estimator of \( \delta_h \) and \( \hat{c}_{kh} \) are the least-squares residuals.

Use is made of the fact that the quadratic variation \( \langle Z, Z \rangle_1 \) is consistently estimated by \( \sum_{k=1}^{T} z_{kh} z_{kh}' / \hat{\lambda}_b \) while \( \Omega_{\psi,1} \) is consistently estimated by \( T \sum_{k=1}^{T} \hat{c}_{kh} z_{kh} z_{kh}' / \left( \hat{\lambda}_b \right) \).

The argument for \( \lambda_0 \| \delta^0 \|^2 \sigma^{-2} \rho \) is less immediate because it involves manipulating the scaling of each of the three estimates. Let \( \vartheta = \| \delta^0 \|^2 \sigma^{-2} \rho \). We use the following estimates for \( \vartheta \) and \( \rho \), respectively,

\[
\hat{\vartheta} = \hat{\rho} \| \delta \|^2 \left( T^{-1} \sum_{k=1}^{T} \hat{c}_{kh}^2 \right)^{-1}, \quad \hat{\rho} = \frac{\left( \hat{\delta}' \left( T_b \right)^{-1} \sum_{k=1}^{T} \hat{c}_{kh} z_{kh} z_{kh}' \delta \right)^2}{\hat{\delta}' \left( T_b \right)^{-1} \sum_{k=1}^{T} \hat{c}_{kh} z_{kh} z_{kh}' \delta},
\]

\[
\hat{\vartheta} = \hat{\rho} \| \delta \|^2 \left( T^{-1} \sum_{k=1}^{T} \hat{c}_{kh}^2 \right)^{-1}, \quad \hat{\rho} = \frac{\left( \hat{\delta}' \left( T_b \right)^{-1} \sum_{k=1}^{T} \hat{c}_{kh} z_{kh} z_{kh}' \delta \right)^2}{\hat{\delta}' \left( T_b \right)^{-1} \sum_{k=1}^{T} \hat{c}_{kh} z_{kh} z_{kh}' \delta},
\]
Whereas we have $\hat{\xi}_i \overset{p}{\rightarrow} \xi_i$ ($i = 1, 2$), the corresponding approximations for $\hat{\vartheta}$ and $\hat{\rho}$ are given by $\hat{\vartheta}/h \overset{p}{\rightarrow} \vartheta$ and $\hat{\rho}/h \overset{p}{\rightarrow} \rho$. To derive the latter two results we used that on the “fast time scale”, Assumption 2.6 implies that the errors have higher volatilities and thus the squared residual $e_{kh}^2$ needs to be multiplied by the factor $h^{1/2}$. Then, $h^{1/2} \sum_{k=1}^{T} e_{kh}^2 \overset{p}{\rightarrow} \sigma^2$. However, before taking the limit as $T \rightarrow \infty$ we can apply a change in variable which results in the extra factor $h$ canceling from the latter two estimates. In addition, our estimates can also be shown to be valid under the standard large-$N$ asymptotics with fixed shifts.

**Proposition 2.5.1.** Under the conditions of Theorem 2.4.2, (2.4.5) holds when using $\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ and $\hat{\vartheta}$ in place of $\xi_1, \xi_2, \rho$ and $\vartheta$, respectively.

The proposition implies that the limiting distribution can be simulated by using plug-in estimates. This allows feasible inference about the break date.

The results are presented in Figure 2.5-2.8 which also plot the classical shrinkage asymptotic distribution from Bai (1997). Here by signal-to-noise ratio we mean $\delta^0/\sigma_e$ which, given $\sigma_e^2 = 1$, equals the break size $\delta^0$. We can summarize the results as follows. The finite-sample distribution shares all of the features characterizing the density of the infeasible continuous record distribution across all break magnitudes and break locations. Furthermore, the density of the feasible version of the continuous record asymptotic distribution provides a good approximation to the infeasible one and thus also to the finite-sample distribution. This holds for both stationary and nonstationary regimes. The latter case corresponds to the following modification of model (2.5.1) where we specify

$$Z_t = 0.5Z_{t-1} + \sigma_{Z,t}e_{Z,t}, \quad \sigma_{Z,t} = \begin{cases} 0.86, & t \leq T^0_b, \\ 1.20, & t > T^0_b \end{cases}, \quad \text{Var}(e_t) = \begin{cases} 1, & t \leq T^0_b \\ 2, & t > T^0_b \end{cases}$$
so that the second moments of both the regressors and the errors roughly duplicates after $T_0$. Figure B.14-B.16 in Chapter 6 suggest interesting observations. First, the density of the finite-sample distribution is never symmetric even when $\lambda_0 = 0.5$. Second, it is always negatively skewed and the mode associated with the end sample point is higher than the mode associated with the starting sample point. Third, the density is never centered at the origin but slightly to the right of it. These features are easy to interpret. There is more variability in the post-break regime and it is more likely that the least-squares estimator overestimates the break date. The feasible density of the continuous record distribution provides a good approximation also in the case of nonstationary regimes. The supplementary material present additional results for a wide variety of models. In all cases, the feasible asymptotic distribution provides a good approximation to the finite-sample distribution.

### 2.6 Highest Density Region-based Confidence Sets

The features of the limit and finite-sample distributions suggest that standard methods to construct confidence intervals may be inappropriate; e.g., two-sided intervals around the estimated break date based on the standard deviations of the estimate. Our approach is rather non-standard and relates to Bayesian methods. In our context, the Highest Density Region (HDR) seems the most appropriate in light of the asymmetry and, especially, the multi-modality of the distribution for small break sizes. When the distribution is unimodal and symmetric, e.g., for large break magnitudes, the HDR region coincides with the standard confidence interval symmetric about the estimate. All that is needed to implement the procedure is an estimate of the density function. Once estimable quantities are plugged-in as explained in Section 2.5, we derive the empirical counterpart of the limiting distribution. Choose some signifi-
cance level $0 < \alpha < 1$ and let $\hat{P}_{T_b}$ denote the empirical counterpart of the probability distribution of $\rho N (\hat{\lambda}_b - \lambda_0^b)$ as defined in Theorem 2.4.2. Further, let $\hat{p}_{T_b}$ denote the density function defined by the Radon-Nikodym equation $\hat{p}_{T_b} = d\hat{P}_{T_b}/d\lambda_L$, where $\lambda_L$ denotes the Lebesgue measure.

**Definition 2.6.1. Highest Density Region:** Assume that the density function $f_Y (y)$ of some random variable $Y$ defined on a probability space $(\Omega_Y, \mathcal{F}_Y, \mathbb{P}_Y)$ and taking values on the measurable space $(\mathcal{Y}, \mathcal{B})$ is continuous and bounded. Then the $(1 - \alpha) 100\%$ Highest Density Region is a subset $S(\kappa_\alpha)$ of $\mathcal{Y}$ defined as $S(\kappa_\alpha) = \{y : f_Y (y) > \kappa_\alpha\}$ where $\kappa_\alpha$ is the largest constant that satisfies $\mathbb{P}_Y (Y \in S(\kappa_\alpha)) \geq 1 - \alpha$.

The concept of HDR and of its estimation has an established literature in statistics. The definition reported here is from Hyndman (1996); see also Samworth and Wand (2010) and Mason and Polonik (2008, 2009).

**Definition 2.6.2. Confidence Sets for $T_b^0$ under a Continuous Record:** Under Assumption 3.1-2.3, 2.4-2.5, 2.6-2.7 and under the “fast time scale”, a $(1 - \alpha) 100\%$ confidence set for $T_b^0$ is a subset of $\{1, \ldots, T\}$ given by

$$C(cv_\alpha) = \{T_b \in \{1, \ldots, T\} : T_b \in S(cv_\alpha)\},$$

where $S(cv_\alpha) = \{T_b : \hat{p}_{T_b} > cv_\alpha\}$ and $cv_\alpha$ satisfies $\sup_{cv_\alpha \in \mathbb{R}^+} \hat{P}_{T_b} (T_b \in S(cv_\alpha)) \geq 1 - \alpha$.

The confidence set $C(cv_\alpha)$ has a frequentist interpretation even though the concept of HDR is often encountered in Bayesian analyses since it associates naturally to the derived posterior distribution, especially when the latter is multi-modal. A feature of the confidence set $C(cv_\alpha)$ under our context is that, at least when the
size of the shift is small, it consists of the union of several disjoint intervals. The appeal of using HDR is that one can directly deal with such features. As the break size increases and the distribution becomes unimodal, the HDR becomes equivalent to the standard way of constructing confidence sets. In practice, one can proceed as follows.

**Algorithm 2. Confidence sets for** $T^0_b$: 1) Estimate by least-squares the break point and the regression coefficients from model (2.2.4); 2) Replace quantities appearing in (2.4.5) by consistent estimators as explained in Section 2.5; 3) Simulate the limiting distribution $\hat{P}_{T_b}$ from Theorem 2.4.2; 4) Compute the HDR of the empirical distribution $\hat{P}_{T_b}$ and include the point $T_b$ in the level $1-\alpha$ confidence set $C(c_{1-\alpha})$ if $T_b$ satisfies the conditions in Definition 2.6.2.

This procedure will not deliver contiguous confidence sets when the size of the break is small. Indeed, we find that in such cases, the overall confidence set for $T^0_b$ consists in general of the union of disjoint intervals if $\hat{T}_b$ is not in the tails of the sample. One is located around the estimate of the break date, while the others are in the pre- and post-break regimes. To provide an illustration, we consider a simple example involving a single draw from a simulation experiment. Figure 2.9 reports the HDR of the feasible limiting distribution of $\rho(\hat{T}_b - T^0_b)$ for a random draw from the model described by (2.5.1) with parameters $\pi^0 = 1$, $\beta^0 = 0$, unit second moments and autoregressive coefficient 0.6 for $Z_t$ and $\sigma^2_e = 1.2$ for the error term. We set $\lambda_0 = 0.35$, 0.5 and $\delta^0 = 0.3$, 0.8, 1.5. The sample size is $T = 100$ and the significance level is $\alpha = 0.05$. Note that the origin is at the estimated break date. The point on the horizontal axis corresponds to the true break date. In each plot, the black intervals on the horizontal axis correspond to regions of high density. The resulting confidence set is their union. Once a confidence region for $\rho(\hat{T}_b - T^0_b)$ is computed,
it is straightforward to derive a 95% confidence set for $T_0^b$. The top panel (left plot) reports results for the case $\delta^0 = 0.3$ and $\lambda_0 = 0.35$ and shows that the HDR is composed of two disjoint intervals. The estimated break date is $\hat{T}_b = 70$ and the implied 95% confidence set for $T_0^b$ is given by $C(c_{v_{0.05}}) = \{1, \ldots, 12\} \cup \{18, \ldots, 100\}$. This includes the true break date $T_0^b$ and the overall length is 95 observations. Table 2.1 reports for each method the coverage rate and length of the confidence sets for this example. The length of Bai’s (1997) confidence interval is 55 but does not include $T_0^b$. Elliott and Müller’s (2007) confidence set, denoted by $U_{T, \text{eq}}$ in Table 2.1, also does not include the true break date at the 90% confidence level, but does so at the 95% and its length is 95.

Figure 2.9 (middle panel) reports results for a larger break size $\delta^0 = 0.8$. The multi-modality is no longer present. When $\lambda_0 = 0.35$, the estimated break date is $\hat{T}_b = 25$ and the length of $C(c_{v_{0.05}})$ is 27 out of 100 observations given by $C(c_{v_{0.05}}) = \{12, \ldots, 38\}$. Relative to Elliott and Müller’s (2007) confidence sets which always cover $T_0^b$ in this example, the set constructed using the HDR is about 30% shorter. Bai’s (1997) method is again shorter than the other methods but it fails to cover the true value when $\lambda_0 = 0.35$. However, it does so when $\lambda_0 = 0.5$ and its length is 18. In the latter case (right plot), our method covers the true break date and the interval has almost the same length whereas Elliott and Müller’s (2007) approach results in an overall length of 35. Our method still provides accurate coverage when raising the break size to $\delta^0 = 1.5$ as can be seen from the bottom panel. When $\lambda_0 = 0.35$ (left panel), the estimated break date is $\hat{T}_b = 36$ and all three methods cover the true break date. The confidence interval from Bai’s (1997) method results in the shortest length since it includes only 8 points whereas our confidence interval includes 9 points and Elliott and Müller’s (2007) method includes 24 points.
This single simulation, by and large, anticipates the small-sample results in the Monte Carlo study reported in the next Section: Bai’s (1997) method results in a coverage probability below the nominal level; our method provides accurate coverage rates and the average length of the confidence set is always shorter than with Elliott and Müller’s (2007) method. It is evident that the confidence set for $T^0_b$ constructed using the HDR provides a useful summary of the underlying probability distribution of the break point estimator. For small break sizes, the HDR captures well the bi- or tri-modality of the density. As we raise the magnitude of the break, the HDR becomes a single interval around the estimated break point, which is a desirable property.

2.7 Small-Sample Properties of the HDR Confidence Sets

We now assess via simulations the finite-sample performance of the method proposed to construct confidence sets for the break date. We also make comparisons with alternative methods in the literature: Bai’s (1997) approach based on the large-$N$ shrinkage asymptotics; Elliott and Müller’s (2007), hereafter EM, method on inverting Nyblom’s (1989) statistic; the Inverted Likelihood Ratio (ILR) approach of Eo and Morley (2015), which essentially involves the inversion of the likelihood-ratio test of Qu and Perron (2007). We omit the technical details of these methods and refer to the original sources or Chang and Perron (2018) for a review and comparisons. The current state of this literature can be summarized as follows. The empirical coverage rates of the confidence intervals obtained from Bai’s (1997) method are often below the nominal level when the magnitude of the break is small. EM’s approach is by far the best in terms of providing an exact coverage rate that is closest to the nominal level. However, the lengths of the confidence sets are larger relative to the other methods, often by a very wide margin. The lengths can be very large (e.g., the whole sample)
even when the size of the break is very large; e.g., in models with serially correlated
errors or with lagged dependent variables. The ILR-based confidence sets display a
coverage probability often above the nominal level and this results in an average length
larger than with Bai’s (1997) method; further, it has a poor coverage probability for
all break sizes in models with heteroskedastic errors and autocorrelated regressors.
These findings suggest that there does not exist a method that systematically provides
over a wide range of empirically relevant models both good coverage probabilities and
reasonable lengths of the confidence sets, especially one that has good properties for
all break sizes, whether large or small.

The results to be reported suggest that our approach has two notable properties.
First, it provides adequate empirical coverage probabilities over all DGPs considered
for any size and/or location of the break in the sample. Second, the lengths of the con-
fidence sets are always shorter than those obtained with EM’s approach. Oftentimes,
the decrease in length is substantial and more so as the size of the break increases. To
have comparable coverage rates, we can compare the lengths of our confidence sets
with those obtained using Bai’s (1997) method only when the size of the break is mo-
derate to large. For those cases, our method delivers confidence sets with lengths only
slightly larger and they become equivalent as the size of the break increases. Also,
our HDR method has, overall, better coverage rates and shorter lengths compared to
ILR.

We consider discrete-time DGPs of the form

\[ y_t = D_t'\pi^0 + Z_t'\beta^0 + Z_t'\delta^0 1_{\{t > T_0\}} + e_t, \quad t = 1, \ldots, T, \]  

(2.7.1)

with \( T = 100 \) and, without loss of generality, \( \pi^0 = 0 \) (except in M4-M5, M7-M9).
We consider ten versions of (4.6.1): M1 involves a break in the mean of an \( i.i.d. \)
series with \( Z_t = 1 \) for all \( t \), \( D_t \) absent, and \( e_t \sim i.i.d. \mathcal{N}(0, 1) \); M2 is the same as M1 but with a simultaneous break in the variance such that \( e_t = \left( 1 + 1_{\{t > T_0\}} \right) u_t \) with \( u_t \sim i.i.d. \mathcal{N}(0, 1) \); M3 is the same as M1 but with stationary Gaussian AR(1) disturbances \( e_t = 0.3 e_{t-1} + u_t \), \( u_t \sim i.i.d. \mathcal{N}(0, 0.49) \); M4 is a partial structural change model with \( D_t = 1 \) for all \( t \), \( \pi^0 = 1 \) and \( Z_t = 0.5 Z_t + u_t \) with \( u_t \sim i.i.d. \mathcal{N}(0, 0.75) \) independent of \( e_t \sim i.i.d. \mathcal{N}(0, 1) \); M5 is similar to M4 but with \( u_t \sim i.i.d. \mathcal{N}(0, 0.49) \) and heteroskedastic disturbances given by \( e_t = v_t |Z_t| \) where \( v_t \) is a sequence of \( i.i.d. \mathcal{N}(0, 1) \) random variables independent of \( \{Z_t\} \); M6 is the same as M3 but with \( u_t \) drawn from a \( t_\nu \) distribution with \( \nu = 5 \) degrees of freedom; M7 is a model with a lagged dependent variable with \( D_t = y_{t-1} \), \( Z_t = 1 \), \( e_t \sim i.i.d. \mathcal{N}(0, 0.49) \), \( \pi^0 = 0.3 \) and \( Z_t^0 1_{\{t > T_0\}} \) is replaced by \( Z_t^0 (1 - \pi^0) \delta^0 1_{\{t > T_0\}} \); M8 is the same as M7 but with \( \pi^0 = 0.8 \) and \( e_t \sim i.i.d. \mathcal{N}(0, 0.04) \); M9 has FIGARCH(1,d,1) errors given by \( e_t = \sigma_t u_t \), \( u_t \sim \mathcal{N}(0, 1) \) and \( \sigma_t = 0.1 + \left( 1 - 0.2 L (1 - L)^d \right) e_t^2 \) where \( d = 0.6 \) is the order of differencing and \( L \) the lag operator, \( D_t = 1 \), \( \pi^0 = 1 \) and \( Z_t \sim i.i.d. \mathcal{N}(1, 1.44) \) independent of \( e_t \). M10 is similar to M5 but with ARFIMA(0.3, \( d \), 0) regressor \( Z_t \) with order of differencing \( d = 0.5 \), \( \text{Var}(Z_t) = 1 \) and \( e_t \sim \mathcal{N}(0, 1) \) independent of \( \{Z_t\} \). We set \( \beta^0 = 1 \) in all models, except in M7-M8 where \( \beta^0 = 0 \).

We use the appropriate limit distribution in each case when applying Bai's (1997), ILR and our method. When the errors are uncorrelated, we simply estimate variances. For model M3, in order to estimate the long-run variance we use for all methods Andrews and Monahan's (1992) AR(1) pre-whitened two-stage procedure to select the bandwidth with a quadratic spectral kernel. Except for M2, we report results for the statistic \( \hat{U}_T_{\text{eq}} \) proposed by EM, which imposes homogeneity across regimes (the results are qualitatively similar using the version \( \hat{U}_T_{\text{neq}} \), which allows heterogeneity across regimes). The methods of Bai (1997), the HDR and ILR
all require an estimate of the break date. We use the least-squares estimate obtained with a trimming parameter $\epsilon = 0.15$. When constructing the confidence set, we apply to all methods the same trimming corresponding to the degrees of freedom of the EM's (2007) statistic. This amounts to eliminating from consideration a few observations at the beginning and end of the sample, i.e., the number of parameters being estimated. We set the significance level at $\alpha = 0.05$, and the break occurs at date $[T\lambda_0]$, where $\lambda_0 = 0.2, 0.35, 0.5$. The results are presented in Table 4.5-2.11 for DGP M1-M10, respectively. The last row in each table includes the rejection probability of a 5%-level sup-Wald test using the asymptotic critical value in Andrews (1993). The sup-Wald rejection probability provides a measure of the magnitude of the break relative to the noise; low (large) values indicating a small (large) break. For models with predictable processes we use the procedure two-step described in Section 2.9.2.

Note that for M9-M10, one cannot apply the result of Theorem 2.4.2 and the associated method to obtain a feasible estimate of the distribution. Thus, we resort to the more general Theorem 2.4.1 which is valid under stochastic variances. The methods used to estimate the distribution is presented in Section 2.9.4.

Overall, the simulation results confirm previous findings about the performance of existing methods. Bai's (1997) method has a coverage rate below the nominal level when the size of the break is small. For example, for M3 with $\lambda_0 = 0.35$ and $\delta^0 = 0.6$, it is below 85% even though the sup-Wald test rejection rate is roughly 70%. When the size of the break is smaller, it systematically fails to cover the true break date with correct probability. These features evidently translate into lengths of the confidence intervals which are relatively shorter than with other methods, but given the differences in coverage rate such comparisons are meaningless. Only for moderate to large shifts is it legitimate to compare our method with that of Bai.
(1997), in which case the confidence intervals are similar in length; our HDR method delivers confidence sets slightly larger for medium-sized breaks (e.g., $\delta^0 = 1.5$) and the differences vanish as the size of the break increases. Overall, our HDR method and that of EM show accurate empirical coverage rates for all DGP considered. The ILR shows coverage rates systematically above the nominal level and, hence, an average length significantly longer than from Bai’s (1997) and our HDR methods in some cases (e.g., M2-M4), at least when the magnitude of the shift is small or moderate. As opposed to our HDR method, the ILR displays poor coverage rates for all break sizes in M5 which includes heteroskedastic errors.

The coverage rates of EM’s method are the most accurate, indeed very close to the nominal level. Turning to the comparisons of the length of the confidence sets, EM’s method almost always displays confidence sets which are larger than those from the other approaches. For example, for M3 with $\lambda_0 = 0.2$ and $\delta^0 = 0.6$, the average length of EM’s confidence set is 78.61 while that of the HDR method is 50.61. Such results are not particular to M3, but remain qualitatively the same across all models. When the size of the break is small to moderate, the lengths of the confidence sets obtained using our HDR method are shorter than those obtained using EM’s with differences ranging from 5% to 40%. The fact that EM’s method provides confidence sets that are larger becomes even more apparent as the size of the break increases. Over all DGPs considered, the average length of the HDR confidence sets are 40% to 70% shorter than those obtained with EM’s approach when the size of the shift is moderate to high. For example, when a lagged dependent variable is present (cf. M8), with $\lambda_0 = 0.5$ and $\delta^0 = 2$, the average length our our HDR confidence set is 6.34 while that of EM is 30.25, a reduction in length of about 75%. The results for M8, a change in mean with a lagged dependent variable and strong correlation, are quite revealing.
EM’s method yields confidence intervals that are very wide, increasing with the size of the break and for large breaks covering nearly the entire sample. This does not occur with the other methods. For instance, when $\lambda_0 = 0.5$ and $\delta^0 = 2$, the average length from the HDR method is 8.34 compared to 93.71 with EM’s. This is in line with the results in Chang and Perron (2018).

Finally, to show that our asymptotic results are valid and still provide good approximations with long-memory volatility, we consider M9 and M10. The results show that Bai’s method is not robust in that its coverage probability is below the nominal level even when the break magnitude is large. In contrast, the HDR-based method performs well and the average length of the confidence set is significantly shorter than that with EM’s or the ILR method, especially when the break is not at mid-sample for the latter.

In summary, the small-sample simulation results suggest that our continuous record HDR-based inference provides accurate coverage probabilities close to the nominal level and average lengths of the confidence sets shorter relative to existing methods. It is also valid and reliable under a wider range of DGPs including long-memory processes. Specifically noteworthy is the fact that it performs well for all break sizes, whether small or large.

2.8 Conclusions

We examined a partial structural change model under a continuous record asymptotic framework. We established the consistency, rate of convergence and asymptotic distribution of the least-squares estimator under very mild assumptions on an underlying continuous-time data generating process. Contrary to the traditional large-$N$ asymptotics, our asymptotic theory is able to provide good approximations and ex-
plain the following features. With the time horizon \([0, N]\) fixed, we can account for the asymmetric informational content provided by the pre- and post-break samples. The time span, the location of the break and the properties of the errors and regressors all jointly play a primary role in shaping the limit distribution of the estimate of the break date. The latter corresponds to the location of the extremum of a function of the (quadratic) variation of the regressors and of a Gaussian centered martingale process over a certain time interval. We derived a feasible counterpart using consistent plug-in estimates and show that it provides accurate approximations to the finite-sample distributions. In particular, the asymptotic and finite-sample distributions are (i) never symmetric unless the break point is located at mid-sample and the regimes are stationary, and (ii) positively (resp., negatively) skewed if the break point occurs in the first (resp., second)-half of the sample. This holds true across different break magnitudes except for very large break sizes in which case the distribution is symmetric. We used our limit theory to construct confidence sets for the break date based on the concept of Highest Density Region. Our method is simple to implement and relies entirely on the derived feasible asymptotic distribution. Overall, it delivers accurate coverage probabilities and relatively short average lengths of the confidence sets across a variety of data-generating mechanisms. Importantly, it does so irrespective of the magnitude of the break, whether large or small, a notoriously difficult problem in the literature.

2.9 Appendix to Chapter 2

2.9.1 Description of the Limiting Process in Theorem 2.4.1

We describe the probability setup underlying the limit process of Theorem 2.4.1. Note that \(Z^*_\Delta e/h^{1/2} = h^{-1/2} \sum_{k=T_b+1}^{T_b} z_{kh}e_{kh}\) if \(T_b \leq T_b^0\). Consider an additional me-
asurable space \((\Omega^*, \mathcal{F}^*)\) and a transition probability \(H (\omega, d\omega^*)\) from \((\Omega, \mathcal{F})\) into \((\Omega^*, \mathcal{F}^*)\). Next, we can define the products \(\tilde{\Omega} = \Omega \times \Omega^*, \tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}^*, \tilde{P} (d\omega, d\omega^*) = P (d\omega) H (\omega, d\omega^*)\). This defines an extension \(\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}\right)\) of the original space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)\). We also consider another filtration \(\{\tilde{\mathcal{F}}_t\}_{t \geq 0}\) which takes the following product form \(\tilde{\mathcal{F}}_t = \cap_{s > t} \mathcal{F}_s \otimes \mathcal{F}^*_s\) where \(\{\mathcal{F}^*_t\}_{t \geq 0}\) is a filtration on \((\Omega^*, \mathcal{F}^*)\). For the transition probability \(H\), we consider the simple form \(H (\omega, d\omega^*) = P^* (d\omega^*)\) for some probability measure \(P^*\) on \((\Omega^*, \mathcal{F}^*)\). This constitutes a “very good” product filtered extension. Next, assume \(\left(\Omega^*, \mathcal{F}^*, (\mathcal{F}^*_t)_{t \geq 0}, P^*\right)\) supports \(p\)-dimensional \(\{\mathcal{F}^*_t\}\)-standard Wiener processes \(W^{i*} (v) (i = 1, 2)\). Finally, we postulate the process \(\Omega_{Ze,t}\) with entries \(\sum Z_{i,j}^2 (v)\) to admit a progressively measurable \(p \times p\) matrix-valued process (i.e., a symmetric “square-root” process) \(\sigma_{Ze}\), satisfying \(\Omega_{Ze} = \sigma_{Ze} \sigma_{Ze}^T\), with the property that \(\|\sigma_{Ze}\|^2 \leq K \|\Omega_{Ze}\|\) for some \(K < \infty\). Define the process \(\mathcal{W} (v) = \mathcal{W}_1 (v)\) if \(v \leq 0\), and \(\mathcal{W} (v) = \mathcal{W}_2 (v)\) if \(v > 0\), where \(\mathcal{W}_1 (v) = \int_{N^0_0}^{N^0_0 + v} \sigma_{Ze,s} dW^{1*}_s\) and \(\mathcal{W}_2 (v) = \int_{N^0_0}^{N^0_0 + v} \sigma_{Ze,s} dW^{2*}_s\) with components \(\mathcal{W}^{(j)} (v) = \sum_{r=1}^{p} \int_{N^0_0}^{N^0_0 + v} \sigma_{Ze,s}^{(j,r)} dW^{1*}_s\) if \(v \leq 0\) and \(\mathcal{W}^{(j)} (v) = \sum_{r=1}^{p} \int_{N^0_0}^{N^0_0 + v} \sigma_{Ze,s}^{(j,r)} dW^{2*}_s\) if \(v > 0\). The process \(\mathcal{W} (v)\) is well defined on the extension \(\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t \geq 0}, \tilde{P}\right)\), and furthermore, conditionally on \(\mathcal{F}\), is a two-sided centered continuous Gaussian process with independent increments and (conditional) covariance

\[
\mathbb{E} \left( \mathcal{W}^{(u)} (v) \mathcal{W}^{(j)} (v) \right) = \Omega^{(u,j)} (v) = \begin{cases} 
\Omega^{(u,j)}_{\mathcal{W},1} (v), & \text{if } v \leq 0 \\
\Omega^{(u,j)}_{\mathcal{W},2} (v), & \text{if } v > 0 
\end{cases}
\]

(2.9.1)

where \(\Omega^{(u,j)}_{\mathcal{W},1} (v) = \int_{N^0_0}^{N^0_0 + v} \Omega_{Ze,s}^{(u,j)} ds\) and \(\Omega^{(u,j)}_{\mathcal{W},2} (v) = \int_{N^0_0}^{N^0_0 + v} \Omega_{Ze,s}^{(u,j)} ds\). Therefore, \(\mathcal{W} (v)\) is conditionally on \(\mathcal{F}\), a continuous martingale with “deterministic” quadratic covariation process \(\Omega_{\mathcal{W}}\). The continuity of \(\Omega_{\mathcal{W}}\) signifies that \(\mathcal{W} (v)\) is not only conditionally Gaussian but also a.s. continuous. Precise treatment of this result can be found in
Section II.7 of Jacod and Shiryaev (2003).

2.9.2 Asymptotic Results for the Model with Predictable Processes

In this section, we present asymptotic results allowing for predictable processes that include a constant and a lagged dependent variable among the regressors. Recall model (2.2.5). Let \( \beta^0 = \left( \mu_0^0, \alpha_0^0, (\delta_{Z,1}^0)^\prime \right)^\prime, \delta^0 = \left( \mu_0^0, \alpha_0^0, (\delta_{Z,2}^0 - \delta_{Z,1}^0)^\prime \right)^\prime, \left( (\beta^0)^\prime, ((\delta^0)^\prime) \right)^\prime \in \Theta_0, \) and \( x_{kh} = ((\mu_1/h/\mu_1^0) h, (\alpha_1/h/\alpha_1^0) Y_{(k-1)} h, \Delta_h D_k', \Delta_h Z_k'). \) In matrix format, the model is \( Y = X \beta^0 + Z_0 \delta^0 + e, \) where now \( X \) is \( T \times (p + q + 2) \) and \( Z_0 = X R, R \triangleq \left( (I_2, 0_{2 \times p})', (0_{(p+q) \times 2}, R) \right)^\prime, \) with \( R \) as defined in Section 2.2.1.

Natural estimates of \( \beta^0 \) and \( \delta^0 \) minimize the following criterion function,

\[
\frac{1}{T} \sum_{k=1}^{T} \left( \Delta_h Y_k - \beta^0 \int_{(k-1)h}^{kh} X_s ds - \delta^0 \int_{(k-1)h}^{kh} Z_s ds \right)^2
\]

\[
= \frac{1}{T} \sum_{k=1}^{T} \left( \Delta_h Y_k - \mu_1^h h - \alpha_1^h \int_{(k-1)h}^{kh} Y_s ds - \pi' \Delta_h D_k \right.
\]

\[
- \delta_{Z,1} \Delta_h Z_k 1 \{k \leq T_b\} - \delta_{Z,2} \Delta_h Z_k 1 \{k > T_b\} \right)^2.
\]

Hence, we define our LS estimator as the minimizer of the following approximation to (2.9.2):

\[
\frac{1}{T} \sum_{k=1}^{T} \left( \Delta_h Y_k - \mu_1^h h - \alpha_1^h \int_{(k-1)h}^{kh} Y_s ds - \pi' \Delta_h D_k \right.
\]

\[
- \delta_{Z,1} \Delta_h Z_k 1 \{k \leq T_b\} - \delta_{Z,2} \Delta_h Z_k 1 \{k > T_b\} \right)^2.
\]

Such approximations are common [cf. Christopeit (1986), Lai and Wei (1983) and Mel’nikov and Novikov (1988) and the more recent work of Galtchouk and Konev]
Define $\Delta h \tilde{Y}_k \triangleq h^{1/2} \Delta h Y_k$ and $\Delta h \tilde{V}_k = h^{1/2} \Delta h V_k \left( \pi^0, \delta_{Z,1}^0, \delta_{Z,2}^0 \right)$ where

$$\Delta h V_k \left( \pi^0, \delta_{Z,1}^0, \delta_{Z,2}^0 \right) \triangleq \begin{cases} (\pi^0)' \Delta h D_k + (\delta_{Z,1}^0)' \Delta h e_k^t, & \text{if } k \leq T_b^0, \\ (\pi^0)' \Delta h D_k + (\delta_{Z,2}^0)' \Delta h e_k^t, & \text{if } k > T_b^0. \end{cases}$$

The small-dispersion format of our model is then

$$\Delta h \tilde{Y}_k = \left( \mu_{10} h + \alpha_{10} \tilde{Y}_{(k-1)h} h \right) 1 \{k \leq T_b^0 \} + \left( \mu_{20} h + \alpha_{20} \tilde{Y}_{(k-1)h} h \right) 1 \{k > T_b^0 \} + \Delta h \tilde{V}_k \left( \pi^0, \delta_{Z,1}^0, \delta_{Z,2}^0 \right).$$

This re-parametrization emphasizes that asymptotically our model describes small disturbances to the approximate dynamical system

$$d\tilde{Y}_t^0/dt = \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_t^0 \right) 1 \{t \leq N_b^0 \} + \left( \mu_2^0 + \alpha_2^0 \tilde{Y}_t^0 \right) 1 \{t > N_b^0 \}.$$

The process $\{\tilde{Y}_t^0\}_{t \geq 0}$ is the solution to the underlying ordinary differential equation. The LS estimate of the break point is then defined as $\hat{T}_b \triangleq \arg \max_{T_b} Q_T (T_b)$, where

$$Q_T (T_b) \triangleq Q_T \left( \hat{\beta} (T_b), \hat{\delta} (T_b), T_b \right) = \hat{\delta}' (Z_2' M Z_2) \hat{\delta},$$

and the LS estimates of the regression parameters are

$$\hat{\theta} \triangleq \arg \min_{\theta \in \Theta_0} h \left( S_T \left( \hat{\beta}, \delta, \hat{T}_b \right) - S_T \left( \hat{\beta}^0, \delta^0, T_b^0 \right) \right),$$

where $S_T$ is the sum of square residuals. With the exception of our small-dispersion assumption and consequent more lengthy derivations, our analysis remains the same as in the model without predictable processes. Hence, the asymptotic distribution of the break point estimator is derived under the same setting as in Section 2.4. We show that the limiting distribution is qualitatively equivalent to that in Theorem
Assumption 2.8. Assumption 2.3 and 2.5 hold. Assumption 3.1, 2.2 and 2.4 now apply to the last \( p \) (resp. \( q \)) elements of the process \( \{Z_t\}_{t \geq 0} \) (resp. \( \{D_t\}_{t \geq 0} \)).

Proposition 2.9.1. Consider model (2.2.5). Under Assumption 2.8: (i) \( \lambda_b \xrightarrow{P} \lambda_0 \); (ii) for every \( \varepsilon > 0 \) there exists a \( K > 0 \) such that for all large \( T, P(T \mid \lambda_b - \lambda_0 > K \|\delta^0\|^{-2} \sigma^2) < \varepsilon. \)

Assumption 2.9. Let \( \delta_h = h^{1/4} \delta^0 \) and for \( i = 1, 2 \) \( \mu_i^h = h^{1/4} \mu_i^0 \) and \( \alpha_i^h = h^{1/4} \alpha_i^0 \), and assume that for all \( t \in (N_b^0 - \epsilon, N_b^0 + \epsilon), \) with \( \epsilon \downarrow 0 \) and \( T^{1-\kappa} \epsilon \to B < \infty, \)

\[ 0 < \kappa < 1/2, \ E \left[ (\Delta_h e_t)^2 \mid \mathcal{F}_{t-h} \right] = \sigma^2_{h,t} \Delta t \text{ P-a.s.,} \]

where \( \sigma_{h,t} \triangleq \sigma_h \sigma_{e,t} \) with \( \sigma_h \triangleq h^{-1/4} \sigma \).

Furthermore, define the normalized residual \( \Delta_h \tilde{e}_t \) as in (2.4.1). We shall derive a stable convergence in distribution for \( Q_T(\cdot, \cdot) \) as defined in Section 2.4. The description of the limiting process is similar to the one presented in the previous section. However, here we shall condition on the \( \sigma \)-field \( \mathcal{G} \) generated by all latent processes appearing in the model. In view of its properties, the \( \sigma \)-field \( \mathcal{F} \) admits a regular version of the \( \mathcal{G} \)-conditional probability, denoted \( H(\omega, d\omega^*) \). The limit process is then realized on the extension \( \left( \tilde{\Omega}, \tilde{\mathcal{F}}, \left\{ \tilde{\mathcal{F}}_t \right\}_{t \geq 0}, \tilde{P} \right) \) of the original filtered probability space as explained in Section 2.9.1. We again introduce a two-sided Gaussian process \( \mathcal{W}_{Ze}(\cdot) \) with a different dimension in order to accommodate for the presence of the predictable regressors in the first two columns of both \( X \) and \( Z \). That is, \( \mathcal{W}_{Ze}(\cdot) \) is a \( p \)-dimensional process which is \( \mathcal{G} \)-conditionally Gaussian and has \( P \)-a.s. continuous sample paths. We then have the following theorem.

Theorem 2.9.1. Consider model (2.9.3). Under Assumption 2.8-2.9: (i) \( \lambda_b \xrightarrow{P} \lambda_0 \); (ii) for every \( \varepsilon > 0 \) there exists a \( K > 0 \) such that for all large \( T, \)

\[ P(T \mid \lambda_b - \lambda_0 > K \|\delta^0\|^{-2} \sigma^2) < \varepsilon. \]
\( \sigma^2 < \varepsilon; \) (iii) under the “fast time scale”:

\[
N \left( \tilde{\lambda}_b - \lambda_0 \right) \overset{\mathcal{L}}{\Rightarrow} \mathcal{N} \left( \frac{- N_0^b \lambda}{\| \delta^0 \|^2 - 2 \sigma^2}, \frac{N - N_0^b}{\| \delta^0 \|^2 - 2 \sigma^2} \right)
\]

argmax \( v \in \left[ - \frac{N_0^b}{\| \delta^0 \|^2 - 2 \sigma^2}, \frac{N - N_0^b}{\| \delta^0 \|^2 - 2 \sigma^2} \right] \left\{ - \left( \delta^0 \right)' \Lambda (v) \delta^0 + 2 \left( \delta^0 \right)' \mathcal{W} (v) \right\}, \quad (2.9.5)
\]

where \( \Lambda (v) \) is a process given by

\[
\Lambda (v) \triangleq \begin{cases} 
\Lambda_1 (v), & \text{if } v \leq 0 \\
\Lambda_2 (v), & \text{if } v > 0
\end{cases}
\]

\[
\Lambda_1 (v) \triangleq \begin{bmatrix} \int_{N_0^b}^{N_0^b + v} ds & \int_{N_0^b}^{N_0^b + v} \tilde{Y}_s ds & 0_{1 \times p} \\
0_{p \times 1} & 0_{p \times 1} & (Z, Z)_1 (v)
\end{bmatrix},
\]

and \( \Lambda_2 (v) \) is defined analogously, where \((Z, Z)_1 (v)\) is the \( p \times p \) predictable quadratic covariation process of the pair \((Z_{(u)}^v, Z_{(j)}^v)\), \( 3 \leq u, j \leq p \) and \( v \leq 0 \). The process \( \mathcal{W} (v) \) is, conditionally on \( \mathcal{F} \), a two-sided centered Gaussian martingale with independent increments.

When \( v \leq 0 \), the limit process \( \mathcal{W} (v) \) is defined as follows,

\[
\mathcal{W}^{(j)} (v) = \begin{cases} 
\int_{N_0^b}^{N_0^b + v} dW_{e,s}, & j = 1, \\
\int_{N_0^b}^{N_0^b + v} \tilde{Y}_s dW_{e,s}, & j = 2, \\
\mathcal{W}^{(j-2)} (v), & j = 3, \ldots, p + 2,
\end{cases}
\]

where \( \mathcal{W}^{(i)} (v) \triangleq \sum_{r=1}^{p} \int_{N_0^b}^{N_0^b + v} \sigma_{Z_{(i)}^v, s} dW_{1} (r) (i = 1, \ldots, p) \) and analogously when \( v > 0 \). That is, \( \mathcal{W}_{Z_e} (v) \) corresponds to the process \( \mathcal{W} (v) \) used for the benchmark model (and so are \( W_{1}^{1 \ast}, W_{2}^{2 \ast} \) and \( \Omega_{Z_e,s} \) below). Its conditional covariance is given by

\[
\mathcal{E} \left( \mathcal{W}^{(u)} (v) \mathcal{W}^{(j)} (v) \right) = \Omega_{\mathcal{W}^{(u)}, \mathcal{W}^{(j)}} (v) = \begin{cases} 
\Omega_{\mathcal{W}^{(u)}, \mathcal{W}^{(j)}} (v), & \text{if } v \leq 0, \\
\Omega_{\mathcal{W}^{(u)}, \mathcal{W}^{(j)}} (v), & \text{if } v > 0
\end{cases}, \quad (2.9.6)
\]
Corollary 2.9.1. \( \Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_b^0}^{N_b^0 + v} \sigma_{e,s}^2 ds \), if \( u, j = 1; \) \( \Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_b^0}^{N_b^0 + v} \tilde{Y}_s^2 \sigma_{e,s}^2 ds \), if \( u, j = 2; \) \( \Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_b^0}^{N_b^0 + v} \tilde{Y}_s^2 \sigma_{e,s}^2 ds \), if \( 1 \leq u, j \leq 2, u \neq j; \) \( \Omega^{(u,j)}_{\mathcal{W},1}(v) = 0 \), if \( u = 1, 2, j = 3, \ldots, p; \) \( \Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_b^0}^{N_b^0 + v} \Omega_{\mathcal{W},e,s} ds \), if \( 3 \leq u, j \leq p + 2 \); and similarly for \( \Omega^{(u,j)}_{\mathcal{W},2}(v) \).

The asymptotic distribution is qualitatively the same as in Theorem 2.4.1. When the volatility processes are deterministic, we have convergence in law under the Skorhokod topology to the same limit process \( \mathcal{W}(\cdot) \) with a Gaussian unconditional law. The case with stationary regimes is described as follows.

**Assumption 2.10.** \( \Sigma^* = \{ \mu, t, \Sigma, t, \sigma_{e,i} \}_{t \geq 0} \) is deterministic and the regimes are stationary.

Let \( W^*_i, i = 1, 2 \), be two independent standard Wiener processes defined on \([0, \infty)\), starting at the origin when \( s = 0 \). Let

\[
\mathcal{Y}(s) = \begin{cases} 
-\frac{|s|}{2} + W^*_1(s), & \text{if } s < 0 \\
-\frac{(\delta^0)'\Lambda_1\delta^0 |s|}{(\delta^0)'\Lambda_1\delta^0} + \left(\frac{(\delta^0)'\Omega_{\mathcal{W},1}\delta^0}{(\delta^0)'\Lambda_1\delta^0}\right)^{1/2} W^*_2(s), & \text{if } s \geq 0.
\end{cases}
\]

**Corollary 2.9.1.** Under Assumption 2.8-2.10,

\[
\frac{(\delta^0)'\Lambda_1\delta^0}{(\delta^0)'\Omega_{\mathcal{W},1}\delta^0} \Rightarrow \argmax_{s \in \left[-\frac{N_b^0}{\|\delta^0\|_2^2\|\Lambda_1\delta^0\|^2}, \frac{N_b^0}{\|\delta^0\|_2^2\|\Lambda_1\delta^0\|^2}\right]} \mathcal{Y}(s). \tag{2.9.7}
\]

In the next two corollaries, we assume stationary errors across regimes. Corollary 2.9.3 considers the basic case of a change in the mean of a sequence of i.i.d. random variables. Let

\[
\mathcal{Y}_{sta}(s) = \begin{cases} 
-\frac{|s|}{2} + W^*_1(s), & \text{if } s < 0 \\
-\frac{(\delta^0)'\Lambda_2\delta^0 |s|}{(\delta^0)'\Lambda_1\delta^0} + \left(\frac{(\delta^0)'\Lambda_2\delta^0}{(\delta^0)'\Lambda_1\delta^0}\right)^{1/2} W^*_2(s), & \text{if } s \geq 0.
\end{cases}
\]
and

\[ \mathcal{V}_{\mu,\text{sta}}(s) = \begin{cases} -\frac{|s|}{2} + W_1^*(s), & \text{if } s < 0 \\ -\frac{|s|}{2} + W_2^*(s), & \text{if } s \geq 0 \end{cases} \]

**Corollary 2.9.2.** Under Assumption 2.8-2.10 and assuming that the second moments of the residual process are stationary across regimes, \( \sigma_{e,s} = \sigma \) for all \( 0 \leq s \leq N \),

\[
\frac{(\delta^0)' \Lambda_1 \delta^0}{\sigma^2} N \left( \hat{\lambda}_b - \lambda_0 \right) \Rightarrow \argmax_{s \in \left[-N_b^0 \sigma^2, N-N_b^0 \sigma^2\right]} \mathcal{V}_{\text{sta}}(s).
\]

**Corollary 2.9.3.** Under Assumption 2.8-2.10, with \( \pi^0 = 0 \), \( \delta_{Z,i}^0 = 0 \), and \( \alpha_i^0 = 0 \) for \( i = 1, 2 \):

\[
\left( \frac{\delta^0}{\sigma} \right)^2 N \left( \hat{\lambda}_b - \lambda_0 \right) \Rightarrow \argmax_{s \in \left[-N_b^0 (\delta^0/\sigma)^2, (N-N_b^0) (\delta^0/\sigma)^2\right]} \mathcal{V}_{\mu,\text{sta}}(s).
\]

**Remark 2.9.1.** The last corollary reports the result for the simple case of a shift in the mean of an i.i.d. process. This case was recently considered by Jiang et al. (2018) under a continuous-time setting in their Theorem 4.2-(b) which is similar to our Corollary 2.9.3. Our limit theory differs in many respects, besides being obviously more general. Jiang et al. (2018) only develop an infeasible distribution theory for the break date estimator whereas we also derive a feasible version. This is because we introduce an assumption about the drift in order to “keep” it in the asymptotics. The limiting distribution is also derived under a different asymptotic experiment (cf. Assumption 2.9 above and the change of time scale as discussed in Section 2.4). A direct consequence is that the estimate of the break fraction is shown to be consistent as \( h \downarrow 0 \) whereas Jiang et al. (2018) do not have such a result.

The results are similar to those in the benchmark model. However, the estimation
of the regression parameters is more complicated because of the identification issues about the parameters associated with predictable processes. Nonetheless, our model specification allows us to construct feasible estimators. Given the small-dispersion specification in (2.9.3), we propose a two-step estimator. In fact, (2.9.4) essentially implies that asymptotically the evolution of the dependent variable is governed by a deterministic drift function given by \( \mu_0^0 + \alpha_0^0 Y^0_0 \) (resp., \( \mu_2^0 + \alpha_2^0 Y^0_0 \)) if \( t \leq N_0^b \) (resp., \( t > N_0^b \)). Thus, in a first step we construct least-squares estimates of \( \mu_i^0 \) and \( \alpha_i^0 \) \((i = 1, 2)\). Next, we subtract the estimate of the deterministic drift from the dependent variable so as to generate a residual component that will be used (after rescaling) as a new dependent variable in the second step where we construct the least-squares estimates of the parameters associated with the stochastic semimartingale regressors.

**Proposition 2.9.2.** Under Assumption 2.8-2.9, as \( h \downarrow 0 \), \( \hat{\theta} \xrightarrow{P} \theta^0 \).

The consistency of the estimate \( \hat{\theta} \) is all that is needed to carry out our inference procedures about the break point \( T_0^b \) presented in Section 4.4. The relevance of the result is that even though the drifts cannot in general be consistently estimated, we can, under our setting, estimate the parameters entering the limiting distribution; i.e., \( \mu_i^0 \) and \( \alpha_i^0 \).

### 2.9.3 Proofs of Theorem 2.4.1-2.4.2

#### 2.9.3.1 Proof of Theorem 2.4.1

**Proof.** Let us focus on the case \( T_b(v) \leq T_b^0 \) (i.e., \( v \leq 0 \)). The change of time scale is obtained by a change in variable. On the old time scale, by Proposition 2.4.1, \( N_b(v) \) varies on the time interval \([N_0^b - |v| h^{1-\kappa}, N_0^b + |v| h^{1-\kappa}]\) with \( v \in [-C, C] \). Lemma 2.4.1 shows that the conditional first moment of \( Q_T(T_b(v)) - Q_T(T_b^0) \) is determined by that of \(-\delta_h^0 (Z^\gamma_\Delta Z_\Delta) \delta_h \pm 2 \delta_h^\gamma (Z_\Delta \epsilon)\). Next, we rescale time with \( s \rightarrow t \triangleq \psi_h^{-1} s \) on \( \mathcal{D}(C) \). This is achieved by rescaling the criterion function \( Q_T(T_b(u)) - Q_T(T_b^0) \) by
the factor $\psi_h^{-1}$. First, note that the processes $Z_t$ and $e_t^*$ [recall (2.2.3) and (2.4.1)] are rescaled as follows on $D(C)$. Let $Z_{\psi,s} \triangleq \psi_h^{-1/2}Z_s$, $W_{\psi,e,s} \triangleq \psi_h^{-1/2}W_{e,s}$ and note that

$$dZ_{\psi,s} = \psi_h^{-1/2} \sigma_{Z,s} dW_{Z,s}, \quad dW_{\psi,e,s} = \psi_h^{-1/2} \sigma_{e,s} dW_{e,s}, \quad \text{with } s \in D(C). \quad (2.9.8)$$

For $s \in [N_0^b - Ch^{1-\kappa}, N_0^b + Ch^{1-\kappa}]$ let $v = \psi_h^{-1}(N_0^b - s)$, and by using the properties of $W_{s,s}$ and the fact that $\sigma_{Z,s}, \sigma_{e,s}$ are $F_s$-measurable, we have

$$dZ_{\psi,t} = \sigma_{Z,t} dW_{Z,t}, \quad dW_{\psi,e,t} = \sigma_{e,t} dW_{e,t}, \quad \text{with } t \in D^*(C). \quad (2.9.9)$$

This can be used into the following quantities for $N_b(v) \in D(C)$. First,

$$\psi_h^{-1}Z'_\Delta Z_\Delta = \sum_{k=T_b(v)+1}^{T_b^0} z_{\psi,kh} \bar{z}_{\psi,kh},$$

which by (2.9.8)-(2.9.9) is such that

$$\psi_h^{-1}Z'_\Delta Z_\Delta = \sum_{k=T_b^0+[v/h]}^{T_b^0} z_{kh} \bar{z}_{kh}, \quad v \in D^*(C). \quad (2.9.10)$$

Using the same argument:

$$\psi_h^{-1}Z'_\Delta \bar{e} = \sum_{k=T_b^0+[v/h]}^{T_b^0} z_{kh} \bar{e}_{kh}, \quad v \in D^*(C). \quad (2.9.11)$$

Now $N_b(v)$ varies on $D^*(C)$. Furthermore, for sufficiently large $T$, Lemma 2.4.1 gives

$$Q_T(T_b) - Q_T(T_b^0) = -\delta_h (Z'_\Delta Z_\Delta) \delta_h \pm 2\delta'_h (Z'_\Delta e) + o_p\left(h^{1/2}\right),$$

and thus, when multiplied by $h^{-1/2}$, we have

$$\overline{Q}_T(T_b) = -\left(\delta^0\right)' Z'_\Delta Z_\Delta \left(\delta^0\right) \pm 2\left(\delta^0\right)' \left(h^{-1/2}Z'_\Delta \bar{e}\right) + o_p(1),$$
since on $D^*(C)$, $e_{kh} \sim \text{i.n.d.} \mathcal{N} \left(0, \sigma^2_{h,k-1} h\right)$, $\sigma_{h,k} = O \left(h^{-1/4}\right)$ $\sigma_{e,k}$ and $\tilde{e}_{kh}$ is the normalized error [i.e., $\tilde{e}_{kh} \sim \text{i.n.d.} \mathcal{N} \left(0, \sigma^2_{e,k-1} h\right)$] defined in (2.4.1). Hence, according to the re-parametrization introduced in the main text, we examine the behavior of

\[
\overline{Q}_T(\theta^*) = -\left(\delta^0\right)' \left( \sum_{k=T_b+1}^{T_b^0} z_{kh} z_{kh}' \right) \delta^0 + 2 \left(\delta^0\right)' \left( h^{-1/2} \sum_{k=T_b+1}^{T_b^0} z_{kh} \tilde{e}_{kh} \right). \tag{2.9.12}
\]

For the first term, a law of large numbers will be applied which yields convergence in probability toward some quadratic covariation process. For the second term, we observe that the finite-dimensional convergence follows essentially from results in Jacod and Protter (2012) (we indicate the precise theorems below) after some adaptation to our context. Hence, we shall then verify the asymptotic stochastic equicontinuity of the sequence of processes $\{\overline{Q}_T(\cdot), T \geq 1\}$. Let us associate to the continuous-time index $t$ a corresponding $D^*(C)$-specific index $t_v$. This means that each $t_v$ identifies a distinct $t$ in $D^*(C)$ through $v$ as define above. More specifically, for each $(\cdot, v) \in D^*(C)$, define the new functions

\[
J_{Z,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} z_{kh} z_{kh}', \quad J_{e,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} z_{kh} \tilde{e},
\]

for $(T_b(v) + 1) h \leq t_v < (T_b(v) + 2) h$. For $v \leq 0$, the lower limit of the summation is $T_b(v) + 1 = T_b^0 + [v/h]$ and thus the number of observations in each sum increases at rate $1/h$. The functions $\{J_{Z,h}(v)\}$ and $\{J_{e,h}(v)\}$ have discontinuous, although càdlàg, paths and thus they belong to $\mathbb{D}(D^*(C), \mathbb{R})$. Since $Z_t^{(j)}$ $(j = 1, \ldots, p)$ is a continuous Itô semimartingale, we have by Theorem 3.3.1 in Jacod and Protter (2012) that $J_{Z,h}(v) \overset{u.c.p.}{\Rightarrow} [Z, Z]_1(v)$, where $[Z, Z]_1(v) \triangleq [Z, Z]_{h \lfloor N^0_b / h \rfloor} - [Z, Z]_{h \lfloor t_v / h \rfloor}$, and recall by Assumption 2.2 that $[Z, Z]_1(v)$ is equivalent to $\langle Z, Z \rangle_1(v)$ where $\langle Z, Z \rangle_1(v) = \langle Z, Z \rangle_{h \lfloor t_v / h \rfloor}$. Next, let $\mathcal{W}_h(v) = h^{-1/2} J_{e,h}(v)$ and $\mathcal{W}_1(v) = J_{N^0_b + v}^{N^0_b + v} \sigma_{Ze,s} dW_s^{1*}$ where $W_s^{1*}$ is defined in Section 2.9.1. By Theorem 5.4.2 in Jacod and Protter (2012) we
have $\mathcal{W}_h (v) \overset{\mathcal{L}}{\rightarrow} \mathcal{W}_i (v)$ under the Skorokhod topology. Note the that both limit processes $[Z, Z]_1 (v)$ and $\mathcal{W}_i (v)$ are continuous. This restores the compatibility of the Skorokhod topology with the natural linear structure of $\mathbb{D} (\mathcal{D}^* (C), \mathbb{R})$. For $v \leq 0$, the finite-dimensional stable convergence in law for $\overline{Q}_T (\cdot)$ then follows: $\overline{Q}_T (\theta^*) \overset{\mathcal{L}}{\rightarrow} \mathbb{E} \left[ z_{kh} e_{kh} \right] \overset{\mathcal{L}}{\rightarrow} \delta^0 + 2 (\delta^0)^' \mathcal{W}_i (v)$, where $\overset{\mathcal{L}}{\rightarrow}$ signifies finite-dimensional stable convergence in law. Similarly, for $v > 0$, $\overline{Q}_T (\theta^*) \overset{\mathcal{L}}{\rightarrow} (Z, Z)_2 (\cdot) \overset{\mathcal{L}}{\rightarrow} \theta^0 + 2 (\delta^0)^' \mathcal{W}_i (v)$.

Next, we verify the asymptotic stochastic equicontinuity of the sequence of processes $\{\overline{Q}_T (\cdot), T \geq 1\}$. For $1 \leq i \leq p$, let $\xi_{i,k} \triangleq \xi_{Z,h,k} \triangleq \mathbb{E} \left[ z_{kh} e_{kh} \right] \overset{\mathcal{L}}{\rightarrow} \mathbb{E} \left[ z_{kh} e_{kh} \right] \mathcal{F}_{(k-1)h}$, and $\xi_{j,k}^* \triangleq \xi_{Z,h,k} - \xi_{h,k}$. For $1 \leq i, j \leq p$, let $\xi_{i,j} \triangleq \mathbb{E} \left[ z_{kh} e_{kh} \right] \mathcal{F}_{(k-1)h}$, and $\xi_{ij} \triangleq \mathbb{E} \left[ z_{kh} e_{kh} \right] \mathcal{F}_{(k-1)h}$, and $\xi_{ij}^* \triangleq \mathbb{E} \left[ z_{kh} e_{kh} \right] \mathcal{F}_{(k-1)h}$. Then, we have the following decomposition for $\overline{Q}_T (\theta^*) \triangleq \overline{Q}_T (\theta^*) + (\delta^0)^' (Z, Z)_1 (\cdot) \delta^0$ (if $v \leq 0$, and defined analogously for $v > 0$),

$$\overline{Q}_T (\theta^*) = \sum_{r=1}^4 \overline{Q}_{r,T} (\theta^*),$$

(2.9.13)

where $\overline{Q}_{1,T} (\theta^*) \triangleq - (\delta^0)^' (\mathbb{E} \left[ z_{kh} e_{kh} \right] \mathcal{F}_{(k-1)h})(\delta^0)$, $\overline{Q}_{2,T} (\theta^*) \triangleq - (\mathbb{E} \left[ z_{kh} e_{kh} \right] \mathcal{F}_{(k-1)h})(\delta^0)$, $\overline{Q}_{3,T} (\theta^*) \triangleq (\delta^0)^' (h^{-1/2} \sum_k \xi_{Z,h,k})$, and $\overline{Q}_{4,T} (\theta^*) \triangleq (\delta^0)^' (h^{-1/2} \sum_k \xi_{Z,h,k})$; where $\sum_k$ stands for $\sum_{k=T_0}^{T_0 + [v/h]}$. We have

$$\sup_{(\theta, v) \in \mathcal{D}^* (C)} \left\| \overline{Q}_{3,T} (\theta^*) \right\| \leq K \left\| \delta^0 \right\| h^{-1/2} \sum_k \left\| \xi_{h,k}^* \right\| \overset{P}{\rightarrow} 0,$$

(2.9.14)

which follows from Jacod and Rosenbaum (2013) given that $\Sigma_{Zv,k} = 0$ identically by Assumption 3.1-(iv). As for $\overline{Q}_{1,T} (\theta, v)$ we prove stochastic equicontinuity directly, using the definition in Andrews (1994). Choose any $\varepsilon > 0$ and $\eta > 0$. Consider any $\theta, v, (\bar{\theta}, \bar{v})$ with $v < 0 < \bar{v}$ (the other cases can be proven similarly) and

\footnotesize
\[5\text{Although in this proof it is not necessary to consider a neighborhood about } \delta^0 \text{ while proving stochastic equicontinuity, this step will be needed to justify our inference methods later. Thus, this proof is more general and may be useful in other contexts.}\]
\[ \bar{\delta} = \delta + c_{p \times 1}, \text{ where } c_{p \times 1} \text{ is a } p \times 1 \text{ vector with each entry equals to } c \in \mathbb{R}, \text{ with } 0 < c \leq \tau < \infty, \text{ then} \]

\[
\left| \mathcal{Q}_{1,T}(\theta^*) - \mathcal{Q}_{1,T}(\bar{\theta}^*) \right| \\
= \left| \bar{\delta}' \left( \sum_{k=T_b^0+1}^{T_b(\bar{\nu})} \zeta_{Z,h,k}^* \right) - \delta' \left( \sum_{k=T_b^0+1}^{T_b^0+1} \zeta_{Z,h,k}^* \right) \right| \\
= \left| \bar{\delta}' \left( \sum_{k=T_b^0+1}^{T_b^0+1} \zeta_{Z,h,k}^* \right) c_{p \times 1} + \delta' \left( \sum_{k=T_b^0+1}^{T_b^0+1} \zeta_{Z,h,k}^* - \sum_{k=T_b^0+1}^{T_b^0+1} \zeta_{Z,h,k}^* \right) \delta \right| \\
\leq K \left( \sum_{k=T_b^0+1}^{T_b^0+1} \left\| \zeta_{Z,h,k}^* \right\| \left\| c_{p \times 1} \right\|^2 + \left\| \sum_{k=T_b^0+1}^{T_b^0+1} \zeta_{Z,h,k}^* \right\| \left\| \delta \right\|^2 \right) \\
\leq K \left( \left( pc^2 \right) \sum_{k=T_b^0+1}^{T_b^0+1} \left\| \zeta_{Z,h,k}^* \right\| + \left\| \sum_{k=T_b^0+1}^{T_b^0+1} \zeta_{Z,h,k}^* \right\| \left\| \delta \right\|^2 \right).
\]

By Itô’s formula \( \left\| \zeta_{Z,h,k}^* \right\| = O \left( h^{3/2} \right) \), and so

\[
\left| \mathcal{Q}_{1,T}(\theta^*) - \mathcal{Q}_{1,T}(\bar{\theta}^*) \right| \leq K \left( c^2 h^{-1} O_p \left( h^{3/2} \right) O \left( \tau \right) + \left\| \delta \right\|^2 h^{-1} O_p \left( h^{3/2} \right) O \left( \tau \right) \right) \leq K \left( c^2 O_p \left( h^{1/2} \right) O \left( \tau \right) + \left\| \delta \right\|^2 O_p \left( h^{1/2} \right) O \left( \tau \right) \right),
\]

which goes to zero uniformly in \( \theta^* \in \Theta \) as \( \tau \to 0 \). Next, consider \( \mathcal{Q}_{2,T}(\theta^*) \) and observe that for any \( r \geq 1 \), standard estimates for Itô semimartingales yields

\[
\mathbb{E} \left( \left\| \zeta_{Z,h,k}^* \right\|^r \right) \leq K_r h^r. \quad (2.9.15)
\]
as demonstrated above, this suffices to guarantee the stable convergence in law of asymptotically stochastic equicontinuous. Since the finite-dimensional convergence

Combining (2.9.14), (2.9.15) and (2.9.16), we conclude that

\[
\mathcal{Q}_{2,T}(\theta^*) \text{ is stochastically equicontinuous. Turning to } \mathcal{Q}_{4,T}(\theta^*),
\]

\[
\left| \mathcal{Q}_{4,T}(\bar{\theta}^*) - \mathcal{Q}_{4,T}(\theta^*) \right| = \left| \tilde{\delta}' \left( h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \zeta_{e,h,k}^* \right) - \delta' \left( h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \zeta_{e,h,k}^* \right) \right|
\]

\[
= \left| \tilde{\delta}' \left( h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \zeta_{e,h,k}^* \right) \right| + \left| \delta' \left( h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \zeta_{e,h,k}^* \right) \right|
\]

\[
\leq K \left( h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \| \zeta_{e,h,k}^* \| \| c_p \| \right) + \left( h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \| \zeta_{e,h,k}^* \| \| \delta \| \right)
\]

\[
\leq K \left( \rho c h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \| \zeta_{e,h,k}^* \| \right) + \left( h^{-1/2} \sum_{k=T_0^0 + [\nu/h]}^{T_0^0 + [\tau/h]} \| \zeta_{e,h,k}^* \| \| \delta \| \right)
\]

By the Burkholder-Davis-Gundy inequality, \( \| \zeta_{e,h,k}^* \| \leq K h^{3/2} \) (recall \( \Sigma_{7.\xi,t} = 0 \) for all \( t \geq 0 \)), so that

\[
\left| \mathcal{Q}_{4,T}(\theta^*) - \mathcal{Q}_{4,T}(\bar{\theta}^*) \right| \leq K \left( c^2 h^{-1/2} h^{-1} h^{3/2} O(\tau) + \| \delta \|^2 h^{-1/2} h^{-1} h^{3/2} O(\tau) \right)
\]

\[
\leq K \left( c^2 O(\tau) + \| \delta \|^2 O(\tau) \right).
\]

Then for every \( \eta > 0 \), with \( B(\tau, (\theta, v)) \) a closed ball of radius \( \tau > 0 \) around \( \theta^* \), the quantity

\[
\limsup_{h \downarrow 0} \mathbb{P} \left[ \sup_{\theta^* \in \Theta : \bar{\theta}^* \in B(\tau, \theta^*)} \left| \mathcal{Q}_{4,T}(\theta^*) - \mathcal{Q}_{4,T}(\bar{\theta}^*) \right| > \eta \right], \quad (2.9.16)
\]

can be made arbitrary less than \( \varepsilon > 0 \) as \( h \downarrow 0 \), by choosing \( \tau \) small enough. Combining (2.9.14), (2.9.15) and (2.9.16), we conclude that \( \{ \mathcal{Q}_{T}(\theta, v), T \geq 1 \} \) is asymptotically stochastic equicontinuous. Since the finite-dimensional convergence was demonstrated above, this suffices to guarantee the stable convergence in law of
the process \( \{Q_T(\theta, v), T \geq 1\} \) toward a two-sided Gaussian limit process with drift \((\delta^0)' [Z, Z](\cdot')\delta^0\), having \(P\)-a.s. continuous sample paths with \(F\)-conditional covariance matrix given in (2.9.1). Because \(N(\hat{\lambda}_0 - \lambda_0) = O_p(1)\) under the new “fast time scale”, and \(D^*(C)\) is compact, then the main assertion of the theorem follows from the continuous mapping theorem for the argmax functional. In view of Section 2.9.3.3, a result which shows the negligibility of the drift term, the proof of Theorem 2.4.1 is concluded. □

2.9.3.2 Proof of Theorem 2.4.2

Proof. By Theorem 2.4.1 and using the property of the Gaussian law of the limiting process,

\[
Q_T(\theta, v) \xrightarrow{L_s} H(v) = \begin{cases} 
- (\delta^0)' [Z, Z]_1 (v) \delta^0 + 2 ( (\delta^0)' \Omega_{\mathcal{W},1} (\delta^0) )^{1/2} W_1^*(v), & \text{if } v \leq 0 \\
- (\delta^0)' [Z, Z]_2 (v) \delta^0 + 2 ( (\delta^0)' \Omega_{\mathcal{W},2} (\delta^0) )^{1/2} W_2^*(v), & \text{if } v > 0.
\end{cases}
\]

However, by a change in variable \(v = \vartheta^{-1} s\) with \(\vartheta = ( (\delta^0)' [Z, Z]_1 \delta^0 )^2 / ( (\delta^0)' \Omega_{\mathcal{W},1} (\delta^0) )\), we can show that

\[
\argmax_{v} \mathcal{H}(v) \overset{d}{=} \argmax_{s} \mathcal{V}(s) = \begin{cases} 
- |s| + W_1^*(s), & \text{if } s < 0 \\
- (\delta^0)' [Z, Z]_2 \delta^0 |s| + ( (\delta^0)' \Omega_{\mathcal{W},2} (\delta^0) )^{1/2} W_2^*(s), & \text{if } s \geq 0,
\end{cases}
\]

where

\[
\mathcal{V}(s) = \begin{cases} 
- |s| + W_1^*(s), & \text{if } s < 0 \\
- (\delta^0)' [Z, Z]_2 \delta^0 |s| + ( (\delta^0)' \Omega_{\mathcal{W},2} (\delta^0) )^{1/2} W_2^*(s), & \text{if } s \geq 0,
\end{cases}
\]
and we have used the facts that \( W_s \equiv W(-s) \), \( W(cs) \equiv |c|^{1/2} W_s \), and for any \( c > 0 \) and any function \( f(s) \), \( \arg\max_s cf(s) = \arg\max_s f(s) \). Thus,

\[
\arg\max_{v \in \left[ -N_b^0, N_b^0 \right]} \mathcal{H}(v) \equiv \arg\max_{s \in \left[ -N_b^0, N_b^0 \right]} \mathcal{H}(s) \equiv \max_{s \in \left[ -N_b^0, N_b^0 \right]} \mathcal{H}(s) = \mathcal{H}\left( \frac{1}{N_b^0} \right),
\]

and finally by the continuous mapping theorem for the argmax functional,

\[
\arg\max_{v \in \left[ -N_b^0, N_b^0 \right]} \mathcal{H}(v) \equiv \arg\max_{s \in \left[ -N_b^0, N_b^0 \right]} \mathcal{H}(s) = \mathcal{H}(s),
\]

This concludes the proof. \( \square \)

### 2.9.3.3 Negligibility of the Drift Term

We are in the setting of Section 2.3-2.4. In Proposition 2.3.1-2.3.3 and 2.4.1 the drift processes \( \mu_{u,t} \) from (2.2.3) are clearly of higher order in \( h \) and so they are negligible. In Theorem 2.4.1, we first changed the time scale and then normalized the criterion function by the factor \( h^{-1/2} \). The change of time scale now results in

\[
dZ_{\psi,s} = \psi_h^{-1/2} \mu_{Z,s} ds + \psi_h^{-1/2} \sigma_{Z,s} dW_{Z,s}, dW_{\psi,e,s} = \psi_h^{-1/2} \sigma_{e,s} dW_{e,s}, \text{ with } s \in \mathcal{D}(C).
\]
Given $s \mapsto t = \psi_h^{-1}s$, we have $\psi_h^{-1/2} \mu_{Z,s} ds = \psi_h^{-1/2} \mu_{Z,s} \psi_h (ds/\psi_h) = \mu_{Z,s} \psi_h \psi_h dt$ with $\vartheta = 1/2$. Then, as in (2.9.9), $dZ_{\psi,t} = \psi^\vartheta_h \mu_{Z,t} dt + \sigma_{Z,t} dW_{Z,t}$ and $dW_{\psi,e,t} = \sigma_{e,t} dW_{e,t}$ with $t \in D^* (C)$. Thus, the change of time scale effectively makes the drift $\mu_{Z,s} ds$ of even higher order. We show a stronger result in that we demonstrate its negligibility even in the case $\vartheta = 0$; hence, we show that the limit law of (2.9.12) remains the same when $\mu, t$ are nonzero. We set for any $1 \leq i \leq p$ and $1 \leq j \leq q + p$, $\mu_{Z,k} = \int_{(k-1)h}^{kh} \mu_{Z,s} ds$, $\mu_{X,k} = \int_{(k-1)h}^{kh} \mu_{X,s} ds$, $z_{0,kh} = \sum_{i=1}^{p} \int_{(k-1)h}^{kh} \sigma_{Z,s}^i dW^i_Z$ and $x_{0,kh} = \sum_{i=1}^{q+p} \int_{(k-1)h}^{kh} \sigma_{X,s}^{i,r} dW^r_X$. Note that $z_{kh} x_{kh} = \mu_{Z,k} \mu_{X,k} + \mu_{Z,k} x_{0,kh} + z_{0,kh} \mu_{X,k} + z_{0,kh} x_{0,kh}$. Recall that $\mu, k$ is $O (h)$ uniformly in $k$, and note that $\mu_{Z,k} x_{0,kh} + \mu_{Z,k} z_{0,kh}$ follows a Gaussian law with zero mean and variance of order $O (h^3)$. Also note that on $D^* (C)$, $T_b^0 - T_b - 1 \asymp 1/h$, where $a_h \asymp b_h$ if for some $c \geq 1$, $b_h/c \leq a_h \leq c b_h$.

Then,

$$
\sum_{k=T_b+1}^{T_b^0} z_{kh} x_{kh} = \sum_{k=T_b+1}^{T_b^0} \mu_{Z,k} \mu_{X,k} + \sum_{k=T_b+1}^{T_b^0} \mu_{Z,k} x_{0,kh} + \sum_{k=T_b+1}^{T_b^0} z_{0,kh} x_{0,kh},
$$

$$
= o (h^{1/2}) + o_p (h^{1/2}) + \sum_{k=T_b+1}^{T_b^0} z_{0,kh} x_{0,kh}.
$$

Therefore, conditionally on $\Sigma^0 = \{\mu, t, \sigma, t\}_{t \geq 0}$, the limit law of

$$
\mathcal{Q}_T (\theta^*) = - (\delta^0)' \left( \sum_{k=T_b+1}^{T_b^0} z_{kh} z_{kh}' \right) \delta^0 + 2 (\delta^0)' \left( h^{-1/2} \sum_{k=T_b+1}^{T_b^0} z_{kh} \bar{e}_{kh} \right),
$$

is the same as the limit law of

$$
- (\delta^0)' \left( \sum_{k=T_b+1}^{T_b^0} z_{0,kh} z_{0,kh}' \right) \delta^0 + 2 (\delta^0)' \left( h^{-1/2} \sum_{k=T_b+1}^{T_b^0} z_{0,kh} \bar{e}_{kh} \right),
$$

which completes the proof of Theorem 2.4.1.
2.9.4 Simulation of the Limiting Distribution in Theorem 2.4.1

We discuss how to simulate the limiting distribution in Theorem 2.4.1 which is slightly different from simulating the limiting distribution in Theorem 2.4.2. However, the idea is similar in that we replace unknown quantities by consistent estimates. First, we replace $N_0^b$ by $\tilde{N}_b$ (cf. Proposition 2.4.1). The ratio $\|\delta_0\|^2 / \sigma^2$ is consistently estimated by $\|\delta\|^2 / (T^{-1} \sum_{k=1}^{T} \tilde{e}_{kh}^2)$ because under the “fast time scale” $h^{1/2} \sum_{k=1}^{T} \tilde{e}_{kh} \overset{p}{\to} \sigma^2$ (cf. Assumption 2.6). Now consider the term

$$\left\{ - (\delta^0)' \left< Z_{\Delta}, Z_{\Delta} \right> (v) \delta^0 + 2 (\delta^0)' \mathcal{W} (v) \right\}.$$ 

For $v \leq 0$, this can be consistently estimated by

$$-T^{1/2} \left[ (\hat{\delta})' \left( \sum_{k=\hat{T}_b+[v/h]}^{\hat{T}_b} \hat{z}_{kh} \hat{z}_{kh}' \right) \hat{\delta} - 2 \hat{\delta}' \hat{\mathcal{W}}_h (v) \right],$$

(2.9.18)

where $\hat{\mathcal{W}}_h$ is a simple-size dependent sequence of Gaussian processes whose marginal distribution is characterized by $h^{1/2} \sum_{k=\hat{T}_b+[v/h]}^{\hat{T}_b} e_{kh}^2 \hat{z}_{kh} \hat{z}_{kh}'$ which is a consistent estimate of $\int_v^0 \Omega_{Z_{\epsilon,s}} ds$. Thus, in the limit $\mathcal{W}_h (v)$ has the same marginal distribution as $\mathcal{W} (v)$, and it follows that the limiting distribution from Theorem 2.4.1 can be simulated. The proposed estimator with (2.9.18) is valid under a continuous-record asymptotic (i.e., under Assumption 2.6 and the adoption of the “fast time scale”). It can also be shown to be valid under a fixed-shifts framework.
Figure 2.1: Distributions with $\rho^2 = 0.2$

The limit probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.2$ and the true fractional break point $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).

Figure 2.2: Distributions with $\rho^2 = 0.3$

The limit probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.3$ and the true fractional break point $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).
**Figure 2.3:** Distributions with $\rho^2 = 0.5$

The limit probability density of $\rho \left( \hat{T}_b - T_0 \right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.5$ and true fractional break date $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).

**Figure 2.4:** Distributions with $\rho^2 = 0.8$

The limit probability density of $\rho \left( \hat{T}_b - T_0 \right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.8$ and the true fractional break date $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).
Figure 2.5: Distributions for model (2.5.1) with $\delta^0 = 0.2$

The probability density of $\rho \left( \tilde{T}_b - T_b^0 \right)$ for model (2.5.1) with break magnitude $\delta^0 = 0.2$ and true break fraction $\lambda_0 = 0.3, 0.5$ and $0.7$ (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0 / \sigma_e = \delta^0$ since $\sigma_e^2 = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.
Figure 2.6: Distributions for model (2.5.1) with $\delta^0 = 0.3$

The probability density of $\rho\left(\hat{T}_b - T^0_b\right)$ for model (2.5.1) with break magnitude $\delta^0 = 0.3$ and true break fraction $\lambda_0 = 0.3, 0.5$ and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0 / \sigma_e = \delta^0$ since $\sigma_e^2 = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.
Figure 2.7: Distributions for model (2.5.1) with $\delta^0 = 0.5$

The probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ for model (5.1) with break magnitude $\delta^0 = 0.5$ and true break fraction $\lambda_0 = 0.3, 0.5$ and $0.7$ (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_e = \delta^0$ since $\sigma^2_e = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.
Figure 2.8: Distributions for model (2.5.1) with $\delta^0 = 1$

The probability density of $\rho \left( \hat{T}_b - T_b^0 \right)$ for model (5.1) with break magnitude $\delta^0 = 1$ and true break fraction $\lambda_0 = 0.3, 0.5$ and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0 / \sigma_e = \delta^0$ since $\sigma^2_e = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.
Figure 2.9: Highest Density Regions

Highest Density Regions (HDRs) of the feasible probability density of $\rho \left( \hat{T}_b - T_0^b \right)$ as described in Section 4.4. The significance level is $\alpha = 0.05$, the true break point is $\lambda_0 = 0.3$ and 0.5 (the left and right panels, respectively) and the break magnitude is $\delta^0 = 0.3$, 0.8 and 1.5 (the top, middle and bottom panels, respectively). The union of the black lines below the horizontal axis is the 95% HDR confidence region.
Table 2.1: Coverage rate and length of the confidence set for the example of Section 4.4

<table>
<thead>
<tr>
<th></th>
<th>$\delta_0 = 0.3$</th>
<th>$\delta_0 = 0.8$</th>
<th>$\delta_0 = 1.5$</th>
<th>$\delta_0 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDR</td>
<td>1 94 1 27 1 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0 55 0 13 1 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_{T,eq}$</td>
<td>1 95 1 37 1 24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDR</td>
<td>1 82 1 14 1 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>1 67 1 18 1 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_{T,eq}$</td>
<td>1 95 1 35 1 14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coverage rate and length of the confidence sets corresponding to the example from Section 4.4. See also Figure 2.9. The significance level is $\alpha = 0.05$. Cov. and Lgth. refer to the coverage rate and average size of the confidence sets (i.e., average number of dates in the confidence sets), respectively. Cov=1 if the confidence set includes $T_0^0$ and Cov=0 otherwise. The sample size is $T = 100$.

Table 2.2: Small-sample coverage rate and length of the confidence set for model M1

<table>
<thead>
<tr>
<th></th>
<th>$\delta_0 = 0.3$</th>
<th>$\delta_0 = 0.6$</th>
<th>$\delta_0 = 1$</th>
<th>$\delta_0 = 1.5$</th>
<th>$\delta_0 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDR</td>
<td>0.956 75.63 0.940 65.39 0.949 35.96 0.969 12.53 0.960 5.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.814 66.67 0.890 41.73 0.931 20.28 0.936 9.22 0.960 7.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_{T,eq}$</td>
<td>0.948 82.64 0.948 59.16 0.948 29.32 0.953 16.25 0.953 11.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.955 83.22 0.954 55.97 0.969 21.65 0.974 8.49 0.983 4.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.202 0.455 0.912 0.999 1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDR</td>
<td>0.958 74.01 0.948 61.50 0.951 33.84 0.954 12.14 0.965 5.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.839 66.12 0.850 41.85 0.901 19.40 0.938 9.18 0.963 5.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_{T,eq}$</td>
<td>0.953 83.32 0.950 61.17 0.950 30.09 0.950 16.15 0.949 11.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.949 83.15 0.960 58.69 0.966 22.94 0.975 8.25 0.985 4.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.192 0.65 0.983 1.000 1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDR</td>
<td>0.901 73.55 0.934 57.34 0.968 31.15 0.975 12.59 0.967 6.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.837 64.44 0.890 41.73 0.931 20.28 0.946 9.42 0.958 5.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\theta}_{T,eq}$</td>
<td>0.950 85.48 0.950 69.84 0.950 38.52 0.950 16.59 0.950 11.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.953 85.71 0.954 66.48 0.967 29.42 0.976 9.214 0.981 4.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.118 0.405 0.878 0.998 1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \beta^0 + \delta^0 1(t > T_{\lambda_0}) + e_t$, $e_t \sim i.i.d. \mathcal{N}(0, 1)$, $T = 100$. Cov. and Lgth. refer to the coverage probability and the average length of the confidence set (i.e., the average number of dates in the confidence set). sup-W refers to the rejection probability of the sup-Wald test using a 5% size with the asymptotic critical value. The number of simulations is 5,000.
Table 2.3: Small-sample coverage rate and length of the confidence set for model M2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.970 86.60</td>
<td>0.937 76.29</td>
<td>0.901 55.59</td>
<td>0.900 33.73</td>
<td>0.934 26.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.854 70.60</td>
<td>0.843 58.27</td>
<td>0.857 40.70</td>
<td>0.894 23.33</td>
<td>0.923 14.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_T$.eq</td>
<td>0.961 88.95</td>
<td>0.961 80.33</td>
<td>0.961 61.15</td>
<td>0.961 39.69</td>
<td>0.964 32.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.980 92.53</td>
<td>0.985 84.06</td>
<td>0.977 58.05</td>
<td>0.974 26.19</td>
<td>0.958 12.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sup-W$</td>
<td>0.121</td>
<td>0.349</td>
<td>0.747</td>
<td>0.988</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.976 89.81</td>
<td>0.961 83.26</td>
<td>0.935 64.87</td>
<td>0.900 38.19</td>
<td>0.934 26.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.823 69.86</td>
<td>0.822 55.87</td>
<td>0.894 38.91</td>
<td>0.898 23.33</td>
<td>0.923 14.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_T$.eq</td>
<td>0.963 89.84</td>
<td>0.963 82.62</td>
<td>0.961 65.87</td>
<td>0.961 43.63</td>
<td>0.964 32.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.990 93.48</td>
<td>0.985 92.53</td>
<td>0.982 68.23</td>
<td>0.979 32.77</td>
<td>0.977 15.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sup-W$</td>
<td>0.121</td>
<td>0.308</td>
<td>0.789</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.978 90.39</td>
<td>0.975 85.89</td>
<td>0.934 70.05</td>
<td>0.954 44.17</td>
<td>0.957 29.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.782 70.24</td>
<td>0.805 56.37</td>
<td>0.831 37.06</td>
<td>0.897 23.19</td>
<td>0.928 14.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_T$.eq</td>
<td>0.968 91.11</td>
<td>0.968 87.62</td>
<td>0.972 78.17</td>
<td>0.968 60.80</td>
<td>0.967 46.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.980 93.32</td>
<td>0.981 91.60</td>
<td>0.978 81.60</td>
<td>0.978 49.04</td>
<td>0.981 22.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sup-W$</td>
<td>0.098</td>
<td>0.262</td>
<td>0.628</td>
<td>0.938</td>
<td>0.995</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \beta^0 + \delta^0 1_{\{t > \lfloor T \lambda_0 \rfloor \}} + e_t$, $e_t = (1 + 1_{\{t > \lfloor T \lambda_0 \rfloor \}}) u_t$, $u_t \sim i.i.d. \mathcal{N}(0, 1)$, $T = 100$. The notes of Table 2.2 apply.

Table 2.4: Small-sample coverage rate and length of the confidence set for model M3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.904 72.44</td>
<td>0.901 58.37</td>
<td>0.919 29.70</td>
<td>0.945 11.29</td>
<td>0.971 5.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.833 66.34</td>
<td>0.834 41.32</td>
<td>0.895 18.63</td>
<td>0.942 8.98</td>
<td>0.969 5.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_T$.eq</td>
<td>0.958 87.16</td>
<td>0.968 71.47</td>
<td>0.958 45.82</td>
<td>0.957 30.73</td>
<td>0.957 28.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.932 79.38</td>
<td>0.944 53.48</td>
<td>0.966 21.98</td>
<td>0.986 8.59</td>
<td>0.993 4.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sup-W$</td>
<td>0.314</td>
<td>0.749</td>
<td>0.990</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.910 70.98</td>
<td>0.902 53.88</td>
<td>0.917 28.07</td>
<td>0.948 11.18</td>
<td>0.973 5.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.849 65.13</td>
<td>0.840 40.43</td>
<td>0.900 18.69</td>
<td>0.949 9.01</td>
<td>0.974 5.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_T$.eq</td>
<td>0.960 87.46</td>
<td>0.961 72.79</td>
<td>0.962 46.44</td>
<td>0.961 31.39</td>
<td>0.961 28.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.942 80.94</td>
<td>0.946 55.20</td>
<td>0.965 23.55</td>
<td>0.983 8.83</td>
<td>0.993 4.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sup-W$</td>
<td>0.308</td>
<td>0.705</td>
<td>0.990</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.905 72.02</td>
<td>0.913 50.61</td>
<td>0.933 25.07</td>
<td>0.947 11.10</td>
<td>0.974 6.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.829 65.56</td>
<td>0.899 41.42</td>
<td>0.932 19.62</td>
<td>0.951 9.20</td>
<td>0.966 5.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_T$.eq</td>
<td>0.962 88.77</td>
<td>0.968 78.61</td>
<td>0.963 57.87</td>
<td>0.968 37.15</td>
<td>0.965 29.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.938 83.24</td>
<td>0.951 63.66</td>
<td>0.972 28.94</td>
<td>0.985 10.18</td>
<td>0.994 5.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sup-W$</td>
<td>0.272</td>
<td>0.595</td>
<td>0.921</td>
<td>0.997</td>
<td>0.999</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \beta^0 + \delta^0 1_{\{t > \lfloor T \lambda_0 \rfloor \}} + e_t$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d. \mathcal{N}(0, 0.49)$, $T = 100$. The notes of Table 2.2 apply.
Table 2.5: Small-sample coverage rate and length of the confidence set for model M4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.915</td>
<td>77.14</td>
<td>0.912</td>
<td>67.14</td>
<td>0.910</td>
<td>60.64</td>
<td>0.901</td>
<td>53.21</td>
<td>0.912</td>
<td>7.15</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_F$ eq</td>
<td>0.850</td>
<td>65.94</td>
<td>0.821</td>
<td>44.07</td>
<td>0.850</td>
<td>20.71</td>
<td>0.878</td>
<td>9.88</td>
<td>0.887</td>
<td>5.96</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.950</td>
<td>85.23</td>
<td>0.951</td>
<td>67.40</td>
<td>0.951</td>
<td>39.87</td>
<td>0.950</td>
<td>23.58</td>
<td>0.955</td>
<td>17.46</td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.961</td>
<td>84.37</td>
<td>0.966</td>
<td>59.94</td>
<td>0.977</td>
<td>26.09</td>
<td>0.986</td>
<td>11.78</td>
<td>0.986</td>
<td>7.14</td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.915</td>
<td>75.53</td>
<td>0.911</td>
<td>58.88</td>
<td>0.905</td>
<td>49.77</td>
<td>0.901</td>
<td>41.14</td>
<td>0.912</td>
<td>7.27</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_F$ eq</td>
<td>0.881</td>
<td>64.69</td>
<td>0.826</td>
<td>42.93</td>
<td>0.849</td>
<td>20.77</td>
<td>0.880</td>
<td>9.92</td>
<td>0.888</td>
<td>5.99</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.948</td>
<td>85.48</td>
<td>0.948</td>
<td>68.95</td>
<td>0.948</td>
<td>41.40</td>
<td>0.948</td>
<td>24.01</td>
<td>0.954</td>
<td>17.57</td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.959</td>
<td>84.67</td>
<td>0.964</td>
<td>61.55</td>
<td>0.973</td>
<td>27.70</td>
<td>0.983</td>
<td>11.79</td>
<td>0.987</td>
<td>7.13</td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.911</td>
<td>74.46</td>
<td>0.931</td>
<td>56.22</td>
<td>0.935</td>
<td>29.22</td>
<td>0.927</td>
<td>12.39</td>
<td>0.929</td>
<td>7.85</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_F$ eq</td>
<td>0.880</td>
<td>64.06</td>
<td>0.870</td>
<td>42.86</td>
<td>0.896</td>
<td>22.11</td>
<td>0.898</td>
<td>10.40</td>
<td>0.887</td>
<td>6.16</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.952</td>
<td>86.80</td>
<td>0.956</td>
<td>75.20</td>
<td>0.952</td>
<td>51.99</td>
<td>0.956</td>
<td>29.92</td>
<td>0.952</td>
<td>19.92</td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.961</td>
<td>86.03</td>
<td>0.964</td>
<td>68.68</td>
<td>0.978</td>
<td>36.34</td>
<td>0.980</td>
<td>13.89</td>
<td>0.985</td>
<td>7.51</td>
</tr>
<tr>
<td></td>
<td>sup-\sup-W</td>
<td>0.186</td>
<td>0.612</td>
<td>0.963</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \pi^0 + Z_t \theta^0 + Z_t \delta^0 1_{(t > \tau_{\lambda_0})} + \epsilon_t$, $X_t = 0.5X_{t-1} + ut$, $ut \sim i.i.d. \mathcal{N}(0, 0.75)$, $\epsilon_t \sim i.i.d. \mathcal{N}(0, 1)$, $T = 100$. The notes of Table 2.2 apply.

Table 2.6: Small-sample coverage rate and length of the confidence set for model M5

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.920</td>
<td>77.03</td>
<td>0.923</td>
<td>70.69</td>
<td>0.930</td>
<td>60.02</td>
<td>0.956</td>
<td>47.06</td>
<td>0.969</td>
<td>35.03</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_F$ eq</td>
<td>0.690</td>
<td>56.73</td>
<td>0.716</td>
<td>41.63</td>
<td>0.783</td>
<td>27.53</td>
<td>0.847</td>
<td>18.32</td>
<td>0.885</td>
<td>12.70</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.926</td>
<td>87.76</td>
<td>0.962</td>
<td>78.32</td>
<td>0.962</td>
<td>63.80</td>
<td>0.962</td>
<td>50.14</td>
<td>0.962</td>
<td>40.82</td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.790</td>
<td>71.07</td>
<td>0.805</td>
<td>59.66</td>
<td>0.824</td>
<td>40.78</td>
<td>0.868</td>
<td>21.59</td>
<td>0.909</td>
<td>11.63</td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.928</td>
<td>76.41</td>
<td>0.925</td>
<td>68.21</td>
<td>0.917</td>
<td>53.33</td>
<td>0.938</td>
<td>41.96</td>
<td>0.946</td>
<td>31.73</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_F$ eq</td>
<td>0.691</td>
<td>55.18</td>
<td>0.720</td>
<td>40.25</td>
<td>0.760</td>
<td>26.90</td>
<td>0.757</td>
<td>17.96</td>
<td>0.826</td>
<td>12.62</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.953</td>
<td>87.76</td>
<td>0.953</td>
<td>78.55</td>
<td>0.981</td>
<td>63.80</td>
<td>0.953</td>
<td>51.51</td>
<td>0.953</td>
<td>41.98</td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.795</td>
<td>71.34</td>
<td>0.804</td>
<td>60.48</td>
<td>0.832</td>
<td>40.12</td>
<td>0.870</td>
<td>20.77</td>
<td>0.903</td>
<td>10.78</td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.915</td>
<td>75.86</td>
<td>0.919</td>
<td>66.79</td>
<td>0.926</td>
<td>52.50</td>
<td>0.945</td>
<td>38.63</td>
<td>0.970</td>
<td>24.86</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_F$ eq</td>
<td>0.707</td>
<td>55.03</td>
<td>0.770</td>
<td>39.77</td>
<td>0.828</td>
<td>26.82</td>
<td>0.862</td>
<td>18.05</td>
<td>0.901</td>
<td>12.68</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.951</td>
<td>88.40</td>
<td>0.952</td>
<td>82.09</td>
<td>0.951</td>
<td>71.84</td>
<td>0.951</td>
<td>60.78</td>
<td>0.950</td>
<td>50.72</td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.795</td>
<td>72.01</td>
<td>0.809</td>
<td>62.75</td>
<td>0.829</td>
<td>45.18</td>
<td>0.870</td>
<td>24.86</td>
<td>0.913</td>
<td>12.62</td>
</tr>
<tr>
<td></td>
<td>sup-\sup-W</td>
<td>0.257</td>
<td>0.517</td>
<td>0.757</td>
<td>0.891</td>
<td>0.950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \pi^0 + Z_t \theta^0 + Z_t \delta^0 1_{(t > \tau_{\lambda_0})} + \epsilon_t$, $\epsilon_t = v_t | Z_t |$, $v_t \sim i.i.d. \mathcal{N}(0, 1)$, $Z_t = 0.5Z_{t-1} + ut$, $ut \sim i.i.d. \mathcal{N}(0, 1)$, $T = 100$. The notes of Table 2.2 apply.
Table 2.7: Small-sample coverage rate and length of the confidence set for model M6

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>HDR</td>
<td>0.918</td>
<td>75.64</td>
<td>0.910</td>
<td>67.46</td>
<td>0.931</td>
<td>48.54</td>
<td>0.947</td>
<td>25.10</td>
<td>0.957</td>
<td>12.50</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.834</td>
<td>70.13</td>
<td>0.824</td>
<td>52.16</td>
<td>0.861</td>
<td>28.69</td>
<td>0.921</td>
<td>14.18</td>
<td>0.948</td>
<td>8.45</td>
<td></td>
</tr>
<tr>
<td>$\tilde{U}_T\cdot eq$</td>
<td>0.959</td>
<td>88.62</td>
<td>0.959</td>
<td>78.87</td>
<td>0.959</td>
<td>58.60</td>
<td>0.960</td>
<td>38.91</td>
<td>0.952</td>
<td>30.15</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.969</td>
<td>86.75</td>
<td>0.959</td>
<td>67.91</td>
<td>0.967</td>
<td>34.13</td>
<td>0.985</td>
<td>15.97</td>
<td>0.995</td>
<td>9.17</td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.245</td>
<td>0.573</td>
<td>0.911</td>
<td>0.997</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td>HDR</td>
<td>0.926</td>
<td>74.78</td>
<td>0.914</td>
<td>64.86</td>
<td>0.924</td>
<td>45.69</td>
<td>0.945</td>
<td>23.57</td>
<td>0.956</td>
<td>12.25</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.851</td>
<td>69.35</td>
<td>0.847</td>
<td>51.17</td>
<td>0.864</td>
<td>30.69</td>
<td>0.964</td>
<td>14.26</td>
<td>0.944</td>
<td>8.47</td>
<td></td>
</tr>
<tr>
<td>$\tilde{U}_T\cdot eq$</td>
<td>0.964</td>
<td>88.82</td>
<td>0.970</td>
<td>73.95</td>
<td>0.981</td>
<td>39.86</td>
<td>0.986</td>
<td>16.06</td>
<td>0.992</td>
<td>9.08</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.972</td>
<td>86.69</td>
<td>0.975</td>
<td>67.91</td>
<td>0.986</td>
<td>34.13</td>
<td>0.986</td>
<td>15.97</td>
<td>0.995</td>
<td>9.17</td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.244</td>
<td>0.559</td>
<td>0.904</td>
<td>0.994</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td>HDR</td>
<td>0.999</td>
<td>78.12</td>
<td>0.921</td>
<td>61.87</td>
<td>0.933</td>
<td>40.66</td>
<td>0.948</td>
<td>20.95</td>
<td>0.961</td>
<td>11.70</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.824</td>
<td>60.94</td>
<td>0.815</td>
<td>51.17</td>
<td>0.878</td>
<td>28.59</td>
<td>0.920</td>
<td>14.26</td>
<td>0.944</td>
<td>8.47</td>
<td></td>
</tr>
<tr>
<td>$\tilde{U}_T\cdot eq$</td>
<td>0.961</td>
<td>89.71</td>
<td>0.960</td>
<td>83.68</td>
<td>0.964</td>
<td>60.26</td>
<td>0.964</td>
<td>39.86</td>
<td>0.992</td>
<td>9.08</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.966</td>
<td>91.48</td>
<td>0.971</td>
<td>82.78</td>
<td>0.984</td>
<td>39.86</td>
<td>0.986</td>
<td>16.06</td>
<td>0.995</td>
<td>10.87</td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.232</td>
<td>0.467</td>
<td>0.804</td>
<td>0.962</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \beta_0 + \delta_0 \{t > \lceil T\lambda_0 \rceil \} + \epsilon_t, \epsilon_t \sim i.i.d., \nu = 5, T = 100$. The notes of Table 2.2 apply.

Table 2.8: Small-sample coverage rate and length of the confidence set for model M7

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>HDR</td>
<td>0.918</td>
<td>75.08</td>
<td>0.913</td>
<td>60.44</td>
<td>0.931</td>
<td>32.30</td>
<td>0.946</td>
<td>12.60</td>
<td>0.965</td>
<td>6.34</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.778</td>
<td>60.94</td>
<td>0.815</td>
<td>38.14</td>
<td>0.885</td>
<td>17.29</td>
<td>0.928</td>
<td>8.53</td>
<td>0.949</td>
<td>5.34</td>
<td></td>
</tr>
<tr>
<td>$\tilde{U}_T\cdot eq$</td>
<td>0.949</td>
<td>84.56</td>
<td>0.950</td>
<td>67.64</td>
<td>0.953</td>
<td>42.95</td>
<td>0.950</td>
<td>29.95</td>
<td>0.955</td>
<td>8.70</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.947</td>
<td>86.99</td>
<td>0.946</td>
<td>63.24</td>
<td>0.956</td>
<td>32.85</td>
<td>0.984</td>
<td>16.20</td>
<td>0.992</td>
<td>10.49</td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.275</td>
<td>0.753</td>
<td>0.991</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td>HDR</td>
<td>0.919</td>
<td>74.16</td>
<td>0.916</td>
<td>58.53</td>
<td>0.931</td>
<td>32.10</td>
<td>0.948</td>
<td>12.95</td>
<td>0.965</td>
<td>6.48</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.799</td>
<td>60.25</td>
<td>0.814</td>
<td>37.94</td>
<td>0.872</td>
<td>17.49</td>
<td>0.919</td>
<td>8.59</td>
<td>0.952</td>
<td>5.35</td>
<td></td>
</tr>
<tr>
<td>$\tilde{U}_T\cdot eq$</td>
<td>0.951</td>
<td>85.01</td>
<td>0.950</td>
<td>67.64</td>
<td>0.953</td>
<td>42.95</td>
<td>0.950</td>
<td>29.95</td>
<td>0.955</td>
<td>30.25</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.946</td>
<td>84.12</td>
<td>0.944</td>
<td>63.99</td>
<td>0.960</td>
<td>33.45</td>
<td>0.984</td>
<td>14.91</td>
<td>0.977</td>
<td>8.71</td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.258</td>
<td>0.700</td>
<td>0.986</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td>HDR</td>
<td>0.912</td>
<td>73.43</td>
<td>0.929</td>
<td>56.18</td>
<td>0.949</td>
<td>31.23</td>
<td>0.956</td>
<td>13.65</td>
<td>0.965</td>
<td>6.96</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.795</td>
<td>59.43</td>
<td>0.864</td>
<td>38.17</td>
<td>0.910</td>
<td>18.52</td>
<td>0.934</td>
<td>8.67</td>
<td>0.954</td>
<td>5.34</td>
<td></td>
</tr>
<tr>
<td>$\tilde{U}_T\cdot eq$</td>
<td>0.950</td>
<td>86.94</td>
<td>0.951</td>
<td>76.52</td>
<td>0.964</td>
<td>55.72</td>
<td>0.955</td>
<td>39.59</td>
<td>0.947</td>
<td>38.80</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.945</td>
<td>83.94</td>
<td>0.953</td>
<td>63.55</td>
<td>0.963</td>
<td>32.41</td>
<td>0.984</td>
<td>24.42</td>
<td>0.982</td>
<td>15.01</td>
<td></td>
</tr>
<tr>
<td>sup-W</td>
<td>0.195</td>
<td>0.546</td>
<td>0.920</td>
<td>0.998</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \delta_0 (1 - \pi_0) \{t > \lceil T\lambda_0 \rceil \} + \pi_0 y_{t-1} + \epsilon_t, \epsilon_t \sim i.i.d., \nu = 5, T = 100$. The notes of Table 2.2 apply.
### Table 2.9: Small-sample coverage rate and length of the confidence sets for model M8

<table>
<thead>
<tr>
<th>$\theta^0$</th>
<th>$\lambda_0 = 0.5$</th>
<th>$\lambda_0 = 0.35$</th>
<th>$\lambda_0 = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta^0 = 1$</td>
<td>0.916</td>
<td>36.05</td>
<td>0.916</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.793</td>
<td>12.87</td>
<td>0.793</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.951</td>
<td>91.64</td>
<td>0.951</td>
</tr>
<tr>
<td>ILR</td>
<td>0.951</td>
<td>46.31</td>
<td>0.951</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td>0.925</td>
<td>33.02</td>
<td>0.925</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.804</td>
<td>13.00</td>
<td>0.804</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.952</td>
<td>91.22</td>
<td>0.952</td>
</tr>
<tr>
<td>ILR</td>
<td>0.949</td>
<td>47.54</td>
<td>0.949</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.992</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td>0.937</td>
<td>34.66</td>
<td>0.937</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.832</td>
<td>13.64</td>
<td>0.832</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.944</td>
<td>89.64</td>
<td>0.944</td>
</tr>
<tr>
<td>ILR</td>
<td>0.946</td>
<td>49.13</td>
<td>0.946</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.935</td>
<td>0.995</td>
<td>0.995</td>
</tr>
</tbody>
</table>

The model is $y_t = \frac{\theta^0 (1 - \pi^0)}{\lambda_0} I_{t > \lfloor T\lambda_0 \rfloor} + \pi^0 y_{t-1} + e_t, e_t \sim i.i.d. \mathcal{N}(0, 0.04)$, $\theta^0 = 0.8, T = 100$. The notes of Table 2.2 apply.

### Table 2.10: Small-sample coverage rate and length of the confidence sets for model M9

<table>
<thead>
<tr>
<th>$\theta^0 = \delta^0$</th>
<th>$\lambda_0 = 0.5$</th>
<th>$\lambda_0 = 0.35$</th>
<th>$\lambda_0 = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta^0 = 0.3$</td>
<td>0.903</td>
<td>61.09</td>
<td>0.903</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>37.86</td>
<td>0.791</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.947</td>
<td>65.23</td>
<td>0.947</td>
</tr>
<tr>
<td>ILR</td>
<td>0.909</td>
<td>72.62</td>
<td>0.909</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.746</td>
<td>0.941</td>
<td>0.941</td>
</tr>
<tr>
<td>$\delta^0 = 0.6$</td>
<td>0.904</td>
<td>60.58</td>
<td>0.904</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>37.70</td>
<td>0.791</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.942</td>
<td>66.27</td>
<td>0.942</td>
</tr>
<tr>
<td>ILR</td>
<td>0.922</td>
<td>72.20</td>
<td>0.922</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.734</td>
<td>0.931</td>
<td>0.931</td>
</tr>
<tr>
<td>$\delta^0 = 1$</td>
<td>0.900</td>
<td>63.17</td>
<td>0.900</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>39.23</td>
<td>0.791</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.934</td>
<td>71.42</td>
<td>0.934</td>
</tr>
<tr>
<td>ILR</td>
<td>0.920</td>
<td>72.68</td>
<td>0.920</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.634</td>
<td>0.884</td>
<td>0.884</td>
</tr>
<tr>
<td>$\delta^0 = 1.5$</td>
<td>0.900</td>
<td>63.17</td>
<td>0.900</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>39.23</td>
<td>0.791</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.934</td>
<td>71.42</td>
<td>0.934</td>
</tr>
<tr>
<td>ILR</td>
<td>0.920</td>
<td>72.68</td>
<td>0.920</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.634</td>
<td>0.884</td>
<td>0.884</td>
</tr>
<tr>
<td>$\delta^0 = 2$</td>
<td>0.900</td>
<td>63.17</td>
<td>0.900</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>39.23</td>
<td>0.791</td>
</tr>
<tr>
<td>$\tilde{U}_T, eq$</td>
<td>0.934</td>
<td>71.42</td>
<td>0.934</td>
</tr>
<tr>
<td>ILR</td>
<td>0.920</td>
<td>72.68</td>
<td>0.920</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.634</td>
<td>0.884</td>
<td>0.884</td>
</tr>
</tbody>
</table>

The model is $y_t = \pi^0 + Z_t \delta^0 + Z_t \delta^0 I_{t > \lfloor T\lambda_0 \rfloor} + e_t, Z_t \sim i.i.d. \mathcal{N}(1, 1.44), \{e_t\}$ follows a FIGARCH(1,0.6,1) process and $T = 100$. The notes of Table 2.2 apply.
Table 2.11: Small-sample coverage rate and length of the confidence set for model M10

<table>
<thead>
<tr>
<th>λ_0</th>
<th>D^0 = 0.3</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>D^0 = 0.6</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>D^0 = 1</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>D^0 = 1.5</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>D^0 = 2</th>
<th>Cov.</th>
<th>Lgth.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HDR</td>
<td>0.952</td>
<td>74.84</td>
<td>0.930</td>
<td>36.02</td>
<td>0.921</td>
<td>13.11</td>
<td>0.916</td>
<td>6.55</td>
<td>0.916</td>
<td>4.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.809</td>
<td>45.33</td>
<td>0.844</td>
<td>17.11</td>
<td>0.864</td>
<td>8.27</td>
<td>0.878</td>
<td>5.08</td>
<td>0.883</td>
<td>3.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U_T.eq</td>
<td>0.959</td>
<td>72.69</td>
<td>0.959</td>
<td>39.81</td>
<td>0.959</td>
<td>24.25</td>
<td>0.959</td>
<td>17.96</td>
<td>0.959</td>
<td>14.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.929</td>
<td>83.23</td>
<td>0.951</td>
<td>69.67</td>
<td>0.971</td>
<td>44.40</td>
<td>0.978</td>
<td>20.76</td>
<td>0.987</td>
<td>10.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.600</td>
<td>0.988</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ_0 = 0.35</td>
<td>HDR</td>
<td>0.934</td>
<td>73.08</td>
<td>0.937</td>
<td>35.37</td>
<td>0.923</td>
<td>13.68</td>
<td>0.920</td>
<td>6.82</td>
<td>0.920</td>
<td>4.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.821</td>
<td>45.70</td>
<td>0.838</td>
<td>17.78</td>
<td>0.867</td>
<td>8.53</td>
<td>0.886</td>
<td>5.22</td>
<td>0.889</td>
<td>3.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U_T.eq</td>
<td>0.964</td>
<td>76.14</td>
<td>0.964</td>
<td>44.61</td>
<td>0.965</td>
<td>27.33</td>
<td>0.965</td>
<td>19.74</td>
<td>0.964</td>
<td>15.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.934</td>
<td>81.32</td>
<td>0.959</td>
<td>62.98</td>
<td>0.977</td>
<td>34.38</td>
<td>0.982</td>
<td>16.73</td>
<td>0.984</td>
<td>9.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.529</td>
<td>0.970</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ_0 = 0.2</td>
<td>HDR</td>
<td>0.941</td>
<td>71.46</td>
<td>0.959</td>
<td>59.03</td>
<td>0.950</td>
<td>15.39</td>
<td>0.926</td>
<td>7.78</td>
<td>0.919</td>
<td>5.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.818</td>
<td>47.82</td>
<td>0.872</td>
<td>20.44</td>
<td>0.878</td>
<td>9.60</td>
<td>0.876</td>
<td>5.64</td>
<td>0.873</td>
<td>3.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U_T.eq</td>
<td>0.971</td>
<td>82.40</td>
<td>0.971</td>
<td>59.03</td>
<td>0.971</td>
<td>39.02</td>
<td>0.971</td>
<td>27.07</td>
<td>0.972</td>
<td>20.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.928</td>
<td>83.26</td>
<td>0.952</td>
<td>70.03</td>
<td>0.964</td>
<td>42.65</td>
<td>0.979</td>
<td>20.15</td>
<td>0.982</td>
<td>10.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sup-W</td>
<td>0.346</td>
<td>0.839</td>
<td>0.981</td>
<td>0.997</td>
<td>0.997</td>
<td>0.999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is \( y_t = \pi^0 + z_t \delta^0 + z_t \delta^0 I_{t > \lfloor T \lambda_0 \rfloor} + e_t, e_t \sim i.i.d. \mathcal{N}(0, 1), z_t \sim \text{ARFIMA}(0.3, 0.6, 0), T = 100. \) The notes of Table 2.2 apply.
Chapter 3

Tests for Forecast Instability and Forecast Failure under a Continuous Record
Asymptotic Framework

3.1 Introduction

Since the seminal contribution of Klein (1969; 1971), economic forecasts had been built upon the presumption that the relationships between economic variables remain stable over time. However, the last decades have been subject to many social-economic episodes and technological advancements that have led economists to reconsider the assumption of model stability. The resonant empirical evidences documented in, among others, Perron (1989) and Stock and Watson (1996) [see also the recent survey by Ng and Wright (2013)] have motivated the development of econometric methods that detect such instabilities—most work directed toward structural changes—and estimate the actual dates at which economic relationships change. Yet, the issue of parameter insatiability is not limited to model estimation. In the forecasting literature, there has been a widespread concordance that the major issue that prevents good forecasts for economic variables is parameter instability—and structural changes as a special case—[cf. Banerjee et al. (2008), Clements and Hendry (1998, 2006), Elliott and Timmermann (2016), Giacomini (2015), Giacomini and Rossi (2015), Inoue and Rossi (2011), Clark and McCracken (2005), Pesaran et al. (2006) and
Rossi (2013a)].

This chapter develops a statistical setting under infill asymptotics to address the issue of testing whether the predictive ability of a given forecast model remains stable over time. Ng and Wright (2013) and Stock and Watson (2003) explain that there has been abundant evidence for which a predictor that has performed well over a certain time period may not perform as well during other subsequent periods. For example, Gilchrist and Zakrajšek (2012) proposed a new credit spread index and showed that a residual component labeled as the excess bond premium—the credit spread adjusted for expected default risk—has considerable predictive content for future economic activity. They documented that this forecasting ability is stronger over the subsample 1985-2010 rather than over the full sample starting from 1973. The latter finding can be attributed to a more developed bond market in the 1985-2010 subsample. Relatedly, Giacomini and Rossi (2010) and Ng and Wright (2013) further examined this finding and found that indeed the predictive ability of commonly used term and credit spreads is unstable and somehow episodic. The latter authors suggested that credit spreads may be more useful predictors of economic activity in a more highly leveraged economy and that recent developments in financial markets translate into credit spreads containing more information than they had previously. We refer to such temporal instability for a given forecasting method as forecast instability or more specifically, as forecast failure. These terminologies are not new to professional forecasters as they were informally introduced by Clements and Hendry (1998) and generalized in econometric terms by Giacomini and Rossi (2009) who interpreted forecast breakdown (or forecast failure) as a situation in which the out-of-sample

\footnote{They reported that structural change tests provide some statistical evidence for a break in a coefficient associated with financial indicators—more specifically the coefficient on the federal funds rate. Given the latter evidence and the well-documented change in the conduct of monetary policy in the late 1970s and the early 1980s, it seems plausible to split the sample in 1985 (see p. 1709 and footnote 11 in their paper).}
performance of a forecast model significantly deteriorates relative to its in-sample performance. Our approach is to formally define forecast instability from the economic forecaster’s perspective.\(^2\) We emphasize that a forecast failure may well result from a short period of instability within the out-of-sample and not necessarily require that the instability be systematic in the sense of persisting throughout the whole out-of-sample period. That is, consistency of a forecast model’s performance with expected performance given the past should hold not only throughout the out-of-sample but also in any sub-sample of the latter. Indeed, many documented episodes of forecast failure seemed to arise from parameter nonconstancy data-generating processes over relatively short time periods compared to the total sample size. Hence, the desire of focusing on statistical tests being able to detect short-lasting instabilities is intuitive: if a test for forecast failure needs the deterioration of the forecasting ability to last for, say, at least half of the total sample in order to have sufficiently high power to reject the null hypotheses, then this test would not perform very well in practice because instability can be short-lasting. Furthermore, the occurrence of recurrent structural instabilities or multiple breaks that compensate each other in the out-of-sample might lead a forecast model to perform, on average, in a similar fashion as in the in-sample period. However, should a forecaster know about those recurrent changes she would conceivably revise its forecast model to adapt to the unstable environment. Hence, we introduce the following definition.

**Definition 3.1.1.** (Forecast Instability)

Forecast Instability refers to a situation of either sustained deterioration or improvement of the predictive ability of a given forecast model relative to the historical

\(^2\)We use the terminology “instability” because not only the deterioration but also the improvement of the performance of a given forecast model over time can provide useful information to the forecaster.
performance that would had led a forecaster to revise or reconsider its forecast model if she had known the occurrence of such instability. The time lengths of these two distinct periods need not bear any relationship.  

Th definition poses at the center the economic forecaster and consequently it is not merely a statistical definition; rather, it is based on an equilibrium concept. Since forecasting constitutes a decision theoretic problem, it should be from the forecaster perspective that a given forecast model is deemed to have failed. It is implicit from the definition to distinguish between forecasting method and model. Two forecasters may share the same forecast model—the relationship between the variable of interest and the predictor—but use different methods (e.g., recursive scheme versus rolling scheme). Thus, instability refers to a given method-model pair. The object of the definition is predictive ability. Since the latter can be measured differently by different loss functions, then the definition applies to a given choice of the loss function. A notable aspect of the definition is the reference to the time span of the historical performance and of the putative period of instability. They need not be related. Consider a given forecasting strategy which has performed well during, say, the Great Moderation (i.e., from mid-1980s up to prior the beginning of the Great Recession in 2007). Assume that during the years 2007-2012 this method endures a time of poor performance and returns to perform well thereafter. According to our definition, this episode constitutes an example of forecast instability. However, if one designs the forecasting exercise in such a way that half of the sample is used for estimation and the remaining half for prediction, then this relatively short period of instability gets “averaged-out” from tests which simply compare the in-sample and out-of-sample averages. Conceivably, such tests would not reject the null hypotheses of no forecast

\footnote{Forecast Failure constitutes a special case of the definition—namely, a sustained deterioration of predictive ability.}
failure while it seems that a forecaster would had revised its strategy during the crisis if she had known about such occurring under-performance in the present and immediate future period. Finally, detection of forecast instability does not necessarily mean that a forecast model should be abandoned. In fact, its performance may have improved over time. Yet, even if forecast instability is induced by performance deterioration, a forecaster might not end up switching to a new predictor. For example, entering a state of high variability might lead to poor performance even if the forecast model is still correct. Hence, our definition uses the term \textit{reconsider}. Continuing with the above example, a forecaster may \textit{reconsider} the choice of the forecasting window since a longer window may now produce better forecasts while keeping the \textit{same} forecast model. In other words, knowledge of forecast instability is important because indicates that care must be exercised to assess the source of the changes.\footnote{Economists have documented episodes of forecast failure in many areas of macroeconomics. In the empirical literature on exchange rates a prominent forecast failure is associated with the Meese and Rogoff’s puzzle [cf. Meese and Rogoff (1983), Cheung et al. (2005), and Rossi (2013b) for an up-to-date account]. In the context of inflation forecasting, forecast failures have been reported by Atkeson and Ohanian (2001) and Stock and Watson (2009). For forecast instability concerning other macroeconomic variables see the surveys of Stock and Watson (2003) and Ng and Wright (2013).}

The theoretical implication is that in this chapter our tests for forecast instability shall be based on the local behavior of the sequence of realized forecast losses. This is opposite to existing tests for forecast instability—and classical structural change tests more generally—which instead rely on a global and retrospective methodology merely comparing the average of in-sample losses with the average of out-of-sample losses. While maintaining approximately correct nominal size, our class of test statistics achieves substantial gains in statistical power relative to previous methods. Furthermore, as the initial timing of the instability moves away from middle sample toward the tail of the out-of-sample, the gains in power become considerable.

In this chapter, we set out a continuous record asymptotic framework for a
forecasting environment where $T$ observations at equidistant time intervals $h$ are made over a fixed time span $[0, N]$, with $N = Th$. These observations are realizations from a continuous-time model for the variable to be forecast and for the predictor. From these discretely observed realizations we compute a sequence of forecasts using either a fixed, recursive or rolling scheme. To this sequence of forecasts there corresponds a continuous-time process which satisfies mild regularity conditions and that under the null hypotheses possesses a continuous sample-path. We exploit this pathwise property to base an hypothesis testing problem on the relative performance of a given forecast model over time. Under the hypotheses we expect the sequence of losses to display a smooth and stable path. Any discontinuous or jump behavior followed by a (possibly short) period of substantial discrepancy from the same path over the in-sample period provides evidence against the hypotheses. Our asymptotic theory involves a continuous record of observations where we let the sample size $T$ grow to infinity by shrinking the sampling interval $h$ to zero with the time span kept fixed at $N$, thereby approaching the continuous-time limit.

Our underlying probabilistic model is specified in terms of continuous Itô semimartingales which are standard building blocks for analysis of macro and financial high-frequency data [cf. Andersen et al. (2001), Andersen et al. (2016), Bandi and Renò (2016) and Barndorff-Nielsen and Shephard (2004)]; the theoretical methodology is thus related to that of Casini and Perron (2017a), Li et al. (2017), Li and Xiu (2016) and Mykland and Zhang (2009).\textsuperscript{5} The framework is not only useful for high-frequency data; in particular, recent work of Casini and Perron (2017a, 2017b) has adopted this continuous-time approach for modeling time series regression models with structural changes fitted to low-frequency data (e.g., macroeconomic data

\textsuperscript{5}Recent work by Li and Patton (2017) extends standard methods for testing predictive accuracy of forecasts to a high-frequency financial setting.
that are sampled at weekly, monthly, quarterly, annual frequency, etc.). They have showed that this continuous-time approach delivers a better approximation to the finite-sample distributions of estimators in structural change models and inference is more reliable than previous methods based on classical long-span asymptotics.

The classical approach to economic forecasting for macroeconomic variables is to formulate models in discrete-time and then base inference on long-span asymptotics where the sample size increases without bound and the sampling interval remains fixed [cf. Diebold and Mariano (1995), Giacomini and White (2006) and West (1996)]. There are crucial distinctions between this classical approach and the setting introduced in this chapter. Under long-span asymptotics, identification of parameters hinges on assumptions on the distributions or moments of the studied processes [cf. the specification of the null hypotheses in Giacomini and Rossi (2009)], whereas within a continuous-time framework, unknown structural parameters are identified from the sample paths of the studied processes. Hence, we only need to assume rather mild pathwise regularity conditions for the underlying continuous-time model and avoid any ergodic or weak-dependence assumption. As in Casini and Perron (2017a), our framework encompasses any time series regression model allowing for general forms of nonstationarity such as heteroskedasticity and serial correlation.

Given a null hypotheses stated in terms of the path properties of the sequence of losses, we propose a test statistic which compares the local behavior of the sequence of surprise losses defined as the difference between the out-of-sample and in-sample losses. More specifically, our maximum-type statistic examines the smoothness of the sequence of surprise losses as the continuous-time limit is approached. Under the hypotheses, the continuous-time analogue of the sequence of losses follows a continuous motion and any deviation from such smooth path is interpreted as evidence
against the hypotheses. The null distribution of the test statistic is non-standard and follows an extreme value distribution. Therefore, our limit theory exploits results from extreme value theory as elaborated by Bickel and Rosenblatt (1973) and Galambos (1987).

We propose two versions of the test statistic: one that is self-normalized and one that uses an appropriate estimator of the asymptotic variance. The test statistic is defined as the maximal deviation between the average surprise losses over asymptotically vanishing time blocks. Further, we consider extensions of each of these statistics which use overlapping rather than non-overlapping blocks. Although they should be asymptotically equivalent, the statistics based on overlapping blocks are more powerful in finite-samples. In a framework where one allows for model misspecification, the problem of nonstationarity such as heteroskedasticity and serial correlation in the forecast losses should be taken seriously. Given the block-based form our test statistics we derive an alternative estimator of the long-run variance of the forecast losses. This estimator differs from the popular estimators of Andrews (1991) and Newey and West (1987) [see Müller (2007) for a review] and it is of independent interest. Finally, we extend results to settings that allow for stochastic volatility, and we conduct a local power analysis and highlight a few differences of our testing framework from the structural change test of Andrews (1993). Related aspects, such as estimating the timing of the instability and covering high-frequency setting with jumps, are being considered in a companion paper.

The rest of the chapter is organized as follows. Section 3.2 introduces the statistical setting, the hypotheses of interest and the test statistics. Section 3.3 derives the asymptotic null distribution under a continuous record. We discuss the estimation of

---

6In nonparametric change-point testing, related works are Wu and Zhao (2007) and Bibinger et al. (2017).
the asymptotic variance in Section 3.4. Some extensions and a local power analysis are presented in Section 3.5. Additional elements that are covered in our companion paper are briefly described in Section 3.6. A simulation study is contained in Section 3.7. Section 4.7 concludes the chapter. The supplemental material to this chapter contains all mathematical proofs and additional simulation experiments.

3.2 The Statistical Environment

Section 3.2.1 introduces the statistical setting with a description of the forecasting problem and the sampling scheme considered throughout. The underlying continuous-time model and its assumptions are introduced in Section 3.2.2. In Section 3.2.3 we set out the testing problem and state the relevant null and alternative hypotheses. The test statistics are presented in Section 3.2.4. Throughout we adopt the following notational conventions. All limits are taken as $T \to \infty$, or equivalently as $h \downarrow 0$, where $T$ is the sample size and $h$ is the sampling interval. All vectors are column vectors and for two vectors $a$ and $b$, we write $a \leq b$ if the inequality holds component-wise. For a sequence of matrices $\{A_T\}$, we write $A_T = o_P(1)$ if each of its elements is $o_P(1)$ and likewise for $O_P(1)$. If $x$ is a non-stochastic vector, $\|x\|$ denotes the its Euclidean norm, whereas if $x$ is a stochastic vector, the same notation is used for the $L^2$ norm. We use $[\cdot]$ to denote the largest smaller integer function and for a set $A$, the indicator function of $A$ is denoted by $1_A$. A sequence $\{u_{kh}\}_{k=1}^T$ is i.i.d. if the $u_{kh}$ are independent and identically distributed. We use $\xrightarrow{P}$, $\Rightarrow$ to denote convergence in probability and weak convergence, respectively. $\mathcal{M}_p^{\text{càdlàg}}$ is used for the space of $p \times p$ positive definite real-valued matrices whose elements are càdlàg. The symbol “≜” is definitional equivalence.
3.2.1 The Forecasting Problem

The continuous-time stochastic process $Z \triangleq (Y, X')$ is defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ and takes value in $Z \subseteq \mathbb{R}^{q+1}$ where $\{Y_t\}_{t \geq 0}$ is the variable to be forecast and $\{X_t\}_{t \geq 0}$ are the predictor variables. The index $t$ is defined as the continuous-time index and we have $t \in [0, N]$, where $N$ is referred to as the time span. In this chapter, $N$ will remain fixed. That is, the unobserved process $Z_t$ evolves within the fixed time horizon $[0, N]$ and the econometrician records $T$ of its realizations, with a sampling interval $h$, at discrete-time points $h, 2h, \ldots, Th$, where accordingly $Th = N$. A continuous record asymptotic framework involves letting the sample size $T$ grow to infinity by shrinking the time interval $h$ to zero at the same rate so that $N$ remains fixed. The index $k$ is used for the observation (or tick) times $k = 1, \ldots, T$.

The objective is to generate a series $\{Y_{(k+\tau)h}\}$ of $\tau$-step ahead forecasts. We shall adopt an out-of-sample procedure whereby splitting the time span $[0, N]$ into an in-sample and out-of-sample window, $[0, N_{in}]$ and $[N_{in} + h, N]$, respectively.\(^7\) The latter two time horizons are supposed to be fixed and therefore within the in-sample (or prediction) window a sample of size $T_m$ is observed whereas within the out-of-sample (or estimation) window the sample is of size $T_n = T - T_m - \tau + 1$. We consider a general framework that allows for the three traditional forecasting schemes: (1) a fixed forecasting scheme with discrete-time observations $h, 2h, \ldots, (T_m - 1)h, T_mh = N_{in}$; (2) a recursive forecasting scheme where at time $kh$ the prediction sample includes observations $h, \ldots, (k - 1)h, kh$; (3) a rolling forecasting scheme where the time

\(^7\)Indeed, $[0, N_{in}]$ corresponds to the in-sample window only for the fixed forecasting scheme to be introduced later—e.g., the rolling scheme only uses the most recent span of data of length $N_{in}$. A minor and straightforward modification to this notation should be applied when the recursive and rolling schemes are considered. However, for all methods $N_{in}$ indicates the artificial separation such that $N_{in} + h$ is the beginning of the out-of-sample period.
span of the rolling window is fixed and of the same length as \( N_{\text{in}} \) (i.e., at time \( kh \) the in-sample window includes observations \( kh - T_{m}h + h, \ldots, (k - 1)h, kh \).\(^8\)

The forecasts may be based on a parametric model whose time-\( kh \) parameter estimates are then collected into the \( q \times 1 \) random vector \( \beta_k \). If no parametric assumption is made, then \( \beta_k \) represents whatever semiparametric or nonparametric estimator used for generating the forecasts. The time-\( kh \) forecast is denoted by \( \hat{f}_k(\beta_k) \triangleq f \left(Z_{kh}, Z_{(k-1)h}, \ldots, Z_{(k-m_f+1)h}; \beta_k\right) \), where \( f \) is some measurable function. The notation indicates that the \( kh \)-time forecast is generated from information contained in a sample of size \( m_f \).\(^9\)

Next, we introduce a loss function \( L(\cdot) \) which serves for evaluating the performance of a given forecast model. More specifically, each out-of-sample loss \( L_{(k+\tau)h} (\beta_k) \triangleq L \left(Y_{(k+\tau)h}, \hat{f}_k(\beta_k)\right) \) constitutes a statistical measure of accuracy of the \( \tau \)-step forecast made at time \( kh \). However, given the objective of detecting potential instability of a certain forecasting method over time, we need additionally to introduce the in-sample losses \( L_{jh} (\beta_k) \triangleq L \left(Y_{jh}, \hat{y}_j(\beta_k)\right) \), where \( \hat{y}_j(\beta_k) \) is an in-sample fitted value with \( j \) varying over the specific in-sample window. That is, for each time-\( kh \) forecast there corresponds a sequence (indexed by \( j \)) of in-sample fitted values \( \hat{y}_j(\beta_k) \).\(^{10}\) Then, the testing problem turns into the detection of any “systematic difference” between the sequence of out-of-sample and in-sample losses; the formal measure of such difference under our context is provided below.

---

\(^8\)Equivalently, the observation times within the rolling widow at the \( k \)th’s observation are \( k - T_{m} + 1, \ldots, k \).

\(^9\)\( m_f \) varies with the forecast scheme; e.g., for the rolling scheme we have \( m_f = T_{m} \) while for the recursive scheme we have \( m_f = k \).

\(^{10}\)We have \( j = \tau + 1, \ldots, T_{m} \) for the fixed scheme, \( j = \tau + 1, \ldots, k \) for the recursive scheme and \( j = k - T_{m} + \tau + 1, \ldots, k \) for the rolling scheme.
3.2.2 The Underlying Continuous-Time Model

The process \( Z \) is a \( \mathbb{R}^{q+1} \)-valued semimartingale on \( (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P}) \) and we further assume that all processes considered in this chapter are càdlàg adapted and possess a \( \mathbb{P} \)-a.s. continuous path on \([0, N]\).\(^{11}\) The continuity property represents a key assumption in our setting and implies that \( Z \) is a continuous Itô semimartingale. The integral form for \( X_t \) is given by,

\[
X_t = x_0 + \int_0^t \mu_{X,s} ds + \int_0^t \sigma_{X,s} dW_{X,s},
\]

(3.2.1)

where \( \{W_{X,t}\}_{t \geq 0} \) is a \( q \times 1 \) Wiener process, \( \mu_{X,s} \in \mathbb{R}^q \) and \( \sigma_{X,s} \in \mathcal{M}_c^{\text{càdlàg}} \) are the drift and spot covariance process, respectively, and \( x_0 \) is \( \mathcal{F}_0 \)-measurable. We incorporate model misspecification into our framework by allowing for a large non-zero drift which adds to the residual process:

\[
Y_t \triangleq y_0 + (\beta^*)' X_{t-} + \int_0^t \mu_{e,s} h^{-\vartheta} ds + e_t, \quad e_t \triangleq \int_0^t \sigma_{e,s} dW_{e,s}
\]

(3.2.2)

where \( \beta^* \in \mathbb{R}^q \), \( \{W_{e,t}\}_{t \geq 0} \) is a standard Wiener process, \( \sigma_{e,s} \in \mathbb{R}_+ \) is its associated volatility, \( \mu_{e,s} \in \mathbb{R} \) and \( y_0 \) is \( \mathcal{F}_0 \)-measurable. In (3.2.2), the last two terms on the right-hand side account for the residual part of \( Y_t \) which is not explained by \( X_{t-} \), where \( X_{t-} = \lim_{s \uparrow t} X_s \). We assume \( \vartheta \in [0, 1/8) \) so that the factor \( h^{-\vartheta} \) inflates the infinitesimal mean of the residual component thereby approximating a setting with arbitrary misspecification.

**Remark 3.2.1.** In (3.2.2), misspecification manifests itself in the form of (time-varying) non-zero conditional mean of the residual process, and in giving rise to serial dependence in the disturbances which in turn leads to dependence in the sequence

\(^{11}\)For accessible treatments of the probabilistic elements used in this section we refer to Aït-Sahalia and Jacod (2014), Jacod and Shiryaev (2003), Jacod and Protter (2012), Karatzas and Shreve (1996) and Protter (2005).
of forecast losses.\textsuperscript{12} Hence, this specification is similar in spirit to the near-diffusion assumption of Foster and Nelson (1996) who studied the impact of misspecification in ARCH models. On the other hand, Casini and Perron (2017a) introduced a “large-drift” asymptotics with $h^{-1/2}$ to deal with non-identification of the drift in their context. Technically, the latter specification implies that as $h$ becomes small the drift features larger oscillations that add to the local Gaussianity of the stochastic part. Casini and Perron (2017a) referred to this specification as small-dispersion assumption. Finally, note that the presence of $h^{-\theta}$ can also be related to the signal plus small Gaussian noise of Ibragimov and Has’minskiĭ (1981) if one sets $\varepsilon_h = h^\theta$ in their model in Section VII.2.

\textbf{Assumption 3.1.} We have the following assumptions: (i) The processes $\{X_t\}_{t \geq 0}$ and $\Sigma^0 \triangleq \{\sigma_{X,t}, \sigma_{e,t}\}_{t \geq 0}$ have $\mathbb{P}$-a.s. continuous sample paths; (ii) The processes $\{\mu_{X,t}\}_{t \geq 0}, \{\mu_{e,t}\}_{t \geq 0}, \{\sigma_{X,t}\}_{t \geq 0}$ and $\{\sigma_{e,t}\}_{t \geq 0}$ are locally bounded; (iii) There exists $0 < \sigma_- < \sigma_+ < \infty$ such that $\mathbb{P}$-a.s. $\inf_{t \in \mathbb{R}^+} \sigma_{V,t}^2 \geq \sigma_-^2$ and $\sigma_+^2 \geq \sup_{t \in \mathbb{R}^+} \sigma_{V,t}^2$ with $V = X, e$; (iv) $\sigma_{X,t} \in \mathcal{M}_{\text{chdl}}^\text{adlag}$ and $\sigma_{e,t} \in \mathcal{M}_{\text{chdl}}^\text{adlag}$ and the conditional variance (or spot covariance) is defined as $\Sigma_{X,t} = \sigma_{X,t} \sigma_{X,t}'$, which for all $t < \infty$ satisfies $\int_0^t \Sigma_{X,s}^{(j,j)} ds < \infty, (j = 1, \ldots, q)$ where $\Sigma_{X,t}^{(j,r)}$ denotes the $(j,r)$-th element of $\Sigma_{X,t}$. Furthermore, for every $j = 1, \ldots, q$, and $k = 1, 2, \ldots, T$, the quantity $h^{-1} \int_{(k-1)h}^{kh} \Sigma_{X,s}^{(j,j)} ds$ is bounded away from zero and infinity, uniformly in $k$ and $h$; (v) The disturbance process $e_t$ is orthogonal (in martingale sense) to $X_t$: $\langle e, X \rangle_t = 0$ identically for all $t \geq 0$.$^{13}$

Part (i) rules out jump processes from our setting. We relax this restriction in our companion paper; see Section 3.6. Part (ii) restricts those processes to be locally

\textsuperscript{12}Asymptotically, these features can be dealt with basic arguments used in the high-frequency financial statistics literature; however, when $h$ is not too small one needs methods that are robust in finite-samples to such misspecification-induced properties. More precisely, we will propose an appropriate estimator of the long-run variance of the sequence of forecast losses in Section 3.4.

\textsuperscript{13}The angle brackets notation $\langle \cdot, \cdot \rangle$ is used for the predictable quadratic variation process.
bounded. These should be viewed as regularity conditions rather than assumptions and are standard in the financial econometrics literature [see Barndorff-Nielsen and Shephard (2004), Li and Xiu (2016) and Li et al. (2017)]; recently, they have been used by Casini and Perron (2017a) in the context of structural change models.

The continuous-time model in (3.2.1)-(3.2.2) is not observable. The econometrician only has access to \( T \) realizations of \( Y_t \) and \( X_t \) with a sampling interval \( h > 0 \) over the horizon \([0, N]\). For each \( h > 0 \), \( Z_{kh} \in \mathbb{R}^{q+1} \) is a random vector step function that jumps only at time 0, \( h \), \( 2h \),..., and so on. The discretized processes \( Y_{kh} \) and \( X_{kh} \) are assumed to be adapted to the increasing and right-continuous filtration \( \{\mathcal{F}_t\}_{t \geq 0} \). The increments of a process \( U \) are denoted by \( \Delta_h U_k = U_{kh} - U_{(k-1)h} \). A seminal result known as Doob-Meyer Decomposition [cf. the original sources are Doob (1953) and Meyer (1967); see also Section III.3 in Protter (2005)] allows us to decompose the semimartingale process \( X_t \) into a predictable part and a local martingale part. Hence, it follows that we can write for \( k = 1, \ldots, T \),

\[
\Delta_h X_k \triangleq \mu_{X,kh} \cdot h + \Delta_h M_{X,k} \quad \text{where the drift} \quad \mu_{X,t} \in \mathbb{R}^q \quad \text{is} \quad \mathcal{F}_{t-h} \quad \text{measurable, and} \quad M_{X,kh} \in \mathbb{R}^q \quad \text{is a continuous local martingale with finite conditional covariance matrix} \quad \mathbb{P}\text{-a.s.} \quad \mathbb{E} \left[ \Delta_h M_{X,k} \Delta_h M_{X,k}^\prime \mid \mathcal{F}_{(k-1)h} \right] = \Sigma_{X,(k-1)h} \cdot h. \]

Turning to equation (3.2.2), the error process \( \{\Delta_h e_k^*, \mathcal{F}_t\} \), with \( \Delta_h e_k^* \triangleq \sigma_{e,(k-1)h} \Delta_h W_{e,k} \), is then a continuous local martingale difference sequence taking its values in \( \mathbb{R} \) with finite conditional variance \( \mathbb{E} \left[ (\Delta_h e_k^*)^2 \mid \mathcal{F}_{(k-1)h} \right] = \sigma_{e,(k-1)h}^2 \cdot h, \mathbb{P}\text{-a.s.} \). Therefore, we express the discretized analogue of (3.2.2) as

\[
\Delta_h Y_k = (\beta^\ast)^\prime \Delta_h X_{k-\tau} + \mu_{e,kh} \cdot h^{1-\vartheta} + \Delta_h e_k, \quad k = \tau + 1, \ldots, T. \quad (3.2.3)
\]

**Remark 3.2.2.** As explained above, we accommodate possible model misspecification by adding the component \( \mu_{e,k} \cdot h^{1-\vartheta} \). In the forecasting literature, often one di-
rectly imposes restrictions on the sequence of losses, say, \( L(e_k) \) where \( e_k = Y_k - \hat{f}_k(\hat{\beta}_k) \)
is a forecast error. There are two main differences from our approach. First, in order to facilitate illustrating our novel framework to the reader, we have chosen, without loss of generality, to express directly the relationship between \( \Delta_h Y_{k+\tau} \) and \( \Delta_h X_k \) while at the same time, allowing for misspecification by including \( \mu_{e,kh} \cdot h^{1-\theta} \). A second distinction from the classical approach is that the latter imposes restrictions on the sequences of losses such as mixing and ergodicity conditions, covariance stationary and so on. In contrast, our infill asymptotics does not require us to impose any ergodic or mixing condition [cf. Casini and Perron (2017a)].

Finally, we have an additional assumption on the path of the volatility process \( \{\sigma^2_{e,t}\}_{t \geq 0} \). This turns out be important because it partly affects the local behavior of the forecast losses.

**Assumption 3.2.** For small \( \eta > 0 \), define the modulus of continuity of \( \{\sigma_{e,t}\}_{t \geq 0} \) on the time horizon \( [0, N] \) by \( \phi_{\sigma,\eta,N} = \sup_{s,t \in [0,N]} \{|\sigma_t - \sigma_s| : |t-s| < \eta\} \). We assume that \( \phi_{\sigma,\eta,\tau_h,N} \leq K_h \eta \) for some sequence of stopping times \( \tau_h \to \infty \) and some \( \mathbb{P}\)-a.s. finite random variable \( K_h \).

The assumption essentially states that \( \phi_{\sigma,\eta,N} \) is locally bounded and \( \{\sigma_{e,t}\}_{t \geq 0} \) is Lipschitz continuous. Lipschitz volatility is a more than reasonable specification for the macroeconomic and financial data to which our analysis is primarily directed. Indeed, the basic case of constant variance \( \sigma^2 \) is easily accommodated by the assumption. Time-varying volatility is also covered provided \( \sigma^2_{e,t} \) is sufficiently smooth. However, the assumption rules out some standard stochastic volatility models often used in finance. We relax that assumption in Section 3.5, so that we can extend our results to, for example, stochastic volatility models driven by a Wiener process.
3.2.3 The Hypotheses of Interest

As time evolves, a forecast model can suffer instability for multiple reasons. However, incorporating model misspecification into our framework necessarily implies that the exact form of the instability is unknown and thus one has to leave it unspecified. This differs from the classical setting for estimation of structural change models [cf. Bai and Perron (1998) and Casini and Perron (2017a)] where (i) the break date is well-defined as it is part of the definition of the econometric problem, and (ii) the form of the instability is explicitly specified through a discrete shift in a regression parameter. In contrast, under our context we remain agnostic regarding both (i) and (ii). There may be multiple dates at which the forecast model suffers instability and they might be interrelated in a complicated way. Forecast instability may manifest itself in several forms, including gradual, smooth or recurrent changes in the predictive relationship between $Y_{(k+\tau)h}$ and $X_{kh}$; certainly, there could also be discrete shifts in $\beta^*$—arguably the most common case in practice—but this is only a possibility in our setting and not an assumption as in structural change models. A forecast failure then reflects the forecaster’s failure to recognize the shift in the predictive power of $X_{kh}$ on $Y_{(k+\tau)h}$. On the other hand, even if one can rule out shifts in $\beta^*$, a forecast instability may be induced by an increase/decrease in the uncertainty in the data which might result, for example, from changes in the unconditional variance of the target variable. In this case, the predictive ability of $X_{kh}$ on $Y_{(k+\tau)h}$, as described for instance by a parameter $\beta$, remains stable while due to an increase in the unconditional variance of $Y_{(k+\tau)h}$ it might become weak and in turn the forecasting power might breakdown. Tests for forecast failure such as those proposed in this chapter and the ones proposed in Giacomini and Rossi (2009) are designed to have power against both of the above
3.2.3.1 The Null and Alternative Hypotheses on Forecast Instability

Define at time \((k + \tau)h\) a surprise loss given by the deviation between the time-\((k + \tau)h\) out-of-sample loss and the average in-sample loss: \(SL_{(k+\tau)h}(\hat{\beta}_k) \triangleq L_{(k+\tau)h}(\hat{\beta}_k) - L_{kh}(\hat{\beta}_k)\), for \(k = T_m, \ldots, T - \tau\), where \(L_{kh}(\hat{\beta}_k)\) is the average in-sample loss computed according to the specific forecasting scheme. One can then define the average of the out-of-sample surprise losses

\[
\bar{SL}_{N_0}(\hat{\beta}_k) \triangleq N_0^{-1} \sum_{k=T_m}^{T-\tau} SL_{(k+\tau)h}(\hat{\beta}_k),
\]

where \(N_0 \triangleq N - N_{in} - h\) denotes the time span of the out-of-sample window. In the classical discrete-time setting, under the hypotheses of no forecast instability one would naturally test whether \(\bar{SL}_{N_0}(\beta^*)\) has zero mean, where \(\beta^*\) is the pseudo-true value of \(\beta\). If the forecasting performance remains stable throughout the whole sample then there should be no systematic surprise losses in the out-of-sample window and thus \(\mathbb{E} \left[ N_0^{-1} \sum_{k=T_m}^{T-\tau} SL_{(k+\tau)h}(\beta^*) \right] = 0\). This reasoning motivated the forecast breakdown test of Giacomini and Rossi (2009). Therefore, under the classical asymptotic setting one exploits time series properties of the process \(SL_{(k+\tau)h}(\beta^*)\) such as ergodicity and mixing together with the representation of the hypotheses by a global moment restriction. By letting the span \(N \to \infty\), this method underlies the classical approach to statistical inference but does not directly extend to an infill

\[14\] Recently, Perron and Yamamoto (2018) proposed to apply modified versions of classical structural break tests to the forecast failure setting. However, their testing framework and hence their null hypotheses are different from ours because they do not fix a model-method pair but only fix the forecast model under the null.

\[15\] By definition \(N_0\) is fixed and should not be confused with \(T_n\), which indicates the number of observations in the out-of-sample window. Indeed, \(N_0 = T_n h\).

\[16\] Global refers to the property that the zero-mean restriction involves the entire sequence of forecast losses.
asymptotic setting. Under continuous-time asymptotics, identification of parameters is achieved by properties of the paths of the involved processes and not by moment conditions. This constitutes the key difference and requires one to recast the above hypotheses into an infill setting thereby making use of assumptions on an underlying continuous-time data-generating mechanism which is assumed to govern the observed data.

We begin with observing that the sequence of losses \( \{L_{kh}(\cdot)\} \) can be viewed as realizations from an underlying continuous-time process \( \{\tilde{L}_t\}_{t \geq 0} \) with \( \tilde{L}_t \triangleq \int_0^t L_s(Y_s, X_{s-}; \beta^*) ds \). That is, \( \tilde{L}_t \) consists of temporally integrated forecast losses where \( L_t \) is the loss at time \( t \) and is defined by some transformation of the target variable \( Y_t \) and of the predictor \( X_{t-} \).\(^{17}\) In order to provide a general theory, we focus on families of loss functions that depend only on the forecast error.\(^ {18}\) We denote this class by \( L_e \) and we say that the loss function \( L(\cdot, \cdot; \cdot) \in L_e \) if \( L_t(Y_t, X_{t-}; \beta) = L_t(e_t; \beta) \) for all \( t \in [0, N] \), where \( e_t = Y_t - \hat{f}_t(\beta) \). The class \( L_e \) comprises the vast majority of loss functions employed in empirical work, including among others the popular Quadratic loss, Absolute error loss and Linex loss. The following examples illustrate how these loss functions are constructed under our setting. For the rest of this section, assume for simplicity \( y_0 = 0, \mu_{e,} = 0 \) and that \( X_t \) is one-dimensional in (3.2.2).

Example 3.2.1. (QL: Quadratic Loss)

The Mean Squared Error or Quadratic loss function is symmetric and is by far the most commonly used by practitioners. Given (3.2.2), we have \( e_t = Y_t - \beta^* X_{t-} \). Then \( L(e) = ae^2 \) or \( L_t(Y_t, X_{t-}; \beta^*) = ae^2_t \) with \( a > 0 \).

\(^{17}\)The definition of \( \tilde{L}_t \) uses that so long as the forecast step \( \tau \) is small and finite one can approximate \( X_{s-\tau h} \) by \( X_{s-} \) for sufficiently small \( h > 0 \).

\(^{18}\)The most popular loss functions used in economic forecasting are within this category [see Elliott and Timmermann (2016) for a recent incisive account of the literature]. Extension to \textit{ad hoc} loss functions requires specific treatment that might vary from case to case.
Example 3.2.2. (LL: Linex Loss)
The Linear-exponential or Linex loss was introduced by Varian (1975) and it is an example of asymmetric loss function. By the same reasoning as in the Quadratic loss case, we have
\[ L(e) = a_1 (\exp (a_2 e) - a_2 e - 1) \]
or
\[ L_t(Y_t, X_{t-}; \beta^*) = a_1 (\exp (a_2 e_t) - a_2 e_t - 1) \]
with \( a_1 > 0, a_2 \neq 0 \).

Below we make very mild pathwise assumptions on the process \( Z \) which imply restrictions on \( \{ \tilde{L}_t \} \). We derive asymptotic results under Lipschitz continuity (in \( t \)) of the coefficients of the system of stochastic differential equations driving the data \( \{ Z_t \} \). We apply the techniques of stochastic calculus to formulate our testing problem. To avoid clutter, we introduce the notation \( g(Y_t, X_{t-}; \beta^*) = L_t(Y_t, X_{t-}; \beta^*) \) and its shorthand \( g(e_t; \beta^*) = L_t(e_t; \beta^*) \).

By Itô Lemma, [cf. Section II.7 in Protter (2005)], under smoothness of \( g(e_t; \beta^*) \),
\[
dL_t(e_t; \beta^*) = \sigma^2 e,t \frac{\partial^2 g(e_t; \beta^*)}{\partial e^2} dt + \sigma e,t \frac{\partial g(e_t; \beta^*)}{\partial e} dW_{e,t}.
\]
Let \( E_\sigma \) denote the expectation conditional on the path \( \{ \sigma_{e,t} \} \). The instantaneous mean of \( dL(e_t; \beta^*) \) is \( E_\sigma [dL(e_t; \beta^*)] = 2^{-1} \sigma^2 e,t \sigma e,t \frac{\partial g(e_t; \beta^*)}{\partial e^2} dW_{e,t} \). Note that the latter is a symbolic abbreviation for
\[
E_\sigma [L_t(e_t; \beta^*) - L_s(e_s; \beta^*)] = \frac{\sigma^2 e,t}{2} \sigma e,t \left[ \frac{\partial^2 g(e_t; \beta^*)}{\partial e^2} \right] (t - s) + o(t - s), \quad \text{as } s \uparrow t.
\]
Since the coefficients of the original system of stochastic equations are Lipschitz continuous in \( t \), one can verify that \( E_\sigma [dL(e_t; \beta^*)] \) is also Lipschitz upon regularity conditions on \( g(\cdot, \beta^*) \) and time-\( t \) information.

We denote by \( \text{Lip}([0, N]) \) the class of Lipschitz continuous functions on \([0, N] \).

---

\( ^{19} \)The notation implicitly assumes that the same loss function is used for estimation and prediction which in turn implies that the subscript \( t \) in \( L_t(e_t; \beta^*) \) can be omitted since it can be understood from that of the argument \( e_t \).
Let \( \{c_t\}_{t \geq 0} \) denote a continuous-time stochastic process that is \( \mathbb{P} \)-a.s. locally bounded and adapted.

**Definition 3.2.1.** The process \( \{c_t\}_{t \geq 0} \) belongs to

\[
\sup_{s, t \in [0, \tau_h \wedge N], t \neq s} |c_t - c_s| < K_h |t - s|,
\]

for some sequence of stopping times \( \tau_h \to \infty \) and some \( \mathbb{P} \)-a.s. finite random variable \( K_h \).

We are in a position to formulate the testing problem in terms of the pathwise property of \( L_t (e_t; \beta^*) \). This implies that the hypotheses are specified in terms of random events which differs from classical hypotheses testing but it is typical under continuous-time asymptotics; see Aït-Sahalia and Jacod (2012) (for many references), Li et al. (2016) and Reiß et al. (2015). We consider the following hypotheses: for any \( L (\cdot; \cdot) \in L_e \),

\[
H_0 : \left\{ \lim_{s \uparrow t} \mathbb{E}_\sigma [L_t (e_t; \beta^*) - L_s (e_s; \beta^*)] \right\} \in Lip ([N_{in} + h, N]), \tag{3.2.5}
\]

which means that we wish to discriminate between the following two events that divide \( \Omega \):

\[
\Omega_0 \equiv \left\{ \omega \in \Omega : \left\{ \lim_{s \uparrow t} \mathbb{E}_\sigma [L_t (e_t (\omega); \beta^*) - L_s (e_s (\omega); \beta^*)] \right\} \in Lip ([N_{in} + h, N]) \right\},
\]

\[
\Omega_1 \equiv \Omega \setminus \Omega_0
\]

The dependence of the hypotheses on \( \omega \) is appropriate because each event \( \omega \) generates a certain path of \( L (e_t (\cdot); \beta^*) \) on \([0, N]\), where \( de_t (\omega) = \sigma_{e,t} (\omega) dW_{e,t} (\omega) \). The hypotheses requires a Lipschitz condition to hold on \([N_{in} + h, N]\), where \( N_{in} \) is the

---

\(^{20}\)Precise assumptions will be stated below.
usual artificial separation date after which the first forecast is made. $N_{in}$ is taken as
given here because the testing problem applies to a specific method-model pair and
$N_{in}$ is part of the chosen forecasting method. From a practical standpoint, it would
be helpful if this separation date is such that the forecast model is stable on $[0, N_{in}]$
[see Casini and Perron (ming) for more details]. The latter property is, however,
unknown a priori by the practitioner. We cover this case in Section 3.6.

**Example 3.2.3. (QL; cont’d)**

For the Quadratic loss $L(e) = ae^2$, Itô Lemma yields $E_{\sigma} \left[ dL_t(e_t; \beta^*) / dt \right] = a\sigma^2_{e,t}$. If
$\sigma_{e,t}$ is Lipschitz continuous, then the hypothesis $H_0$ holds.

**Example 3.2.4. (LL; cont’d)**

From Itô Lemma,

$$dL_t(e_t; \beta^*) = a_1 \left\{ a_2 \left[ 2^{-1} a_2^2 \sigma^2_{e,t} \exp(a_2 e_t) dt + (\exp(a_2 e_t) - 1) \sigma_{e,t} dW_{e,t} \right] - 1 \right\}.$$ 

Consequently, by Itô Isometry [cf. Section 3.3.2 in Karatzas and Shreve (1996) or
Lemma 3.1.5 in Øksendal (2000)]

$$E_{\sigma} \left[ dL(e_t; \beta^*) / dt \right] = a_1 \left( a_2^2 \left( \sigma^2_{e,t} / 2 \right) \right) \exp \left( a_2^2 \left( \int_0^t \sigma^2_{e,s} ds / 2 \right) \right),$$

and hypotheses $H_0$ is seen to hold under Lipschitz continuity of $\sigma_{e,t}$.

21Recall that composition of Lipschitz functions is Lipschitz and that under our context
$\exp \left( a_2 \left( \int_0^t \sigma^2_{e,s} ds / 2 \right) \right)$ is Lipschitz because (i) $\sigma^2_{e,s}$ is locally bounded and Lipschitz, and (ii) $t \leq N$
and $N$ remains fixed.

We have reduced the forecast instability problem to examination of the local
properties of the path of $L_t$. However, we still have to face the question on how to
use the data to test $H_0$ in practice. Even if we could observe $\tilde{L}_t$, it would not be clear
how to formulate a testing problem on the stability of $L_t$ by using path properties of $\tilde{L}_t$. The reason is that $\tilde{L}_t$ is always absolutely continuous by definition, and thus it would provide little information on the large deviations of the forecast error $e_t$. In order to study the local behavior of $L_t$ one needs to consider the small increments of $L_t$ close to time $t$. Leaving the definition of $\tilde{L}_t$ aside for a moment, observe that $\mathbb{P}$-a.s. continuity of $Z_t$ is equivalent to having the relationship between $Y_t$ and $X_t$ holding for any infinitesimal interval of time. For the basic parametric linear model: $dY_t = \beta^* dX_t + de_t$. Then, the forecast loss is $L(de_t)$, which is difficult to interpret in rigorous probabilistic terms. However, we can consider its discrete-time analogue. We normalize the forecast error by the factor $\psi_h = h^{1/2}$ and redefine $L_{\psi, kh} (\Delta_h e_k; \beta^*) \triangleq L_{kh} \left( \psi_h^{-1} \Delta_h e_k; \beta^* \right)$.\(^{22}\) Then, for all $k$, the mean of $L_{\psi, kh} (\Delta_h e_k; \beta^*)$—conditional on $\sigma_{e, kh}$—depends on the parameters of the model and its local behavior can be used as a proxy for the local behavior of the infinitesimal mean of $dL_t (e_t; \beta^*)$. If the corresponding structural parameters of the continuous-time data-generating process satisfy a Lipschitz continuity in $t$, then—knowing $\sigma_{e, kh}$—also $\mathbb{E}_{\sigma} [L_{\psi, kh} (\Delta_h e_k; \beta^*)]$ should be Lipschitz in the continuous-time limit. Under the hypotheses $H_0$ there should be no break in $\mathbb{E}_{\sigma} [L_{\psi, kh} (\Delta_h e_k; \beta^*)]$ and an appropriately defined right local average of $L_{\psi, kh} (\Delta_h e_k; \beta^*)$ should not differ too much from its left local average. That is, one can test for forecast instability by using a two-sample t-test over asymptotically vanishing time blocks.

**Example 3.2.5.** (QL; cont’d)

Conditional on $\{ \sigma_t \}_{t \geq 0}$, $\Delta_h e_k \sim \mathcal{N} \left( 0, \sigma_{e, (k-1)h}^2 \cdot h \right)$. Thus, $\mathbb{E}_{\sigma} [L_{\psi, kh} (\Delta_h e_k; \beta^*)] = a \sigma_{e, (k-1)h}^2$. If $\sigma_{e, t}$ is Lipschitz continuous, then the hypothesis $H_0$ holds.

**Example 3.2.6.** (LL; cont’d)

\(^{22}\)Alternatively, $L_{\psi, kh} (\Delta_h Y_k, \Delta_h X_k; \beta^*) = L_{kh} \left( \psi_h^{-1} (\Delta_h Y_k - \beta^* \Delta_h X_k) \right)$. 


Similar to the Quadratic loss case, we have

$$\mathbb{E}_\sigma [dL(e_t; \beta^*) / dt] = a_1 (a_2^2 (\sigma^2_{e,t} / 2)) \exp \left( a_2^2 \left( \int_0^t \sigma^2_{e,s} ds \right) / 2 \right),$$

Again, the hypotheses $H_0$ is satisfied if $\sigma_{e,t}$ is Lipschitz.

Both examples demonstrate that pathwise assumptions on the data-generating process implies restrictions on the properties of the sequence of loss functions. For the QL example, if there is a structural break at the observation $k = T_b$, then this would result in the mean of $L_{\psi,kh} (\Delta_h e_k; \beta^*)$ shifting to a new level after time $T_b h$. Given that the same reasoning extends to the sequence of surprise losses, one may consider to construct a test statistic on the basis of the local behavior of the surprise losses over time. If there is no instability in the predictive ability of a certain model, then the sequence of out-of-sample surprise losses should display a certain stability. Under the framework of Giacomini and Rossi (2009), this stability is interpreted in a retrospective and global sense as a zero-mean restriction on the sequence over the entire out-of-sample. In contrast, under our continuous-time setting, this stability manifests itself as a continuity property of the path of the continuous-time counterpart of the sequence.

### 3.2.4 The Test Statistics

By inspection of the null hypotheses in (3.2.5), it is evident that a considerable number of forms of instabilities are allowed. These may result from discrete shifts in a model's structural parameter and/or in structural properties of the processes considered such as conditional and unconditional moments and so on. This first set of nonstationarities relates to the popular case of structural changes which are designed to be detected with high probability by the structural break tests of, among
others, Andrews (1993) and Andrews and Ploberger (1994), Bai and Perron (1998) and Elliott and Müller (2006) in univariate settings and of Qu and Perron (2007) in multivariate settings. However, a forecast instability may be generated by many other forms of nonstationarities against which such classical tests for structural breaks are not designed for and consequently they might have little power against. For example, consider the case of smooth changes in model parameters, or in the unconditional variance of \( Y_{kh} \). Even more serious would be the presence of recurrent smooth changes in the marginal distribution of the predictor since in this case the above-mentioned tests are likely to falsely reject \( H_0 \) too often [cf. Hansen (2000)]. Thus, the null hypotheses of no forecast instability calls for a new statistical hypotheses testing framework. Ideally, in this context one needs a test statistic that retains power against any discontinuity, jump and recurrent switch at any point in the out-of-sample and for any magnitude of the shift. We propose a test statistic which aims asymptotically at distinguishing any discontinuity from a regular Lipschitz continuous motion. We introduce a sequence of two-sample t-tests over asymptotically vanishing adjacent time blocks. This should lead to significant gains in power whenever on fixed time intervals the out-of-sample losses exhibit instabilities of any form such as breaks, jumps and relatively large deviations. Such gains are likely to occur especially when instabilities take place within a small portion of the sample relative to the whole time span—a common case in practice that has characterized many episodes of forecast failure in economics.

Interestingly, for the Quadratic loss function we can exploit properties of the local quadratic variation and propose a self-normalized test statistic. Thus, we separate the discussion on the Quadratic loss from that on general loss functions. Let

\[
SL_{\psi,(k+\tau)h} (\hat{\beta}_k) \triangleq L_{\psi,(k+\tau)h} (\hat{\beta}_k) - L_{\psi,kh} (\hat{\beta}_k), \quad k = T_m, \ldots, T - \tau.
\]

Next, we partition
the out-of-sample into \( m_T \triangleq \lfloor T_n/n_T \rfloor \) blocks each containing \( n_T \) observations. Let
\[
B_{h,b} \triangleq n_T^{-1} \sum_{j=1}^{n_T} SL_{\psi,T_{m+\tau+bT}+j-1} \left( \hat{\beta}_{T_{m+bT}+j-1} \right)
\]
and
\[
\mathbb{E}_\sigma [dL(e_t; \beta^*) / dt] = a_1 \left( a_2 \left( \frac{\sigma_{e,t}^2}{2} \right) \right) \exp \left( a_2^2 \left( \int_0^t \sigma_{e,s}^2 ds \right) / 2 \right),
\]
for \( b = 0, \ldots, \lfloor T_n/n_T \rfloor - 1 \).

### 3.2.4.1 Test Statistics under Quadratic Loss

We propose the following statistic
\[
B_{\text{max},h} (T_n, \tau) \triangleq \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| \frac{B_{h,b+1} - B_{h,b}}{B_{h,b+1}} \right|.
\]
The quantity \( B_{h,b} \) is a local average of the surprise losses within the block \( b \). We have partitioned the out-of-sample window into \( m_T \) blocks of asymptotically vanishing length \( [bn_T h, (b + 1) n_T h] \). We consider an asymptotic experiment in which the number of blocks \( m_T \) increases at a controlled rate to infinity while the per-block sample size \( n_T \) grows without bound at a slower rate than the out-of-sample size \( T_n \). The appeal of the \( B_{\text{max},h} (T_n, \tau) \) statistic is that a large deviation \( B_{h,b+1} - B_{h,b} \) suggests the existence of either a discontinuity or non-smooth shift in the surprise losses close to time \( bn_T h \) and thus it provides evidence against \( H_0 \). We comment on the nature of the normalization \( B_{h,b+1} \) in the denominator of \( B_{\text{max},h} \) below, after we introduce a version of \( B_{\text{max},h} \) statistic which uses all admissible overlapping blocks of length \( n_T h \):

\[
MB_{\text{max},h} (T_n, \tau) \triangleq \max_{i=n_T, \ldots, T_n-n_T} \left| \left( n_T^{-1} \sum_{j=i-n_T+1}^i SL_{\psi,T_{m+\tau+j-1}} \left( \hat{\beta}_{T_{m+j-1}} \right) - n_T^{-1} \sum_{j=i+1}^{i+n_T} SL_{\psi,T_{m+\tau+j-1}} \left( \hat{\beta}_{T_{m+j-1}} \right) \right) / B_{h,i} \right|,
\]
where $B_{h,i} = n_T^{-1} \sum_{j=i+1}^{i+n_T} L_{\psi,T_m+\tau+j-1} \left( \bar{\beta}_{T_m+j-1} \right)$. Since under the alternative hypotheses the exact location of the change-point—or possibly the locations of the multiple change-points—within the block might actually affect the power of the $B_{\text{max},h}$-based test in small samples, we indeed find in our simulation study that the test statistic $MB_{\text{max},h}$ which uses overlapping blocks is more powerful especially when the instability arises in forms other than the simple one-time structural change. Thus, the power of the $B_{\text{max},h}$ test is slightly sensible to the actual location of the change-point within the block, with higher power achieved when the change-point is close to either the beginning or the end of the block. In contrast, the statistical power of $MB_{\text{max},h}$ is uniform over the location of the change-point in the sample. The latter property is not shared by the exiting test of Giacomini and Rossi (2009) given that its power tends to be substantially lower if the instability is not located at about mid sample.

An important characteristic of both $B_{\text{max},h}$ and $MB_{\text{max},h}$ is that they are self-normalized; no asymptotic variance appears in their definition. The reason for why $B_{h,b+1}$ appears in the denominator of, for example, $B_{\text{max},h}$ is that even though $B_{h,b+1}$ constitutes a more logical self-normalizing term, it might be close to zero in some cases. This would occur under Quadratic loss if, for example, $\sigma_{e,t} = \sigma_e$ for all $t \geq 0$. This is not true for the factor $B_{h,b+1}$.

In addition, observe that allowing for misspecification naturally leads one to deal carefully with artificial serial dependence in the forecast losses in small samples. Thus, we consider a version of the statistics $B_{\text{max},h}$ and $MB_{\text{max},h}$ that are normalized by their asymptotic variance:

$$Q_{\text{max},h} (T_n, \tau) \triangleq \nu_L^{-1} \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor-2} |B_{h,b+1} - B_{h,b}|,$$
and similarly,

\[ \text{MQ}_{\text{max}, h}(T_n, \tau) \]

\[ \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} SL_{\psi, T_m + \tau + j - 1} (\tilde{\beta}_{T_m + j - 1}) - n_T^{-1} \sum_{j=i-n_T+1}^{i} SL_{\psi, T_m + \tau + j - 1} (\tilde{\beta}_{T_m + j - 1}) \right|. \]

The quantity \( \nu_L \) standardizes the test statistic so that under the null hypotheses we obtain a distribution-free limit. This can be useful because given the fully nonstationary setting together with the possible consequences of misspecification in finite-samples, standardization by the square-root of the asymptotic variance \( \nu_L^2 \) might lead to a more precise empirical size in small samples. We relegate theoretical details on \( \nu_L \) as well as on its estimation to Section 3.4 where we also present a discussion about its relation with the choice of the number of blocks.

### 3.2.4.2 Test Statistics under General Loss Function

For general loss \( L \in L_e \), we propose the following statistic,

\[ G_{\text{max}, h}(T_n, \tau) \triangleq \max_{b=0, \ldots, \lceil T_n / n_T \rceil - 2} \left| \frac{B_{h,b+1} - B_{h,b}}{\sqrt{D_{h,b+1}}} \right|, \]

where \( B_{h,b}, B_{h,b+1} \) are defined as in the quadratic case and

\[ D_{h,b+1} \triangleq n_T^{-1} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m + \tau + (b+1)n_T + j - 1)h} (\tilde{\beta}_{T_m + (b+1)n_T + j - 1}) - \mathcal{L}_{\psi,b} (\tilde{\beta}) \right)^2, \]

with \( \mathcal{L}_{\psi,b} (\tilde{\beta}) \triangleq n_T^{-1} \sum_{j=1}^{n_T} L_{\psi,(T_m + \tau + bn_T + j - 1)h} (\tilde{\beta}_{T_m + bn_T + j - 1}). \) The interpretation of \( G_{\text{max}, h} \) is essentially the same as of \( B_{\text{max}, h} \), the only difference arising from the denominator \( \sqrt{D_{h,b+1}} \) that estimates the within-block variance. A version that uses all
overlapping blocks is

\[ MG_{\text{max},h}(T_n, \tau) \triangleq \max_{i=n_T, \ldots, n_T - n_T} \frac{n_T^{-1} \sum_{j=i+1}^{i+n_T} SL_{\psi,T_m+\tau+j-1} (\beta_{T_m+j-1}) - n_T^{-1} \sum_{j=i-n_T+1}^{i} SL_{\psi,T_m+\tau+j-1} (\beta_{T_m+j-1})}{\sqrt{D_{h,i}}}, \]

where \( D_{h,i} \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( L_{\psi,(T_m+\tau+j-1)h} (\beta_{T_m+j-1}) - T_{\psi,i} (\beta) \right)^2, \) with

\[ T_{\psi,i} (\beta) \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} L_{\psi,(T_m+\tau+j-1)h} (\beta_{T_m+j-1}). \]

As argued above, it is useful to consider versions of the statistic \( B_{\text{max},h} \) and \( MB_{\text{max},h} \) that are normalized by their asymptotic variance:

\[ Q_{\text{max},h}^G(T_n, \tau) \triangleq \nu_L^{-1} \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b+1} - B_{h,b}|, \]

and similarly,

\[ MQ_{\text{max},h}^G(T_n, \tau) \triangleq \nu_L^{-1} \max_{i=n_T, \ldots, n_T - n_T} \frac{n_T^{-1} \sum_{j=i+1}^{i+n_T} \sum_{j=i-n_T+1}^{i} SL_{\psi,T_m+\tau+j-1} (\beta_{T_m+j-1})}{\sqrt{D_{h,i}}}. \]

### 3.3 Continuous Record Distribution Theory for the Test Statistics

#### 3.3.1 Asymptotic Distribution under the Null Hypotheses

We begin with a set of assumptions. Assumption 3.5 below is a finite-moment condition on the sequence of rescaled forecast losses and on its first-order derivative. It has a similar scope to A4 in Giacomini and Rossi (2009). Assumption 3.6 is similar to A5 in Giacomini and Rossi (2009) and it imposes the first-order derivative of the forecast losses to be constant over time. It trivially holds when one employs the same
loss function for estimation and evaluation. Assumption 3.8 demands the existence of a consistent estimator for \( \beta^* \) at the parametric rate \( \sqrt{T} \) and it encompasses many estimation procedures. In model (3.2.3), the popular least-squares method will satisfy the condition [cf. Barndorff-Nielsen and Shephard (2004) and Li et al. (2017)].

**Assumption 3.3.** The process \( \{ Y_t - (\beta^*)' X_t \}_{t \in [0, N]} \) takes value in an open set \( \mathcal{E} \subseteq \mathbb{R} \), and \( \beta^* \) takes value in a compact parameter space \( \Theta \subset \mathbb{R}^{\dim(\beta)} \).

**Assumption 3.4.** For any \( L \in \mathcal{L} \) we assume \( L : \mathcal{E} \times \Theta \to \mathbb{R} \) is a measurable function and \( L \in C^{2,2} \) (i.e., twice continuously differentiable in both arguments). For every open set \( B \) that contains \( \beta^* \) there exists \( C < \infty \) such that for all \( k \geq 1 \),

\[
\sup_{\beta \in B} \left\| \partial^2 L_{\psi,kh}(\Delta_h e_k; \beta) / \partial \beta \partial \beta' \right\| < C.
\]

**Assumption 3.5.** We have

\[
\sup_{k=1, \ldots, T} \mathbb{E}_\sigma \left\| (L_{\psi,kh}(\Delta_h e_k; \beta^*), \partial L_{\psi,kh}(\Delta_h e_k; \beta^*) / \partial \beta)' \right\|^{4+\varepsilon} < \infty,
\]

for \( \varepsilon > 0 \).

**Assumption 3.6.** For all \( k \geq 1 \), \( \mathbb{E}_\sigma [\partial L_{\psi,kh}(e; \cdot) / \partial e] = \mathcal{K} \), for some \( \mathcal{K} < \infty \).

**Assumption 3.7.** For all \( k \geq 1 \), \( |\partial L_{\psi,kh}(e; \cdot) / \partial e| \) is bounded on bounded sets.

**Assumption 3.8.** There exists a sequence \( \{ \hat{\beta}_k \}_{k=T_m}^{T-\tau} \) such that \( \| \hat{\beta}_k - \beta^* \| = O_p \left( 1 / \sqrt{T} \right) \) uniformly over \( k = T_m, \ldots, T - \tau \).

Our asymptotic results are valid under the following conditions on the auxiliary sequence \( n_T \).

**Condition 2.** The sequence \( \{ n_T \} \) satisfies for some \( \epsilon > 0 \),

\[
n_T \to \infty \quad \text{as} \quad T \to \infty \quad \text{and} \quad T^* n_T^{-1} + n_T^{3/2} h \sqrt{\log(T)} \to 0. \quad (3.3.1)
\]
Condition 2 imposes a lower bound and an upper bound on the growth condition of the sequence \( \{ n_T \} \). The first part of (3.3.1) requires \( n_T \) to grow to infinity at any faster rate than \( T^\epsilon \) with \( \epsilon > 0 \), which we interpret as saying that the number of observations \( n_T \) in each block cannot be too small. The second part of (3.3.1) provides an upper bound on the growth of \( n_T \) and relates to Assumption 3.2 concerning the smoothness of \( \{ \sigma_{e,t} \}_{i \geq 0} \) thereby ensuring that, for example, the random oscillations of \( B_{\text{max},h}(T_n, \tau) \) can be controlled. As we shall explain in the simulation study of Section 3.7, we recommend to set \( n_T \propto T^{2/3-\epsilon} \) for small \( \epsilon > 0 \).

### 3.3.1.1 Asymptotic Distribution Under Quadratic Loss Function

**Theorem 3.3.1.** Let \( \gamma_{m_T} = \left[ 4 \log (m_T) - 2 \log (\log (m_T)) \right]^{1/2} \). Recall \( m_T = \lfloor T_n / n_T \rfloor \).

Assume Assumption 3.1-3.2, 3.3-3.8, and Condition 2 hold. Let \( \mathcal{V} \) denote a random variable defined by
\[
\mathbb{P}(\mathcal{V} \leq v) = \exp \left(-\pi^{-1/2} \exp(-v)\right).
\]
Under \( H_0 \), we have
1. \( \sqrt{\log (m_T)} \left( 2^{-1/2} n_T^{1/2} B_{\text{max},h}(T_n, \tau) - \gamma_{m_T} \right) \Rightarrow \mathcal{V} ;
2. 2^{-1/2} \sqrt{\log (m_T)} n_T^{1/2} MB_{\text{max},h}(T_n, \tau) - 2 \log (m_T) - \frac{1}{2} \log \log (m_T) - \log 3 \Rightarrow \mathcal{V} .

**Corollary 3.3.1.** Under the same assumptions of the previous theorem, we have \( \sqrt{\log (m_T)} \left( n_T^{1/2} \nu^{-1}_L Q_{\text{max},h}(T_n, \tau) - \gamma_{m_T} \right) \Rightarrow \mathcal{V} \) and
\[
\sqrt{\log (m_T)} \left( n_T^{1/2} \nu^{-1}_L M Q_{\text{max},h}(T_n, \tau) - 2 \log (m_T) - \frac{1}{2} \log \log (m_T) - \log 3 \Rightarrow \mathcal{V} ,
\]
where \( \mathcal{V} \), \( m_T \) and \( \gamma_{m_T} \) are defined as in the previous theorem.

Theorem 3.3.1 shows that the asymptotic null distribution of our test statistics follows an extreme value distribution whose critical values can be computed directly. In nonparametric change-point analysis, Wu and Zhao (2007) and Bibinger et al. (2017) have derived an extreme value null distribution for tests statistics which share a similar form to ours. As it is stated, the tests statistics are not yet feasible because
the asymptotic variances $\nu_L^2$ is unknown. However, we can find statistical consistent estimators which can be used in place of $\nu_L^2$ to make the test feasible. We relegate the treatment of its consistent estimation to Section 3.4.

### 3.3.1.2 Asymptotic Distribution Under General Loss Function

**Theorem 3.3.2.** Under the same assumptions of the previous theorem and with $\mathcal{V}$, $m_T$ and $\gamma_{m_T}$ defined analogously, we have under $H_0$,

(i) $2^{-1/2}\sqrt{\log(m_T)} \left( n_T^{1/2} G_{\text{max},h}(T, \tau) - \gamma_{m_T} \right) \Rightarrow \mathcal{V};$

(ii) $2^{-1/2}\sqrt{\log(m_T)} n_T^{1/2} M_{\text{max},h}(T, \tau) - 2 \log(m_T) - \frac{1}{2} \log \log(m_T) - \log 3 \Rightarrow \mathcal{V}.$

**Corollary 3.3.2.** Under the same assumptions of the previous theorem, we have under $H_0$, $\sqrt{\log(m_T)} \left( n_T^{1/2} \nu_L^{-1} Q_{\text{max},h}^G(T, \tau) - \gamma_{m_T} \right) \Rightarrow \mathcal{V}$ and

$$\sqrt{\log(m_T)} n_T^{1/2} \nu_L^{-1} M_{\text{max},h}^G(T, \tau) - 2 \log(m_T) - \frac{1}{2} \log \log(m_T) - \log 3 \Rightarrow \mathcal{V}.$$  

### 3.4 Estimation of the Asymptotic Variance

The purpose of this section is to show how to construct an asymptotically valid estimator of the variance $\nu_L^2$ that enters the definition of our test statistics. This is an important aspect that together with the selection of the block length might affect statistical inferences based on the proposed tests in finite-samples. Allowing for misspecification is customary in the forecasting literature, and as a consequence this may result in forecast losses that artificially exhibit heteroskedasticity and serial dependence in small samples.

### 3.4.1 Estimation of the Asymptotic Variance

We begin with the case of stationary forecast losses, including constant $\nu_L$ as a special case.
3.4.1.1 Stationary Forecast Losses

Recall that our test statistics are related to a maximum over blocks of data. Thus, for i.i.d. forecast losses one can use the following estimator for \( \nu_L \) in \( Q_{\text{max}, h}(T_n, \tau) \):

\[
\hat{\nu}_{Q1,b}^2 \triangleq \frac{2}{n_T} \sum_{j=1}^{n_T} \left( SL_{\psi, T_m + \tau + bn_T + j - 1} \left( \beta_{T_m + bn_T + j - 1} \right) - SL_{\psi, b} \right)^2,
\]

where \( SL_{\psi, b} \triangleq n_T^{-1} \sum_{j=1}^{n_T} SL_{\psi, T_m + \tau + bn_T + j - 1} \left( \beta_{T_m + bn_T + j - 1} \right) \). The estimator \( \hat{\nu}_{Q1,b}^2 \) normalizes the difference in the out-of-sample forecast losses between the \( b + 1 \) and \( b \) blocks. The statistic \( Q_{\text{max}, h}(T_n, \tau) \) then results in

\[
Q_{\text{max}, h}(T_n, \tau) = \max_{b=0, \ldots, \lfloor T_n / n_T \rfloor - 2} \left| (B_{h,b+1} - B_{h,b}) / \hat{\nu}_{Q1,b+1} \right|.
\]

For the overlapping blocks case, the estimator is

\[
\hat{\nu}_{MQ1,i}^2 \triangleq 2n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( SL_{\psi, T_m + \tau + j - 1} \left( \beta_{T_m + j - 1} \right) - SL_{\psi, i} \right)^2,
\]

where \( SL_{\psi, i} \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} SL_{\psi, T_m + \tau + j - 1} \left( \beta_{T_m + j - 1} \right) \) so that we can write

\[
MQ_{\text{max}, h}(T_n, \tau) \triangleq \max_{i=n_T, \ldots, T_n - n_T} \left| \hat{\nu}_{MQ1,i}^{-1} n_T^{-1} \left( \sum_{j=i+1}^{i+n_T} SL_{\psi, T_m + \tau + j - 1} \left( \beta_{T_m + j - 1} \right) \right) - \sum_{j=i-n_T+1}^{i} SL_{\psi, T_m + \tau + j - 1} \left( \beta_{T_m + j - 1} \right) \right|.
\]

Both \( \hat{\nu}_{Q1,b+1}^2 \) and \( \hat{\nu}_{MQ1,i}^2 \) apply a natural block-wise normalization in order to guarantee a distribution-free limit under \( H_0 \). However, it is useful to consider estimators that use all of the observations in the out-of-sample period. Thus, one exploits covariance stationarity of the sequence of forecast losses. Let \( \Phi_{0.75} = 0.647 \ldots \) denote the third
quartile of the standard normal distribution and define

\[
\hat{\nu}_{2,h} \triangleq \frac{\sqrt{T_n}}{2(m_T - 1)} \sum_{b=1}^{m_T - 1} |B_{h,b} - B_{h,b-1}|
\]

\[
\hat{\nu}_{3,h} \triangleq \frac{\sqrt{T_n}}{2(m_T - 1)} \left( \sum_{b=1}^{m_T - 1} |B_{h,b} - B_{h,b-1}|^2 \right)^{1/2}
\]

\[
\hat{\nu}_{4,h} \triangleq \frac{\sqrt{T_n}}{2\Phi_{0.75}} \text{median} (|B_{h,b} - B_{h,b-1}|), \quad 1 \leq b \leq m_T - 1.
\]

Note that \(\hat{\nu}_{2,h}, \hat{\nu}_{3,h}\) and \(\hat{\nu}_{4,h}\) can be used to implement both \(Q_{\max,h}(T_n, \tau)\) and \(MQ_{\max,h}(T_n, \tau)\).\(^{23}\) \(\hat{\nu}_{3,h}\) is related to Carlstein’s (1986) subseries variance estimate in the context of strong mixing processes and it was also used by Wu and Zhao (2007). Each of the estimators \(\hat{\nu}_{2,h}, \hat{\nu}_{3,h}\) and \(\hat{\nu}_{4,h}\) allows for dependence but requires stationarity. The simulation study in Wu and Zhao (2007) suggests that \(\hat{\nu}_{4,h}\) is more robust whereas \(\hat{\nu}_{2,h}\) and \(\hat{\nu}_{3,h}\) are less precise when there are large instabilities or jumps. For two sequences \(\{a_k\}\) and \(\{b_k\}\), we write \(a_k \asymp b_k\) if for some \(c \geq 1\), \(b_k/c \leq a_k \leq cb_k\) for all \(T\). The following theorem is similar to Theorem 3 in Wu and Zhao (2007) and in particular, part (ii) states that if \(n_T \asymp T_n^{1/3}\) then \(\hat{\nu}_{3,h}^2\) achieves the optimal MSE \(O\left(n_T^{-2/3}\right)\).

**Condition 3.** The sequence \(\{n_T\}\) satisfies

\[n_T \rightarrow \infty \quad \text{as} \quad T \rightarrow \infty \quad \text{and} \quad \sqrt{T_n n_T^{-1} \log (T_n)} + n_T T_n^{-2/3} (\log (T))^{1/3} \rightarrow 0.\]

\((3.4.1)\)

**Theorem 3.4.1.** In addition to the assumptions of Theorem 3.3.1, assume that Cov \(\left(\hat{L}_{\psi,kh} (\beta^*), \hat{L}_{\psi,(k-j)h} (\beta^*)\right)\) depends on \(j\) but not on \(kh\). Then, under \(H_0\),

(i) Let \(n_T \asymp T_5^{5/8}\). Then, \(\hat{\nu}_{2,h}, \hat{\nu}_{4,h} = \nu_L + O_P \left( T_n^{-1/16} \log (T_n) \right)\); (ii) Let \(n_T \asymp T_n^{1/3}\). Then \(\mathbb{E} \left( \left[ \hat{\nu}_{3,h}^2 - \nu_L^2 \right]^2 \right) = O \left( T_n^{-2/3} \right)\).

Under covariance-stationarity, given Theorem 3.4.1, the results of Corollary 3.3.1-3.3.2 are applicable after replacing \(\nu_L\) by an appropriate consistent estimator.

\(^{23}\)They can be also applied to the test statistics \(Q_{\max,h}^G(T_n, \tau)\) and \(MQ_{\max,h}^G(T_n, \tau)\) with \(G_{h,b}\) in place of \(B_{h,b}\).
3.4.1.2 Heterogeneous Forecast Losses

We now consider estimation of the asymptotic variance in the case the forecast losses are heterogeneous. The estimator $\hat{\nu}$ that we introduce below depends on the specific loss function and thus it can be used for replacing $\nu_L$ in Corollary 3.3.1-3.3.2. Non-stationarity implies that $\sigma^2_{e,t}$ is time-varying and thus the results of Theorem 3.4.1 are not applicable due to the presence of many extra parameters that account for the time-varying structure. To deal with this issue we propose a novel block-wise self-normalization technique which simultaneously addresses two issues. First, the block-wise self-normalization ensures that the difference in forecast losses between two adjacent blocks are asymptotically independent across non-adjacent blocks and that within each block the losses are standardized so that time-varying variances cancel out. Second, by computing an average—over all blocks—of the self-normalized difference in forecast losses we account for possible serial dependence. We derive asymptotic results within a general framework based on the strong invariance principle for stationary processes developed in Wu (2007) and extended to modulated stationary processes by Zhao and Li (2013).

For each block $b = 0, \ldots, m_T - 2$, let

$$A_{h,b} (\beta) \triangleq n_T^{-1} \sum_{j=1}^{n_T} \left( L_{\psi,T_m+\tau+(b+1)n_T+j-1} (\beta_{T_m+(b+1)n_T+j-1}) \right),$$

$$V_{h,b} (\beta) \triangleq n_T^{-1} \sum_{j=1}^{n_T} \left( L_{\psi,T_m+\tau+(b+1)n_T+j-1} (\beta) - \bar{L}_{\psi,b} (\beta) \right)^2,$$

where $\bar{L}_{\psi,b} (\beta) = n_T^{-1} \sum_{j=1}^{n_T} L_{\psi,T_m+\tau+(b+1)n_T+j-1} (\beta)$ and define the statistic

$$\zeta_{h,b} (\beta) \triangleq \sqrt{n_T} \left( A_{h,b} (\beta) - A_{h,b-1} (\beta) \right) / \sqrt{V_{h,b}}.$$

Finally, an average—over all blocks $m_T$—of the per-block self-normalized statistics
ζ_{h,b}'s is used to define an estimator of the asymptotic variance:
\[ \hat{\nu}_L^2 \triangleq 2^{-1} (m_T - 1)^{-1} \sum_{b=0}^{m_T-1} \zeta_{h,b}^2. \]

Let \( \sigma_{L,kh}^2 \triangleq \operatorname{Var}(L_{\psi,kh}(\beta^*)) \). We also need to introduce the following quantities,
\[ F_{h,b}^* \triangleq \left| \sigma_{L,(T_m + \tau + (b+2)n_T)h} \right|, \quad J_{h,b}^* \triangleq \sigma_{L,(T_m + \tau + (b+2)n_T)h}^2, \]
\[ \Sigma_{h,b}^* \triangleq \sum_{j=1}^{n_T} \sigma_{L,(T_m + \tau + (b+1)n_T + j)h}^2, \quad \tilde{\Sigma}_{h,b}^* \triangleq \left( \sum_{j=1}^{n_T} \sigma_{L,(T_m + \tau + (b+1)n_T + j)h}^4 \right)^{1/2}. \]

**Theorem 3.4.2.** Under Condition 3 we have \( \hat{\nu}_L^2 - \nu_L^2 = O_P\left(r_h^{-1}\right) \), where \( r_h = O_P\left(T_h^2/(\log(T_h))^2\right) \) with \( \epsilon \in (0, 1/4) \) such that \( r_h \to \infty \).

The theorem simply states that \( \hat{\nu}_L \) is consistent for \( \nu_L \) and therefore the asymptotic results of Section 3.3 continue to hold when we replace \( \nu_L \) by \( \hat{\nu}_L \).

### 3.5 Continuous Semimartingale Volatility and Asymptotic Local Power

#### 3.5.1 Asymptotic Results under Continuous Semimartingale Volatility

In this section we relax the Lipschitz condition on \( \sigma_{e,t} \) and extend the results for the quadratic loss case from Theorem 3.3.1 to stochastic volatility models driven by a Wiener process. Consequently, this relaxation enables one to utilize the tests proposed in this chapter in setting involving high-frequency financial variables. More specifically, we assume that \( \sigma_{e,t} \) is an Itô continuous semimartingale that is almost surely bounded and strictly positive adapted process. We replace Assumption 3.2 by the following.

**Assumption 3.9.** Under \( H_0 \) the process \( \{\sigma_{e,t}\}_{t \geq 0} \) satisfies \( \phi_{\sigma,\eta,\tau_h,N} \leq K_h \eta^\kappa \) for some \( \kappa > 0 \), some sequence of stopping times \( \tau_h \to \infty \) and some \( \mathbb{P}\text{-a.s.} \) finite random
variable $K_h$.

The assumption implies that $\sigma_{e,t}$ belongs to a rather large class of volatility processes usually considered in financial econometrics. The parameter $\kappa$ plays a key role in the testing framework of this section and we refer to it as the regularity exponent. When $\kappa = 1$ we recover the case of Lipschitz volatility considered in the previous sections while the standard stochastic volatility model without jumps correspond to $\kappa = 1/2 - \epsilon$ for a sufficiently small $\epsilon > 0$. Next, we have a slightly different version of Condition 2.

**Condition 4.** The sequence $\{n_T\}$ satisfies for some $\epsilon > 0$,

$$n_T \to \infty \text{ as } T \to \infty \quad \text{and} \quad T^n T^{-1} + \sqrt{n_T (n_T h)^{\kappa}} \sqrt{\log(T)} \to 0. \quad (3.5.1)$$

For Itô continuous semimartingale volatility $\sigma_{e,t}$ the condition suggests $n_T \propto T^{1/2-\epsilon}$ for small $\epsilon > 0$. Let $\Gamma_t \triangleq \mathbb{E}_{\sigma} \left[ dL_t (e_t; \beta^*) / dt \right]$.\(^{24}\) The more general framework considered here requires us to consider the following null hypotheses: under quadratic loss,

$$H_0 : \{\Gamma_t\}_{t \in [N_{in} + h, N]} \in C(\kappa, K_h), \quad (3.5.2)$$

where $C(\kappa, K_h)$ is a class of continuous functions on $[N_{in} + h, N]$,

$$C(\kappa, K_h) \triangleq \left\{ \{\Gamma_t\}_{t \in [N_{in} + h, N]} : \sup_{s,t \in [N_{in} + h, N], |t-s| < \eta} |\Gamma_t - \Gamma_s| \leq K_h \eta^\kappa \right\},$$

where $\kappa > 0$ and $K_h$ is given in Assumption 3.9. Thus, we wish to discriminate

\(^{24}\)For example, for the quadratic loss with $\mu_{e,t} = 0$ the notation reduces to $\Gamma_t = \sigma_{e,t}^2$.\)
between $H_0$ and

$$H_1 : \exists \lambda \in [N_{in} + h, N] \quad \text{with} \quad \{\Gamma_t(\omega)\}_{t \in [N_{in} + h, N]} \in J_\lambda(\kappa, K_h, d_h),$$

(3.5.3)

where

$$J_\lambda(\kappa, K_h, d_h) \triangleq \left\{ \{\Gamma_t\}_{t \in [N_{in} + h, N]} : \{\Gamma_t - \Delta \Gamma_t\}_{t \in [N_{in} + h, N]} \in C(\kappa, K_h) ; |\Delta \Gamma_\lambda| \geq d_h \right\},$$

$\Delta \Gamma_\lambda = \Gamma_\lambda - \lim_{s \uparrow \lambda} \Gamma_s$ and $\{d_h\}$ is a decreasing sequence. The following theorem extends Theorem 3.3.1 to the current setting.

**Theorem 3.5.1.** Let $m_T, \gamma_{m_T}$ and $\mathcal{V}$ as defined in Theorem 3.3.1. Assume the assumptions of Theorem 3.3.1 hold with Assumption 3.2 replaced by Assumption 3.9. Under Condition 4 and a quadratic loss, the same results of Theorem 3.3.1 hold.

### 3.5.2 Asymptotic Local Power

In this section we consider the behavior of $MQ_{\max,h}$ under a sequence of local alternatives.

**Assumption 3.10.** We have the same assumptions as in Theorem 3.5.1 and assume (i) in model (3.2.2) we replace $\beta^*$ by $\beta_t = \beta^* + \mu_{\beta,t}/(\log(T_n)n_T)^{1/4}$ where $\mu_{\beta,t} \in \mathbb{R}^q$ is $\mathbb{P}$-a.s. locally bounded and adapted process; (ii) we set $\mu_{e,t} = 0$ for all $t \geq 0$; (iii) we replace Assumption 3.8 by $\|\tilde{\beta}_k - \beta^*\| = \mu_{\beta,kh}/(\log(T_n)n_T)^{1/4} + O\mathbb{P}\left(T^{-1/2}\right)$ uniformly in $k$. 
Part (iii) is a consequence of part (i) as it can be easily verified. Let

\[ \tilde{M}_{Q_{\max}}(T_n, \tau) \triangleq \nu_L^{-1} \max_{i=n_T, \ldots, T_n-n_T} n_T^{-1} \left| \sum_{j=i+1}^{i+n_T} \left( SL_{\psi, T_m+\tau+j-1} \left( \widehat{\beta}_{T_m+j-1} \right) - 2\zeta_{\mu,j,+} \right) \right| \]

\[ -n_T^{-1} \sum_{j=i-n_T+1}^{i} \left( SL_{\psi, T_m+\tau+j-1} \left( \widehat{\beta}_{T_m+j-1} \right) - 2\zeta_{\mu,j,-} \right) \],

where

\[ \zeta_{\mu,j,+} \triangleq \mu_{\beta_j(T_m+\tau+j-1)h} \sum X_i(T_m+\tau+i-1)h \mu_{\beta_j(T_m+\tau+j-1)h} / (\log (T_n) n_T)^{1/2} \]

\[ \zeta_{\mu,j,-} \triangleq \mu_{\beta_j(T_m+\tau+j-1)h} \sum X_i(T_m+\tau+i-n_T-1)h \mu_{\beta_j(T_m+\tau+j-1)h} / (\log (T_n) n_T)^{1/2} . \]

**Theorem 3.5.2.** Under Assumption 3.10,

\[ \sqrt{\log (m_T) \left( n_T^{1/2} \nu_L^{-1} \right)} \tilde{M}_{Q_{\max}}(T_n, \tau) - 2 \log (m_T) - \frac{1}{2} \log \log (m_T) - \log 3 \Rightarrow \mathcal{V}, \]

where \( \mathcal{V} \), and \( m_T \) are defined as in Theorem 3.3.1.

**Remark 3.5.1.** (i) The theorem suggests that under the local alternatives \( \beta_t = \beta^* + \mu_{\beta,t} / (\log (T_n) n_T)^{1/4} \) there is a bias term arising from the presence of \( \zeta_{\mu} \). This bias term does not vanish asymptotically and results in shifting the center of the distribution. Moreover, it depends on the second moments of the regressors and on the function \( \mu_{\beta} \); (ii) The theorem illustrates the sensitivity of the asymptotic power to the form of the alternative. We can attempt to compare Theorem 3.5.2 with the local power result regarding the sup-Wald test of Andrews (1993). Unlike Theorem 4 in Andrews (1993), our result suggests that the location of the instability should not play any special role and the power should not be sensitive to whether the break in predictive ability occurs at middle sample or toward the tail of the sample. This follows because of the local nature of our test statistic and contrasts with classical
tests for parameter instability and structural change since their performance hinges on the location of the break [see Deng and Perron (2008), Kim and Perron (2009) and Perron and Yamamoto (2018) for additional results on the power of classical structural break tests]. However, the magnitude of the break—here shrinking at rate \((\log (T_n) n_T)^{1/4}\)—under our specification of the local alternatives is larger than the one considered by Andrews (1993)—which shrinks at rate \(1/\sqrt{T}\). This implies a trade-off between location and magnitude of the break, and it is consistent with the evidence provided in our simulation study; (iii) Although not shown here, the local power of the tests is the same when a subset of the vector \(\beta\) is not subject to shift.

Theorem 3.5.2 can be used to show that our test possesses nontrivial power against alternatives for which the parameter \(\beta_t\) is time-varying and non-smooth.

**Corollary 3.5.1.** Suppose the assumptions of the previous theorem hold with \(\beta_t = \beta^* + c\mu_{\beta,t}/(\log (T_n) n_T)^{1/4}\), where \(c \in \mathbb{R}\). If \(\mu_{\beta,\cdot}\) and/or \(\{\mu_{\beta,t} \cdot \sigma_{X,t}\}_{t \geq 0}\) is non-smooth, we have

\[
\lim_{c \to \infty} \lim_{h \downarrow 0} \mathbb{P}
\left( \sum_{T_n \leq v \leq T_n + h} \frac{1}{M_{Q_{\max,h}}(T_n, \tau)} - 2 \log (m_T) - \frac{1}{2} \log \log (m_T) - \log 3 > c_{V_{1-\alpha}} \right)
\]

where \(c_{V_{1-\alpha}}\) is the level \((1 - \alpha)\) critical value of the distribution of \(V\) and \(\alpha \in (0, 1)\).

### 3.6 Extensions

A number of extensions is treated in our companion paper Casini (2018a). As explained above, it would be useful to ensure that there are no instabilities in the in-sample period \([0, N_{in}]\). We propose a procedure that involves a pre-test about instability on \([0, N_{in}]\) for a given \(N_{in}\) chosen by the forecaster. Instabilities in the in-sample \([0, N_{in}]\)
are much easier to be detected relative to instabilities in the out-of sample because they do not face the so-called “contamination effect”. The latter arises, for example under the recursive and rolling scheme, when the instability originally occurring in the out-of-sample eventually enters the moving in-sample window [cf. Casini and Perron (ming) and Perron and Yamamoto (2018)]. The consequence is that existing tests face substantial power losses. This property is not shared by our test statistics because of their local nature. Our procedure works very well and we show through simulations that instabilities occurring in the in-sample only or occurring both in the in-sample and in the out-of-sample simultaneously, lead easily to rejection of the null hypotheses relative to instabilities occurring in the out-of-sample only—as we consider here.

A second issue is that, in this chapter, we have considered processes that have a continuous sample path under the null hypotheses. Thus, it is of interest to extend the results to a setting that involves jump processes which are important in high-frequency financial data. This can be achieved by using techniques that are able to separate the continuous part from the discontinuous part of a Itô semimartingale [see e.g. Li et al. (2017) and Li and Xiu (2016)].

Another important issue is the estimation of the time at which forecast instability occurs. Once the null hypotheses has been rejected, a forecaster may take into consideration the possibility of revising the forecasting method and/or model. Hence, it becomes crucial to learn some information about the timing of the instability. For example, consider the case of a one-time structural change in a parameter of the data-generating process at time $T^0_b = [T \lambda^0_b]$, where $T^0_b$ is the break point and $\lambda^0_b \in (0, 1)$ is the fractional break date. Once $H_0$ is rejected, a forecaster would benefit from knowing that the forecast method originally employed is found to statistically either
under or over-perform over part of the sample after $T^0_b$ relative to the part prior $T^0_b$. Then, a forecaster would entertain the possibility of modifying the forecast model in order to generate future forecasts for $Y_t$. Not only the forecast model might be revised but most importantly, knowledge of beginning of the instability at $T^0_b$ can be further exploited to design the forecasting method for the future forecasts. It would be inappropriate, for instance, to use a rolling scheme where the rolling window used to construct the forecast include observations prior to $T^0_b$ since those observations provide little informational content for the purpose of predicting $Y_t$ after the change-point $T^0_b$. On the other hand, this line of reasoning is justified by this particular example and indeed in practice many issues arise when dealing with the timing of the insatiability in our context. For example, the exact form of the insatiability may be unknown. Under the latter scenarios, there is no clear-cut break date $T^0_b$ that can be defined. Thus, it is less obvious how a forecaster should proceed in those cases. Nonetheless, one can meaningfully think about the timing of the instability by not just attempting to estimate $T^0_b$—which is not clear how it is defined—but rather attempting to detect the initial date in the sample after which the forecasts become unstable as well as to detect the last date after which the forecasts remain stable relative to the in-sample period. Since our test statistics are local in nature, one can introduce a procedure which sequentially tests the hypotheses $H_0$ in regions of the sample where $H_0$ has not yet been rejected. One then records the number of times for which $H_0$ is rejected and estimates the corresponding change-point dates. After ordering these change-point dates, one has finally access to useful information such has the initial timing of the forecast instability and the last part of the out-of-sample period which remains stable. Such information can arguably be advantageous to the forecaster.
3.7 Small-Sample Evaluation

We now examine the empirical size and power of our proposed tests and compare them to those of Giacomini and Rossi (2009), abbreviated GR (2009). In particular, we consider both the uncorrected and corrected version of the \( t_{stat}^{T_m,T_n,T_n,\tau} \) statistic of GR (2009).\(^{25}\) Size and power properties for the Quadratic loss with fixed scheme are reported in Section 3.7.1 and Section 3.7.2, respectively. Chapter 6 includes corresponding simulation studies for the recursive and rolling schemes and for the Linex loss; these results are not reported here because they are qualitatively equivalent.

Overall, one can draw the following conclusions from our simulation study. In terms of size control, the statistics \( B_{max,h} \) and \( Q_{max,h} \) are comparable with the corrected version \( t_{stat,c}^{T_m,T_n,T_n,\tau} \) proposed by GR (2009).\(^{26}\) Moreover, the test \( MQ_{max,h} \) that uses overlapping blocks is also comparable in terms of size. The same is not true for \( MB_{max,h} \) because often it seems to be somewhat liberal. Turning to the power comparison, each of our test statistics \( B_{max,h} \), \( Q_{max,h} \) and \( MQ_{max,h} \) displays significant power gains over the \( t_{stat}^{T_m,T_n,T_n,\tau} \) statistics especially as the period of instability (i) is comparatively short relative to the total sample size and/or (ii) is not located at middle sample. In the latter circumstances, the gains in power are, uniformly over different data-generating processes and over parameter break magnitudes, on the order of 30-40%.

Throughout, we restrict attention to one-step ahead forecast horizon (i.e., \( \tau = 1 \)), and we use the same loss function for estimation and evaluation. We use our

\(^{25}\)We use a superscript \( c \) to indicate the corrected version: \( t_{stat,c}^{T_m,T_n,T_n,\tau} \).

\(^{26}\)As shown by GR (2009), the uncorrected version of \( t_{stat}^{T_m,T_n,T_n,\tau} \) can be oversized for models that induce serial dependence in the forecast losses. The authors then proposed a finite-sample correction and did not consider \( t_{stat}^{T_m,T_n,T_n,\tau} \) further in their power analysis. Similarly, GR (2009) showed that just using classical structural break tests in this context is not very helpful as they might have statistical power equals to the size in some cases. Moreover, simulations in Perron and Yamamoto (2018) confirmed that, under rolling and recursive scheme, structural break tests suffer power losses which can be attributed to a so-called “contamination effect” arising when the instability enters the in-sample window [see also Casini and Perron (ming)].
asymptotic results as an approximation for the case where \( h = 1 \) in our theoretical model in (3.2.3) and consider discrete-time DGPs. In models with serially correlated losses (i.e., S2 and S6 below) for the statistics \( Q_{\text{max},h} \) and \( MQ_{\text{max},h} \) we employ the long-run variance estimator from Theorem 3.4.2. With regards to the tests of GR (2009) we use the appropriate version of \( t^{\text{stat}} \) and of \( t^{\text{stat,c}} \).27

**Remark 3.7.1.** Implementation of our tests statistics requires to choose the number of blocks \( m_T \) — satisfying Condition 2. The finite-sample properties can be sensitive to the choice of \( m_T \). This is confirmed in our numerical study, where assigning larger values to \( m_T \) than the smallest one allowed by the condition may result in oversized tests. Therefore, we recommend practitioners to set \( m_T \) equal to the smallest integer as allowed by Condition 2. This is the strategy we have adopted in the Monte Carlo study of this section, and as we will show, it results in approximately correct size and good power across different data-generating mechanisms.

### 3.7.1 Empirical Size

We consider discrete-time DGPs of the form

\[
Y_t = \mu + \beta X_{t-1} + e_t, \quad t = 1, \ldots, T, \tag{3.7.1}
\]

for various in-sample and out-of-sample sizes and with a total sample size ranging from \( T = 100 \) to \( T = 500 \). Note that (4.6.1) is a special case of the theoretical model with a sampling interval \( h = 1 \). We consider six versions of (4.6.1), where the first and second specification (S1 and S2 below) are calibrated to the Phillips curve model of U.S. inflation from Staiger et al. (1997): S1 involves \( \mu = 2.73, \beta = -0.44 \), and where \( \{X_t\} \) and \( \{e_t\} \) are independent sequences of zero-mean i.i.d. Gaussian disturbances

---

27As recommended by GR (2009) we set the truncation lag of their HAC estimator equal to \( T_n^{1/3} \); we also use the truncation lag \( \left[ 0.75T_n^{1/3} \right] \).
with unit variance; S2 is the same as S1 but with ARCH errors \( e_t = \sigma_{e,t} u_t, \sigma_{e,t} = 1 + 0.5e_{t-1}^2 \) with \( u_t \sim \mathcal{N}(0,1) \); S3 specifies \( \{X_t\} \) to follow a zero-mean Gaussian AR(1) with autoregressive coefficient 0.4, \( \beta = 1 \) and \( e_t \sim \mathcal{N}(0,0.49) \) independent of \( X_t \); S5 is a model with a lagged dependent variable \( X_{t-1} = Y_{t-1}, \mu = 0, \beta = 0.3 \) and \( e_t \sim \mathcal{N}(0,0.49) \); S6 involves serially correlated disturbances \( e_t = 0.3e_{t-1} + u_t, u_t \sim \mathcal{N}(0,1) \).

Table 3.1-3.2 report the rejection rates for significance levels \( \alpha = 0.05 \) and 0.10 for model S1-S2. Results for the other DGPs can be found in the appendix. We first focus on i.i.d. forecast losses (i.e., models S1 and S3-S5). Both \( B_{\max,h} \) and \( Q_{\max,h} \) are well-sized. As the sample size increases their performance improves and we note that their rejection frequencies are closer to the nominal level when the in-sample size is one half of the total sample. In model S1, when the in-sample size is 0.25\( T \), \( B_{\max,h} \) and \( Q_{\max,h} \) tend to be slightly conservative while the opposite occurs when in-sample size is 0.75\( T \). The version of \( B_{\max,h} \) that uses overlapping blocks (\( MB_{\max,h} \)) can be quite liberal (cf. models S1 and S3). In contrast, \( MQ_{\max,h} \) seems to control the size well, though it tends to be slightly liberal but that depends on the relative size of the in-sample and out-of-sample windows. We observe that there is no clear pattern in size performance for our test statistics as we raise the sample size \( T \). The reason is straightforward: as we raise \( T \) we also need to adjust the choice of \( m_T \) (the number of blocks) in accordance with Condition 2. This explains why, for example, in Table 3.1, top panel, the empirical size of \( Q_{\max,h} \) for \( (T_m = 100, T_n = 100) \) is better than for \( (T_m = 150, T_n = 150) \). Turning to the \( t^{stat} \) statistics of Giacomini and Rossi (2009), the uncorrected version performs better than the corrected version since the latter systematically displays an empirical size 2-3% below the nominal level. We can conclude that in models with i.i.d. errors the statistics \( B_{\max,h}, Q_{\max,h}, MQ_{\max,h} \) and
$t^{\text{stat}}$ are comparable in terms of empirical size, whereas $MB_{\max,h}$ and $t^{\text{stat,c}}$ tend to over-reject and under-reject, respectively.

Let us now turn to models with serially correlated losses. When the disturbances follow an ARCH process, (cf. model S2, Table 3.2), we observe that both statistics that do not use overlapping blocks, $B_{\max,h}$ and $Q_{\max,h}$, show reasonable size control. The same feature applies to $MQ_{\max,h}$ while $MB_{\max,h}$ displays rejection rates that are systematically above the significance level. It also appears that the corrected version of the statistic of GR (2009) has now size regularly below the nominal level. In contrast, the uncorrected version $t^{\text{stat}}$ seems to control size well. When the errors follow an autoregressive process (cf. model S6), $t^{\text{stat}}$ and $MB_{\max,h}$ are arbitrarily oversized for all sample sizes. $MQ_{\max,h}$ and $t^{\text{stat,c}}$ possess rejection rates frequently below the desired nominal level. The statistic that shows the best empirical sizes across different $T$ is $Q_{\max,h}$.

Overall, our analysis on the size properties of the tests suggests that when the DGP involves i.i.d. errors it is fair to compare $B_{\max,h}$, $Q_{\max,h}$, $MQ_{\max,h}$ and $t^{\text{stat}}$ whereas the rejection rates of $MB_{\max,h}$ and $t^{\text{stat,c}}$ tend to deviate systematically from the nominal level. When there are autocorrelated errors, it is difficult to compare $t^{\text{stat}}$ and $MB_{\max,h}$ with the other statistics because the former can be highly oversized. The statistics that appear to perform better in terms of approximate size control uniformly over different data-generating mechanisms are $Q_{\max,h}$ and $MQ_{\max,h}$.

### 3.7.2 Empirical Power

We report the small sample power of the tests under various sources of forecast instability. We consider several sample sizes $T$ as well as several designs varying for the distribution of the total sample between in-sample and out-of-sample window. The break date—or the date of the first change-point when more complicated designs
are used—is denoted by \( T^0_b = T\lambda_0 \), where \( \lambda_0 \in (0, 1) \) is the fractional break date. We shall bring special attention to the location of \( T^0_b \) in the sample as well as to the duration of the instability (i.e., \( T - T^0_b \)). We shall see that both factors are actually important for the performance of the methods proposed by Giacomini and Rossi (2009) while our test statistics being local in nature possess essentially uniform power over distinct locations \( T^0_b \). Furthermore, our definition of forecast instability does not demand any relationship between the stable and unstable period and thus it is useful to examine the differences in power properties when a one-time change-point is present relative to when short-lasting instabilities arise.

We consider both discrete shifts—a structural break—and recurrent changes in a parameter: model P1a (break in a regression coefficient): \( Y_t = 2.73 - 0.44X_{t-1} + \delta X_{t-1} \mathbf{1}\{t > T^0_b\} + e_t \), where \( X_{t-1} \sim \text{i.i.d.} \mathcal{N}(0, 1) \) and \( e_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \); model P1b: it is the same as model P1a but with \( X_{t-1} \sim \text{i.i.d.} \mathcal{N}(1, 1) \); model P2: \( Y_t = X_{t-1} + \delta X_{t-1} \mathbf{1}\{t > T^0_b\} + e_t \), where \( X_{t-1} \) is a Gaussian AR(1) with autoregressive coefficient 0.4 and unit variance, and \( e_t \sim \text{i.i.d.} \mathcal{N}(0, 0.49) \); model P3 (recurrent break in mean): \( Y_t = \beta_t + e_t \), where \( \beta_t \) switches between \( \delta \) and 0 every \( p \) periods and \( e_t \sim \text{i.i.d.} \mathcal{N}(0, 0.64) \); model P4 (single break in variance): \( Y_t = 0.5X_{t-1} + (1 + \delta \mathbf{1}\{t > T^0_b\}) e_t \), where \( X_{t-1} \sim \text{i.i.d.} \mathcal{N}(1, 1) \) and \( e_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \); model P5 (recurrent break in variance): \( Y_t = \mu + (1 + \beta_t) e_t \), where \( \beta_t \) switches between \( \delta \) and 0 every \( p \) periods and \( e_t \sim \text{i.i.d.} \mathcal{N}(0, 0.49) \); model P6 (lagged dependent variable): \( Y_t = \delta \mathbf{1}\{t > T^0_b\} + 0.3Y_{t-1} + e_t \), where \( X_{t-1} \sim \text{i.i.d.} \mathcal{N}(0, 0.49) \); model P7 (ARCH disturbances): \( Y_t = 2.73 - 0.44X_{t-1} + \delta X_{t-1} \mathbf{1}\{t > T^0_b\} + e_t \), where \( X_{t-1} \sim \text{i.i.d.} \mathcal{N}(0, 1.5) \) and \( e_t = \sigma_t u_t \), \( \sigma_t^2 = 0.5 + 0.5e_{t-1}^2 \), \( u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \); model P8 (autocorrelated errors): \( Y_t = 1 + X_{t-1} + \delta X_{t-1} \mathbf{1}\{t > T^0_b\} + e_t \), where \( X_{t-1} \sim \text{i.i.d.} \mathcal{N}(0, 1.4) \) and \( e_t = 0.4e_{t-1} + u_t \), \( u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \). For models that do not involve recurrent
changes we also consider power comparisons when the instability lasts only for some period of time as opposed to the post-$T_b^0$ period. This requires replacing $\mathbf{1}\{t > T_b^0\}$ in models P1-P2, P4 and P6-P8 with $\mathbf{1}\{T_b^0 < t \leq T_b^0 + p\}$ where $p$ is the number of consecutive observations in which the forecast model is unstable. The value of $p$ depends on the sample size $T$. For example, when $T = 100$ we set $p = 10$; when $T = 200$ we set $p = 20$ and so on.\(^{28}\) The case of short-term instability is the most prevalent in empirical work because it is very unlikely that a professional forecaster would use a poor-performing predictor or forecast model for many consecutive years (e.g., the whole out-of-sample).

Figure 3.1-3.9 in the appendix plot the power functions for models P1a, P4 and P7. Figure C.1-C.13 in Chapter 7 plot the power functions for the remaining DGPs. They include several sample sizes ranging from $T = 100$ to $T = 500$, several in-sample and out-of-sample sizes as well as different locations $\lambda_0$ of the breaks. We begin with considering general instabilities first and then move to short-term instabilities. Figure 3.1-3.2 reports the results for model P1a. When $T = 100$, 150 Figure 3.1 shows that our tests have good power against model P1a while the tests of GR (2009) seem to be less powerful. For example, when the break date is at $T_0 = 0.8T$ our tests display reasonable power. However, both $t^{\text{stat}}$ statistics of GR (2009) perform significantly worse and the associated power curve is bounded away from one even for a very large break size $\delta = 3$. This feature disappears when we raise the sample size to $T \geq 200$ and maintain the break date at $T_0 = 0.8T$; see Figure 3.2. The latter figure also shows that for large sample sizes and instabilities that last for more than 50% of the out-of-sample (top panels) all tests have good power even though the $t^{\text{stat}}$ statistics

\(^{28}\)Note that for $(T_m = 50, T_n = 50)$ the value $p = 10$ corresponds to a period of instability lasting for one-fifth of the out-of-sample; thus, the duration of the instability is nontrivial and consistent with forecasting applications. See the notes to each figure for the other values of $p$. The title of a figure corresponding to a short-lasting instability is labeled “short-term instability”.
of GR (2009) have slightly higher power. The power turns to be essentially the same when \( \lambda_0 = 0.8 \) (i.e., the instability only lasts for 40% of the out-of-sample). For model P2, Figure C.3 plots the power functions for \( T = 100, 200 \) and \( \lambda_0 = 0.7, 0.8 \). Except for the pair \( (T = 200, \lambda_0 = 0.7) \) (cf. top-right panel) for which our tests and the \( t_{\text{stat}} \)-type tests display roughly the same power, it is clear that our tests are more powerful than the \( t_{\text{stat}} \) tests (both corrected and uncorrected version). The power gains are substantial and range from 20% to 40%. Moreover, as for model P1a and P1b when the instability lasts for less than 50% of the out-of-sample (cf. \( \lambda_0 = 0.8 \); bottom-left panel) the statistics \( B_{\text{max},h} \) and \( Q_{\text{max},h} \) achieve trivial power already for a break magnitude \( \delta = 1.5 \) whereas the \( t_{\text{stat}} \) tests of GR (2009) display rejection rates below 60% even when \( \delta = 2 \) and yet below 70% when \( \delta = 2.5 \); that is, their power function does not attain unit power even for very large break magnitudes. These properties characterize all models with i.i.d. errors and extend to model with lagged dependent variables as predictors (cf. model P7; Figure C.10 in the Supplement).

Let us now turn to recurrent breaks in the mean. For recurrent breaks we implement the statistics \( M_{B_{\text{max},h}} \) and \( M_{Q_{\text{max},h}} \) that use overlapping blocks. Figure C.4 plots the power curves for model P3. All tests have power and their performance is essentially the same. Figure 3.3 corresponds to model P4 (single break in the variance) and shows that when the instability begins in the second half of the out-of-sample (cf. \( \lambda_0 = 0.8 \); bottom panels) our tests \( M_{B_{\text{max},h}} \) and \( M_{Q_{\text{max},h}} \) achieves good power while the \( t_{\text{stat}} \)-type tests have little power that does not attain unity even for a large break magnitude \( \delta = 1.5 \). When there are recurrent breaks in the variance as in model P5, Figure C.6 shows that the our tests \( M_{B_{\text{max},h}} \) and \( M_{Q_{\text{max},h}} \) and the \( t_{\text{stat}} \)-type tests have all good power and their performance is analogous.

Let us now consider models with either ARCH errors or autocorrelated errors.
Observe that the latter models both imply that the forecast losses are serially corre-
lated. Figure 3.6 shows that when the errors follow an ARCH(1) process the statistic
$Q_{\text{max},h}$ based on the asymptotic variance estimator $\hat{\nu}_L^2$ performs well in terms of em-
pirical power. In contrast, the $t_{\text{stat}}$-type tests fail as their power is non-monotonic,
ever reaches 20% and it decreases to zero as the magnitude of the break rises.\textsuperscript{29} We
note that the version of $\hat{\nu}_L$ that uses more blocks is less precises. The same results
hold true when the disturbances are autocorrelated; see Figure C.11.

Finally, we consider short-term instabilities in Figure 3.6-3.9. It is straightforward to recognize a general pattern: the tests of GR (2009) have little power whereas
our tests possess good empirical power against all data-generating processes, break
locations and sample sizes. Furthermore, the small sample power properties are uni-
form over the location of the instability and over the relative size of the in-sample and
out-of-sample windows. The latter property is important in practice because forecast
instabilities are frequently short-lived.

To sum up, our test statistics perform well in controlling the size, even though
the versions that use overlapping blocks are somewhat liberal. For our tests, empirical
size being close to the significance level is a feature that holds over different DGPs
and sample sizes. Turning to power comparison, there is clear evidence that our tests
are reliable in that they have good power against different form of instabilities. There
appears to be substantial power gains relative to existing methods especially when
the instability (i) is short-lasting and/or (ii) is located toward the tail of the out-
of-sample. These properties characterize both statistics using non-overlapping and
\textsuperscript{29}We actually implemented the $t_{\text{stat}}$ by using either Andrews’s (1991) or Newey and West’s (1987)
estimator of the long-run variance. We also experimented different choices for the truncation lag.
The results, however, were unchanged. We suspect that this property depends on the estimation
of the long-run variance in our forecasting context which can be challenging due to small sample
sizes and to the presence of breaks. The same issues were found in Martins and Perron (2016) and
overlapping blocks.

3.8 Conclusions

We have formalized the concepts of forecast instability and forecast failure. Our definition poses at the center the economic forecaster and emphasizes the importance of the time duration of the instability. We assume the data arise as an outcome of an underlying system of stochastic differential equations which then implies that we can approximate the sequence of forecast losses by a continuous-time stochastic process. We have built a testing framework based on the local pathwise properties of that process and have adopted an infill asymptotics to derive the null distribution of the test statistics. The null distribution follows an extreme value distribution. Our results can be used to test whether the predictive ability of a given forecast model changes over time and can be applied in forecasting exercises involving either low-frequency as well as high-frequency macroeconomic and financial variables. The simulation study confirms that there are substantial power gains especially when the instability (i) is short-lasting and/or (ii) is located toward the tail of the out-of-sample. Our framework allows for misspecification, different types of parameter instability and arbitrary forms of nonstationarity such as heteroskedasticity and serial correlation. Our continuous-time specification and associated continuous record asymptotic scheme can provide a promising complementary framework to the classical approach for forecasting in economics.
### 3.9 Appendix to Chapter 3

#### 3.9.1 Tables

Table 3.1: Empirical small sample size of forecast instability tests based on model S1

<table>
<thead>
<tr>
<th></th>
<th>(T_m)</th>
<th>(T_n)</th>
<th>(\ell_{\text{stat}})</th>
<th>(\ell_{\text{stat},c})</th>
<th>(B_{\text{max},h})</th>
<th>(Q_{\text{max},h})</th>
<th>(M_{\text{max},h})</th>
<th>(M_{Q_{\text{max}},h})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha = 0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T = 100)</td>
<td>25</td>
<td>75</td>
<td>0.052</td>
<td>0.038</td>
<td>0.044</td>
<td>0.037</td>
<td>0.112</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.063</td>
<td>0.030</td>
<td>0.019</td>
<td>0.011</td>
<td>0.078</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>25</td>
<td>0.051</td>
<td>0.046</td>
<td>0.056</td>
<td>0.050</td>
<td>0.096</td>
<td>0.095</td>
</tr>
<tr>
<td>(T = 200)</td>
<td>50</td>
<td>150</td>
<td>0.047</td>
<td>0.036</td>
<td>0.029</td>
<td>0.032</td>
<td>0.136</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.052</td>
<td>0.035</td>
<td>0.032</td>
<td>0.030</td>
<td>0.110</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>50</td>
<td>0.078</td>
<td>0.028</td>
<td>0.060</td>
<td>0.058</td>
<td>0.059</td>
<td>0.058</td>
</tr>
<tr>
<td>(T = 300)</td>
<td>75</td>
<td>225</td>
<td>0.045</td>
<td>0.041</td>
<td>0.040</td>
<td>0.049</td>
<td>0.106</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>150</td>
<td>0.054</td>
<td>0.034</td>
<td>0.061</td>
<td>0.057</td>
<td>0.145</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>75</td>
<td>0.072</td>
<td>0.028</td>
<td>0.104</td>
<td>0.095</td>
<td>0.092</td>
<td>0.076</td>
</tr>
<tr>
<td>(T = 400)</td>
<td>100</td>
<td>300</td>
<td>0.050</td>
<td>0.048</td>
<td>0.054</td>
<td>0.068</td>
<td>0.129</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.056</td>
<td>0.042</td>
<td>0.063</td>
<td>0.063</td>
<td>0.122</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>100</td>
<td>0.059</td>
<td>0.030</td>
<td>0.069</td>
<td>0.067</td>
<td>0.108</td>
<td>0.068</td>
</tr>
<tr>
<td>(\alpha = 0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T = 100)</td>
<td>25</td>
<td>75</td>
<td>0.154</td>
<td>0.112</td>
<td>0.099</td>
<td>0.142</td>
<td>0.186</td>
<td>0.163</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.102</td>
<td>0.087</td>
<td>0.115</td>
<td>0.096</td>
<td>0.111</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>25</td>
<td>0.137</td>
<td>0.071</td>
<td>0.053</td>
<td>0.067</td>
<td>0.128</td>
<td>0.130</td>
</tr>
<tr>
<td>(T = 200)</td>
<td>50</td>
<td>150</td>
<td>0.103</td>
<td>0.106</td>
<td>0.095</td>
<td>0.117</td>
<td>0.130</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.116</td>
<td>0.096</td>
<td>0.100</td>
<td>0.096</td>
<td>0.157</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>50</td>
<td>0.114</td>
<td>0.076</td>
<td>0.128</td>
<td>0.150</td>
<td>0.103</td>
<td>0.093</td>
</tr>
<tr>
<td>(T = 300)</td>
<td>75</td>
<td>225</td>
<td>0.108</td>
<td>0.110</td>
<td>0.077</td>
<td>0.109</td>
<td>0.167</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>150</td>
<td>0.103</td>
<td>0.094</td>
<td>0.116</td>
<td>0.106</td>
<td>0.204</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>75</td>
<td>0.135</td>
<td>0.132</td>
<td>0.142</td>
<td>0.118</td>
<td>0.194</td>
<td>0.163</td>
</tr>
<tr>
<td>(T = 400)</td>
<td>100</td>
<td>300</td>
<td>0.098</td>
<td>0.108</td>
<td>0.108</td>
<td>0.108</td>
<td>0.192</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.105</td>
<td>0.091</td>
<td>0.087</td>
<td>0.139</td>
<td>0.167</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>100</td>
<td>0.112</td>
<td>0.079</td>
<td>0.114</td>
<td>0.109</td>
<td>0.166</td>
<td>0.112</td>
</tr>
</tbody>
</table>

The table reports the rejection probabilities of 100\(\alpha\)%-level tests proposed in the chapter and those proposed by Giacomini and Rossi (2009) [(abbreviated GR (2009)] for model S1. For all methods we use the fixed forecasting scheme. \(T = T_m + T_n\), where \(T\) is the total sample size, \(T_m\) is the size of the in-sample window and \(T_n\) is the size of the out-of-sample window. \(m_T\) is set equal to the smallest integer allowed by Condition 2. Based on 5,000 replications.
Table 3.2: Empirical small sample size of forecast instability tests based on model S2

<table>
<thead>
<tr>
<th>$\alpha = 0.05$</th>
<th>$T_m$</th>
<th>$T_n$</th>
<th>$T_{stat}$</th>
<th>$T_{stat,c}$</th>
<th>$B_{max,h}$</th>
<th>$Q_{max,h}$</th>
<th>$MB_{max,h}$</th>
<th>$MQ_{max,h}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 100$</td>
<td>25</td>
<td>75</td>
<td>0.049</td>
<td>0.019</td>
<td>0.086</td>
<td>0.098</td>
<td>0.090</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>0.069</td>
<td>0.024</td>
<td>0.058</td>
<td>0.072</td>
<td>0.083</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>25</td>
<td>0.039</td>
<td>0.016</td>
<td>0.081</td>
<td>0.111</td>
<td>0.092</td>
<td>0.091</td>
</tr>
<tr>
<td>$T = 200$</td>
<td>50</td>
<td>150</td>
<td>0.049</td>
<td>0.025</td>
<td>0.076</td>
<td>0.072</td>
<td>0.138</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>0.057</td>
<td>0.036</td>
<td>0.070</td>
<td>0.073</td>
<td>0.106</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>50</td>
<td>0.075</td>
<td>0.020</td>
<td>0.055</td>
<td>0.070</td>
<td>0.082</td>
<td>0.070</td>
</tr>
<tr>
<td>$T = 300$</td>
<td>75</td>
<td>225</td>
<td>0.050</td>
<td>0.029</td>
<td>0.058</td>
<td>0.036</td>
<td>0.102</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>150</td>
<td>0.059</td>
<td>0.032</td>
<td>0.077</td>
<td>0.072</td>
<td>0.144</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>75</td>
<td>0.065</td>
<td>0.025</td>
<td>0.096</td>
<td>0.103</td>
<td>0.152</td>
<td>0.123</td>
</tr>
<tr>
<td>$T = 400$</td>
<td>100</td>
<td>300</td>
<td>0.054</td>
<td>0.032</td>
<td>0.061</td>
<td>0.041</td>
<td>0.123</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0.051</td>
<td>0.035</td>
<td>0.065</td>
<td>0.048</td>
<td>0.111</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>100</td>
<td>0.068</td>
<td>0.031</td>
<td>0.067</td>
<td>0.063</td>
<td>0.115</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Model S2. We use the estimator $\nu_L$ from Theorem 3.4.2 for the asymptotic variance of $Q_{max,h}$ and $MQ_{max,h}$. For the statistics $t_{stat}$ and $t_{stat,c}$ we use the Newey-West estimator with truncation lags $\left\lfloor T_n^{1/3} \right\rfloor$ as recommended by Giacomini and Rossi (2009). The notes of Table 3.1 applies.
3.9.2 Figures

3.9.2.1 General Instability

**Figure 3.1:** Power functions for model P1a with $T = 100$ and $T = 150$

Power functions for model P1a: $Y_t = 2.73 - 0.44X_{t-1} + \delta X_{t-1} \mathbb{1}\{t > T_0^b\} + e_t$, where $X_{t-1} \sim \text{i.i.d.} \mathcal{N}(0, 1)$, $e_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$, and $T_0^b = T\lambda_0$. $T = 100$ (left panels) and $T = 150$ (right panels). $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = 0.4T$ while out-of-sample size is $T_n = 0.6T$. The green and blue broken lines correspond to $B_{\text{max}, h}$ and $Q_{\text{max}, h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure 3.2: Power functions for model P1a with $T = 200$ and $T = 300$
Power functions for model P1a. $T = 200$ (left panels) and $T = 300$ (right panels). The notes of Figure 3.1 apply.
Figure 3.3: Power functions for model P4 with $T = 200$ and $T = 300$

Power functions for model P4 (single break in variance): $Y_t = 0.5X_{t-1} + (1 + \delta 1 \{t > T^0\})e_t$ where $X_{t-1} \sim \text{i.i.d.}\mathcal{N}(1, 1)$ and $e_t \sim \text{i.i.d.}\mathcal{N}(0, 1)$. $T = 200$ (left panels) and $T = 300$ (right panels). $\lambda_0 = 0.6$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = 0.3T$ while out-of-sample size is $T_n = 0.7T$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure 3.4: Power functions for model P4 with $T = 400$ and $T = 500$

Power functions for model P4. $T = 400$ (left panels) and $T = 500$ (right panels). $\lambda_0 = 0.8$ (top panels) and $\lambda_0 = 0.9$ (bottom panels). The notes of Figure 3.3 apply.
Figure 3.5: Power functions for model P7 with $T = 200$ and $T = 300$

Power functions for model P7 (ARCH errors): $Y_t = 2.73 - 0.44X_{t-1} + \delta X_{t-1} \mathbf{1}\{t > T^0_b\} + \epsilon_t$, where $X_{t-1} \sim \mathcal{N}(0, 1.5)$ and $\epsilon_t = \sigma_t u_t$, $\sigma_t^2 = 0.5 + 0.5 \epsilon_{t-1}^2$, $u_t \sim \mathcal{N}(0, 1)$. $T = 200$ (left panels) and $T = 300$ (right panels). $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). $T_m = 0.5T$ and $T_n = 0.5T$. The light-blue and blue broken lines correspond to a version of $Q_{\max,h}$ that uses $\hat{\nu}_L$ but with different choices of $m_T$ (for the light-blue broken line we increase the number of blocks by one relative to the recommended value of $m_T$).
3.9.2.2 Short-Term Instability

Power functions for model P1a with short-term instability and with $T = 100$ and 150.

Power functions for model P1a with short-term instability: $Y_t = 2.73 - 0.44X_{t-1} + \delta X_{t-1}1 \{T_0^b < t \leq T_0^b + p\} + e_t$, where $X_{t-1} \sim \text{i.i.d.} \mathcal{N}(0, 1)$, $e_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$, and $T_0^b = T\lambda_0$. We set $(T, p) = \{(100, 20), (150, 25)\}$. $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). $T_m = 0.4T$ and $T_n = 0.6T$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.

Figure 3.6: Power functions for model P1a with short-term instability and with $T = 100$ and 150.
Figure 3.7: Power functions for model P1a with $T = 200$ and $T = 300$

Power functions for model P1a. We set $(T, p) = \{(200, 20), (300, 30)\}$. The notes of Figure 3.6 apply.
Figure 3.8: Power functions for model P4 with $T = 200$ and $T = 300$

Power functions for model P4 (single break in variance) with short-term instability: $Y_t = 0.5X_{t-1} + (1 + \delta 1 \{T_b^0 < t \leq T_b^0 + p\}) e_t$ where $X_{t-1} \sim \text{i.i.d.} \mathcal{N}(1, 1)$ and $e_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$. We set $(T, p) = \{(200, 30), (300, 30)\}$. $\lambda_0 = 0.6$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). $T_m = 0.3T$ and $T_n = 0.7T$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t^{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure 3.9: Power functions for model P4 with short-term instability and with $T = 400$ and $T = 500$

Power functions for model P4 (single break in variance) with short-term instability. We set $(T, p) = \{(400, 30), (500, 30)\}$. The notes of Figure 3.8 apply.
Chapter 4

Generalized Laplace Inference in Multiple Change-Points Models\textsuperscript{1}

4.1 Introduction

In the context of the multiple change-points model of Bai and Perron (1998), we develop inference methods for the change-point dates which build upon an asymptotic distribution theory derived under the classical long-span asymptotic framework for a class of Generalized Laplace (GL) estimators. The distinguishing feature of the class of GL estimators is that they are defined by an integration-based rather than an optimization-based method, the latter typically characterizing classical extremum estimators. The idea behind our method traces back to Laplace (1774), who first suggested to interpret a certain transformation of a least-squares criterion function as a statistical belief over a parameter of interest. Hence, a Laplace estimator is defined similarly to a Bayesian estimator although the former relies on a statistical criterion function rather than a parametric likelihood. Consequently, the GL estimator is interpreted as a classical (non-Bayesian) estimator and the inference methods proposed in this chapter retain a frequentist interpretation. More specifically, the GL estimators of the change-point dates are constructed as a function of integral transformations of the least-squares criterion. In a first step, we use the approach of Bai and Perron (1998) to evaluate the least-squares criterion function at all candidate

\textsuperscript{1}This chapter is based on joint work with Pierre Perron.
break dates. We then apply a certain transformation to obtain a proper distribution over the parameter of interest, referred to as the Quasi-posterior. For a given choice of a loss function and of a prior density, the estimator is then defined either explicitly as, for example, the mean or median of the (weighted) Quasi-posterior or implicitly as the minimizer of a smooth convex optimization problem.

The underlying asymptotic framework considered is the long-span shrinkage asymptotics of Bai (1997), Bai and Perron (1998) and also Perron and Qu (2006) who considerably relaxed some conditions, where the magnitude of the parameter shift is sample-size dependent and approaches zero as the sample size increases to infinity. Early contributions on this approach are Hinkley (1971), Bhattacharya (1987), and Yao (1987)—for estimating break points—and Hawkins (1977), Picard (1985), Kim and Siegmund (1989), Andrews (1993), Horváth (1993), Andrews and Ploberger (1994)—for testing for structural breaks. See also the reviews of Csörgő and Horváth (1997), Perron (2006), Casini and Perron (ming) and references therein.

Our goal is to develop GL inference, as alternatives to the ones based on the least-squares estimator, which have better small-sample properties, namely lower Mean Absolute Error (MAE) and Root-Mean Squared Error (RMSE), and confidence sets with accurate coverage probabilities and relatively short lengths for a wide range of break sizes, whether smaller or large. These properties are not fully shared by existing methods which work well for either small or large breaks, but not for both.

The asymptotic distribution of the Generalized Laplace estimator is derived via a local parameter defined as a normalized deviation from the true fractional break date $\lambda_0^b$. The normalization factor corresponds to the rate of convergence of the original (extremum) least-squares estimator of $\lambda_0^b$ as established by Bai and Perron (1998). Under a shrinking-shift setting, this rate of convergence was shown to be $T \| \delta_T \|^2$,
where \( T \) is the sample size, \( \delta_T \) is the shrinking magnitude of the parameter shift satisfying the restriction \( \delta_T = v_T \delta_0 \) with \( v_T \to 0 \) and \( T^{1/2 - \vartheta} v_T \to \infty \) for some \( \vartheta \in (0, 1/2) \). The GL estimator of the fractional break date \( \lambda_0^b \) attains the same convergence rate as the least-squares estimator of e.g. Bai (1997) and Bai and Perron (1998). However, it imposes a further condition on \( \vartheta \), which is now required to lie in the interval \((0, 1/4)\). This minor additional condition is needed because the GL estimator is defined through integration-based estimation and involves “smoothing” the least-squares criterion function.

The asymptotic distribution of the GL estimator depends on a sample-size dependent smoothing parameter sequence applied to the least-squares criterion function. We derive two distinct limiting distributions that correspond to different smoothing of the criterion function, with the rate of convergence being the same in both cases.\(^2\) We establish a dichotomy of the GL estimator. In one case, the estimator displays the same limiting law as the asymptotic distribution of the least-squares estimator as derived in Bai and Perron (1998) [see also Hinkley (1971), Picard (1985) and Yao (1987)]. In a second case, the limiting distribution is characterized by a ratio of integrals over functions of Gaussian processes and resembles the limiting distribution of Bayesian change-point estimators. This latter approximation is exploited for the purpose of constructing confidence sets for the break dates. We use the concept of highest density regions (HDR) introduced by Casini and Perron (2017a) which best summarizes the properties of the probability distribution of interest. Compared to the asymptotic method of Bai (1997), this procedure yields prediction sets for the break date which better accounts for the uncertainty over the parameter space because it effectively combines information already present in the point-wise least-squares estimate of the

\(^2\)A limiting distribution that depends on an input parameter was also shown in Jun et al. (2015), who introduced Laplace-type estimators for a class of cube-root consistent estimators [cf. Kim and Pollard (1990)].
break date with a statistical measure of the uncertainty in the least-squares criterion function.

The dual nature of the asymptotic distribution of the GL estimator constitutes a significant theoretical result. On one hand, one may expect that an estimator defined as a transformation of a statistical criterion function should have a limiting distribution similar to an estimator defined as the extremum of the same criterion function. In our context, the first approximation result for the class of GL estimators demonstrates a first-order asymptotic theoretical equivalence of the Laplace estimator to the standard least-squares estimator of the break date. On the other hand, forming a Quasi-posterior through a transformation of the criterion function leads one to associate the Laplace estimator to a Bayes-type estimator. Hence, a second approximation result states that the Laplace estimator admits a limit law equivalent to the corresponding Bayesian change-point estimators. This is notably useful because it allows the development of inference methods using a Bayesian approach yet retaining a classical (frequentist) interpretation.

Laplace’s seminal insight has been applied successfully in many disciplines and extended in the statistics literature to give rise to a technique known as Laplace approximation. Such approximations adopt integration-based methods in order to solve statistical extremum problems. In econometrics, Chernozhukov and Hong (2003) introduced Laplace-type estimators as an alternative to classical (regular) extremum estimators in several microeconometric problems such as censored median regression, nonlinear instrumental variable (IV) and many others; see also Forneron and Ng (2017) for a review and comparisons. Their main motivation was developing an estimation method able to solve the curse of dimensionality inherent to the computation of such estimators. In contrast, the class of GL estimators in structural change
models serves distinct multiple purposes. First and foremost, despite a considerable amount of research, inference about the break dates presents several challenges and it is difficult for a single method to provide satisfactory performance uniformly over different data-generating mechanisms and break magnitudes. The GL inference better accounts for the uncertainty over the parameter space by combining different sources of information. Thus, it proves to be reliable and accurate besides maintaining a classical interpretation. Second, its hybrid definition allows for the adoption of Bayesian approaches within a classical (frequentist) framework which seems to be theoretically and practically relevant due to the non-regularities of the structural change problem.

Turning to the problem of constructing confidence sets for the break date, the standard asymptotic method for the linear regression model was proposed in Bai (1997), while Elliott and Müller (2007) proposed to invert the locally best invariant test of Nyblom (1989). Moreover, Eo and Morley (2015) suggested to invert the likelihood-ratio statistic of Qu and Perron (2007). These alternatives to the standard asymptotic methods of Bai (1997) were mainly motivated by finite-sample results indicating that the exact coverage rates of the confidence intervals obtained from Bai’s (1997) method are often below the nominal level when the magnitude of the break is small. It has been shown that the method of Elliott and Müller (2007) delivers the most accurate coverage rates but the average length of the confidence sets is significantly larger than that other methods. In recent work, Casini and Perron (2017a, 2017b) developed a continuous record asymptotic theory and proposed inference methods based on the least-squares or the GL estimates. However, their results apply only to linear models with a single change-point without trending regressors.

The confidence sets for the break dates constructed from the GL inference that we develop result in exact coverage rates close to the nominal level and short length
of the confidence sets. This holds true even when the magnitude of the break is small, a case in which Bai’s (1997) method delivers coverage probabilities far below the nominal level. This more accurate coverage probability is accompanied by an overall length of the confidence set that is comparable with that from Bai’s (1997) method. Compared to the OLS-CR method of Casini and Perron (2017a), it is larger for medium-sized breaks while it is as short for large breaks.

The rest of the chapter is organized as follows. We first focus on the single change-point case. Section 4.2 presents the statistical setting. We develop the asymptotic theory in Section 4.3 and the inference methods in Section 4.4. Results for multiple change-points models are given in Section 4.5. Section 4.6 presents results from a Monte Carlo study. Section 4.7 concludes. All proofs are included in the supplementary material.

### 4.2 The Model and its Assumptions

We start with introducing the formal setup for our analysis. Section 4.2.1 introduces the structural change model with a single change-point and Section 4.2.2 reviews the least-squares estimation method for the break date. In Section 4.2.3 we present the relevant assumptions. The following notation is used throughout. We denote the transpose of a matrix \( A \) by \( A' \). The sample size is \( T \) and all limits in this chapter are taken as \( T \to \infty \). We use \( \| \cdot \| \) to denote the Euclidean norm of a linear space, i.e., \( \|x\| = (\sum_{i=1}^{p} x_i^2)^{1/2} \) for \( x \in \mathbb{R}^p \). For a matrix \( A \) we use the vector-induced norm, i.e., \( \|A\| = \sup_{x \neq 0} \|Ax\| / \|x\| \). All vectors are column vectors. For two vectors \( a \) and \( b \), we write \( a \leq b \) if the inequality holds component-wise. We use \( \lfloor \cdot \rfloor \) to denote the largest smaller integer function. Boldface is used for sets. We use \( \xrightarrow{p} \) to denote convergence in probability and convergence in distribution, respectively.
\( \mathbb{C}_b(E) \) \([\mathbb{D}_b(E)]\) is the collection of bounded continuous (càdlàg) functions from \( E \) to \( \mathbb{R} \). Weak convergence on either \( \mathbb{C}_b(E) \) or \( \mathbb{D}_b(E) \) is denoted by \( \Rightarrow \). The symbol “≜” stands for definitional equivalence.

4.2.1 The Structural Change Model with a Single Change-point

We consider a sample of observations \( \{(y_t, w_t, z_t) : t = 1, \ldots, T\} \), defined on a filtered probability space \((\Omega, \mathcal{F}, \mathbb{P})\), on which, all of the random elements introduced in what follows are defined. The model is

\[
y_t = w'_t \phi^0 + z'_t \delta^0_1 + e_t, \quad (t = 1, \ldots, T^0_b)
\]

\[
y_t = w'_t \phi^0 + z'_t \delta^0_2 + e_t, \quad (t = T^0_b + 1, \ldots, T)
\]

(4.2.1)

where \( y_t \) is an observed scalar dependent variable, \( w_t \) and \( z_t \) are observed regressors of dimension, \( p \) and \( q \), respectively, and \( e_t \) is an unobserved error term. The true parameter vectors \( \phi^0, \delta^0_1 \) and \( \delta^0_2 \) are unknown and we define \( \delta^0 \triangleq \delta^0_2 - \delta^0_1 \), with \( \delta^0 \neq 0 \) so that a structural change at date \( T^0_b \) has taken place. It is useful to re-parametrize the model as follows. Letting \( x_t \triangleq (w'_t, z'_t)' \) and \( \beta^0 \triangleq ((\phi^0)\', (\delta^0_1)\')' \), we can rewrite the model as

\[
y_t = x'_t \beta^0 + e_t, \quad (t = 1, \ldots, T^0_b)
\]

\[
y_t = x'_t \beta^0 + z'_t \delta^0 + e_t \quad (t = T^0_b + 1, \ldots, T)
\]

(4.2.2)

More generally, we can define \( z_t \) as a linear transformation of \( z_t \triangleq D' x_t \), where \( D \) is a \((p + q) \times q\) matrix with full column rank. A pure structural change model in which all regression parameters are subject to change corresponds to \( D = I_{(p+q) \times q} \), whereas a partial structural change model arises when \( D = (0_{q \times p}, I_{q \times q})' \). In order to facilitate the derivations, we reformulate model (4.2.2) in matrix format. Let \( Y = (y_1, \ldots, y_T)' \), \( X = (x_1, \ldots, x_T)' \), \( e = (e_1, \ldots, e_T)' \), \( X_1 = (x_1, \ldots, x_{T_b}, 0, \ldots, 0)' \),
$X_2 = (0, \ldots, 0, x_{T_0+1}, \ldots, x_T)'$ and $X_0 = (0, \ldots, 0, x_{T_0+1}, \ldots, x_T)'$. Further, define $Z_1, Z_2$ and $Z_0$ in a similar way: $Z_1 = X_1D$, $Z_2 = X_2D$ and $Z_0 = X_0D$. We omit the dependence of the matrices $X_i$ and $Z_i$ ($i = 1, 2$) on $T_b$. Then, (4.2.2) is equivalent to

$$Y = X\beta + Z_0\delta + e. \quad (4.2.3)$$

### 4.2.2 The Least-squares Criterion Function

Let $\theta^0 \triangleq (\phi^0, \delta_1^0, \delta_2^0)$ denote the true value of the parameter vector $\theta$. The break date least-squares (LS) estimator $\hat{T}_{b}^{\text{LS}}$ is the minimizer of the sum of squares residuals [denoted $S_T(\theta, T_b)$] from (4.2.3). The parameter $\theta$ can be concentrated out of the criterion function, which results in a criterion function depending only on $T_b$:

$$\hat{\theta}^{\text{LS}}(T_b) = \arg\min_\theta S_T(\theta, T_b), \quad \hat{T}_{b}^{\text{LS}} = \arg\min_{1 \leq T_b \leq T} S_T\left(\hat{\theta}^{\text{LS}}(T_b), T_b\right).$$

We also have

$$\arg\min_{1 \leq T_b \leq T} S_T\left(\hat{\theta}^{\text{LS}}(T_b), T_b\right) = \arg\max_{T_b}\hat{\delta}^{\text{LS}}(T_b) \left(Z'_2MZ_2\right) \hat{\delta}^{\text{LS}}(T_b) \quad (4.2.4)$$

where $M_X \triangleq I - X(X'X)'$, $\hat{\delta}_{T_b}$ is the least-squares estimator of $\delta^0$ obtained by regressing $Y$ on $X$ and $Z_2$ and the statistic $Q_T\left(\hat{\delta}^{\text{LS}}(T_b), T_b\right)$ is the numerator of the sup-Wald statistic. The Laplace-type inference to be introduced in the next section builds on the least-squares criterion function $Q_T(\delta(T_b), T_b)$.

### 4.2.3 Assumptions

**Assumption 4.1.** $T_b^0 = \lfloor T\lambda_b^0 \rfloor$, where $\lambda_b^0 \in \Gamma^0 \subset (0, 1)$.

**Assumption 4.2.** With $\{\mathcal{F}_t, t = 1, 2, \ldots\}$ a sequence of increasing $\sigma$-fields, $\{z_t u_t, \mathcal{F}_t\}$ forms a $L^r$-mixingale sequence with $r = 2 + \nu$ for some $\nu > 0$. That is, there
exist nonnegative constants \( \{\varrho_{1,t}\}_{t \geq 1} \) and \( \{\varrho_{2,j}\}_{j \geq 0} \) such that \( \varrho_{2,j} \to 0 \) as \( j \to \infty \) and for all \( t \geq 1 \) and \( j \geq 0 \), and we have for \( r \geq 1 \): (i) \( \| \mathbb{E}(z_t u_t | \mathcal{F}_{t-j}) \|_r \leq \varrho_{1,t} \varrho_{2,j} \), (ii) \( \| z_t u_t - \mathbb{E}(z_t u_t | \mathcal{F}_{t-j}) \|_r \leq \varrho_{1,t} \varrho_{2,j+1} \). In addition, (iii) \( \max_j \varrho_{1,t} < C_1 < \infty \) and (iv) \( \sum_{j=0}^{\infty} j^{1+\nu} \varrho_{2,j} < \infty \) for some \( \nu > 0 \), (v) \( \| z_t \|_{2r} < C_2 < \infty \) and \( \| u_t \|_{2r} < C_3 < \infty \) for some \( C_1, C_2, C_3 > 0 \).

Assumption 4.3. There exists an \( l_0 > 0 \) such that for all \( l > l_0 \), the minimum eigenvalues of \( H_l = \frac{1}{l} \sum_{T-l}^{T} x_t x_t' \), \( H^*_l = \frac{1}{l} \sum_{T_0-l}^{T_0} x_t x_t' \) and \( H^{**}_l = \frac{1}{l} \sum_{T_0+l}^{T_0+l} x_t x_t' \) are bounded away from zero. In addition, these matrices are invertible when \( l \geq p+q \) and have stochastically bounded norms uniformly in \( l \).

Assumption 4.4. \( T^{-1} X'X \xrightarrow{p} \Sigma_{XX} \), where \( \Sigma_{XX} \) is some positive definite matrix. If \( x_t \) is a stochastic regressor, then \( \sup_t \mathbb{E} \| x_t \|^{4+\nu} \leq C < \infty \).

The above assumptions are standard in the literature and similar to those in Perron and Qu (2006) who relaxed those used by Bai (1997) and Bai and Perron (1998). We refer to these papers for a comprehensive discussion. It is well-known that a statistically consistent estimator for \( T^0_b \) does not exist because only the fractional break date \( \lambda^0_b \) can be consistently estimated. The fractional change-point estimator \( \hat{\lambda}^{LS}_b \) has a \( T \)-rate of convergence, which is faster than the standard \( \sqrt{T} \)-rate pertaining to regular estimators. The corresponding result for the break date estimator \( \hat{T}^{LS}_b \) only states that, as the sample size increases, \( \hat{T}^{LS}_b \) remains within a bounded distance from \( T^0_b \). However, this does not affect the estimation problem of the regression coefficients \( \theta^0 \), for which \( \hat{\theta}^{LS} \) is a regular estimator; i.e., \( \sqrt{T} \)-consistent and asymptotically normally distributed. The latter properties are a consequence of a standard theoretical result in change-point analysis: the estimation of the regression parameters is asymptotically independent from estimation of the change-point, and given the fast rate of convergence of \( \hat{\lambda}^{LS}_b \), the regression parameters are essentially estimated as if
the change-point were known.

More complex is the derivation of the asymptotic distribution of the fractional break point estimator \( \hat{\lambda}_b^{LS} \). From Hinkley (1971), the limiting distribution is complicated even for an i.i.d. Gaussian process with a change in the mean. Therefore, in order to make progress it is necessary to consider a shrinkage asymptotic setting in which the size of the shift converges to zero as the sample size \( T \) goes to infinity. This approach was initiated by Picard (1985) and Yao (1987) and then extended by Bai (1997) to general linear models.

### 4.3 Generalized Laplace Estimation

We define the GL estimator in Section 4.3.1 and discuss its usefulness in Section 4.3.2. In Section 4.3.3 we describe the asymptotic framework under which we derive the limiting distribution together with some assumptions. The large-sample results are presented in Section 4.3.4.

#### 4.3.1 The Class of Laplace Estimators

The class \( \mathcal{L}(\theta, T_b) \) of Generalized Laplace (GL) estimators in structural change models relies on the original least-squares criterion function \( Q_T(\delta(T_b), T_b) \). Our parameter of interest is \( \lambda_0^b = T_0^b / T \). Given the criterion function \( Q_T(\delta(T_b), T_b) \), the Quasi-posterior \( p_T(\lambda_b) \) is defined by the following exponential transformation,

\[
p_T(\lambda_b) \triangleq \frac{\exp (Q_T(\delta(\lambda_0^b), \lambda_b)) \pi(\lambda_b)}{\int_{\Gamma^0} \exp (Q_T(\delta(\lambda_0^b), \lambda_b)) \pi(\lambda_b) d\lambda},
\]

where \( \pi(\cdot) \) is a weighting function. It is evident that \( p_T(\lambda_b) \) defines a proper distribution over the parameter space \( \Gamma^0 \triangleq (0, 1) \). The Quasi-posterior mean, median and quantiles are straightforward to define given \( p_T(\lambda_b) \). However, such quantities
may also be implicitly defined as solutions of smooth convex optimization problems for a given loss function. This is how we shall formally define the $\mathcal{L}(\theta, T_b)$-class of estimators, restricting attention to convex loss functions $l_T(\cdot)$. Common examples of $l_T(\cdot)$ include (a) $l_T(r) = a_T |r|^m$, the polynomial loss function (the squared loss function is obtained when $m = 2$ while the absolute deviation loss function corresponds to $m = 1$); (b) $l_T(r) = a_T (\tau - 1 (r \leq 0)) r$, the check loss function; where $a_T$ is a divergent sequence. We can then define the Expected Risk function, under the density $p_T(\cdot)$ and the loss $l_T(\cdot)$ as $\mathcal{R}_{l,T}(s) \triangleq \mathbb{E}_{p_T}[l_T(s - \tilde{\lambda}_b)]$, where $\tilde{\lambda}_b$ is a random variable with distribution $p_T$, and $\mathbb{E}_{p_T}$ denotes expectation taken under $p_T$. Using (4.3.1) we can write,

$$
\mathcal{R}_{l,T}(s) \triangleq \int_{\Gamma_0} l_T(s - \lambda_b) p_T(\lambda_b) d\lambda_b = \int_{\Gamma_0} l_T(s - \lambda_b) \left( \frac{\exp(Q_T(\delta(T_b), T_b)) \pi(\lambda_b)}{\int_{\Gamma_0} \exp(Q_T(\delta(T_b), T_b)) \pi(\lambda_b) d\lambda_b} \right) d\lambda_b.
$$

(4.3.2)

The Laplace-type estimator $\hat{\lambda}^{GL}_b$ shall be interpreted as a decision rule that, given the information contained in the Quasi-posterior $p_T$, is least unfavorable according to the loss function $l_T$ and the prior density $\pi$. This leads us to define $\hat{\lambda}^{GL}_b$ as the minimizer of the expected risk function in (4.3.2), i.e., $\hat{\lambda}^{GL}_b \triangleq \arg\min_{s \in \Gamma_0} [\mathcal{R}_{l,T}(s)]$. The choice of the loss and of the weight function $\pi(\cdot)$ hinges on the statistical problem addressed. In the structural change problem, a natural choice for the Quasi-prior $\pi$ is the density of the asymptotic distribution of $\hat{\lambda}^{LS}_b$. This requires to replace the population quantities appearing in that distribution by consistent plug-in estimates—cf. Bai and Perron (1998)—and derive its density via simulations as in Casini and Perron (2017a). The attractiveness of the Quasi-posterior in (4.3.1) is that it provides additional information about the parameter of interest $\lambda^0_b$ beyond what is already included in the point estimate $\hat{\lambda}^{LS}_b$. We discuss in more detail why the GL approach
is useful in change-point models in Section 4.3.2. Then, inference procedures based, for example, on the mean (or median) of the Quasi-posterior density function should yield better inference. Observe that the Generalized Laplace estimator $\lambda_{GL}$ results in the mean (median) of the Quasi-posterior upon the choice of the squared (absolute deviation) loss function. This approach will result in more accurate inference even in cases with high uncertainty in the data as we shall document in the small-sample Monte Carlo study in Section 4.6.

**Assumption 4.5.** Let $l_T (r) \triangleq l (a_T r)$, with $a_T$ being a positive divergent sequence. $L$ denotes the set of functions $l : \mathbb{R} \to \mathbb{R}_+$ that satisfy (i) $l (r)$ is defined on $\mathbb{R}$, with $l (r) \geq 0$ and $l (r) = 0$ if and only if $r = 0$; (ii) $l (r)$ is continuous at $r = 0$ but is not identically zero; (iii) $l (\cdot)$ is convex and $l (r) \leq 1 + |r|^m$ for some $m > 0$.

**Assumption 4.6.** $\pi : \mathbb{R} \to \mathbb{R}_+$ is a continuous, uniformly positive density function satisfying $\pi^0 \triangleq \pi (\lambda^0 b) > 0$, and for some finite $C_\pi < \infty$, $\pi^0 < C_\pi$. Also, $\pi (\lambda b) = 0$ for all $\lambda b / \in \Gamma^0$, and $\pi$ is twice continuously differentiable with respect to $\lambda b$ at $\lambda^0 b$.

Part (1)-(3) of Assumption 4.5 are similar to those in Bickel and Yahav (1969), Ibragimov and Has’minskiï (1981) and Chernozhukov and Hong (2003). Practical reasons offer justification for the convexity assumption on the loss function $l_T (\cdot)$.

The dominant restriction in part (3) is conventional and assumes implicitly that the loss function has been scaled by some constant. What is important is that the growth of the function $l_T (r)$ as $|r| \to \infty$ is slower than $\exp (\epsilon |r|)$ for any $\epsilon > 0$. Assumption 4.6 on the prior is satisfied for any reasonable choice.

The large-sample properties of the $\mathcal{L} (\theta, T_b)$-class are studied under the shrinkage asymptotic setting of Bai (1997) and Bai and Perron (1998). Thus, we need the following assumption.
Assumption 4.7. Let $\delta T_0 \triangleq v_T \delta_0$ where $v_T > 0$ is a scalar satisfying $v_T \to 0$ as $T \to \infty$ and $T^{1/2-\vartheta} v_T \to \infty$ for some $\vartheta \in (0, 1/4)$.

Assumption 4.7 requires the magnitude of the break to shrink to zero at any slower rate than $1/\sqrt{T}$. However, the specific rate differs from that in Bai (1997) and Bai and Perron (1998). They require $\vartheta \in (0, 1/2)$ whereas Assumption 4.7 specifies $\vartheta \in (0, 1/4)$. The reason is merely technical; the asymptotics of the Laplace-type estimator involve “smoothing” the criterion function, and thus one needs to guarantee that $\hat{\lambda}_b$ approaches $\lambda_0^b$ at a sufficiently fast rate. Under the shrinkage asymptotics, Proposition 1 in Bai (1997) states that $T \| \delta T \|^2 \left( \hat{\lambda}_b^{LS} - \lambda_0^b \right) = O_P(1)$.

4.3.2 Discussion about the GL Approach

We use Figure 4.1-4.2 to illustrate the main idea behind the usefulness of the GL method. They present plots of the density of the distribution of the least-squares estimate of the break date for the simple model $y_t = \phi_0 + z_t' (\delta_1^0 + \delta_0^{1} \{ t > T_0^b \}) + e_t$ where \{zt\} follows an ARMA(1,1) process and \{et\} is a sequence of i.i.d. Gaussian disturbances. The distributions presented are: the exact finite-sample distribution, Bai’s (1997) classical large-$T$ limit distribution, the infeasible version of Casini and Perron’s (2017a) continuous record limit distribution and its feasible version constructed using plug-in estimates. Noteworthy is the non-standard features of the finite-sample distribution when the break magnitude is small which include multi-modality, fat tails and asymmetry—the latter if the change-point is not at mid-sample. The central mode corresponds to the estimated change-point while the two modes are in the tails near the starting and the end sample points, respectively; when the break magnitude is small the least-squares estimator tends to locate the break in the tails since the evidence of a break is weak. In addition, it is evident that the continuous record asymptotic theory provides a better approximation to the finite-sample distribution.
than the classical large-\(T\) asymptotic distribution.

The GL method is useful because it weights the information from the least-squares criterion function with the information from the prior density—which, here, is set equal to the density of the classical large-\(T\) asymptotic distribution of the least-squares estimate. Note that the least-squares objective function is quite flat when the magnitude of the break is small and so the least-squares estimate is imprecise, while the resulting Quasi-posterior or e.g. the median of the Quasi-posterior may lead to better estimates.

4.3.3 Normalized Version of the GL Estimator

In order to develop the asymptotic results, we introduce an input “parameter” sequence \(\{\gamma_T\}\) whose properties are specified below. We assume that \(\lambda_0^T \in \Gamma^0 \subset (0, 1)\) is the unknown extremum of \(\tilde{Q}(\theta^0, \lambda_b) = \mathbb{E}[Q_T(\theta^0, \lambda_b)]\) and that \(\theta^0 \triangleq \left( (\phi^0)', (\delta_1^0)', (\delta_2^0)' \right) \in S \subset \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q\). Our analysis is within a vanishing neighborhood of \(\theta^0\). For any \(\theta \in S\), let \(\lambda_b^0(\theta)\) be an arbitrary element of \(\Gamma^0(\theta) \triangleq \left\{ \lambda_b \in \Gamma^0 : \tilde{Q}(\theta, \lambda_b) = \sup_{\lambda_b \in \Gamma^0} \tilde{Q}(\theta, \lambda_b) \right\}\).

Provided a uniqueness condition is assumed (see below), \(\Gamma^0(\theta)\) contains a single element, \(\lambda_b^0\). Further, let \(\overline{Q}_T(\theta, \lambda_b) \triangleq Q_T(\theta, \lambda_b) - Q_T(\theta, \lambda_b^0), \overline{Q}_T^0(\theta, \lambda_b) \triangleq \mathbb{E}[Q_T(\theta, \lambda_b) - Q_T(\theta, \lambda_b^0)]\), and \(G_T(\theta, \lambda_b) = \overline{Q}_T(\theta, \lambda_b) - \overline{Q}_T^0(\theta, \lambda_b)\). The exact expressions for \(G_T, \overline{Q}_T^0\) and \(\overline{Q}_T\) are provided in the preliminary section of the mathematical appendix. The Generalized Laplace estimator \(\hat{\lambda}_{b^{GL}}(\theta)\) can be defined as the minimizer of a
normalized version of $\mathcal{R}_{l,T}(s)$:

$$
\Psi_{l,T}(s; \theta) = \int_{\mathcal{R}^0} l(s - \lambda_b) \frac{\exp \left( \left( \gamma_T / \left( T \| \delta_T \| \right)^2 \right) Q_T(\theta, \lambda_b) \right) \pi(\lambda_b)}{\int_{\mathcal{R}^0} \exp \left( \left( \gamma_T / \left( T \| \delta_T \| \right)^2 \right) Q_T(\theta, \lambda_b) \right) \pi(\lambda_b) d\lambda_b} d\lambda_b.
$$

(4.3.3)

The choice of $\{\gamma_T\}$ gives rise to different limiting distributions. Our analysis is local in nature and thus we shall write $\check{\lambda}_{b}^{GL}(\hat{\theta}) \triangleq \check{\lambda}_{b}^{GL,*}(r_T(\hat{\theta} - \theta^0))$, in terms of some local parameters $v$ and $\tilde{v}$ assumed to belong to some compact set $V \subset \mathbb{R}^{p+2q}$, where $r_T$ is the convergence rate of $\hat{\theta} - \theta^0$. With this notation, $\check{\lambda}_{b}^{GL}(\hat{\theta}) = \check{\lambda}_{b}^{GL,*}(\tilde{v}, v)$ minimizes

$$
\Psi_{l,T}(s; \tilde{v}, v) = \int_{\mathcal{R}^0} l(s - \lambda_b) \times
$$

$$
\frac{\exp \left( \left( \gamma_T / \left( T \| \delta_T \| \right)^2 \right) (G_T(\theta^0 + \tilde{v}/r_T, \lambda_b) + Q_T^0(\theta^0 + v/r_T, \lambda_b)) \right) \pi(\lambda_b)}{\int_{\mathcal{R}^0} \exp \left( \left( \gamma_T / \left( T \| \delta_T \| \right)^2 \right) (G_T(\theta^0 + \tilde{v}/r_T, \lambda_b) + Q_T^0(\theta^0 + v/r_T, \lambda_b)) \right) \pi(\lambda_b) d\lambda_b} d\lambda_b.
$$

(4.3.4)

As mentioned in the previous section, the regression parameters $\theta^0$ are estimated as if the true break date $T_b^0$ were known. This implies that $\hat{\theta}$ belongs to the class of regular estimators which are $\sqrt{T}$-consistent and asymptotically normal. Thus, we set $r_T = \sqrt{T}$ hereafter.

The main theoretical result of this section concerns the large-sample properties of $\check{\lambda}_{b}^{GL}$. Since $\check{\lambda}_{b}^{GL}$ is defined implicitly as an external estimator, its large-sample properties can be derived as follows. We first show, for each pair $(v, \tilde{v})$ with $v, \tilde{v} \in V$, the convergence of the marginal distributions of the sample function $\Psi_{l,T}(s; v, \tilde{v})$ to the marginal distributions of the random function

$$
\Psi_{l}^0(s) = \int_{\mathcal{R}} l(s - u) \frac{\mathcal{V}(u)}{\int_{\mathcal{R}} \mathcal{V}(v) dv} du,
$$
where the limit process $\Psi_l^0 (s)$ does not depend on $v$ nor $\tilde{v}$. Next, we show that the family of probability measures in $C_b (K)$, with $K \triangleq \{ s \in \mathbb{R} : |s| \leq K$ and $K < \infty \}$, generated by the contractions of $\Psi_{l,T} (s; \tilde{v}, v)$ on $K$ is dense uniformly in $(v, \tilde{v})$. Finally, we examine the oscillations of the minimizers of the sample criterion $\Psi_{l,T} (s; v, \tilde{v})$.

It is important to note that the results derived in this section are considerably more general than what is required for the structural change model. The reason is that the change-point model is recovered as a special case corresponding to $\Psi_{l,T} (s) = \Psi_{l,T} (s; 0, 0)$. That is, defining the Laplace estimator in a $1/r_T$-neighborhood of the slope parameter vector $\theta^0$ is not strictly necessary and one can essentially develop the same analysis with $\theta$ fixed at its true value $\theta^0$. This observation relies on the properties of (orthogonal) least-squares projections and would not apply, for example, to the least absolute deviation (LAD) estimator of the break date [cf. Bai (1995)] for which $\Psi_{l,T} (s; \tilde{v}, v)$ should instead be considered.³ The key intuition which allows studying the weak convergence of $\Psi_{l,T} (\cdot; \tilde{v}, v)$ for fixed $(\tilde{v}, v)$ relies on the property that the limit process does not depend on $v$ nor $\tilde{v}$. Nonetheless, we establish theoretical results under this more general setting since they may be useful for future work.

³The same issue is present when estimating structural changes in the quantile regression model [cf. Oka and Qu (2010)] and in the instrumental variable model [cf. Hall et al. (2010) and Perron and Yamamoto (2014; 2015)].
$G_T \left( \theta^0 + \bar{v}/r_T, \lambda^0_b \left( v \right) + u/\psi_T \right)$. Apply a simple substitution in (4.3.4) to yield,

$$
\Psi_{t,T} (s; \bar{v}, v) = \int_{\Gamma_T} l(s - u) \frac{\exp \left( \left( \gamma_T/T \right) \delta_T \right) \left( \tilde{G}_{T,V} (u, \bar{v}) + Q_{T,V} (u) \right)}{\int_{\Gamma_T} \exp \left( \left( \gamma_T/T \right) \delta_T \right) \left( \tilde{G}_{T,V} (w, \bar{v}) + Q_{T,V} (w) \right)} \pi_{T,v} (u) \, du,
$$

where $\Gamma_T \triangleq \{ u \in \mathbb{R} : \lambda^0_b + u/\psi_T \in \Gamma^0 \}$.

**Assumption 4.8.** \{\{z_t, e_t\}\} is second-order stationary within each regime such that $\mathbb{E} (z_t z'_t) = V_1$ and $\mathbb{E} (e_t^2) = \sigma^2_1$ for $t \leq T^0_b$ and $\mathbb{E} (z_t z'_t) = V_2$ and $\mathbb{E} (e_t^2) = \sigma^2_2$ for $t > T^0_b$.

**Assumption 4.9.** For $r \in [0, 1]$, $(T^0_b)^{-1/2} \sum_{t=1}^{\lfloor r T^0_b \rfloor} z_t e_t \Rightarrow \mathcal{G}_1 (r)$ and $(T - T^0_b)^{-1/2} \sum_{t=1}^{T^0_b + \lfloor r (T - T^0_b) \rfloor} z_t e_t \Rightarrow \mathcal{G}_2 (r)$, $z_t e_t \Rightarrow \mathcal{G}_2 (r)$, where $\mathcal{G}_i (\cdot)$ is a multivariate Gaussian process on $[0, 1]$ with zero mean and covariance $\mathbb{E} \left[ \mathcal{G}_i (u), \mathcal{G}_i (s) \right] = \min \{ u, s \}$, $\Sigma_i$ ($i = 1, 2$), and

$$
\Sigma_1 \triangleq \lim_{T \to \infty} \mathbb{E} \left[ (T^0_b)^{-1/2} \sum_{t=1}^{\lfloor r T^0_b \rfloor} z_t e_t \right]^2, \quad \Sigma_2 \triangleq \lim_{T \to \infty} \mathbb{E} \left[ (T - T^0_b)^{-1/2} \sum_{t=1}^{\lfloor r T^0_b \rfloor} z_t e_t \right]^2.
$$

Furthermore, for any $0 < r_0 < 1$ with $r_0 \neq \lambda_0$, $T^{-1} \sum_{t=\lfloor r_0 T \rfloor + 1}^{\lfloor \lambda_0 T \rfloor} z_t z'_t \overset{\mathbb{P}}{\to} (\lambda_0 - r_0) V_1$, and $T^{-1} \sum_{t=\lfloor r_0 T \rfloor + 1}^{\lfloor \lambda_0 T \rfloor} z_t z'_t \overset{\mathbb{P}}{\to} (r_0 - \lambda_0) V_2$ so that $\lambda_-$ and $\lambda_+$ (the minimum and maximum eigenvalues of the last two matrices) satisfy $0 < \lambda_- \leq \lambda_+ < \infty$.

Assumptions 4.8-4.9 are equivalent to A9 in Bai (1997) and A7 in Bai and Perron (1998). More specifically, Assumption 4.9 requires that, within each regime, an Invariance Principle holds for $\{z_t e_t\}$. Let $\zeta_t \triangleq z_t e_t$. For $T_b \leq T^0_b$ let $g (\zeta_t; u) \triangleq (\delta^0)' \sum_{t=\lfloor r_0 T \rfloor + [u/\psi_T]}^{\lfloor \lambda_0 T \rfloor} \zeta_t$ and

$$
\bar{g} (\zeta_t; u, \bar{v}, v; \psi_T, r_T) \triangleq \sqrt{\psi_T} \left( \delta^0 + \bar{v}/r_T \right)' \sum_{t=\lfloor \lambda_0 T \rfloor + [u/\psi_T]}^{\lfloor \lambda_0 T \rfloor + [u/\psi_T]} \zeta_t.
$$
Define analogously \( g(\zeta_t; u) \) and \( \tilde{g}(\zeta_t; u, \tilde{v}, v; \psi_T, r_T) \) for the case \( T_b > T_b^0 \). We now present some technical assumptions that are necessary for the derivation of the asymptotic results for the GL estimators.

**Assumption 4.10.** For some neighborhood \( \Theta^0 \subset S \) of \( \theta^0 \), (i) for all \( \lambda_b \neq \lambda_b^0 \), \( \tilde{Q}(\theta_0, \lambda_b) < \tilde{Q}(\theta^0, \lambda^0_b) \); (ii) for any \( v, \tilde{v}_1, \tilde{v}_2 \in V \) and \( u, s \in \mathbb{R} \),

\[
\Sigma(u, s) \triangleq \lim_{T \to \infty} \mathbb{E}[\tilde{g}(\zeta_t; u, \tilde{v}_1, v; \psi_T, r_T) \tilde{g}(\zeta_t; s, \tilde{v}_2, v; \psi_T, r_T)],
\]

**Assumption 4.11.** The random variable \( \xi^0_l \triangleq \xi(\lambda^0_b) \) is uniquely defined by \( \Psi_l(\xi^0_l) \triangleq \inf_s \Psi_l(s) = \int_{\mathbb{R}} l(s - u) \left( \exp(\mathcal{V}(u)) / (\int_{\mathbb{R}} \exp(\mathcal{V}(w)) dw) \right) du \), where

\[
\mathcal{V}(s) \triangleq \begin{cases} 
2 \left( (\delta^0)' \Sigma_1 \delta^0 \right)^{1/2} W_1(-s) - |s| (\delta^0)' V_1 \delta^0, & \text{if } s \leq 0 \\
2 \left( (\delta^0)' \Sigma_2 \delta^0 \right)^{1/2} W_2(s) - s (\delta^0)' V_2 \delta^0, & \text{if } s > 0,
\end{cases}
\]

and \( W_1, W_2 \) are two independent standard Wiener processes defined on \( [0, \infty) \).
4.3.4 Asymptotic Results for Generalized Laplace Estimates

In practice, the squared loss function is often employed. Hence, it is useful to first present the theoretical results for the case when the GL estimator reduces to the Quasi-posterior mean. This allows us to keep the theoretical results tractable and provide the main intuition without the need of complex notation. Furthermore, the case of the Quasi-posterior mean is instructive since we can compare our results with the corresponding results concerning the least-squares as well as the Bayesian change-point estimators. Theorem 4.3.1 presents the large-sample results for the Quasi-posterior mean. Corresponding results for general loss functions satisfying Assumption 4.5 are given in Theorem 4.3.2. In deriving the asymptotic distribution of the GL estimator $\lambda^{\text{GL}}_b$ we need to consider its limiting behavior as $\theta$ lies within a shrinking neighborhood of $\theta^0$, i.e., $\hat{\theta} = \theta^0 + v/\gamma$ for some $v \in V$. This gives rise to further notations [cf. $\tilde{\lambda}^{\text{GL}}_b(\theta)$ and $\tilde{\lambda}^{\text{GL},*}_b(\tilde{v}, v)$ defined below]. We remark that these notations are used only for the development of the asymptotic results concerning $\lambda^{\text{GL}}_b$ and do not mean that, for example, $\lambda^{\text{GL}}_b(\theta)$ or $\tilde{\lambda}^{\text{GL},*}_b(\tilde{v}, v)$ is a different estimator from $\tilde{\lambda}^{\text{GL}}_b$.

4.3.4.1 The Asymptotic Distribution of the Quasi-posterior Mean

Proceeding as above we define $\tilde{\lambda}^{\text{GL}_b}_b(\theta) \triangleq \tilde{\lambda}^{\text{GL},*}_b(\gamma T (\theta - \theta^0))$, where

$$\tilde{\lambda}^{\text{GL},*}_b(\tilde{v}, v) \triangleq \frac{\int_{\Gamma_0} \lambda_b \exp \left( \left( \gamma T / T \| \delta_T \|_2^2 \right) \right) (G_T (\theta^0 + \tilde{v}/\gamma, \lambda_b) + Q^T_0 (\theta^0 + v/\gamma, \lambda_b)) \pi (\lambda_b) d\lambda_b}{\int_{\Gamma_0} \exp \left( \left( \gamma T / T \| \delta_T \|_2^2 \right) \right) (G_T (\theta^0 + \tilde{v}/\gamma, \lambda_b) + Q^T_0 (\theta^0 + v/\gamma, \lambda_b)) \pi (\lambda_b) d\lambda_b},$$

and $v, \tilde{v}$ each belong to some compact set $V \subset \mathbb{R}^{p+2q}$. For each $v \in V$, we consider $\tilde{\lambda}^{\text{GL},*}_b(\cdot, v)$ as a random process with paths in $\mathbb{D}_b(V)$. The same reasoning explained above applies for the Quasi-posterior mean: we focus on the weak convergence of
\( \lambda^*_b (\cdot, v) \) for fixed \( v \) since the limit process is independent of \( v \) and constant as a function of \( \tilde{v} \). More precisely, we will show that for \( \lambda^0_{b,T} (v) = \lambda^0_{b,T} (\theta^0 + v/r_T) \) and diverging sequences \( \{ \gamma_T \} \) and \( \{ r_T \} \), the sequence \( a_T \left( \lambda^*_{b}(\tilde{v}, v) - \lambda^0_{b,T} (v) \right) \) converges in distribution in \( \mathbb{D}_b (\mathbb{V}) \) for each \( v \) to a limit process whose properties do not depend on \( v \) nor \( \tilde{v} \). Introduce the local parameter \( u = \psi_T \left( \lambda_b - \lambda^0_{b,T} (v) \right) \) and apply a simple substitution in (4.3.8) to deduce that,

\[
\psi_T \left( \lambda^*_{b}(\tilde{v}, v) - \lambda^0_{b,T} (v) \right) = \int_{\mathbb{R}} u \exp \left( \left( \gamma_T / \left( T \| \delta_T \|^2 \right) \right) \left( \bar{G}_{T,v} (u, \tilde{v}) + Q_{T,v} (u) \right) \right) \pi_{T,v} (u) \, du,
\]

where again we have used the notation \( \pi_{T,v} (u) = \pi \left( \lambda^0_{b,T} (v) + u/\psi_T \right) \), \( Q_{T,v} (u) = Q^0_T \left( \theta^0 + v/r_T , \lambda^0_{b,T} (v) + u/\psi_T \right) \) and \( \bar{G}_{T,v} (u, \tilde{v}) = G_T \left( \theta^0 + \tilde{v}/r_T , \lambda^0_{b,T} (v) + u/\psi_T \right) \).

The limit of the GL estimator depends on the limit of the process

\[
\left( \gamma_T / \left( T \| \delta_T \|^2 \right) \right) \left( \bar{G}_{T,v} (u, \tilde{v}) + Q_{T,v} (u) \right).
\]

As part of the proof of Theorem 4.3.1, we show that the sequence of processes \( \left\{ \bar{G}_{T,v} (u, \tilde{v}) , T \geq 1 \right\} \) converges weakly in \( \mathbb{D}_b (\mathbb{R} \times \mathbb{V}) \) to a Gaussian limit process \( \mathcal{W} \) which does not vary with \( \tilde{v} \), whereas \( Q_{T,v} (\cdot) \) is approximated by a (deterministic) drift process which takes negative values and is independent of \( v \) and flat in \( \tilde{v} \).

In anticipation of the results, we make a few comments about the notation for the weak convergence of processes on the space of bounded càdlàg functions \( \mathbb{D}_b \). Let \( \mathbb{V} \subset \mathbb{R}^{p+2q} \) be a compact set. Let \( W_T (u, \tilde{v}, v) \) denote an arbitrary sample process with bounded càdlàg paths evaluated at the local parameters \( u \in \mathbb{R} \), and \( v, \tilde{v} \in \mathbb{V} \). For each fixed \( v \in \mathbb{V} \), we shall write \( W_T (u, \tilde{v}, v) \Rightarrow \mathcal{W} (u, \tilde{v}, v) \) in \( \mathbb{D}_b (\mathbb{R} \times \mathbb{V}) \) whenever the process \( W_T (\cdot, \cdot, v) \) converges weakly to \( \mathcal{W} (\cdot, \cdot, v) \), where \( \mathcal{W} (\cdot, \cdot, v) \)
also belong to \( \mathbb{D}_b(\mathbb{R} \times \mathbb{V}) \). As a shorthand, we shall omit the argument \( u(\bar{v}) \) if the limit process does not depend on \( u(\bar{v}) \). The same notational conventions are used for the case when \( W_T \) is only a function of \((\bar{v}, v)\). In Theorem 4.3.1 the convergence occurs in \( \mathbb{D}_b(\mathbb{V}) \) and holds for every \( v \in \mathbb{V} \). In the statement of the theorem we write this as convergence in \( \mathbb{D}_b \). The same convention is used for the other results of this section.

**Condition 5.** As \( T \to \infty \) there exist a positive finite number \( \kappa_\gamma \) such that \( \gamma_T/T \| \delta_T \|^2 \to \kappa_\gamma \).

**Theorem 4.3.1.** Assume \( l(\cdot) \) is the squared loss function. Under Assumption 4.1-4.4 and 4.5-4.11, and Condition 5,

\[
T \| \delta_T \|^2 \left( \tilde{\lambda}^{GL}_b - \lambda^0_b \right) \Rightarrow \int \frac{u \exp \left( \mathcal{W}(u) - A^0(u) \right) du}{\int \exp \left( \mathcal{W}(u) - A^0(u) \right) du},
\]

in \( \mathbb{D}_b \), where \( \mathcal{W}(\cdot) \) is a two-sided Gaussian process with covariance \( \Sigma \) given in Assumption 4.10.

Theorem 4.3.1 states that the asymptotic distribution of the GL estimator is a ratio of integrals of functions of tight Gaussian processes. We shall compare this result with the limiting distribution of the Bayesian change-point estimator of Ibragimov and Has’minskii (1981). They considered a simple diffusion process with a change-point in the deterministic drift. The limiting distribution of their Bayesian estimator can be found in equation (2.17) on pp. 338 of Ibragimov and Has’minskii (1981). One can observe that the limiting distribution of the GL estimate from Theorem 4.3.1 for the case of a break in the mean for model (4.2.1) is essentially the same as the one appearing in Ibragimov and Has’minskii (1981). The process \( \mathcal{W}(u) - A^0(u) \) in Theorem 4.3.1 is simply replaced by the process \( Z_0(s) \) appearing in the limiting
distribution of Ibragimov and Has’minskii (1981); see their equation (2.13) on pp. 334:

\[
Z_0(s) = \begin{cases} 
W_1(s) - |s|/2, & \text{if } s < 0 \\
W_2(s) - |s|/2, & \text{if } s \geq 0,
\end{cases}
\]

where \( W_i(s), i = 1, 2 \) are two independent standard Wiener processes starting at 0. Hence, while the Generalized Laplace estimator conserves a classical (frequentist) interpretation, it is first-order equivalent in law to a corresponding Bayes-type estimator.

We now present a result concerning the dichotomy of the limiting distribution of the GL estimator. The following proposition shows that, under a different condition on the smoothing sequence parameter \( \{\gamma_T\} \), the GL estimator achieves a distinct limiting distribution.

**Condition 6.** As \( T \to \infty \), \( T \|\delta_T\|^2 / \gamma_T = o(1) \).

**Proposition 4.3.1.** Assume \( l(\cdot) \) is the squared loss function. Under Assumption 4.1-4.4 and 4.5-4.11, and Condition 6, \( T \|\delta_T\|^2 \left( \lambda^{GL}_b - \lambda^0_{b,T} \right) \Rightarrow \arg \max_{s \in \mathbb{R}} \mathcal{V}(s) \) in \( \mathbb{D}_b \).

**Corollary 4.3.1.** Define \( \Xi_e \equiv (\delta^0)' \Sigma_2 \delta^0 / (\delta^0)' \Sigma_1 \delta^0 \) and \( \Xi_Z \equiv (\delta^0)' V_2 \delta^0 / (\delta^0)' V_1 \delta^0 \). Under Assumption 4.1-4.4 and 4.5-4.11, and Condition 6,

\[
\left( (\delta_T' V_1 \delta_T)^2 / \delta_T' \Sigma_1 \delta_T \right) \left( \tilde{T}^{GL}_b - T^0_{b,T} \right) \xrightarrow{d} \arg \max_{s \in \mathbb{R}} \mathcal{V}^*(s),
\]
in \( \mathbb{D}_b \) where

\[
\mathcal{N}(s) \triangleq \begin{cases} 
W_1(-s) - |s|/2, & \text{if } s \leq 0 \\
\Xi^{1/2}W_2(s) - \Xi Z s/2, & \text{if } s > 0.
\end{cases}
\]

Corollary 4.3.1 together with Theorem 4.3.1 shows that when enough smoothing is applied, the GL estimator is (first-order) asymptotically equivalent to the least-squares or Maximum Likelihood estimator [cf. Bai (1997) and Yao (1987), respectively]. The intuition for this result is straightforward: when the criterion function is subject to sufficient smoothing then the sequence of Quasi-posterior probability densities converges to the generalized dirac probability measure concentrated at the argmax of the limit criterion function. This is in analogy to a well-known result in statistics [cf. Corollary 5.11 in Robert and Casella (2004)], stating that in a parametric statistical experiment indexed by a parameter \( \theta \in \Theta \), the maximum likelihood estimator \( \hat{\theta}_{T}^{ML} \) is a limit of a Bayes estimator as the smoothing parameter \( \gamma \) diverges to infinity, i.e., using obvious notation:

\[
\hat{\theta}_{T}^{ML} = \arg \max_{\theta \in \Theta} L_T(\theta) = \lim_{\gamma \to \infty} \frac{\int_{\Theta} \theta \exp(\gamma L_T(\theta)) \pi(\theta) d\theta}{\int_{\Theta} \exp(\gamma L_T(\theta)) \pi(\theta) d\theta}.
\]

4.3.4.2 The Asymptotic Distribution for General Loss Functions

We return to the general case of loss functions satisfying Assumption 4.5. Theorem 4.3.2 shows that \( T \|\delta_T\|^2 \left( \hat{\lambda}_{b}^{GL} - \lambda_{b}^0 \right) \) is (first-order) asymptotically equivalent to the random variable \( \xi_{l}^0 \) determined by

\[
\Psi_{l} \left( \xi_{l}^0 \right) \triangleq \inf_{r} \Psi_{l}(r) = \inf_{r \in \mathbb{R}} \left\{ \frac{1}{\kappa \gamma} \int_{\mathbb{R}} l(r - u) \frac{\exp(\varphi(u) - \Lambda^0(u))}{\int \exp(\varphi(u) - \Lambda^0(u)) du} du \right\}.
\]
Theorem 4.3.2. Under Assumption 4.1-4.4 and 4.5-4.11, and Condition 5, for \( l \in L \),
\[
T \| \delta_T \|^2 \left( \lambda_{GL}^l - \lambda_0^l \right) \Rightarrow \xi_l^0,
\]
where the random variable \( \xi_l^0 \) is determined by equation (4.3.11).

The existence and uniqueness of \( \xi_l^0 \) holds from Assumption 4.11. Let \( p_0^* \triangleq \frac{\exp(W(u) - A_0(u))}{\int \exp(W(u) - A_0(u))du} \). If one interprets \( p_0^* \) as a true posterior density function, then \( \xi_l^0 \) would naturally be viewed as a Bayesian estimator for the loss function \( l_T(\cdot) \). In particular, in analogy to the above comparison with the Bayesian estimator of Ibragimov and Has’minskii (1981), one can interpret the GL estimator as a Quasi-Bayesian estimator. While this is by itself a theoretically interesting result, we actually exploit it in order to construct more reliable inference methods about the date of a structural change. Under the least-absolute deviation loss, the GL estimator converges in distribution to the median of \( p_0^* \). Certainly, we shall use Theorem 4.3.1-4.3.2 but not Proposition 4.3.1 since the latter results in the same confidence intervals of Bai (1997) and of Bai and Perron (1998). After some investigation, we found that both estimation and inference under the least-absolute loss works well and this is what will be used in our simulation study. We shall see that statistical inference based on the GL class can be more reliable since the ratio of integrals-type limiting distribution provides a more accurate description of the uncertainty over the parameter space.

4.4 Inference based on GL Estimators

In this section, we discuss inference procedures about the break date based on the large-sample results of the previous section. Inference under general loss functions based on Theorem 4.3.2 is what we recommend to use in practice and it is supported by our simulation study below. The dichotomy of the limiting distribution arising for Quasi-posterior mean can be accommodated by a rate-adaptive inference—adaptive
to the choice of \( \{\gamma_T\} \). However, this is beyond the scope of the chapter, and we thus omit the details—which remain available upon request.

Since the limiting distribution from Theorem 4.3.2 involves certain population quantities, we begin by assuming that they can be replaced by statistically consistent estimators. They are easy to construct [cf. Bai (1997) and Bai and Perron (1998); see also Section 4.6].

**Assumption 4.12.** There exist sequences of estimators \( \hat{\lambda}_{0,T}, \hat{\delta}_T, \hat{\Xi}_{Z,T}, \) and \( \hat{\Xi}_{e,T} \) such that
\[
\hat{\lambda}_{0,T} = \lambda_0 + o_p(1), \quad \hat{\delta}_T = \delta^0 + o_p(1), \quad \hat{\Xi}_{Z,T} = \Xi_Z + o_p(1) \quad \text{and} \quad \hat{\Xi}_{e,T} = \Xi_e + o_p(1).
\]
Furthermore, for all \( u, s \in \mathbb{R} \) and any \( c > 0 \), there exist covariation processes \( \hat{\Sigma}_{i,T}(\cdot) \) \( (i = 1, 2) \) that satisfy (i) \( \hat{\Sigma}_{1,T}(u, s) = \Sigma_1^0(u, s) + o_p(1) \) and \( \hat{\Sigma}_{2,T}(u, s) = \Sigma_2^0(u, s) + o_p(1) \), (ii) \( \hat{\Sigma}_{i,T}(u - s, u - s) = \hat{\Sigma}_{i,T}(u, u) + \hat{\Sigma}_{i,T}(s, s), i = 1, 2 \), (iii) \( \hat{\Sigma}_{i,T}(cu, cu) = c^2 \hat{\Sigma}_{i,T}(u, u), i = 1, 2 \), (iv) \( \mathbb{E}\left\{\sup_{\|u\|=1} \hat{\Sigma}_{i,T}^2(u, u)\right\} = O(1), i = 1, 2 \).

Let \( \{\hat{\gamma}_T\} \) be a (sample-size dependent) sequence of two-sided zero-mean Gaussian processes characterized by \( \hat{\Sigma}_T \). Construct the process \( \hat{\gamma}_T \) by replacing the population quantities in \( \hat{\gamma} \) by their corresponding estimators from the first part of Assumption 4.12 and further, replace \( \hat{\gamma} \) by \( \hat{\gamma}_T \). Assumption 4.12-(i) basically implies that the finite-dimensional limit law of \( \{\hat{\gamma}_T\} \) is the same as the finite-dimensional law of \( \gamma \) while parts (ii)-(iii) are needed for the integrability of the transform \( \exp(\hat{\gamma}_T(\cdot)) \).

Part (iv) is needed for the proof of asymptotic stochastic equicontinuity of \( \{\hat{\gamma}_T\} \). Let us introduce the following sample quantity:
\[
\hat{\xi}_T \triangleq \int_{\mathbb{R}} \int_{\mathbb{R}} u \exp\left(\hat{\gamma}_T(u)\right) \exp\left(\hat{\gamma}_T(v)\right) du dv.
\]

For a given choice of the input sequence \( \{\gamma_T\} \), the distribution \( \hat{\xi}_T \) can be evaluated numerically.
Proposition 4.4.1. Let \( l \in L \). Under Assumption 4.12, \( \hat{\xi}_T \) converges in distribution to the limiting distribution in Theorem 4.3.1.

The asymptotic distribution theory of the GL estimator may be exploited in several ways for inference about the break date. As emphasized by Casini and Perron (2017a), the finite-sample distribution of the break date least-squares estimator displays significant non-standard features. Hence, a conventional two-sided confidence interval may not result in a prediction set with reliable statistical properties across all break magnitudes and break locations. Thus, we use the concept of Highest Quasi-posterior Density (HQPD) regions, defined analogously to the Highest Density Region (HDR); cf. Hyndman (1996). The concept of Highest Density Region (HDR) was first introduced for inference in structural change models in Casini and Perron (2017a).

Definition 4.4.1. Highest Density Region: Assume that the density function \( f_Y(y) \) of a random variable \( Y \) defined on a probability space \((\Omega_Y, \mathcal{F}_Y, \mathbb{P}_Y)\) and taking values on the measurable space \((\mathcal{Y}, \mathcal{B})\) is continuous and bounded. The \((1 - \alpha)\) 100% Highest Density Region is a subset \( S(\kappa_\alpha) \) of \( \mathcal{Y} \) defined as \( S(\kappa_\alpha) = \{y : f_Y(y) \geq \kappa_\alpha\} \) where \( \kappa_\alpha \) is the largest constant that satisfies \( \mathbb{P}_Y(Y \in S(\kappa_\alpha)) \geq 1 - \alpha \).

For \( s = T \|\delta_T\|^2 \left( \tilde{\lambda}_b^{LS} - \lambda_b^0 \right) \), the asymptotic distribution theory of Bai (1997) suggests a belief \( \pi(s) \) over \( s \in \mathbb{R} \). This belief function can be used as a Quasi-prior for \( \lambda_b \) in the definition of the Quasi-posterior \( p_T(\lambda_b) \). Let \( \mu(\lambda_b) \) denote some density function defined by the Radon-Nikodym equation \( \mu(\lambda_b) = dp_T(\lambda_b)/d\lambda_L \), where \( \lambda_L \) denotes the Lebesgue measure. The following algorithm describes how one can construct a Quasi-Bayesian confidence set for \( T_b^0 \).

\footnote{See also Samworth and Wand (2010) and Mason and Polonik (2008, 2009) for more recent developments.}
Algorithm 3. GL HQDR-based Confidence Sets for $T_b^0$:

(1) Estimate by least-squares the break date and the regression coefficients from model (4.2.3);

(2) Set the Quasi-prior $\pi(\lambda_b)$ equal to the probability density of the limiting distribution from Proposition 3 in Bai (1997);

(3) Construct the Quasi-posterior given in (4.3.1);

(4) Obtain numerically the density $\mu(\lambda_b)$ as explained above and label it by $\hat{\mu}(\lambda_b)$;

(5) Compute the Highest Quasi-Posterior Density (HQPD) region of the probability distribution $\hat{\pi}_T(\lambda_b)$ and include the point $T_b$ in the level $(1 - \alpha)$% confidence set $C_{\text{HQPD}}(cv_\alpha)$ if $T_b$ satisfies the conditions in Definition 4.4.1.

If a general Quasi-prior $\pi(\lambda_b)$ is used, one simply begins directly with step 3.

In principle, any Quasi-prior $\pi(\lambda_b)$ satisfying Assumption 4.6 can be used. Note that $C_{\text{HQPD}}(cv_\alpha)$ retains a frequentist interpretation, since no prior probability nor parametric likelihood function of the data is required to compute it.

4.5 Models with Multiple Change-Points

Following Bai and Perron (1998), the multiple linear regression model with $m$ change-points is

$$y_t = w_t' \phi^0 + z_t' \delta^0_j + e_t, \quad \left( t = T_{j-1}^0 + 1, \ldots, T_j^0 \right)$$

for $j = 1, \ldots, m + 1$, where by convention $T_0^0 = 0$ and $T_{m+1}^0 = T$. There are $m$ unknown break points $(T_1^0, \ldots, T_m^0)$ and consequently $m + 1$ regimes each corresponding to a distinct parameter value $\delta^0_j$. The purpose is to estimate the unknown regression coefficients together with the break points when $T$ observations on $(y_t, w_t, z_t)$ are available. Many of the theoretical results pertaining to multiple breaks models...
follow directly from the single break case. The most important result concerning the limiting distribution is that the break points are asymptotically distinct and thus the limit distribution theory for the single break date extends readily to multiple breaks. More complicated is the actual computation of the estimates of the break dates which has been addressed by Bai and Perron (2003) who proposed an efficient algorithm based on the principle of dynamic programming; see also Hawkins (1976).

Let $T_i \triangleq \lfloor T \lambda_i \rfloor$ ($i = 1, \ldots, m$) and $\theta \triangleq (\phi', \delta'_1, \ldots, \delta'_m)'$. The class $\mathcal{L}(\theta, T_i; 1 \leq i \leq m)$ of GL estimators in multiple change-points models relies on the original least-squares criterion function

$$Q_T(\delta(\lambda_i), \lambda_i) = \sum_{i=1}^{m+1} \sum_{t=T_{i-1}}^{T_i} (y_t - w_t' \phi - z_t' \delta_i)^2,$$

where $\lambda_i \triangleq (\lambda_i; 1 \leq i \leq m)$ is a $m \times 1$ vector and constitutes the parameter of interest. In order to state the large-sample properties of the $\mathcal{L}(\theta, T_i)$-class we need to introduce the shrinkage theoretical framework of Bai and Perron (1998).

**Assumption 4.13.** Assumption A1-A5 of Bai and Perron (1998) hold. Let $\Delta^0_i = \delta^0_{i+1} - \delta^0_i$ and assume $\Delta_{T,i} = v_T \Delta^0_i$, where $v_T > 0$ is a scalar satisfying $v_T \to 0$ and $T^{1/2 - \varphi} v_T \to \infty$ for some $\varphi \in (0, 1/4)$. In addition, $\mathbb{E} \|z_t\|^2 < C$ and $\mathbb{E} \|e_t\|^{2/\varphi} < C$ for some $C < \infty$ and all $t$.

**Assumption 4.14.** Let $\Delta T^0_i = T^0_i - T^0_{i-1}$. For $i = 1, \ldots, m + 1$, uniformly in $s \in [0, 1]$,

(a) $(\Delta T^0_i)^{-1} \sum_{t=T_{i-1}}^{T_i} [s^{\Delta T^0_i}] z_t z_t' \overset{p}{\to} sV_i$, $(\Delta T^0_i)^{-1} \sum_{t=T_{i-1}}^{T_i} [s^{\Delta T^0_i}] e_t^2 \overset{p}{\to} s\sigma_i^2$, and

$$\left(\Delta T^0_i\right)^{-1} \sum_{t=T_{i-1}}^{T_i} [s^{\Delta T^0_i}] \sum_{r=T_{i-1}}^{T_{i+1}} [s^{\Delta T^0_i}] \mathbb{E}(z_t z_t' u_t u_r) \overset{p}{\to} s\Sigma_i;$$
(b) \( (\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0}^{T_{i+1}^0} z_t u_t \overset{p}{\to} \mathcal{G}_i(s) \) where \( \mathcal{G}_i(s) \) is a multivariate Gaussian process on \([0, 1]\) with mean zero and covariance \( \mathbb{E} [\mathcal{G}_i(s) \mathcal{G}_i(u)] = \min \{s, u\} \Sigma_i \).

Next, for \( i = 1, \ldots, m \), define \( \Xi_{Z,i} = (\Delta_0^0)^t V_{i+1} \Delta_0^0 / (\Delta_0^0)^t V_i \Delta_0^0, \) \( \Xi_{\varepsilon,i} = (\Delta_0^0)^t \Sigma_i \Delta_0^0 / (\Delta_0^0)^t \Sigma_i \Delta_0^0 \), and let \( W_1^{(i)}(s) \) and \( W_2^{(i)}(s) \) be independent Wiener processes defined on \([0, \infty)\), starting at 0 when \( s = 0 \). Note that \( W_1^{(i)}(s) \) and \( W_2^{(i)}(s) \) are also independent over \( i \). Finally, define

\[
\mathcal{Y}^{(i)}(s) \triangleq \begin{cases} 
2 \left( (\Delta_0^0)^t \Sigma_i \Delta_i \right)^{1/2} W_1^{(i)}(-s) - |s|(\Delta_0^0)^t V_i \Delta_i, & \text{if } s \leq 0 \\
2 \left( (\Delta_0^0)^t \Sigma_{i+1} \delta_0 \right)^{1/2} W_2^{(i)}(s) - s(\Delta_0^0)^t V_{i+1} \Delta_i, & \text{if } s > 0.
\end{cases} \tag{4.5.1}
\]

We now extend the notation of Section 4.3 to the present context. By redefining the Quasi-posterior \( p(\lambda_b) \) in terms of the parameter vector \( \lambda_b \), we have the definition of the GL estimator as the minimizer of the associated risk function [recall (4.3.2)], \( \hat{\lambda}^{\text{GL}}_b = \arg \min_{s \in I^0} [\mathcal{R}_{i,T}(s)], \) where now \( I^0 = B_1 \times \ldots \times B_m \), with \( B_i \) a compact subset of \((0, 1)\). The sets \( B_i \) are disjoint and satisfy \( \sup_{\lambda_i \in B_i} \inf_{\lambda_i \in B_i+1} \) for all \( i \).

**Assumption 4.15.** Assumption 4.5-4.6 hold with obvious modifications to allow for the multidimensional parameter \( \lambda_b \in I^0 \). Furthermore, Assumption 4.10 holds where now in part (i) \( \lambda_b \) replaces \( \lambda_b \), and in part (ii) \( \Sigma^{(i)}(\cdot, \cdot) \) \((1 \leq i \leq m + 1)\) replaces \( \Sigma(\cdot, \cdot) \) and is defined analogously for each regime.

Note that Assumption 4.11 implies that \( \xi_{i,i}^0 \triangleq \xi(\lambda_i^0) \) is uniquely defined by

\[
\Psi_i(\xi_{i,i}^0) \triangleq \inf_{u} \Psi_{i,i}(s) = \int_{\mathbb{R}} l(s - u) \left( \exp \left( \mathcal{Y}^{(i)}(u) \right) / \left( \int_{\mathbb{R}} \exp \left( \mathcal{Y}^{(i)}(w) \right) dw \right) \right) du,
\]

where \( \mathcal{Y}^{(i)}(u) = \mathcal{Y}^{(i)}(u) - \Lambda_i^0(u) \), with \( \mathcal{Y}^{(i)}(u) \) being a two-sided Gaussian process characterized by the covariance function \( \Sigma^{(i)}(\cdot, \cdot) \) and \( \Lambda_i^0(s) = |s|(\Delta_0^0)^t V_i \Delta_0^0 \) if \( s \leq 0 \) or \( \Lambda_i^0(s) = s(\Delta_0^0)^t V_{i+1} \Delta_0^0 \) if \( s > 0 \). The GL estimator is defined as the minimizer of

\[
\mathcal{R}_{i,T} \triangleq \int_{I^0} l(s - \lambda_b) \frac{\exp (Q_T(\delta(\lambda_b), \lambda_b)) \pi(\lambda_b)}{\int_{I^0} \exp (Q_T(\delta(\lambda_b), \lambda_b)) \pi(\lambda_b) d\lambda_b} d\lambda_b.
\]
The analysis is now in terms of the $m \times 1$ local parameter $u$ with components $u_i = T \| \Delta_{T,i} \|^2 \left( \lambda_i - \lambda^0_{i,T} (v) \right)$, with $\lambda^0_{i,T} (v) = \lambda^0_{i,T} (\theta^0 + v/r_T)$.

Theorem 4.5.1-4.5.2 and Proposition 4.5.1 extend corresponding results from Theorem 4.3.1-4.3.2 and Proposition 4.3.1, respectively, to multiple change-points. The key observation is that asymptotically the behavior of the GL estimator only matters in a small neighborhood of each $T^0_i$. Since each such neighborhood increases at rate $1/v_T$ while $T$ increases to infinity at a faster rate, these neighborhoods are asymptotically distinct and the limiting distribution is then similar to that in the single break case. This is the same argument underlying the analysis of Bai and Perron (1998) and of Ibragimov and Has’minskiı (1981). In particular, the limiting distribution of Proposition 4.5.1 corresponds to that from Proposition 5 in Bai and Perron (1998) whereas the limiting distribution in Theorem 4.5.1 should be compared with Theorem VII.2.3 in Ibragimov and Has’minskiı (1981).\footnote{More specifically, see pp. 335-336 of Ibragimov and Has’minskiı (1981) and the discussion before their Theorem 2.1.}

The same comments as those in Section 4.3 apply.

**Condition 7.** For $1 \leq i \leq m$ there exist positive finite numbers $\kappa_{\gamma,i}$ such that $\gamma_T / T \| \Delta_{T,i} \|^2 \rightarrow \kappa_{\gamma,i}$.

**Theorem 4.5.1.** Assume $l (\cdot)$ is the squared loss function. Under Assumptions 4.13-4.15 and Condition 7,

$$
T \| \Delta_{T,i} \|^2 \left( \hat{\lambda}^0_{i} - \lambda^0_{i} \right) \Rightarrow \frac{\int u \exp \left( \mathcal{H}^{(i)} (u) - A^0_{i} (u) \right) du}{\int \exp \left( \mathcal{H}^{(i)} (u) - A^0_{i} (u) \right) du},
$$

in $\mathbb{D}_b$.

**Condition 8.** For all $1 \leq i \leq m$, $T \| \Delta_{T,i} \|^2 / \gamma_T = o (1)$.\footnote{More specifically, see pp. 335-336 of Ibragimov and Has’minskiı (1981) and the discussion before their Theorem 2.1.}
Proposition 4.5.1. Assume \( l(\cdot) \) is the squared loss function. Under Assumptions 4.13-4.15 and Condition 8, \( T \| \Delta_{T,i} \|^2 \left( \hat{\lambda}_{i}^{\text{GL}} - \lambda_{0}^{i} \right) \Rightarrow \arg \max_{s \in \mathbb{R}} \mathcal{V}^{(i)}(s) \) in \( \mathbb{D}_b \) where \( \mathcal{V}^{(i)}(s) \) is defined in (4.5.1).

Turning to the general case of loss functions satisfying Assumption 4.5. Theorem 4.5.2 shows that the random quantity \( T \| \delta_{T} \|^2 \left( \hat{\lambda}_{i}^{\text{GL}} - \lambda_{0}^{i} \right) \) is (first-order) asymptotically equivalent to the random variable \( \xi_{0,l,i}^{i} \) determined by

\[
\Psi_{l} \left( \xi_{0,l,i}^{i} \right) \triangleq \inf_{r} \Psi_{l,i}(r) = \arg \inf_{r \in \mathbb{R}} \left\{ \int_{\mathbb{R}} l(r-u) \frac{\exp \left( \mathcal{W}^{(i)}(u) - \Lambda_{0}^{i}(u) \right)}{\int \exp \left( \mathcal{W}^{(i)}(u) - \Lambda_{0}^{i}(u) \right) du} \, du \right\}
\]  
(4.5.3)

Theorem 4.5.2. Under Assumptions 4.13-4.15 and Condition 7, for \( l \in \mathcal{L} \),

\[
T \| \Delta_{T,i} \|^2 \left( \hat{\lambda}_{i}^{\text{GL}} - \lambda_{0}^{i} \right) \Rightarrow \arg \max_{s \in \mathbb{R}} \mathcal{V}^{(i)}(s),
\]

where the random variable \( \xi_{0,l,i}^{i} \) is determined by equation (4.5.3).

A direct consequence of the results of this section is that statistical inference for the break dates \( T_{0}^{i} (i = 1, \ldots, m) \) can be carried out using the same methods for the single break case as described in Section 4.4.

4.6 Finite-Sample Evaluation of GL Method

The purpose of this section is twofold. Section 4.6.1 assesses the accuracy of the GL estimate of the change-point while Section 4.6.2 evaluates the small-sample properties of the confidence sets proposed.
4.6.1 Precision of the Change-point Estimate

We include in the study the following estimators of the change-point $T^0_b$: the least-squares estimator (OLS), the GL estimator under a least-absolute loss function with a long-span prior (GL-LN) [i.e., the prior set to be equal to the density of the limit distribution stated in Bai (1997), labelled LN, since it is based on a large span, say $N$, asymptotic framework]; the GL estimator under a least-absolute loss function with a uniform prior (GL-Uni). We compare the mean absolute error (MAE), standard deviation (Std), root-mean-squared error (RMSE), and the 25% and 75% quantiles.

We consider DGPs that take the following form:

$$y_t = D_t \alpha^0 + Z_t \beta^0 + Z_t \delta^0 1_{\{t > T^0_b\}} + e_t, \quad t = 1, \ldots, T, \quad (4.6.1)$$

with a sample size $T = 100$. Three versions of (4.6.1) are investigated: M1 involves a break in the mean which corresponds to $Z_t = 1$, $D_t$ absent, and $e_t \sim i.i.d. \mathcal{N}(0, 1)$; M2 is similar to M1 but with zero-mean stationary Gaussian AR(1) disturbances $\{e_t\}$ with autoregressive coefficient 0.3 and unit innovation variance; M3 is a dynamic model with $D_t = y_{t-1}$, $Z_t = 1$, $e_t \sim i.i.d. \mathcal{N}(0, 0.5)$ and $\alpha^0 = 0.6$. We set $\beta^0 = 1$ in M1-M2 and $\beta^0 = 0$ in M3. We consider fractional change-points $\lambda^0 = 0.3$ and 0.5, and break magnitudes $\delta^0 = 0.3, 0.4, 0.6$ and 1.

Table 4.1-4.3 present the results. When the magnitude of the break is small, the LS estimator displays quite large MAE. This absolute bias increases as the change-point point moves toward the tails. In contrast, the GL estimator shows substantially lower MAE when the size of the break is small or moderate, especially when the change-point point is at mid-sample. The GL estimator has smaller variance as well as lower RMSE than the LS estimator. In model M3 with $\lambda^0 = 0.3$, the GL and LS estimators have similar properties; this occurs more generally when breaks are large.
(i.e., $\delta^0 = 1$, bottom panel) since the LS estimator is quite precise. Notably, the distribution of GL-LN concentrates a higher fraction of the mass around the mid-sample relative to the finite-sample distribution of the LS estimate. This is mainly due to the fact that the Quasi-posterior essentially does not share the marked trimodality of the finite-sample distribution [cf. Casini and Perron (2017a)]. When the break magnitude is small, the objective function is quite flat with a small peak at the LS estimate. The Quasi-posterior has higher mass close to the LS estimate—which corresponds to the middle mode—and accordingly lower in the tails. The GL estimator that uses the uniform prior (GL-Uni) has similar properties to the LS estimator, indeed it is more precise, though by a relatively small margin. This reflects the fact that the highly non-standard features of the change-point problem implies that different priors yield GL estimates with different properties. Under a flat prior, the GL estimate uses information only from the LS objective function and it is thus similar to the LS estimate.

4.6.2 Properties of the GL Confidence Sets

We now assess the performance of the suggested inference procedures about the break date. We compare it with the following popular existing methods: Bai’s (1997) approach, Elliott and Müller’s (2007) approach based on inverting a sequence of locally best invariant tests using Nyblom’s (1989) statistic, the inverted likelihood-ratio (ILR) method of Eo and Morley (2015) which inverts the likelihood-ratio test of Qu and Perron (2007) and the HDR method proposed in Casini and Perron (2017a) based on continuous record asymptotics, labelled OLS-CR. These methods have been discussed in detail in Casini and Perron (2017a) and in Chang and Perron (2018). We can summarize their properties as follows. The confidence intervals obtained from Bai’s (1997) method display empirical converge rates often below the nominal level
when the size of the break is small. In general, Elliott and Müller’s (2007) approach achieves the most accurate coverage rates but the average length of the confidence sets is always substantially larger relative to other methods. In addition, this approach faces a drawback in models with serially correlated errors or lagged dependent variables, whereby the length of the confidence set approaches the whole sample as the magnitude of the break increases. The ILR has coverage rates often above the nominal level and an average length significantly longer than the OLS-CR method when the magnitude of the shift is small. The simulation study in Casini and Perron (2017a) suggests that the HDR method based on the continuous record asymptotics strikes the best balance between accurate coverage probabilities and length of the confidence sets. In this Monte Carlo study we shall see that the confidence sets derived from the GL inference display, on average, coverage rates that are higher than those from Bai’s (1997) method and close to the nominal level. This implies that the GL inference is comparable in terms of coverage probability with the other methods and the average length of the confidence sets is shorter than that from Elliott and Müller’s (2007) approach, though it tends to be larger for medium-sized breaks that are close to mid-sample. The HDR-based method of Casini and Perron (2017a) is not available for models with multiple breaks or models with trending regressors. For the models considered, it provides good coverage rates and its average lengths when breaks are large are equivalent to those from the GL method.

We consider the same DGPs as in the previous sub-section. When the errors are uncorrelated (i.e., M2-M3) we simply estimate variances rather than long-run variances. The least-squares estimation method is employed with a trimming parameter $\epsilon = 0.15$ and we use the required degrees of freedom adjustment for the statistic $\hat{U}_T$ of Elliott and Müller (2007). To construct the OLS-CR method, we follow the steps
outlined in Casini and Perron (2017a). To implement Bai’s (1997) method we use the usual steps described in Bai (1997) and Bai and Perron (1998). We implement the GL estimator using a least-absolute loss with the long-span prior and thus we actually use the results of Theorem 4.3.2. For model M1, the estimate of the long-run variance is given by a heteroskedasticity and autocorrelation (HAC) estimator where we use for all methods Andrews and Monahan’s (1992) AR(1) pre-whitened two-stage procedure to select the bandwidth with a quadratic spectral kernel. We consider the version $\hat{U}_T(T_m)_{eq}$ proposed by Elliott and Müller (2007) that allows for heterogeneity across regimes for all models; using the restricted version when applicable leads to similar results. Finally, the last row of each panel includes the rejection probability of a 5%-level sup-Wald test using the asymptotic critical value of Andrews (1993); it serves as a statistical measure about the magnitude of the break.

Overall, the results in Table 4.4-4.6 confirm previous findings on the performance of existing methods. Bai’s (1997) method has a coverage rate below the nominal level when the size of the break is small. For example, in model M2, with $\lambda_0 = 0.5$ and $\delta^0 = 0.8$ (cf. Table 4.5, top panel) Bai’s (1997) method has a coverage probability below 82% even though the Sup-Wald test rejects roughly for 70% of the samples. When the size of the break is even smaller, Bai’s (1997) method systematically fails to cover the true break date with correct probability. In contrast, the method of Elliott and Müller (2007) yields very accurate empirical coverage rates. However, the average length of the confidence intervals from the latter method is systematically much larger than those from all other methods across all DGPs, break sizes and break locations. For large break sizes Bai’s (1997) delivers good coverage rates and the shortest average length among all methods. The ILR method displays in general the largest length for small breaks. For medium-large break sizes the ILR method
has similar features to the GL and OLS-CR method.

Turning to the GL method, it displays good coverage rates across different break magnitudes and tends to have the shortest lengths among all methods for small breaks. We note that its average lengths tend to be somewhat large in model M1 and for medium breaks more generally. For example, in model M3 with $\lambda_0 = 0.5$ and $\delta^0 = 0.8$, the average length from the GL method is 74.61 whereas it is 57.23 from the OLS-CR method. However, in model M2 and M3, as the break magnitude increases the average lengths from the GL method become very close to those from the OLS-CR method and thus the GL method strikes a good balance between approximate coverage probability and average lengths.

4.7 Conclusions

We developed large-sample results for a class of Generalized Laplace estimators in multiple change-points models. As far as implementation is concerned, the GL class implicitly uses the least-squares method of Bai and Perron (1998) whereas inference about the break date is different as it relies on large-sample results for the GL estimators. We showed that the GL estimate of the break date is in general more accurate in small samples than the least-squares one, a feature that can be reconciled with the property that the GL estimator extracts more information from the objective function. Further, inference methods about the change-point dates relying on this class are characterized by more accurate coverage probabilities while yielding comparable average lengths of the confidence sets compared to other methods. The GL estimator is interpreted as a classical (non-Bayesian) estimator and the inference methods proposed in this chapter retain a frequentist interpretation. The limiting distribution of the GL estimator is given by a ratio of integrals over functi-
ons of Gaussian processes and thus essentially equivalent to the limiting distribution of Bayesian change-point estimators. Under the squared loss function, we presented a dichotomy of the asymptotic distribution of the GL estimator which depends on an input (smoothing) parameter. Under an appropriate choice of the input parameter, the GL estimator can display the same limiting law as the asymptotic distribution of the least-squares or Maximum Likelihood estimator proposed by Bai and Perron (1998) and Yao (1987), respectively.

4.8 Appendix to Chapter 4

Table 4.1: Small-sample accuracy of the estimates of the break point $T^0_t$ for model M1

<table>
<thead>
<tr>
<th>$\delta^0$</th>
<th>OLS</th>
<th>GL-LN</th>
<th>GL-Uni</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_0 = 0.3$</td>
<td>19.60</td>
<td>18.02</td>
<td>17.56</td>
</tr>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>16.63</td>
<td>13.98</td>
<td>13.78</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>19.93</td>
<td>7.27</td>
<td>17.11</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>19.95</td>
<td>7.30</td>
<td>17.10</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>37</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>66</td>
<td>52</td>
<td>61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\delta^0 = 0.4$</th>
<th>OLS</th>
<th>GL-LN</th>
<th>GL-Uni</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.3$</td>
<td>15.99</td>
<td>16.32</td>
<td>14.11</td>
</tr>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>13.50</td>
<td>4.39</td>
<td>11.76</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>17.79</td>
<td>6.89</td>
<td>17.10</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>17.78</td>
<td>6.90</td>
<td>17.10</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>40</td>
<td>47</td>
<td>45</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>60</td>
<td>53</td>
<td>55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\delta^0 = 0.6$</th>
<th>OLS</th>
<th>GL-LN</th>
<th>GL-Uni</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.3$</td>
<td>9.60</td>
<td>12.12</td>
<td>8.74</td>
</tr>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>8.96</td>
<td>4.22</td>
<td>7.88</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>13.05</td>
<td>6.29</td>
<td>11.46</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>13.04</td>
<td>6.29</td>
<td>11.45</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>45</td>
<td>47</td>
<td>45</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>55</td>
<td>53</td>
<td>55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\delta^0 = 1$</th>
<th>OLS</th>
<th>GL-LN</th>
<th>GL-Uni</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.3$</td>
<td>3.49</td>
<td>6.09</td>
<td>3.63</td>
</tr>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>3.55</td>
<td>3.92</td>
<td>3.79</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>6.29</td>
<td>7.50</td>
<td>6.65</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>6.29</td>
<td>7.30</td>
<td>6.56</td>
</tr>
<tr>
<td>$Q_{0.25}$</td>
<td>48</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td>$Q_{0.75}$</td>
<td>52</td>
<td>51</td>
<td>55</td>
</tr>
</tbody>
</table>

The model is $y_t = \delta^0 1_{\{t > |T\lambda_0|\}} + \epsilon_t$, $\epsilon_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$, $T = 100$. The columns refer to Mean Absolute Error (MAE), standard deviation (Std), Root Mean Squared Error (RMSE) and the 25% and 75% empirical quantiles. OLS is the least-squares estimator; GL-LN is the GL estimator under a least-absolute loss function with the long-span prior; GL-Uni is the GL estimator under a least-absolute loss function with a uniform prior. The number of simulations is 3,000.
The probability density of the LS estimator for the model:
\[ y_t = \mu_0 + Z_t \delta_0^1 \{ t > \lfloor T \lambda_0 \rfloor \} + e_t, \quad Z_t = 0.3 Z_{t-1} + u_t - 0.1 u_{t-1}, \quad u_t \sim \text{i.i.d.} N(0, 1), \quad e_t \sim \text{i.i.d.} N(0, 1), \{ u_t \} \text{ independent from } \{ e_t \}, \quad T = 100 \] with \( \delta_0 = 0.3 \) and \( \lambda_0 = 0.25 \) and 0.5 (the left and right panel, respectively). The blue solid (resp., green broken) line is the density of the infeasible (reps., feasible) continuous record asymptotic distribution of CP, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line break is the density of the finite-sample distribution.

Table 4.2: Small-sample accuracy of the estimates of the break point \( T_0^0 \) for model M2

| \( \delta_0 \) | \( \lambda_0 = 0.3 \) | MAE | Std | RMSE | \( Q_{0.25} \) | \( Q_{0.75} \) | MAE | Std | RMSE | \( Q_{0.25} \) | \( Q_{0.75} \) |
|----|----------|---|---|---|---|---|---|---|---|---|
| 0.3 | OLS      | 21.96 | 22.85 | 28.29 | 27 | 69 | 18.21 | 21.46 | 21.45 | 32 | 69 |
|    | GL-LN    | 18.18 | 10.17 | 20.29 | 42 | 53 | 6.47 | 9.16 | 9.16 | 46 | 54 |
|    | GL-Uni   | 20.72 | 21.05 | 26.46 | 28 | 64 | 16.04 | 19.54 | 19.56 | 35 | 65 |
| 0.4 | OLS      | 19.74 | 21.96 | 26.29 | 26 | 64 | 16.36 | 20.04 | 20.02 | 35 | 65 |
|    | GL-LN    | 16.77 | 10.42 | 19.44 | 40 | 52 | 6.54 | 9.03 | 9.29 | 46 | 54 |
|    | GL-Uni   | 18.75 | 20.67 | 24.98 | 27 | 61 | 14.93 | 18.44 | 18.43 | 37 | 62 |
| 0.6 | OLS      | 14.32 | 19.41 | 21.39 | 26 | 49 | 12.73 | 16.88 | 16.88 | 40 | 60 |
|    | GL-LN    | 13.56 | 10.39 | 16.71 | 35 | 49 | 6.19 | 8.77 | 8.77 | 46 | 54 |
|    | GL-Uni   | 13.57 | 18.01 | 19.99 | 26 | 45 | 12.02 | 16.04 | 16.01 | 40 | 58 |
| 1   | OLS      | 6.91  | 12.07 | 12.34 | 27 | 34 | 6.77 | 10.73 | 10.72 | 47 | 54 |
|    | GL-LN    | 8.24  | 7.92  | 11.10 | 34 | 40 | 4.34 | 6.58 | 6.60 | 47 | 52 |
|    | GL-Uni   | 6.84  | 11.85 | 12.05 | 27 | 33 | 6.63 | 10.44 | 10.44 | 46 | 53 |

The model is \( y_t = \delta_0^1 \{ t > \lfloor T \lambda_0 \rfloor \} + e_t, \quad e_t \sim \text{i.i.d.} N(0, 1), \quad T = 100 \). The notes of Table 4.4 apply.
Figure 4.2: The probability density of the LS estimator with $\delta^0 = 1.5$

The comments in Figure 4.1 apply but with a break magnitude $\delta^0 = 1.5$.

Table 4.3: Small-sample accuracy of the estimates of the break point $T^0_b$ for model M3

<table>
<thead>
<tr>
<th>$\delta^0$</th>
<th>MAE</th>
<th>Std</th>
<th>RMSE</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.75}$</th>
<th>MAE</th>
<th>Std</th>
<th>RMSE</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.75}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda^0 = 0.3$</td>
<td>OLS</td>
<td>12.99</td>
<td>17.77</td>
<td>19.38</td>
<td>27</td>
<td>45</td>
<td>16.55</td>
<td>20.11</td>
<td>20.09</td>
<td>34</td>
</tr>
<tr>
<td>GL-LN</td>
<td>13.75</td>
<td>17.22</td>
<td>19.89</td>
<td>28</td>
<td>50</td>
<td>15.79</td>
<td>19.16</td>
<td>19.15</td>
<td>36</td>
<td>66</td>
</tr>
<tr>
<td>GL-Uni</td>
<td>12.75</td>
<td>14.93</td>
<td>18.09</td>
<td>29</td>
<td>48</td>
<td>13.74</td>
<td>17.11</td>
<td>17.51</td>
<td>37</td>
<td>61</td>
</tr>
<tr>
<td>$\lambda^0 = 0.5$</td>
<td>OLS</td>
<td>10.30</td>
<td>15.76</td>
<td>16.98</td>
<td>27</td>
<td>39</td>
<td>13.56</td>
<td>17.51</td>
<td>17.51</td>
<td>38</td>
</tr>
<tr>
<td>GL-LN</td>
<td>10.21</td>
<td>14.78</td>
<td>16.03</td>
<td>28</td>
<td>41</td>
<td>13.76</td>
<td>17.54</td>
<td>17.52</td>
<td>38</td>
<td>61</td>
</tr>
<tr>
<td>GL-Uni</td>
<td>9.41</td>
<td>12.54</td>
<td>14.71</td>
<td>29</td>
<td>41</td>
<td>12.13</td>
<td>15.80</td>
<td>15.79</td>
<td>40</td>
<td>59</td>
</tr>
<tr>
<td>$\delta^0 = 0.4$</td>
<td>OLS</td>
<td>9.24</td>
<td>14.64</td>
<td>15.29</td>
<td>27</td>
<td>37</td>
<td>8.85</td>
<td>13.95</td>
<td>13.95</td>
<td>45</td>
</tr>
<tr>
<td>GL-LN</td>
<td>11.64</td>
<td>17.31</td>
<td>18.42</td>
<td>27</td>
<td>42</td>
<td>9.72</td>
<td>12.83</td>
<td>12.83</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>GL-Uni</td>
<td>10.80</td>
<td>15.69</td>
<td>16.91</td>
<td>27</td>
<td>41</td>
<td>8.90</td>
<td>13.04</td>
<td>13.03</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>$\delta^0 = 0.6$</td>
<td>OLS</td>
<td>3.45</td>
<td>6.40</td>
<td>6.44</td>
<td>29</td>
<td>32</td>
<td>3.55</td>
<td>6.16</td>
<td>6.15</td>
<td>48</td>
</tr>
<tr>
<td>GL-LN</td>
<td>4.10</td>
<td>7.64</td>
<td>7.64</td>
<td>28</td>
<td>31</td>
<td>3.92</td>
<td>6.89</td>
<td>6.09</td>
<td>48</td>
<td>51</td>
</tr>
<tr>
<td>GL-Uni</td>
<td>3.87</td>
<td>7.15</td>
<td>7.71</td>
<td>28</td>
<td>31</td>
<td>3.79</td>
<td>6.58</td>
<td>6.50</td>
<td>48</td>
<td>51</td>
</tr>
</tbody>
</table>

The model is $y_t = \delta^0 1_{\{t > T^0 \lambda^0 \}} + \alpha^0 y_{t-1} + \epsilon_t$, $\epsilon_t \sim \text{i.i.d.} \mathcal{N}(0, 0.5)$, $\alpha^0 = 0.6$, $T = 100$. The notes of Table 4.1 apply.
Table 4.4: Small-sample coverage rates and lengths of the confidence sets for model M1

<table>
<thead>
<tr>
<th>$\lambda_0$</th>
<th>OLS-CR</th>
<th>Bai (1997)</th>
<th>$\bar{U}<em>T(T</em>{ln})$</th>
<th>GL-LN</th>
<th>sup-W</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0 = 0.4$</td>
<td>0.922</td>
<td>0.812</td>
<td>0.959</td>
<td>0.942</td>
<td>0.384</td>
</tr>
<tr>
<td>77.32</td>
<td>0.933</td>
<td>58.12</td>
<td>0.862</td>
<td>76.14</td>
<td>59.32</td>
</tr>
<tr>
<td>49.46</td>
<td>0.946</td>
<td>28.75</td>
<td>0.928</td>
<td>41.68</td>
<td>0.958</td>
</tr>
<tr>
<td>22.51</td>
<td>0.950</td>
<td>13.78</td>
<td>0.950</td>
<td>14.44</td>
<td>51.72</td>
</tr>
<tr>
<td>14.28</td>
<td>0.950</td>
<td>8.16</td>
<td>0.950</td>
<td>7.15</td>
<td>0.930</td>
</tr>
<tr>
<td>$\delta_0 = 0.8$</td>
<td>0.934</td>
<td>0.862</td>
<td>0.950</td>
<td>0.948</td>
<td>0.916</td>
</tr>
<tr>
<td>49.46</td>
<td>0.946</td>
<td>28.75</td>
<td>0.928</td>
<td>41.68</td>
<td>0.958</td>
</tr>
<tr>
<td>22.51</td>
<td>0.950</td>
<td>13.78</td>
<td>0.950</td>
<td>14.44</td>
<td>51.72</td>
</tr>
<tr>
<td>14.28</td>
<td>0.950</td>
<td>8.16</td>
<td>0.950</td>
<td>7.15</td>
<td>0.930</td>
</tr>
<tr>
<td>$\delta_0 = 1.2$</td>
<td>0.946</td>
<td>0.950</td>
<td>0.976</td>
<td>0.948</td>
<td>1.000</td>
</tr>
<tr>
<td>21.78</td>
<td>0.950</td>
<td>14.44</td>
<td>0.976</td>
<td>49.46</td>
<td>1.000</td>
</tr>
<tr>
<td>17.78</td>
<td>0.950</td>
<td>7.15</td>
<td>0.976</td>
<td>27.10</td>
<td>1.000</td>
</tr>
<tr>
<td>$\delta_0 = 1.6$</td>
<td>0.938</td>
<td>0.950</td>
<td>0.976</td>
<td>0.958</td>
<td>0.992</td>
</tr>
<tr>
<td>10.48</td>
<td>0.950</td>
<td>14.79</td>
<td>0.976</td>
<td>51.72</td>
<td>1.000</td>
</tr>
<tr>
<td>8.16</td>
<td>0.950</td>
<td>7.15</td>
<td>0.976</td>
<td>27.10</td>
<td>1.000</td>
</tr>
<tr>
<td>5.45</td>
<td>0.950</td>
<td>7.15</td>
<td>0.976</td>
<td>27.10</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The model is $y_t = \delta_0 1_{\{t > T_{\lambda_0}\}} + e_t$, $e_t \sim i.i.d. N(0, 1)$, $T = 100$. Cov. and Lgth. refer to the coverage probability and the average length of the confidence set (i.e., the average number of dates in the confidence set). sup-W refers to the rejection probability of the sup-Wald test using a 5% asymptotic critical value. The number of simulations is 3,000.

Table 4.5: Small-sample coverage rates and lengths of the confidence sets for model M2

<table>
<thead>
<tr>
<th>$\lambda_0$</th>
<th>OLS-CR</th>
<th>Bai (1997)</th>
<th>$\bar{U}<em>T(T</em>{ln})$</th>
<th>GL-LN</th>
<th>sup-W</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0 = 0.4$</td>
<td>0.922</td>
<td>0.812</td>
<td>0.959</td>
<td>0.942</td>
<td>0.384</td>
</tr>
<tr>
<td>77.32</td>
<td>0.933</td>
<td>58.12</td>
<td>0.862</td>
<td>76.14</td>
<td>59.32</td>
</tr>
<tr>
<td>49.46</td>
<td>0.946</td>
<td>28.75</td>
<td>0.928</td>
<td>41.68</td>
<td>0.958</td>
</tr>
<tr>
<td>22.51</td>
<td>0.950</td>
<td>13.78</td>
<td>0.950</td>
<td>14.44</td>
<td>51.72</td>
</tr>
<tr>
<td>14.28</td>
<td>0.950</td>
<td>8.16</td>
<td>0.950</td>
<td>7.15</td>
<td>0.930</td>
</tr>
<tr>
<td>$\delta_0 = 0.8$</td>
<td>0.934</td>
<td>0.862</td>
<td>0.950</td>
<td>0.948</td>
<td>0.916</td>
</tr>
<tr>
<td>49.46</td>
<td>0.946</td>
<td>28.75</td>
<td>0.928</td>
<td>41.68</td>
<td>0.958</td>
</tr>
<tr>
<td>22.51</td>
<td>0.950</td>
<td>13.78</td>
<td>0.950</td>
<td>14.44</td>
<td>51.72</td>
</tr>
<tr>
<td>14.28</td>
<td>0.950</td>
<td>8.16</td>
<td>0.950</td>
<td>7.15</td>
<td>0.930</td>
</tr>
<tr>
<td>$\delta_0 = 1.2$</td>
<td>0.946</td>
<td>0.950</td>
<td>0.976</td>
<td>0.948</td>
<td>1.000</td>
</tr>
<tr>
<td>21.78</td>
<td>0.950</td>
<td>14.44</td>
<td>0.976</td>
<td>49.46</td>
<td>1.000</td>
</tr>
<tr>
<td>17.78</td>
<td>0.950</td>
<td>7.15</td>
<td>0.976</td>
<td>27.10</td>
<td>1.000</td>
</tr>
<tr>
<td>$\delta_0 = 1.6$</td>
<td>0.938</td>
<td>0.950</td>
<td>0.976</td>
<td>0.958</td>
<td>0.992</td>
</tr>
<tr>
<td>10.48</td>
<td>0.950</td>
<td>14.79</td>
<td>0.976</td>
<td>51.72</td>
<td>1.000</td>
</tr>
<tr>
<td>8.16</td>
<td>0.950</td>
<td>7.15</td>
<td>0.976</td>
<td>27.10</td>
<td>1.000</td>
</tr>
<tr>
<td>5.45</td>
<td>0.950</td>
<td>7.15</td>
<td>0.976</td>
<td>27.10</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The model is $y_t = \delta_0 1_{\{t > T_{\lambda_0}\}} + e_t$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d. N(0, 1)$, $T = 100$. The notes of Table 4.4 apply.
Table 4.6: Small-sample coverage rates and lengths of the confidence sets for model M3

<table>
<thead>
<tr>
<th></th>
<th>$\delta^U = 0.4$</th>
<th>$\delta^U = 0.8$</th>
<th>$\delta^U = 1.2$</th>
<th>$\delta^U = 1.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS-CR</td>
<td>0.954</td>
<td>80.29</td>
<td>0.952</td>
<td>57.23</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.781</td>
<td>55.85</td>
<td>0.902</td>
<td>26.23</td>
</tr>
<tr>
<td>$\bar{U}_T (T_m) \neq $</td>
<td>0.958</td>
<td>81.28</td>
<td>0.957</td>
<td>55.34</td>
</tr>
<tr>
<td>ILR</td>
<td>0.946</td>
<td>78.04</td>
<td>0.959</td>
<td>45.98</td>
</tr>
<tr>
<td>GL-LN</td>
<td>0.970</td>
<td>87.70</td>
<td>0.967</td>
<td>74.61</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.407</td>
<td>0.931</td>
<td>0.953</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\delta^U = 0.3$</th>
<th>$\delta^U = 0.8$</th>
<th>$\delta^U = 1.2$</th>
<th>$\delta^U = 1.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS-CR</td>
<td>0.968</td>
<td>83.69</td>
<td>0.951</td>
<td>54.13</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.795</td>
<td>64.06</td>
<td>0.896</td>
<td>26.33</td>
</tr>
<tr>
<td>$\bar{U}_T (T_m) \neq $</td>
<td>0.960</td>
<td>86.42</td>
<td>0.953</td>
<td>59.13</td>
</tr>
<tr>
<td>ILR</td>
<td>0.956</td>
<td>80.30</td>
<td>0.944</td>
<td>47.35</td>
</tr>
<tr>
<td>GL-LN</td>
<td>0.976</td>
<td>85.96</td>
<td>0.954</td>
<td>72.21</td>
</tr>
<tr>
<td>sup-W</td>
<td>0.232</td>
<td>0.884</td>
<td>0.999</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The model is $y_t = \delta^U 1_{(t \geq \lceil T \lambda_0 \rceil)} + \alpha^U y_{t-1} + \epsilon_t, \epsilon_t \sim i.i.d. \mathcal{N}(0, 0.5), \alpha^U = 0.6, T = 100$. The notes of Table 4.4 apply.
Appendix A

Supplement to Chapter 1: Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models

A.1 Implementation of $\hat{J}_T$ HAC in GMM, IV and Structural Change Models

This section reviews HAC estimation in GMM, IV and Nonlinear LS problems.

A.1.1 GMM

We begin with the GMM setup [see Hansen (1982)]. For a $k$-vector $\beta^*$ of unknown parameters, we have the moment condition $E m_t (\beta^*) = 0$ and $m_t (\beta)$ is a $p$-vector of functions of the data and parameters, $k \geq p$. The GMM estimator $\hat{\beta}$ is the solution of $\min_{\beta} m_T (\beta) \hat{W}_T m_T (\beta)$, where $m_T (\beta) = \sum_{t=1}^{T} m_t (\beta) / T$ is the vector of sample moments $m_t (\beta)$ and $\hat{W}_T$ is (possibly) random, symmetric weighting matrix. The asymptotic covariance matrix of $\hat{\beta}$ is given by

$$H_T = (L_T' W_T L_T)^{-1} L_T W_T J_T W_T L_T (L_T' W_T L_T)^{-1},$$
where \( L_T = \sum_{t=1}^{T} \mathbb{E} m_{t, \beta} (\beta_*) / T \) and \( m_{t, \beta} (\beta) \) is the \( p \times k \) matrix of partial derivatives of \( m_t (\beta) \), \( W_T \) is nonrandom matrix such that \( \hat{W}_T \overset{p}{\to} W_T \), and

\[
J_T = \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E} \left( m_t (\beta_*) m_s (\beta_*)' \right) / T.
\]

Consistent estimation of \( H_T \) boils down to consistent estimation of \( J_T \) since estimation of \( L_T \) and \( W_T \) is straightforward. \( \hat{W}_T \) is a natural estimator of \( W_T \) while under regularity conditions \( L_T - \sum_{t=1}^{T} m_{t, \beta} (\hat{\beta}) / T \overset{p}{\to} 0 \). In place of classical HAC estimators we now estimate \( J_T \) by

\[
\hat{J}_T = \sum_{k=-T+1}^{T-1} K_1 (b_{1, T} k) \hat{\Gamma} (k), \quad \text{where} \quad \hat{\Gamma} (k) \triangleq \frac{n_T}{T - n_T} \sum_{r=0}^{[(T-n_T)/n_T]} \hat{c}_{T} (rn_T / T, k),
\]

(A.1.1)

where

\[
\hat{c}_{T} (rn_T / T, k) \triangleq \begin{cases} (Tb_{2, T})^{-1} \sum_{s=k+1}^{T} K_2 \left( \frac{((r+1)n_T-(s+k)/2) / T}{b_{2, T}} \right) \hat{m}_s \hat{m}_s' - k, & k \geq 0 \\ (Tb_{2, T})^{-1} \sum_{s=-k+1}^{T} K_2 \left( \frac{((r+1)n_T-(s-k)/2) / T}{b_{2, T}} \right) \hat{m}_s \hat{m}_s' - k, & k < 0 \end{cases}
\]

and \( \hat{m}_s = m_s (\hat{\beta}) \). We can implement \( \hat{J}_T \) with data-dependent methods for selecting \( b_{1, T} \) and \( b_{2, T} \), and choose \( K_1 \) and \( K_2 \) on the basis of the optimality results of Section 1.4. For the kernel \( K_1 \) one can use the QS kernel while for the kernel \( K_2 \) one can choose \( K_2 = 6x (1 - x) \) for \( 0 \leq x \leq 1 \) and 0 otherwise as suggested in Section 1.4.

From the results in Section 1.5, we know that

\[
\hat{b}_{1, T} = 0.6828 \left( \hat{\phi} (2) T \hat{b}_{2, T} \right)^{-1/5}
\]

\[
\hat{b}_{2, T} (u_r) = 1.7781 \left( D_{1, \theta} (u_r) \right)^{-1/5} \left( \hat{D}_2 (u_r) \right)^{1/5} T^{-1/5}, \quad u_r = rn_T / T,
\]
where

\[ D_{1,\theta}(u) = \frac{1}{\pi} \left( 1 + (1.8 (-4\pi \sin (4\pi u))) \right) \left( 1.8 (-4\pi \sin (4\pi u)) \right) \]

\[ + \frac{1}{\pi} \left( 1 + (1.8 (-4\pi \sin (4\pi u))) \right) \left( 1.8 (-16\pi^2 \cos (4\pi u)) \right) , \]

and the expressions for \( \hat{\phi}(2) \) and \( \hat{D}_2(u_r) \) are given in the same section.

**A.1.2 IV**

Consider the linear model

\[ y_t = x_t' \beta_0 + e_t \ (t = 1, \ldots, T), \]

where \( \beta_0 \in \Theta \subset \mathbb{R}^p \), \( y_t \) is an observation on the dependent variable, \( x_t \) is a \( p \)-vector of regressors and \( e_t \) is an unobservable disturbance which may exhibit heteroskedasticity and/or autocorrelation. Suppose the regressor is endogenous: \( E(x_t e_t) \neq 0 \). The IV estimator \( \hat{\beta}_{IV} \) is given by

\[ \hat{\beta}_{IV} = \left( Z'X \right)^{-1} Z'Y, \]

where \( Y = (y_1, \ldots, y_T)' \), \( X = (x_1, \ldots, x_T)' \) and \( Z = (z_1, \ldots, z_T)' \) where \( z_t \) is a \( p \)-vector of instruments. The asymptotic variance of the IV estimator is given by

\[ \text{Var} \left( \sqrt{T} \left( \hat{\beta}_{IV} - \beta_0 \right) \right) = Q_{XX}^{-1} J_T Q_{XX}^{-1}, \]

where \( Q_{XX} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} z_t x_t' \) and \( J_T = T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} E \left( e_t z_t (e_t z_t)' \right) \). A natural estimator of \( Q_{XX} \) is \( T^{-1} \sum_{t=1}^{T} z_t x_t' \). \( J_T \) can be estimated by the \( \tilde{J}_T \) as given in (A.1.1) where \( \tilde{m}_s \) is replaced by \( \tilde{e}_t z_t \) where \( \tilde{e}_t = y_t - x_t' \hat{\beta}_{IV} \). The asymptotic variance of the two-stages least-squares (2SLS) estimator is more complex but it still requires the same estimate \( \tilde{J}_T \) of \( J_T \) so that the same method can be applied.

**A.2 Mathematical Appendix**

In some of the proofs below \( \overline{\beta} \) is understood to be on the line segment joining \( \hat{\beta} \) and \( \beta_0 \). We discard the degrees of freedom adjustment \( T/(T - p) \) from the derivations since asymptotically it does not play any role. Similarly, we use \( T/n_T \) in place of \( (T - n_T)/n_T \) in the expression for \( \hat{\Gamma}(k) \). Some parts of the proofs of the results in
Section 1.3-1.5 follow the arguments in Andrews (1991).

A.2.1

Proof of the Results of Section 1.2.1

A.2.1.1 Proof of Theorem 1.2.1

We adapt the arguments in the proof of Theorem 2.2 in Dahlhaus (1996) to our context. Suppose $T_u \notin T$. Without loss of generality, assume $T^0_{j-1} < T_u < T^0_j$ for some $1 \leq j \leq m + 1$. Then,

$$f_{j,T}(u, \omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \int_{-\pi}^{\pi} \exp(i\eta s) A_{j,[Tu-s/2],T}(\eta) A_{j,[Tu+s/2],T}(\eta) d\eta,$$

and

$$f_j(u, \omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \int_{-\pi}^{\pi} \exp(i\eta s) A_j(u, \eta) A_j(u, -\eta) d\mu.$$

We have

$$\int_{-\pi}^{\pi} |f_T(u, \omega) - f(u, \omega)|^2 d\omega = \int_{-\pi}^{\pi} \left| \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \int_{-\pi}^{\pi} \exp(i\eta s) \left( A^0_{j,[Tu-s/2],T}(\eta) A^0_{j,[Tu+s/2],T}(\eta) - A_j(u, \eta) A_j(u, -\eta) \right) d\eta \right|^2 d\omega$$

$$= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |c_{s,j}|^2 + o(1),$$

where $c_{s,j} = \int_{-\pi}^{\pi} \exp(i\eta s) G_j(s/2T, \eta) d\eta$ and

$$G_j\left(\frac{s}{2T}, \eta\right) = A_j\left(u - \frac{s}{2T}, \eta\right) A_j\left(u + \frac{s}{2T}, -\eta\right) - A_j(u, \eta) A_j(u, -\eta),$$

with $A_1(u, \mu) = A_1(0, \mu)$ for $u < 0$ and $A_{m+1}(u, \mu) = A_{m+1}(1, \mu)$ for $u > 1$. By well-known results on Fourier coefficients [cf. Bary (1964), Chapter 2.3], $|c_{s,j}| \leq Cs^{-\vartheta}$ and thus $\sum_{s=n}^{\infty} |c_{s,j}|^2 = O\left(n^{1-2\vartheta}\right)$. Let $\Delta_s(\omega) = \sum_{r=0}^{s-1} \exp(-i\omega r)$. Applying
summation by parts yields

\[
\sum_{s=0}^{n-1} |c_{s,j}|^2
= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{s=0}^{n-1} \exp(-i(\omega - \eta) s) G_j \left( \frac{s}{2T}, \omega \right) G_j \left( \frac{s}{2T}, \eta \right) d\omega d\eta \\
\leq \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| -\sum_{s=0}^{n-1} \left[ G_j \left( \frac{s}{2T}, \omega \right) G_j \left( \frac{s}{2T}, \eta \right) - G_j \left( \frac{s-1}{2T}, \omega \right) G_j \left( \frac{s-1}{2T}, \eta \right) \right] \Delta_s (\eta - \omega) \right| d\omega d\eta \\
+ G_j \left( \frac{n-1}{2T}, \omega \right) G_j \left( \frac{n-1}{2T}, \eta \right) \Delta_n (\eta - \omega) \left| d\omega d\eta \right|
= O \left( \frac{n \ln n}{T^\vartheta} \right).
\]

A similar bound holds for \( \sum_{s=-n}^{\infty} |c_{-s,j}|^2 \). The result for \( Tu \notin T \) follows choosing an \( n \) appropriately. Next, suppose \( Tu \in T \) and \( u = T_j^0/T \). Then, we have

\[
f_{j,T} (u, \omega) \equiv \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \int_{-\pi}^{\pi} \exp(i\eta s/2) A_{j,Tu-3|s|/2,T} (\eta) A_{j,Tu-|s|/2,T} (\eta) d\eta
\]

and

\[
f_j (u, \omega) \equiv \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s/2) \int_{-\pi}^{\pi} \exp(i\eta s) A_j (u, \eta) A_j (u, \eta) d\eta.
\]
Proceeding as above,

\[
\int_{-\pi}^{\pi} |f_T(u, \omega) - f(u, \omega)|^2 d\omega
\]

\[
= \int_{-\pi}^{\pi} \left| \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \left[ \int_{-\pi}^{\pi} \exp(\i\eta s/2) A_{j,\lfloor uT-3|s|/2\rfloor, T}(\eta) A_{j,\lfloor uT-|s|/2\rfloor, T}(\eta) d\eta \right. \right. \\
\left. \left. \exp(i\eta s) A_j(u, \eta) A_j(u, \eta) d\eta \right| d\omega
\]

\[
= \int_{-\pi}^{\pi} \left| \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp(-i\omega s) \right|^2 d\omega \\
\left[ \int_{-\pi}^{\pi} \exp(\i\eta s/2) \left( A_{j,\lfloor uT-3|s|/2\rfloor, T}(\eta) A_{j,\lfloor uT-|s|/2\rfloor, T}(\eta) - A_j(u, \eta) A_j(u, \eta) \right) d\eta \right]^2 d\omega
\]

\[
= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |c_{s,j}|^2 + o(1),
\]

with \( c_{s,j} = \int_{-\pi}^{\pi} \exp(\i\eta s/2) G_j(s/2T, \eta) d\eta \) and

\[
G_j\left(\frac{s}{2T}, \eta\right) = A_j\left(u - \frac{3|s|}{2T}, \eta\right) A_j\left(\frac{|s|}{2T}, -\eta\right) - A_j(u, \eta) A_j(u, -\eta).
\]

Using the definition of \( \Delta_s(\omega) \) and the above-mentioned properties of \( c_{s,j} \) which continue to hold, summation by parts and the continuity of \( A_j(u, -\eta) \) then imply

\[
\sum_{s=0}^{n-1} |c_{s,j}|^2 = O(n \ln n / T^g).
\]

Since the same bound applies to \( \sum_{s=n}^{\infty} |c_{s,j}|^2 \), we can choose an appropriate \( n \) to yield the result for \( Tu \in \mathcal{T} \).

**A.2.2 Proof of the Results of Section 1.3**

**A.2.2.1 Proof of Proposition 1.3.1**

The results about the bias and variance for the scalar case follow from Dahlhaus (2012). Using a standard bias-variance argument, we have

\[
\tilde{c}_T(u_0, k) - c(u_0, k) =
\]
\( o_p(1) \). If \( Tb_{2,T}^q \rightarrow \eta \in (0, \infty) \), the asymptotic MSE of \( \tilde{c}_T(u_0, k) \) is given by

\[
\lim_{T \to \infty} \text{MSE} (\tilde{c}_T(u_0, k)) = \frac{\eta}{4} \left( \int_0^1 x^2 K_2(x) \, dx \right)^2 \left[ \frac{\partial^2}{\partial^2 u} \text{vec} (c(u_0, k)) \right]' W \left[ \frac{\partial^2}{\partial^2 u} \text{vec} (c(u_0, k)) \right] + \int_0^1 K_2^2(x) \, dx \, \text{tr} W \sum_{l=-\infty}^\infty \text{vec} (c(u_0, l)) \\
\times \left[ \text{vec} (c(u_0, l))' + \text{vec} (c(u_0, l+2k))' \right].
\]

The latter suggests that if \( Tb_{2,T}^q \rightarrow \eta \in (0, \infty) \), then \( \tilde{c}_T(u_0, k) - c(u_0, k) = O_p \left( \sqrt{Tb_{2,T}} \right) \) for all \( u_0 \in (0, 1) \). \( \square \)

**A.2.2.2 Proof of Theorem 1.3.1**

We first prove the result for the scalar case and then extend it to the vector case.

**Lemma A.2.1.** Suppose \( p = 1 \), \( K_1(\cdot) \in K_1 \), Assumption 1.2 holds with \( \vartheta = 1 \), \( b_{1,T}, b_{2,T} \rightarrow 0 \), \( n_T \rightarrow \infty \), \( n_T/T \rightarrow 0 \) and \( 1/Tb_{1,T}b_{2,T} \rightarrow 0 \). We have:

(i) \( \lim_{T \to \infty} Tb_{1,T}b_{2,T} \text{Var} \left( J_T \right) = 4\pi^2 \int K_1^2(y) \, dy \int_0^1 K_2^2(x) \, dx \left( \int_0^1 f(u, 0) \, du \right)^2. \)

(ii) If \( 1/Tb_{1,T}b_{2,T} \rightarrow 0 \), \( n_T/Tb_{1,T} \rightarrow 0 \) and \( b_{2,T}^2/b_{1,T}^2 \rightarrow 0 \) for some \( q \in [0, \infty) \) for which \( K_{1,q} \), \( \int_0^1 f^{(q)}(u, 0) \, du \in [0, \infty) \), then

\[
\lim_{T \to \infty} b_{1,T}^{-q} \left[ \mathbb{E} \left( J_T - J_T \right) \right] = -2\pi K_{1,q} \int_0^1 f^{(q)}(u, 0) \, du.
\]

(iii) If \( n_T/Tb_{1,T}^q \rightarrow 0 \), \( b_{2,T}^2/b_{1,T}^2 \rightarrow 0 \) and \( Tb_{1,T}^2b_{2,T} \rightarrow \gamma \in (0, \infty) \) for some \( q \in [0, \infty) \) for which \( K_{1,q} \), \( \int_0^1 f^{(q)}(u, 0) \, du \in [0, \infty) \), then

\[
\lim_{T \to \infty} \text{MSE} \left( Tb_{1,T}b_{2,T}, J_T, 1 \right) = 4\pi^2 \left[ K_{1,q}^2 \left( \int_0^1 f^{(q)}(u, 0) \, du \right)^2 + \int K_1^2(y) \, dy \int K_2^2(x) \, dx \left( \int_0^1 f(u, 0) \, du \right)^2 \right].
\]
Proof of Lemma A.2.1. We begin with part (i). Note that for any fixed non-negative \( \tau_1, \tau_2 \in \mathbb{R} \),

\[
\text{Cov} \left( V_s V_{s-\tau_1}, V_l V_{l-\tau_2} \right) = \mathbb{E} \left[ \left( V_s V_{s-\tau_1} - \mathbb{E} (V_s V_{s-\tau_1}) \right) \left( V_l V_{l-\tau_2} - \mathbb{E} (V_l V_{l-\tau_2}) \right) \right]
\]

\[
= \mathbb{E} (V_s V_{s-\tau_1} V_l V_{l-\tau_2}) - \Gamma_{s/T}(\tau_1) \Gamma_{l/T}(\tau_2)
- \Gamma_{s/T}(\tau_1) \Gamma_{l/T}(\tau_2) - \Gamma_{l/T}(l-s) \Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)
- \Gamma_{(l-\tau_2)/T}(l-s-\tau_2) \Gamma_{l/T}(l-s+\tau_1)
+ \Gamma_{s/T}(\tau_1) \Gamma_{l/T}(\tau_2) + \Gamma_{l/T}(l-s) \Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)
+ \Gamma_{(l-\tau_2)/T}(l-s-\tau_2) \Gamma_{l/T}(l-s+\tau_1)
\]

\[
= \kappa_{V,s}(\tau_1, l-s, l-s-\tau_2) + \Gamma_{l/T}(l-s) \Gamma_{(l-\tau_2)/T}(l-s-\tau_2+\tau_1)
+ \Gamma_{(l-\tau_2)/T}(l-s-\tau_2) \Gamma_{l/T}(l-s+\tau_1)
\]

For large \( T \) we have

\[
\Gamma_{(l-\tau_2)/T}(k) - \Gamma_{l/T}(k) = O_P \left( \frac{|l/T - (l-\tau_2)/T| + T^{-1}}{T} \right) = O_P \left( \frac{\tau_2}{T} \right),
\]

and \( \Gamma_{(s-\tau_1)/T}(k) = \Gamma_{s/T}(k) + O_P(\tau_1/T) \) for all \( k \). Let us apply the changes in variable \( w = l-s \) and \( v = l \), then

\[
\sum_{s=\tau_1+1}^T \sum_{l=\tau_2+1}^T \text{Cov} \left( V_{s/T} V_{(s-\tau_1)/T}, V_{l/T} V_{(l-\tau_2)/T} \right)
\]

\[
= \sum_{s=\tau_1+1}^T \sum_{l=\tau_2+1}^T \left[ \kappa_{V,s}(\tau_1, l-s, l-s-\tau_2) \right]
+ \sum_{v=\tau_2+1}^T \sum_{w=-\tau_2+1}^{T-\tau_2+1} \left[ \Gamma_{v/T}(w) \Gamma_{v/T}(w+\tau_2-\tau_1) + \Gamma_{v/T}(w-\tau_2) \Gamma_{v/T}(w+\tau_1) \right]
+ \sum_{v=\tau_2+1}^T \sum_{w=-\tau_2+1}^{T-\tau_2+1} \left[ \Gamma_{v/T}(w) O_P(\tau_2/T) + O_P(\tau_2/T) \Gamma_{v/T}(w+\tau_1) \right]. \quad (A.2.1)
\]
Since $\kappa_{V,s}$ is absolutely summable, the term involving $\kappa_{V,s}(\tau_1, l-s, l-s-\tau_2)$ vanishes. A bound for the last term in (A.2.1) is

$$\sum_{v=\tau_2+1}^{T} \sum_{w=-T+\tau_2+1}^{T-\tau_2-1} \Gamma_{v/T}(w) O_p(\tau_2/T)$$

$$\leq O_p(\tau_2/T) \sum_{v=\tau_2+1}^{T} \sum_{w=-T+\tau_2+1}^{T-\tau_2-1} \sup_{(v/T)\in[0,1]} \Gamma_{v/T}(w)$$

$$\leq O_p(\tau_2/T) \sum_{v=\tau_2+1}^{T} O_p(1)$$

$$\leq O_p(\tau_2/T) O_p(T) = O_p(1), \quad (A.2.2)$$

where we have used Assumption 1.2-(i). The argument for the term involving $O_p(\tau_2/T) \Gamma_{v/T}(w+\tau_1)$ is analogous. We next evaluate the covariance of $\tilde{c}_T(t/T, k)$. For any $1 \leq t_1, t_2 \leq T$ and (without loss of generality) non-negative integers $\tau_1, \tau_2 \in \mathbb{R}$, apply the following changes in variables $w = l - s$ and $v = l$, so that

$$Tb_{2,T} \text{Cov} [\tilde{c}_T(t_1/T, \tau_1), \tilde{c}_T(t_2/T, \tau_2)]$$

$$= Tb_{2,T} \left( \frac{1}{Tb_{2,T}} \right)^2 \sum_{s=\tau_1+1}^{T} \sum_{v=\tau_2+1}^{T} K_2 \left( \frac{(t_1 - (s + \tau_1/2))/T}{b_{2,T}} \right) K_2 \left( \frac{(t_2 - (v + \tau_2/2))/T}{b_{2,T}} \right)$$

$$\times \text{Cov} (V_sV_{s-\tau_1}, V_lV_{l-\tau_2})$$

$$= \frac{1}{Tb_{2,T}} \sum_{w=-T+\tau_2+1}^{T-\tau_2-1} \sum_{v=\tau_2+1}^{T} K_2 \left( \frac{(t_1 - (v - w + \tau_1/2))/T}{b_{2,T}} \right) K_2 \left( \frac{(t_2 - (v + \tau_2/2))/T}{b_{2,T}} \right)$$

$$\times \left\{ [\Gamma_{v/T}(w) \Gamma_{v/T}(w + \tau_2 - \tau_1) + \Gamma_{v/T}(w - \tau_2) \Gamma_{v/T}(w + \tau_1)] \right\} + A_T,$$
where

\[
A_T \triangleq \frac{1}{T b_{2,T}} \sum_{w=-T+\tau_2+1}^{T-\tau_1+1} \sum_{v=v_2+1}^{T} K_2 \left( \frac{(t_1 - (v - w + \tau_1 / 2)) / T}{b_{2,T}} \right) K_2 \left( \frac{(t_2 - (v + \tau_2 / 2)) / T}{b_{2,T}} \right) \times \left\{ \left[ \Gamma_{v/T} (w) \mathcal{O}_\tau (\tau_2 / T) + \mathcal{O}_\tau (\tau_2 / T) \Gamma_{v/T} (w + \tau_1) \right] \right\}.
\]

Using (A.2.2), we have \( A_T = \mathcal{O}_\tau (1 / T b_{2,T}) \). Then, using the change of variable \( z = v / T b_{2,T} \),

\[
T b_{2,T} \text{ Cov} \left[ \tilde{c}_T \left( \frac{t_1}{T}, \tau_1 \right), \tilde{c}_T \left( \frac{t_2}{T}, \tau_2 \right) \right]
\]

\[
= \frac{1}{T b_{2,T}} \sum_{w=-T+\tau_2+1}^{T-\tau_1+1} \sum_{v=v_2+1}^{T} K_2 \left( \frac{(t_1 - v + w - \tau_1 / 2 + v - v) / T}{b_{2,T}} \right) K_2 \left( \frac{(t_2 - z T b_{2,T} - \tau_2 / 2) / T}{b_{2,T}} \right) \times \left\{ \left[ \Gamma_{z T b_{2,T}} (w) \Gamma_{z T b_{2,T}} (w + \tau_2 - \tau_1) + \Gamma_{z T b_{2,T}} (w - \tau_2) \Gamma_{z T b_{2,T}} (w + \tau_1) \right] \right\} + A_T
\]

\[
= \frac{1}{T b_{2,T}} \sum_{w=-T+\tau_2+1}^{T-\tau_1+1} \sum_{v=v_2+1}^{T} K_2 \left( \frac{(t_1 - v + w - \tau_1 / 2 + v - v) / T}{b_{2,T}} - z \right) K_2 \left( \frac{(t_2 - \tau_2 / 2) / T}{b_{2,T}} - z \right) \times \left\{ \left[ \Gamma_{z T b_{2,T}} (w) \Gamma_{z T b_{2,T}} (w + \tau_2 - \tau_1) + \Gamma_{z T b_{2,T}} (w - \tau_2) \Gamma_{z T b_{2,T}} (w + \tau_1) \right] \right\} + A_T
\]

\[
= \frac{1}{T b_{2,T}} \sum_{w=-T+\tau_2+1}^{T-\tau_1+1} \sum_{v=v_2+1}^{T} K_2 \left( \frac{(t_1 + w - \tau_1 / 2) / T}{b_{2,T}} - z \right) K_2 \left( \frac{(t_2 - \tau_2 / 2) / T}{b_{2,T}} - z \right) \times \left\{ \left[ \Gamma_{z T b_{2,T}} (w) \Gamma_{z T b_{2,T}} (w + \tau_2 - \tau_1) + \Gamma_{z T b_{2,T}} (w - \tau_2) \Gamma_{z T b_{2,T}} (w + \tau_1) \right] \right\} + A_T. \tag{A.2.3}
\]
Thus, with $u = t_1 / T$ and $v = t_2 / T$, the limit of the first term of (A.2.3) is equal to

$$\int_0^1 K_2^2 (x) \, dx \left\{ \sum_{w = -\infty}^{\infty} \left[ \Gamma_u (w) \Gamma_v (w + \tau_2 - \tau_1) + \Gamma_u (w - \tau_2) \Gamma_v (w + \tau_1) \right] \right\}$$

$$= \int_0^1 K_2^2 (x) \, dx \left\{ \sum_{w = -\infty}^{\infty} \left[ \Gamma_u (w) \Gamma_v (w + \tau_2 - \tau_1) + \Gamma_u (-w + \tau_2) \Gamma_v (-w - \tau_1) \right] \right\}.$$

When $\tau_1 = \tau_2 = k$ and $t = t_1 = t_2$, we have

$$Tb_{2, T} \text{Var} (\tilde{c}_T (t/T, k))$$

$$= \int_0^1 K_2 (x)^2 \, dx \left\{ \sum_{w = -\infty}^{\infty} \left[ \Gamma_u (w) \Gamma_u (w) + \Gamma_u (w - k) \Gamma_u (w + k) \right] \right\}$$

$$= \int_0^1 K_2 (x)^2 \, dx \left\{ \sum_{h = -\infty}^{\infty} \left[ \Gamma_u (h) \Gamma_u (h) + \Gamma_u (h) \Gamma_u (h + 2k) \right] \right\},$$

where $u = t / T$ and we have used the change in variable $h = w - k$. Next, we consider $\text{Cov} \left[ \tilde{\Gamma} (\tau_1), \tilde{\Gamma} (\tau_2) \right]$. Note that,

$$Tb_{2, T} \text{Cov} \left[ \tilde{\Gamma} (\tau_1), \tilde{\Gamma} (\tau_2) \right]$$

$$\to \int_0^1 K_2^2 (x) \, dx \int_0^1 \int_0^1 \left\{ \sum_{h = -\infty}^{\infty} \left[ \Gamma_u (h) \Gamma_u (h + \tau_2 - \tau_1) + \Gamma_v (-h + \tau_2) \Gamma_v (-h - \tau_1) \right] \right\} dvdu.$$
The latter can be used to evaluate \( \text{Var} \left[ \sum_{k=-T}^{T-1} K_1 (b_{1,T}k) \tilde{\Gamma} (k) \right] \) as follows

\[
Tb_{1,T}b_{2,T} \text{Var} \left[ \sum_{k=-T}^{T-1} K_1 (b_{1,T}k) \tilde{\Gamma} (k) \right] \\
= 2b_{1,T} \sum_{k=-T}^{T-1} \sum_{j=0}^{T-1} K_1 (b_{1,T}k) K_1 (b_{1,T}j) \\
\times \left( \frac{n_T}{T} \right)^{2 T/n_T} \frac{1}{Tb_{2,T}} \sum_{r=0}^{T} \sum_{b=0}^{T} K_2 \left( \frac{(rn_T + 1) - (s + k/2)}{Tb_{2,T}} \right) K_2 \left( \frac{(bn_T + 1) - (l + j/2)}{Tb_{2,T}} \right) \\
\times \{ \kappa_{V,s} (k, w, w - j) + \sum_{w=-\infty}^{\infty} [\Gamma_s (w) \Gamma_s (w + j - k) + \Gamma_l (-w + j) \Gamma_l (-w - k)] \} + o_P (1).
\]

We first show that the term involving \( \Gamma_l (-w + j) \Gamma_l (-w - k) \) vanishes in the limit.

Using a change in variables \( z_1 = j + k \) and \( z = w - j \), this becomes

\[
2b_{1,T} \sum_{j=0}^{T-1} \sum_{z_1=j}^{T-1+j} K_1 (b_{1,T} (z_1 - j)) K_1 (b_{1,T}j) \left( \frac{n_T}{T} \right)^{2 T/n_T} \frac{1}{Tb_{2,T}} \sum_{r=0}^{T} \sum_{b=0}^{T} K_2 \left( \frac{((rn_T + 1) - (s + (z_1 - j)/2)) / T}{b_{2,T}} \right) \\
\times \left( \frac{(bn_T + 1) - (l + j/2)) / T}{b_{2,T}} \right) \\
\times \{ \sum_{z=-\infty}^{\infty} [\Gamma_{(bn_T+1)/T} (-z) \Gamma_{(bn_T+1)/T} (- (z + z_1))] \} \text{.} \]
Making the change in variable $z_2 = j b_{1,T}$, (A.2.5) can be expressed as,

\[
2 b_{1,T} \sum_{T+1}^{(T-1)/b_{1,T}} \sum_{T-1}^{+z_2/b_{1,T}} K_1 \left( b_{1,T} (z_1 - z_2/b_{1,T}) \right) K_1 \left( z_2 \right) \left( \frac{n_T}{T} \right)^2 \sum_{r=0}^{T-n_T} \sum_{b=0}^{T/n_T} \frac{1}{b_{2,T} T} \sum_{s=-z_1-2z_2/b_{1,T}+1}^{z_2/b_{1,T}+1} \sum_{l=-z_2/b_{1,T}+1}^{z_2/b_{1,T}+1} \frac{K_2 \left( \left( (rn_T + 1) - (s + z_1/2 - z_2/2b_{1,T}) \right) / T \right)}{b_{2,T}}
\]

\[
\times \left\{ \sum_{T-w=0}^{T} \left[ \Gamma_s (w) \Gamma_s (w + u_1) \right] \right\},
\]

which converges to zero because the range of summation over $z_1$ tends to infinity.

Next, let us consider the term of (A.2.4) involving $\Gamma_s (w) \Gamma_s (w + j - k)$. With the changes in variables $u_1 = j - k$ and $u_2 = j$, this becomes

\[
2 b_{1,T} \sum_{u_2=0}^{T-1} \sum_{u_1=u_2}^{T-1+u_2} K_1 \left( b_{1,T} (u_2 - u_1) \right) K_1 \left( b_{1,T} u_2 \right) \left( \frac{n_T}{T} \right)^2 \sum_{r=0}^{T-n_T} \sum_{b=0}^{T/n_T} \frac{1}{b_{2,T} T} \sum_{s=-u_2-u_1+1}^{u_2+1} \sum_{l=-u_2+1}^{u_2+1} \frac{K_2 \left( \left( (bn_T + 1) - (l + u_2/2) \right) / T \right)}{b_{2,T}}
\]

\[
\times \left\{ \sum_{w=-T}^{T} \left[ \Gamma_s (w) \Gamma_s (w + u_1) \right] \right\}.
\]

Apply the change in variable $z = b_{1,T} u_2$ and consider the lattice points $z_n = nb_{1,T}$, where $n = 1, \ldots, T$. As $T \to \infty$, the distance between the lattice points $z_n = nb_{1,T}$ converges to zero and the highest lattice point converges to infinity. Hence, (A.2.6)
can be expressed as,

\[ 2 \sum_{z_n=0}^{b_1 T} \sum_{u_1=zn/b_1 T} K_1 (-b_1 T u_1 + z_n) K_1 (z_n) \]

\[ \times \left( \frac{n_T}{T} \right) \frac{T}{T b_2 T} \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \frac{1}{T b_2 T} \sum_{s=zn/b_1 T - u_1 +1}^{T} \sum_{l=zn/b_1 T + 1}^{T} \]

\[ K_2 \left( \frac{(rn_T + 1) - (s + (z_n/b_1 T - u_1)/2)}{b_2 T} \right) \]

\[ \times K_2 \left( \frac{(bn_T + 1) - (l + z/2b_1 T))}{b_2 T} \right) \left\{ \sum_{u=T}^{T} [\Gamma_s(w) \Gamma_u(w + u_1)] \right\}, \]

and its limit is

\[ 2 \int_0^\infty K_1 (y)^2 dy \int_0^1 K_2 (x)^2 dx \int_0^1 \sum_{u_1=\infty}^{\infty} \sum_{u=-\infty}^{\infty} [\Gamma_u(w) \Gamma_u(w + u_1)] du \]

\[ = 4\pi^2 \int_0^\infty K_1 (y)^2 dy \int_0^1 K_2 (x)^2 dx \left( \int_0^1 f(u, 0) du \right) \left( \int_0^1 f(a, 0) da \right). \]

This proves the result of part (i). We now move to part (ii). We begin with the following relationship,

\[ \mathbb{E} \left( \tilde{J}_T - J_T \right) = \sum_{k=-T+1}^{T-1} K_1 (b_1 T k) \mathbb{E} \left( \tilde{\Gamma}_k \right) - \left( \int_0^1 c(u, 0) + 2 \sum_{k=1}^{T-1} \int_0^1 c(u, k) du \right). \]
Using Proposition 1.3.1, we have for any \(-T + 1 \leq k \leq T - 1\),

\[
\mathbb{E} \left( \frac{n_T}{T} \sum_{r=0}^{T/n_T} \bar{c}_T (rn_T/T, k) - \int_0^1 c(u, k) \, du \right)
\]

\[
= \frac{n_T}{T} \sum_{r=0}^{T/n_T} (c(rn_T/T, k))
\]

\[
+ \frac{1}{2} b_{2,T} \int_0^1 x^2 K_2(x) \, dx \int_0^1 \frac{\partial^2}{\partial^2 u} c(u, k) \, du + o \left( b_{2,T}^2 \right) + O \left( \frac{1}{b_{2,T} T} \right)
\]

\[
- \int_0^1 c(u, k) \, du
\]

\[
= \frac{n_T}{T} \sum_{r=0}^{T/n_T} (c(rn_T/T, k)) - \int_0^1 c(u, k) \, du
\]

\[
+ \frac{1}{2} b_{2,T} \int_0^1 x^2 K_2(x) \, dx \int_0^1 \frac{\partial^2}{\partial^2 u} c(u, k) \, du + o \left( b_{2,T}^2 \right) + O \left( \frac{1}{T b_{2,T}} \right)
\]

\[
= O \left( \frac{n_T}{T} \right) + \frac{1}{2} b_{2,T} \int_0^1 x^2 K_2(x) \, dx \int_0^1 \frac{\partial^2}{\partial^2 u} c(u, k) \, du + o \left( b_{2,T}^2 \right) + O \left( \frac{1}{T b_{2,T}} \right),
\]

where the last equality follows from the convergence of approximations to Riemann sums. This leads to
\[ \bar{b}^{-q}_{1,T} \mathbb{E} \left( \tilde{J}_T - J_T \right) \]
\[ = - \bar{b}^{-q}_{1,T} \sum_{k = -T+1}^{T} (1 - K_1(b_{1,T}k)) \int_0^1 c(u, k) \, du \]
\[ + \frac{1}{2} \frac{b_{2,T}^2}{b_{1,T}^q} \int_0^1 x^2 K_2(x) \sum_{k = -T+1}^{T} K_1(b_{1,T}k) \int_0^1 \frac{\partial^2}{\partial^2 u} c(u, k) \, du \]
\[ + O \left( \frac{1}{T b_{1,T}^q} \right) + O \left( \frac{n_T}{T b_{1,T}^q} \right) \]
\[ = - \bar{b}^{-q}_{1,T} \sum_{k = -T+1}^{T} (1 - K_1(b_{1,T}k)) \int_0^1 c(u, k) \, du \]
\[ - \frac{1}{2} \frac{b_{2,T}^2}{b_{1,T}^q} \int_0^1 x^2 K_2(x) \frac{b_{1,T}^q}{b_{1,T}} \sum_{k = -T+1}^{T} (1 - K_1(b_{1,T}k)) \int_0^1 \frac{\partial^2}{\partial^2 u} c(u, k) \, du \]
\[ + \frac{1}{2} \frac{b_{2,T}^2}{b_{1,T}^q} \int_0^1 x^2 K_2(x) \sum_{k = -T+1}^{T} \int_0^1 \frac{\partial^2}{\partial^2 u} c(u, k) \, du \]
\[ + O \left( \frac{1}{T b_{1,T}^q b_{2,T}} \right) + O \left( \frac{n_T}{T b_{1,T}^q} \right) \]
\[ = - \bar{b}^{-q}_{1,T} \sum_{k = -T+1}^{T} (1 - K_1(b_{1,T}k)) \int_0^1 c(u, k) \, du \]
\[ - \frac{1}{2} \frac{b_{2,T}^2}{b_{1,T}^q} \int_0^1 x^2 K_2(x) O(1) + \frac{1}{2} \frac{b_{2,T}^2}{b_{1,T}^q} \int_0^1 x^2 K_2(x) O(1) \]
\[ + O \left( \frac{1}{T b_{1,T}^q b_{2,T}} \right) + O \left( \frac{n_T}{T b_{1,T}^q} \right) \]

since \( \left| \sum_{k = -\infty}^{\infty} |k|^q \int_0^1 \left( \frac{\partial^2}{\partial^2 u} c(u, k) \right) \, du \right| < \infty \). Therefore,

\[ \lim_{T \to \infty} \bar{b}^{-q}_{1,T} \mathbb{E} \left( \tilde{J}_T - J_T \right) = -2\pi K_{1,q} \int_0^1 f^{(q)}(u, 0) \, du, \]

because \( b_{2,T}^2/b_{1,T}^q \to 0 \). It remains to show part (iii). Note that

\[ Tb_{1,T} b_{2,T} = Tb_{1,T} b_{2,T} b_{1,T}^q/b_{1,T} = b_{1,T}^{-2q} / \left( 1/T b_{1,T}^q b_{2,T} \right) = b_{1,T}^{-2q} / (1/ (\gamma + o(1))). \]
Hence, using part (i)-(ii), we deduce the desired result, namely,

\[
\lim_{T \to \infty} \text{MSE} \left( T b_{1,T} b_{2,T}, \tilde{J}_T, 1 \right) = \\
\lim_{T \to \infty} b_{1,T}^{-2} \mathbb{E} \left[ (\tilde{J}_T - J_T)^2 \right] (\gamma + o(1)) + \lim_{T \to \infty} T b_{1,T} b_{2,T} \text{Var} (\tilde{J}_T) = \\
4\pi^2 \left[ K_{1,q}^2 \left( \int_0^1 f(q)(u, 0) du \right)^2 + \int K_1^2 (y) dy \int_0^1 K_2^2 (x) dx \left( \int_0^1 f(u, 0) du \right)^2 \right].
\]

□

Proof of Theorem 1.3.1. We can now complete the proof of the theorem. We begin with part (i). We provide the expression for the asymptotic covariance between the \((i, l)\) and \((m, n)\) elements of \(\tilde{J}_T\):

\[
T b_{1,T} b_{2,T} \text{Cov} \left[ \sum_{k=-T+1}^{T-1} K_1 (b_{1,T} k) \tilde{\Gamma}^{(i,l)} (k), \sum_{j=-T+1}^{T-1} K_1 (b_{1,T} j) \tilde{\Gamma}^{(m,n)} (j) \right]
= 2b_{1,T} \sum_{k=0}^{T-1} \sum_{j=0}^{T-1} K_1 (b_{1,T} k) K_1 (b_{1,T} j) \left( \frac{n_T}{T} \right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} \sum_{h=j+1}^{T} K_2 \left( \frac{((rn_T + 1) - (s + k/2))}{b_{2,T}} \right) K_2 \left( \frac{((bn_T + 1) - (h + j/2))}{b_{2,T}} \right)
\times \left\{ \kappa^{(i,l,m,n)}_{V,s} (k, h-s, h-s-j) \right. \\
\left. + \left[ \Gamma^{(i,m)}_{h/T} (h-s) \Gamma^{(l,n)}_{h/T} (h-s+j-k) + \Gamma^{(i,n)}_{h/T} (s-h-j) \Gamma^{(l,m)}_{h/T} (s-h+k) \right] \right\}
+ o_p(1),
\]

where the \(o_p(1)\) term follows from (A.2.2). As for the scalar case, the term involving \(\kappa^{(i,l,m,n)}_{V,s} (k, h-s, h-s-j)\) is negligible. The limit of the term involving \(\Gamma^{(i,m)}_{h/T} (h-s) \Gamma^{(l,n)}_{h/T} (h-s+j-k)\) is, according to the derivations for part (i) of Lemma A.2.1,

\[
4\pi^2 \int K_1 (y)^2 dy \int_0^1 K_2 (x)^2 dx \left( \int_0^1 f^{(i,m)} (u, 0) du \right) \left( \int_0^1 f^{(l,n)} (v, 0) dv \right). \quad (A.2.8)
\]
Similarly, the limit of the term involving \( \Gamma^{(i,n)}_{h/T} (s - h - j) \Gamma^{(l,m)}_{h/T} (s - h + k) \) is the same as (A.2.8) but with \( m \) and \( n \) interchanged. The commutation-tensor product formula arises from the fact that the asymptotic covariances between \( \bar{J}^{(i,j)}_T \) and \( \bar{J}^{(m,n)}_T \) for \( i, j, m, n \leq p \) are of the same form as the covariances between \( X_i X_j \) and \( X_m X_n \), where \( X = (X_1, \ldots, X_p)' \sim \mathcal{N} (0, \Sigma) \). The formula then follows from \( \text{Var} \left( \text{vec} \left( XX' \right) \right) = \text{Var} \left( X \otimes X \right) = (I + C_{pp}) \Sigma \otimes \Sigma \).

Part (ii) of the theorem follows from the scalar case with minor changes. Since part (iii) simply uses part (i)-(ii), it follows that

\[
\lim_{T \to \infty} \text{MSE} \left( Tb_{1,T} b_{2,T}, \bar{J}_T, W \right) = \lim_{T \to \infty} \gamma b_{1,T}^{-2q} \text{E} \left( \bar{J}_T - J_T \right) W \text{E} \left( \bar{J}_T - J_T \right) + \lim_{T \to \infty} Tb_{1,T} b_{2,T} \text{tr} W \text{Var} \left( \text{vec} \left( \bar{J}_T \right) \right),
\]

converges to the desired limit. \( \square \)

**A.2.2.3 Proof of Theorem 1.3.2**

Under Assumption 1.2, \( \left\| F_0 (0, 0) \right\| < \infty \). In view of \( K_{1,0} = 0 \), Theorem 1.3.1-(i,ii) [with \( q = 0 \) in part (ii)] implies \( \bar{J}_T - J_T = o_p (1) \). Noting that \( \bar{J}_T - \bar{J}_T = o_p (1) \) if and only if \( b' \bar{J}_T b - b' \bar{J}_T b = o_p (1) \) for arbitrary \( b \in \mathbb{R}^p \) we shall provide the proof only for the scalar case. We first show that \( \sqrt{T} b_{1,T} b_{2,T} \left( \bar{J}_T - \bar{J}_T \right) = O_p (1) \) under Assumption 1.3. Let \( \bar{J}_T (\beta) \) denote the estimator that uses \( \{ V_{i,T} (\beta) \} \). A mean-value expansion of \( \bar{J}_T (\beta) \) (\( = \bar{J}_T \)) about \( \beta_0 \) yields

\[
\sqrt{T} b_{1,T} \left( \bar{J}_T - \bar{J}_T \right) = b_{1,T} \frac{\partial}{\partial \beta'} \bar{J}_T (\beta) \sqrt{T} (\beta - \beta_0)
= b_{1,T} \sum_{k=-T+1}^{T-1} K_1 (b_{1,T} k) \frac{\partial}{\partial \beta'} \Gamma (k) |_{\beta = \beta} \sqrt{T} (\beta - \beta_0),
\]

for some \( \beta \) on the line segment joining \( \hat{\beta} \) and \( \beta_0 \). Note that also \( \bar{c} (r n_T / T, k) \) depends on \( \beta \) although we have omitted it. We have for \( k \geq 0 \) (the case \( k < 0 \) is similar and
omitted),

\[
\left\| \frac{\partial}{\partial \beta'} \hat{c}(rn_T/T, k) \right\|_{\beta=\bar{\beta}} \tag{A.2.10}
\]

\[
= \left\| (Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2 \left( \frac{(r+1)n_T - (s+k/2)}{Tb_{2,T}} \right) \times \left( V_s(\beta) \frac{\partial}{\partial \beta} V_{s-k}(\beta) + \frac{\partial}{\partial \beta'} V_s(\beta) V_{s-k}(\beta) \right) \right\|_{\beta=\bar{\beta}}
\]

\[
\leq 2 \left( (Tb_{2,T})^{-1} \sum_{s=1}^{T} K_2^2 \left( \frac{(r+1)n_T - (s+k/2)}{Tb_{2,T}} \right) \sup_{s \geq 1} \sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta'} V_s(\beta) \right\|^2 \right)^{1/2}
\]

\[
\times \left( (Tb_{2,T})^{-1} \sum_{s=1}^{T} K_2^2 \left( \frac{(r+1)n_T - (s+k/2)}{Tb_{2,T}} \right) \sup_{s \geq 1} \sup_{\beta \in \Theta} \left\| \frac{\partial}{\partial \beta} V_s(\beta) \right\|^2 \right)^{1/2}
\]

\[
= \mathcal{O}_P(1),
\]

where we have used the boundedness of the kernel \( K_2 \), Assumption 1.3-(ii,iii) and Markov’s inequality to each term in parentheses; also \( \sup_{s \geq 1} \mathbb{E} \sup_{\beta \in \Theta} \| V_s(\beta) \|^2 < \infty \) under Assumption 1.3-(ii,iii) by a mean-value expansion and

\[
(Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2^2 (((r+1)n_T - (s+k/2))/Tb_{2,T}) \to \int_0^1 K_2^2(x) \, dx < \infty.
\]

Then, (A.2.9) becomes

\[
b_{1,T} \sum_{k=-T+1}^{T-1} K_1(b_{1,T} k) \frac{\partial}{\partial \beta'} \hat{\Gamma}(k) \mid_{\beta=\bar{\beta}} \sqrt{T} (\tilde{\beta} - \beta_0)
\]

\[
\leq b_{1,T} \sum_{k=-T+1}^{T-1} K_1(b_{1,T} k) \frac{n_T}{T} \sum_{r=0}^{T/n_T} \mathcal{O}_P(1) \mathcal{O}_P(1)
\]

\[
= \mathcal{O}_P(1),
\]

where the last equality uses \( b_{1,T} \sum_{k=-T+1}^{T-1} K_1(b_{1,T} k) \to \int |K_1(x)| \, dx < \infty \). This concludes the proof of part (i) of Theorem 1.3.2 because \( \sqrt{T} b_{1,T} \to \infty \) by assumption.
The next step is to show that \( \sqrt{Tb_{1,T}} (\hat{J}_T - \tilde{J}_T) = o_P(1) \) under the assumptions of Theorem 1.3.2-(ii). A second-order Taylor expansion gives

\[
\sqrt{Tb_{1,T}} (\hat{J}_T - \tilde{J}_T) = \left[ \sqrt{b_{1,T}} \frac{\partial}{\partial \beta'} \tilde{J}_T (\beta_0) \right] \sqrt{T} (\hat{\beta} - \beta_0) + \frac{1}{2} \sqrt{T} (\hat{\beta} - \beta_0)' \left[ \sqrt{b_{1,T}} \frac{\partial^2}{\partial \beta \partial \beta'} \tilde{J}_T (\beta) / \sqrt{T} \right] \sqrt{T} (\hat{\beta} - \beta_0)
\]

\[\triangleq G_T \sqrt{T} (\hat{\beta} - \beta_0) + \frac{1}{2} \sqrt{T} (\hat{\beta} - \beta_0)' H_T \sqrt{T} (\hat{\beta} - \beta_0).\]

Proceeding as in (A.2.10) but now using Assumption 1.4,

\[
\left\| \frac{\partial^2}{\partial \beta \partial \beta'} \hat{\beta} (rn/T, k) \right\|_{\beta = \hat{\beta}} = \left\| (Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2 \left( \frac{(r+1)n_T - (s+k/2)}{b_{2,T}} \right) \left( \frac{\partial^2}{\partial \beta \partial \beta'} V_s (\hat{\beta}) V_{s-k} (\hat{\beta}) \right) \right\|_{\beta = \hat{\beta}} = O_P(1)
\]

and thus,

\[
\| H_T \| \leq \left( \frac{b_{1,T}}{T} \right)^{1/2} \sum_{k=-T+1}^{T-1} \sup_{\beta \in \Theta} \left\| \frac{\partial^2}{\partial \beta \partial \beta'} \hat{\beta} (k) \right\| \leq \left( \frac{b_{1,T}}{T} \right)^{1/2} \sum_{k=-T+1}^{T-1} |K_1 (b_{1,T}k)| O_P(1) \leq \left( \frac{1}{Tb_{1,T}} \right)^{1/2} b_{1,T} \sum_{k=-T+1}^{T-1} |K_1 (b_{1,T}k)| O_P(1) = o_P(1),
\]

since \( Tb_{1,T} \to \infty \). Next, we want to show that \( G_T = o_P(1) \). Following Andrews (1991) (cf. the last paragraph of p. 852), we apply the results of Theorem 1.3.1-(i,ii) to \( \tilde{J}_T \) where the latter is constructed using \((V_t, \partial V_t / \partial \beta') - E(\partial V_t / \partial \beta')\)' rather than just with \( V_t \). The first row and column of the off-diagonal elements of this \( \tilde{J}_T \) (written
by Theorem 1.3.1-(i,ii), each expression above is $O_p(1)$. Since

$$G_T = \sqrt{b_{1,T} (A_1 + A_2)} + \sqrt{b_{1,T}} \sum_{k=-T+1}^{T-1} K_1 (b_{1,T} k) \frac{n_T}{T} \sum_{r=0}^{T/n_T} 1 \frac{1}{T b_{2,T}^2} \times \sum_{s=k+1}^{T} K_2 \left( \frac{(r+1) n_T - (s+k/2) / T}{b_{2,T}} \right) (V_s + V_{s-k}) \mathbb{E} \left( \frac{\partial}{\partial \beta} V_s \right)$$

$$\triangleq \sqrt{b_{1,T} (A_1 + A_2)} + A_3 \mathbb{E} \left( \frac{\partial}{\partial \beta} V_s \right),$$

it remains to show that $A_3$ is $o_p(1)$. Note that

$$\mathbb{E} (A_3^2) \leq b_{1,T} \sum_{k=-T+1}^{T-1} \sum_{j=-T+1}^{T-1} |K_1 (b_{1,T} k) K_1 (b_{1,T} j)| 4 \left( \frac{n_T}{T} \right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \frac{1}{T b_{2,T}^2} \frac{1}{T b_{2,T}} \sum_{s=1}^{T} \sum_{l=1}^{T} K_2 \left( \frac{(r+1) n_T - (s+k/2) / T}{b_{2,T}} \right) \times K_2 \left( \frac{(b+1) n_T - (l+j/2) / T}{b_{2,T}} \right) |\mathbb{E} (V_s V_l)|,$$

and that $\mathbb{E} (V_s V_l) = c(u, h) + O(T^{-1})$ uniformly in $h = s - l$ with $u = s/T$. Since
\[ \sum_{h=-\infty}^{\infty} \sup_{u \in [0, 1]} |c(u, h)| < \infty, \]

\[ \mathbb{E} \left( A_3^2 \right) \leq \frac{1}{T b_{1,T} b_{2,T}} \left( b_{1,T} \sum_{k=-T+1}^{T-1} |K_1(b_{1,T} k)| \right)^2 \int_0^1 K_2^2(x) \, dx \]

\[ \int_0^1 \sum_{h=-\infty}^{\infty} |c(u, h)| \, du = o(1). \]

This implies \( G_T = o_p(1) \). It follows that \( \sqrt{T b_{1,T}} \left( \hat{J}_T - \bar{J}_T \right) = o_p(1) \) which concludes the proof of part (ii) because \( \sqrt{T b_{1,T} b_{2,T}} \left( \hat{J}_T - \bar{J}_T \right) = O_p(1) \) by Theorem 1.3.1-(iii).

Finally, we need to consider part (iii). Let

\[ \xi_T \triangleq T b_{1,T} \left( \text{vec} \left( \hat{J}_T - \bar{J}_T \right)' W_T \text{vec} \left( \hat{J}_T - \bar{J}_T \right) - \text{vec} \left( \tilde{J}_T - \bar{J}_T \right)' W_T \text{vec} \left( \tilde{J}_T - \bar{J}_T \right) \right). \]

By part (i)-(ii) we know that \( \sqrt{T b_{1,T}} \left( \hat{J}_T - \bar{J}_T \right) = O_p(1) \) and \( \sqrt{T b_{1,T}} \left( \tilde{J}_T - \bar{J}_T \right) = o_p(1) \). This implies

\[ T b_{1,T} \left( \text{vec} \left( \hat{J}_T - \bar{J}_T \right)' W_T \text{vec} \left( \hat{J}_T - \bar{J}_T \right) - \text{vec} \left( \tilde{J}_T - \bar{J}_T \right)' W_T \text{vec} \left( \tilde{J}_T - \bar{J}_T \right) \right) \overset{P}{\to} 0. \]

Then, using Assumption 1.5, \( \xi_T = o_p(1) \) and since \( |\xi_T| \) is bounded we have \( \mathbb{E}(\xi_T) \to 0 \) by Lemma A1 in Andrews (1991). \( \square \)

### A.2.3 Proof of the Results of Section 1.4

#### A.2.3.1 Proof of Proposition 1.4.1

We first need to show that \( \sqrt{T b_{2,T}} (\tilde{c}_T (rn_T/T, k) - \bar{c}(rn_T/T, k)) = o_p(1) \). From (A.2.10),

\[ \left\| \frac{\partial}{\partial \beta'} \tilde{c}_T (rn_T/T, k) \right\|_{\beta = \bar{\beta}} = O_p(1), \]
uniformly in $r$. A mean-value Taylor expansion gives
\[
\sqrt{Tb_{2,T}} (\hat{c}_T (rn_T/T, k) - \bar{c}_T (rn_T/T, k)) \\
= \sqrt{b_{2,T}} \frac{\partial}{\partial \beta} \hat{c}_T (rn_T/T, k) \bigg|_{\beta = \hat{\beta}} \sqrt{T} (\hat{\beta} - \beta_0) \\
\leq \sqrt{b_{2,T}} \sup_{r \geq 1} \left\| \frac{\partial}{\partial \beta} \hat{c} (rn_T/T, k) \right\| \bigg|_{\beta = \hat{\beta}} \sqrt{T} (\hat{\beta} - \beta_0) \\
= \sqrt{b_{2,T}} O_P (1) = o_P (1).
\]
Thus, where
\[
\xi_T = \text{vec} (\hat{c}_T (rn_T/T, k) - \bar{c} (rn_T/T, k))' \tilde{W}_T \text{vec} (\hat{c}_T (rn_T/T, k) - \bar{c} (rn_T/T, k)) \\
\xrightarrow{\mathbb{P}} 0.
\]
Since $\xi_T$ is a bounded sequence, $\mathbb{E} (\xi_T) \xrightarrow{\mathbb{P}} 0$. Hence, given that $\tilde{W}_T \xrightarrow{\mathbb{P}} \tilde{W}$, we have $\text{MSE} (1, \hat{c}_T (u_0, k), \tilde{W}_T) = \text{MSE} (1, \bar{c}_T (u_0, k), \tilde{W}) + o_P (1)$. Without loss of generality we can focus on the scalar case. By using the results of Proposition 1.3.1, the MSE of $\hat{c}_T (u_0, k)$ for any $u_0 \in (0, 1)$ and any integer $k$, is given by
\[
\mathbb{E} [\hat{c}_T (u_0, k) - c (u_0, k)]^2 \\
= \frac{1}{4} b_{2,T}^4 \left( \int_0^1 x^2 K_2 (x) \, dx \right)^2 \left( \frac{\partial^2}{\partial^2 u} c (u_0, k) \right)^2 \\
+ \frac{1}{Tb_{2,T}} \int_0^1 K_2^2 (x) \, dx \sum_{l=-\infty}^{\infty} c (u_0, l) [c (u_0, l) + c (u_0, l + 2k)] \\
+ o \left( b_{2,T}^4 \right) + o (1/ (b_{2,T} T)) \\
\triangleq g (K_2, b_{2,T}) + o \left( b_{2,T}^4 \right) + o (1/ (b_{2,T} T)).
\]
Then \( g(K_2, b_{2,T}) = 4^{-1}b_{2,T}^{-1}H(K_2)D_1(u_0) + (Tb_{2,T})^{-1} F(K_2)D_2(u_0). \) The minimum of \( g(K_2, b_{2,T}) \) in \( b_{2,T} \) is determined by the equation

\[
\frac{\partial}{\partial b_{2,T}} g(K_2, b_{2,T}) = b_{2,T}^3H(K_2)D_1(u_0) - \frac{1}{Tb_{2,T}^2} F(K_2)D_2(u) = 0.
\]

The minimum is achieved at \( b_{2,T}^{\text{opt}} = \left[ H(K_2)D_1(u_0) \right]^{-1} (F(K_2)D_2(u))^{1/5} T^{-1/5}. \) Next, we minimize \( g(K_2, b_{2,T}^{\text{opt}}) \) with respect to the class of kernels \( K_2 : \mathbb{R} \rightarrow [0, \infty] \) that are centered at \( x = 1/2 \) with

\[
\int_{\mathbb{R}} K_2(x) \, dx = 1, \tag{A.2.11}
\]

\[
K_2(x) = K_2(1 - x). \tag{A.2.12}
\]

We use similar arguments as in Chapter 7.5 of Priestley (1981) and in Dahlhaus and Giraitis (1998)]. Let

\[
\sqrt{K_{2\sigma}}(x) = \frac{1}{\sqrt{\sigma}} \left( K_2 \left( \frac{x - 1/2}{\sigma} + \frac{1}{2} \right) \right)^{1/2}, \quad \text{where } \sigma \in (0, \infty).
\]

We have \( F(K_{2\sigma}) = (1/\sigma) \, F(K_2) \) and \( H(K_{2\sigma}) = \sigma^4 H(K_2) \) (with the integrals in the definition of \( F \) and \( H \) extended to \( \mathbb{R} \) and with the variable of integration \( x \) subtracted by 1/2). Then, \( b_{2,K_{2\sigma},T}^{\text{opt}} = \sigma^{-1} b_{2,T}^{\text{opt}} \) where \( b_{2,K_{2\sigma},T}^{\text{opt}} \) is the optimal bandwidth associated with the kernel \( K_{2\sigma} \). Also, \( g(K_{2\sigma}, b_{2,K_{2\sigma},T}^{\text{opt}}) = g(K_2, b_{2,T}^{\text{opt}}) \). We can thus restrict our attention to \( K_2 \) satisfying

\[
\int_{\mathbb{R}} \left( x - \frac{1}{2} \right)^2 K_2(x) \, dx = \int_{\mathbb{R}} \left( x - \frac{1}{2} \right)^2 K_{2,T}^{\text{opt}}(x) \, dx, \tag{A.2.13}
\]
where \( K_2^{\text{opt}}(x) = 6x(1-x) \) and \( K_2^{\text{opt}}(x) = 0 \) for \( x \notin [0, 1] \). Therefore, we have to show that, for any \( K_2 \) that satisfies (A.2.11)-(A.2.12),
\[
\int_{\mathbb{R}/[0,1]} K_2^2(x) \, dx + \int_0^1 K_2^2(x) \, dx = \int_{\mathbb{R}} K_2^2(x) \, dx \geq \int_{\mathbb{R}} \left(K_2^{\text{opt}}(x)\right)^2 \, dx = \int_0^1 \left(K_2^{\text{opt}}(x)\right)^2 \, dx.
\]
This is implied by
\[
\int_0^1 K_2^2(x) \, dx \geq \int_0^1 \left(K_2^{\text{opt}}(x)\right)^2 \, dx.
\]
Let \( K_2(x) = K_2^{\text{opt}}(x) + \varepsilon(x), x \in \mathbb{R} \), where \( \varepsilon \neq 0 \). Since \( \int_{\mathbb{R}} \varepsilon^2(x) \, dx \geq 0 \) and \( K_2^{\text{opt}} \) vanishes outside \([0, 1]\), it is sufficient to prove that \( \int_0^1 \left(K_2^{\text{opt}}(x)\varepsilon(x)\right) \, dx \geq 0 \) because
\[
\int_0^1 K_2^2(x) \, dx = \int_0^1 \left(K_2^{\text{opt}}(x) + \varepsilon(x)\right)^2 \, dx \geq \int_0^1 \left(K_2^{\text{opt}}(x)\right)^2 + 2 \int_0^1 \left(K_2^{\text{opt}}(x)\varepsilon(x)\right) \, dx.
\]
By (A.2.11) we have \( \int_{\mathbb{R}} \varepsilon(x) \, dx = 0 \) while we have \( \int_{\mathbb{R}} \varepsilon(x)(x^2 - x) \, dx = 0 \) in view of
\[
0 = \int_{\mathbb{R}} \left(K_2(x) - K_2^{\text{opt}}(x)\right) \left(x - \frac{1}{2}\right)^2 \, dx = \int_{\mathbb{R}} \left(K_2(x) - K_2^{\text{opt}}(x)\right) \left(x^2 - x\right) \, dx + \frac{1}{4} \int_{\mathbb{R}} \varepsilon(x) \, dx = \int_{\mathbb{R}} \left(K_2(x) - K_2^{\text{opt}}(x)\right) \left(x^2 - x\right) \, dx.
\]
Note that \( \int_{\mathbb{R}} \left(K_2(x) - K_2^{\text{opt}}(x)\right) \left(x^2 - x\right) \, dx \) and \( (x^2 - x) = x(x - 1) \). Therefore, we deduce
\[
6 \int_{\mathbb{R}/[0,1]} x(1-x)\varepsilon(x) \, dx + 6 \int_0^1 x(1-x)\varepsilon(x) \, dx = 0.
\]
Rearranging the last expression it gives,
\[ \int_0^1 K_2^{\text{opt}}(x) \varepsilon(x) \, dx = 6 \int_{\mathbb{R}/[0,1]} x (x - 1) \varepsilon(x) \, dx \geq 0 \]
because \( \varepsilon(x) \geq 0 \) and \( x (x - 1) \geq 0 \) for \( x \not\in [0, 1] \). □

### A.2.3.2 Proof of Theorem 1.4.1

Without loss of generality we provide the proof for the scalar case. If \( T_{b_{1,T}, b_{2,T}} \to \gamma \in (0, \infty) \) for some \( q \in [0, \infty) \) for which \( K_{1,q} \), \( \left| \int_0^1 f^{(q)}(u, 0) \, du \right| \in [0, \infty) \), then

\[
\lim_{T \to \infty} \text{MSE} \left( T_{b_1, T_{b_2, T}}, J_T(b_{1,T,K_1}), W_T \right) = 4\pi^2 \times \\
\left[ \gamma K_{1,q}^2 \left( \int_0^1 f^{(q)}(u, 0) \, du \right)^2 + \int K_1^2(y) \, dy \int_0^1 (K_2(x))^2 \, dx \left( \int_0^1 f(u, 0) \, du \right)^2 \right].
\]

Assume \( q = 2 \) so that \( T_{b_{1,T}, b_{2,T}} \to \gamma \). Then, \( T_{b_{1,T,K_1}, b_{2,T}} \to \gamma / \left( \int K_1^2(x) \, dx \right)^5 \) and

\[ T_{b_1, T_{b_2, T}} = T_{b_1, T_{b_1, K_1}, b_{2, T}} \int K_1^2(x) \, dx. \]

Therefore, given \( K_{1,2} < \infty \),

\[
\lim_{T \to \infty} \text{MSE} \left( T_{b_1, T_{b_2, T}}, J_T(b_{1,T,K_1}), W_T \right) \\
= 4\gamma \pi^2 \left( \int_0^1 f^{(q)}(u, 0) \, du \right)^2 \left[ K_{1,2}^2 \left( \int K_1^2(y) \, dy \right)^4 \int_0^1 (K_2(x))^2 \, dx \right],
\]

and

\[
\lim_{T \to \infty} \left( \text{MSE} \left( T_{b_1, T_{b_2, T}}, J_T(b_{1,T,K_1}), W_T \right) - \text{MSE} \left( T_{b_1, T_{b_2, T}}, J_{T_{\text{QS}}}(b_{1,T}), W_T \right) \right) \\
= 4\gamma \pi^2 \left( \int_0^1 f^{(q)}(u, 0) \, du \right)^2 \int_0^1 (K_2(x))^2 \, dx \left[ K_{1,2}^2 \left( \int K_1^2(y) \, dy \right)^4 \right. - \left. (K_{1,2})^2 \right].
\]
Let $\widetilde{K}_1(\cdot)$ and $\widetilde{K}^{QS}_1(\cdot)$ denote the spectral window generators of $K_1(\cdot)$ and $K^{QS}_1(\cdot)$, respectively. They have the following properties: $K_{1,2} = \int_{-\infty}^{\infty} \omega^2 \widetilde{K}_1(\omega) \, d\omega$, $K_1(0) = \int_{-\infty}^{\infty} \widetilde{K}_1(\omega) \, d\omega$, and $\int_{-\infty}^{\infty} K^2_2(x) \, dx = \int_{-\infty}^{\infty} \widetilde{K}_1(\omega) \, d\omega$. As in Andrews (1991), the result of the theorem follows if we can show the following inequality,

$$K^2_{1,2} \left( \int K^2_2(x) \, dx \right)^4 \geq \left( K^{QS}_{1,2} \right)^2$$

for all $K_1(\cdot) \in \widetilde{K}_1$. (A.2.14)

Priestley (1981, Ch. 7.5) showed that $\widetilde{K}^{QS}_1(\cdot)$ minimizes

$$\int_{-\infty}^{\infty} \omega^2 \widetilde{K}_1(\omega) \, d\omega \left( \int_{-\infty}^{\infty} \widetilde{K}^2_1(\omega) \, d\omega \right)^2.$$ (A.2.15)

subject to (a) $\int_{-\infty}^{\infty} \widetilde{K}_1(\omega) \, d\omega = 1$, (b) $\widetilde{K}_1(\omega) \geq 0$, $\forall \omega \in \mathbb{R}$, and (c) $\widetilde{K}_1(\omega) = \widetilde{K}_1(-\omega)$, $\forall \omega \in \mathbb{R}$, where $K^{QS}_1(\omega) = (5/8\pi)(1 - \omega^2/c^2)$ for $|\omega| \leq c$ and $K^{QS}_1(\omega) = 0$ otherwise for $c = 6\pi/5$. Note that the inequality (A.2.14) holds if and only if $\widetilde{K}^{QS}_1(\cdot)$ minimizes (A.2.15). This proves the inequality of the theorem. Strict inequality holds when $K^{QS}_1(x) \neq K_1(x)$ with positive Lebesgue measure. □

### A.2.3.3 Proof of Corollary 1.4.1

Note that $T^{2q+1}b_{2,T}^{2q} = \left( T \gamma^{2q+1} b_{2,T} \right)^{-1/(2q+1)} T b_{1,T} b_{2,T} = \left( \gamma^{-1/(2q+1)} + o(1) \right) T b_{1,T} b_{2,T}$. Thus,

$$\lim_{T \to \infty} \text{MSE} \left( T^{2q+1} b_{2,T}^{2q}, \hat{f}_T, W_T \right) = \gamma^{-1/(2q+1)} 4\pi^2 \left[ K^2_{1,2} \text{vec} \left( \int_0^1 f(q)(u, 0) \, du \right)^T W \text{vec} \left( \int_0^1 f(q)(u, 0) \, du \right) \right.$$

$$+ \int K^2_1(y) \, dy \int_0^1 K^2_2(x) \, dx$$

$$\left. \text{tr} W (I - C_{pp}) \left( \int_0^1 f(u, 0) \, du \right) \otimes \left( \int_0^1 f(v, 0) \, dv \right) \right].$$ (A.2.16)
Minimizing this with respect to $\gamma$ gives

$$\lim_{T \to \infty} \text{MSE} \left( T\frac{2q}{2q+1} b_{2,T}^{\frac{2q+1}{2q}}, \hat{J}_T, W_T \right)$$

$$= \gamma^{-1/(2q+1)} 4\pi^2 \left[ \gamma K_{1,q}^2 \text{vec} \left( \int_0^1 f^{(q)}(u, 0) \, du \right)' W \text{vec} \left( \int_0^1 f^{(q)}(u, 0) \, du \right) \right]$$

$$+ \int K_1^2(y) \, dy \int_0^1 K_2^2(x) \, dx \quad \text{tr} W (I - C_{pp}) \left( \int_0^1 f(u, 0) \, du \right) \otimes \left( \int_0^1 f(v, 0) \, dv \right).$$

(A.2.17)

or

$$\gamma^{opt} = \frac{1}{2q} \int K_1^2(y) \, dy \int K_2^2(x) \, dx \text{tr} W (I + C_{pp}) \left( \int_0^1 f(u, 0) \, du \right) \otimes \left( \int_0^1 f(v, 0) \, dv \right)$$

$$= \left( qK_{1,q}^2 \phi(q) \right)^{-1} \left( \int K_1^2(y) \, dy \int_0^1 K_2^2(x) \, dx \right).$$

Note that $\gamma^{opt} > 0$ provided that $0 < \phi(q) < \infty$ and $W$ is positive semidefinite. Hence, $\{b_{1,T}\}$ is optimal in the sense that $T b_{1,T}^{2q+1} b_{2,T} \to \gamma^{opt}$ if and only if $b_{1,T} = b_{1,T}^{opt} + o \left( (Tb_{2,T})^{-1/(2q+1)} \right)$. □

A.2.4 Proofs of the Results of Section 1.5

A.2.4.1 Proof of Theorem 1.5.1

Without loss of generality, we assume that $V_t$ is a scalar. The constant $C < \infty$ may vary from line to line. We begin with the proof of part (ii) because it becomes then simpler to prove part (i). By Theorem 1.3.2-(ii),

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}} \left( \hat{J}_T (b_{\theta_1,T}, b_{\theta_2,T}) - J_T \right) \to O_P(1).$$
It remains to establish the second result of Theorem 1.5.1-(ii). Let
\[
r \in \left( \max \left\{ \frac{(8b - 5 - 2q)}{8(b - 1)}, \frac{(b - 1/2)}{b - 1}, \frac{q}{(l - 1)} \right\}, \min \left\{ \frac{3q}{4} + \frac{9}{8}, 3 \right\} \right),
\]
and \(S_T = \left\lfloor b^r \theta_1 \right\rfloor\). We will use the following decomposition
\[
\hat{J}_T (\hat{b}_{1,T}, \hat{b}_{2,T}) - \hat{J}_T (\theta_1, \theta_2, T) = (\hat{J}_T (\hat{b}_{1,T}, \hat{b}_{2,T}) - \hat{J}_T (\theta_1, \theta_2, T)) + (\hat{J}_T (\theta_1, \theta_2, T) - \hat{J}_T (\theta_1, \theta_2, T)).
\]

\[N_1 \triangleq \{ -S_T, -S_T + 1, \ldots, -1, 1, \ldots, S_T - 1, S_T \}\]
\[N_2 \triangleq \{ -T + 1, \ldots, -S_T - 1, S_T + 1, \ldots, T - 1 \}\]

Let us consider the first term above,
\[
T^{8q/10(2q+1)} \left( \hat{J}_T (\hat{b}_{1,T}, \hat{b}_{2,T}) - \hat{J}_T (\theta_1, \theta_2, T) \right)
= T^{8q/10(2q+1)} \sum_{k \in N_1} \left( K_1 (\hat{b}_{1,T}k) - K_1 (\theta_1, \theta_2, T) \right) \hat{\Gamma} (k)
+ T^{8q/10(2q+1)} \sum_{k \in N_2} K_1 (\hat{b}_{1,T}k) \hat{\Gamma} (k)
- T^{8q/10(2q+1)} \sum_{k \in N_2} K_1 (\theta_1, \theta_2, T) \hat{\Gamma} (k)
\triangleq A_{1,T} + A_{2,T} - A_{3,T}.
\]
We first show that \( A_{1,T} \xrightarrow{p} 0 \). Let \( A_{1,T} \) denote \( A_{1,T} \) with the summation restricted over positive integers \( k \). We can use the Liptchitz condition on \( K_1(\cdot) \in K_3 \) to yield,

\[
|A_{1,T}| \leq T^{8q/10(2q+1)} \sum_{k=1}^{S_T} C_2 |\hat{b}_{1,T} - b_{\theta_1,T}| k |\hat{\Gamma}(k)|
\]

(A.2.20)

\[
\leq C \sqrt{n_T} |\hat{\phi}(q)^{1/(2q+1)} - \phi_0^{1/(2q+1)}|
\]

\[
\left( \hat{\phi}(q) \phi_0 \right)^{-1/(2q+1)} \hat{b}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} k |\hat{\Gamma}(k)|,
\]

for some \( C < \infty \). By Assumption 1.6-(ii) \( (\sqrt{n_T} |\hat{\phi}(q) - \phi_0| = O_p(1)) \) and using the delta method, it suffices to show that \( B_{1,T} + B_{2,T} + B_{3,T} \xrightarrow{p} 0 \), where

\[
B_{1,T} = b_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} k |\hat{\Gamma}(k) - \Gamma_T(k)|
\]

(A.2.21)

\[
B_{2,T} = b_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} k |\Gamma_T(k)|,
\]

\[
B_{3,T} = b_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} k |\Gamma_T(k)|,
\]

with \( \Gamma_T(k) \triangleq (n_T/T) \sum_{r=0}^{T/n_T} c(rn_T/T, k) \). By a mean-value expansion, we have

\[
B_{1,T} \leq b_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} k \left| \left( \frac{\partial}{\partial \beta} \hat{\Gamma}(k) \right)_{\beta = \hat{\beta}} \right| \sqrt{T} \left( \hat{\beta} - \beta_0 \right)
\]

(A.2.22)

\[
\leq C b_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} T^{-1/2} \left( T b_{\theta_2,T}^{2q/(2q+1)} n_T^{-1/2} \right)^{2r/(2q+1)} n_T^{-1/2}
\]

\[
\times \sup_{k \geq 1} \left| \left( \frac{\partial}{\partial \beta} \hat{\Gamma}(k) \right)_{\beta = \hat{\beta}} \right| \sqrt{T} \left( \hat{\beta} - \beta_0 \right)
\]

\[
\leq C b_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} T^{-1/2} \left( T^{2r/(2q+1)} n_T^{-1/2} \right)^{2r/(2q+1)} n_T^{-1/2}
\]

\[
\times \sup_{k \geq 1} \left| \left( \frac{\partial}{\partial \beta} \hat{\Gamma}(k) \right)_{\beta = \hat{\beta}} \right| \sqrt{T} \left( \hat{\beta} - \beta_0 \right) \xrightarrow{p} 0
\]
since \( n_T/T^{1/2} \to \infty \), \( r < 3q/4 + 9/8 \), \( \sqrt{T} \left\| \hat{\beta} - \beta_0 \right\| = O_p(1) \), and

\[
\sup_{k \geq 1} \left\| (\partial/\partial \beta) \hat{\Gamma}(k) \right\|_{\beta=\hat{\beta}} = O_p(1),
\]

using (A.2.10) and Assumption 1.3-(ii,iii). In addition,

\[
\mathbb{E} \left( B_{2,T}^2 \right) \leq \mathbb{E} \left( \hat{b}_{2,T}^{-2/(2q+1)} T^{(8q-10)/5(2q+1)} n_T^{-1} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} k j \left| \hat{\Gamma}(k) - \Gamma_T(k) \right| \left| \hat{\Gamma}(j) - \Gamma_T(j) \right| \right) \]

\[
\leq \hat{b}_{2,T}^{-2/(2q+1)-1} T^{(8q-10)/5(2q+1)-1} S_T^4 \sup_{k \geq 1} \text{Var} \left( \hat{\Gamma}(k) \right) \]

\[
\leq \hat{b}_{2,T}^{-2/(2q+1)-1} T^{(8q-10)/5(2q+1)-1} \left( T b_{2,T} \right)^{4r/(2q+1)} \sup_{k \geq 1} \text{Var} \left( \hat{\Gamma}(k) \right) \]

\[
\leq T^{1/5 \cdot n-2/5(2q+1)} T^{(8q-10)/5(2q+1)-1} T^{4r/(2q+1)} T^{-4r/5(2q+1)} \]

\[
\sup_{k \geq 1} \text{Var} \left( \hat{\Gamma}(k) \right) \]

\[
\to 0,
\]

given that \( \sup_{k \geq 1} T b_{2,T} \text{Var} \left( \hat{\Gamma}(k) \right) = O(1) \) using Proposition 1.3.1 with \( r < 3 \). Assumption 1.6-(iii) and \( \sum_{k=1}^{\infty} k^{1-l} < \infty \) for \( l > 2 \) yield

\[
B_{3,T} \leq \hat{b}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} n_T^{-1/2} C_3 \sum_{k=1}^{\infty} k^{1-l} \]

\[
\leq T^{(-13-2q)/10(2q+1)} C_3 \sum_{k=1}^{\infty} k^{1-l} \to 0,
\]

where we have used the fact that \( 1/n_T = o \left( T^{-1/2} \right) \). Combining (A.2.20)-(A.2.24) we deduce that \( A_{1,1,T} \xrightarrow{p} 0 \). The same argument applied to \( A_{1,T} \) where the summation now also extends over negative integers \( k \) gives \( A_{1,T} \xrightarrow{p} 0 \). Next, we show that \( A_{2,T} \xrightarrow{p} 0 \). Again, we use the notation \( A_{2,1,T} \) (resp., \( A_{2,2,T} \)) to denote \( A_{2,T} \) with the summation over positive (resp., negative) integers. Let \( A_{2,1,T} = L_{1,T} + L_{2,T} + L_{3,T} \),
where

\[ L_{1,T} = T^{8q/10(2q+1)} \sum_{k=S_T+1}^{T-1} K_1 \left( L_{1,T} k \right) \left( \hat{T} (k) - T (k) \right) \]  
\[ L_{2,T} = T^{8q/10(2q+1)} \sum_{k=S_T+1}^{T-1} K_1 \left( L_{2,T} k \right) \left( \hat{T} (k) - T (k) \right) \]  
\[ L_{3,T} = T^{8q/10(2q+1)} \sum_{k=S_T+1}^{T-1} K_1 \left( L_{3,T} k \right) \left( \hat{T} (k) - T (k) \right). \]

We apply a mean-value expansion, use \( \sqrt{T} (\hat{\beta} - \beta_0) = O_p (1) \) as well as (A.2.10) to obtain

\[ |L_{1,T}| = T^{8q/10(2q+1) - 1/2} \sum_{k=S_T+1}^{T-1} C_1 \left( |L_{1,T} k| \right)^{-b} \left| \left( \frac{\partial}{\partial \beta} \hat{T} (k) \right) |_{\beta=\hat{\beta}} \sqrt{T} (\hat{\beta} - \beta_0) \right| \]  
\[ \begin{align*} & = T^{8q/10(2q+1) - 1/2 + 4b/5(2q+1)} \sum_{k=S_T+1}^{T-1} C_1 k^{-b} \left| \left( \frac{\partial}{\partial \beta} \hat{T} (k) \right) |_{\beta=\hat{\beta}} \sqrt{T} (\hat{\beta} - \beta_0) \right| \\
& = T^{8q/10(2q+1) - 1/2 + 4b/5(2q+1) + 4r(1-b)/5(2q+1)} \left| \left( \frac{\partial}{\partial \beta} \hat{T} (k) \right) |_{\beta=\hat{\beta}} \sqrt{T} (\hat{\beta} - \beta_0) \right| \\
& = T^{8q/10(2q+1) - 1/2 + 4b/5(2q+1) + 4r(1-b)/5(2q+1)} O_p (1) O_p (1), \end{align*} \]

which goes to zero since \( r > (8b - 5 - 2q) / 8 (b - 1) \). Let us now consider \( L_{2,T} \). We have

\[ |L_{2,T}| = T^{(8q-1)/10(2q+1)} \sum_{k=S_T+1}^{T-1} C_1 \left( |L_{2,T} k| \right)^{-b} \left| \hat{T} (k) - \Gamma (k) \right| \]  
\[ \begin{align*} & = C_1 \left( q K_{1,q} \hat{\phi} (q) \right)^{b/(2q+1)} T^{8q/10(2q+1) + b/(2q+1) - 1/2} b_{2,T}^{b/(2q+1) - 1/2} \left( \sum_{k=S_T+1}^{T-1} k^{-b} \right) \\
& \times \sqrt{T} b_{2,T} \left| \hat{T} (k) - \Gamma (k) \right| \end{align*} \]
Note that

\[
E \left( T^{8q/10(2q+1)+b/(2q+1)-1/2} \sum_{k=S_T}^{T-1} k^{-b} \sqrt{Tb_{2,T}} \left| \tilde{\Gamma} (k) - \Gamma_T (k) \right| \right)^2 \quad (A.2.28)
\]

\[
\leq T^{8q/5(2q+1)+2b/(2q+1)-1} \sum_{k=S_T}^{T-1} k^{-b} \sqrt{Tb_{2,T} \left( \text{Var} \left( \tilde{\Gamma} (k) \right) \right)^{1/2}}^2
\]

\[
= T^{8q/5(2q+1)+2b/(2q+1)-1} S_T^{2(1-b)} O(1) + 0,
\]

since \( r > (b - 1)/ (b - 1) \) and \( Tb_{2,T} \text{Var} \left( \tilde{\Gamma} (k) \right) = O(1) \) as above. Combining equations (A.2.27) and (A.2.28) yields \( L_{2,T} \overset{p}{\rightarrow} 0 \), since \( \hat{\phi} (q) = O_p (1) \). Let us turn to \( L_{3,T} \). By Assumption 1.6-(iii) and \( |K_1 (\cdot)| \leq 1 \), we have,

\[
|L_{3,T}| \leq T^{8q/10(2q+1)} \sum_{k=S_T}^{T-1} C_3 k^{-l} \leq T^{(8q-1)/10(2q+1)} C_3 S_T^{1-l}
\]

\[
\leq C_3 T^{8q/10(2q+1)} T^{-4r(l-1)/5(2q+1)} \rightarrow 0,
\]

since \( r > q/ (l - 1) \). In view of (A.2.25)-(A.2.29) we deduce that \( A_{2,1,T} \overset{p}{\rightarrow} 0 \). Applying the same argument to \( A_{2,2,T} \), we have \( A_{2,T} \overset{p}{\rightarrow} 0 \). Using similar arguments, one has \( A_{3,T} \overset{p}{\rightarrow} 0 \). It remains to show that

\[
T^{8q/5(2q+1)} \left( \tilde{J}_T (b_{\theta_1,T}, \tilde{b}_{2,T}) - \tilde{J}_T (b_{\theta_1,T}, b_{\theta_2,T}) \right) \overset{p}{\rightarrow} 0.
\]
Let $\tilde{c}_{\theta_2,T} (rn_T/T, k)$ denote the estimator that uses $\hat{b}_{\theta_2,T}$ in place of $\hat{b}_{2,T}$. We have for $k \geq 0$,

$$
\tilde{c}_T (rn_T/T, k) - \tilde{c}_{\theta_2,T} (rn_T/T, k)
= (Tb_{\theta_2,T})^{-1} \sum_{s=k+1}^{r} \frac{1}{b_{2,T} ((r+1)n_T/T)} \left( K_2 \left( \frac{(r+1)n_T - (s+k/2)}{b_{2,T} ((r+1)n_T/T)} \right) - K_2 \left( \frac{(r+1)n_T - (s+k/2)}{b_{2,\theta_2,T}} \right) \right)
\times \tilde{V}_s \tilde{V}_s' - k
+ O_P \left( 1/Tb_{2,\theta_2,T} \right).
$$

Given Assumption 1.6-(v,vi) and the delta method, we have for $u = (r+1)n_T/T$:

$$
K_2 \left( \frac{(r+1)n_T - (s+k/2)}{b_{2,T} (u)} \right) - K_2 \left( \frac{(r+1)n_T - (s+k/2)}{b_{\theta_2,T} (u)} \right)
\leq C_4 \left( \frac{1}{Tb_{2,T} (u)} - \frac{1}{Tb_{\theta_2,T} (u)} \right)
= C_4 \left( \frac{3H \left( K_{2,\text{opt}} \right) D_1 (u)}{4F \left( K_{2,\text{opt}} \right)} \right)^{1/5} T^{1/5} \left( \frac{1}{\tilde{D}_2^{1/5} (u)} - \frac{1}{D_2^{1/5} (u)} \right) O_P (1/Tb_{\theta_2,T})
= C_4 \left( Tb_{\theta_2,T} (u) \right)^{-1/2} \left( \frac{3H \left( K_{2,\text{opt}} \right) D_1 (u)}{4F \left( K_{2,\text{opt}} \right)} \right)^{1/5} T^{1/5} \sqrt{Tb_{\theta_2,T} (u)}
\times \left( \frac{1}{\tilde{D}_2^{1/5} (u)} - \frac{1}{D_2^{1/5} (u)} \right) O_P (1/Tb_{\theta_2,T})
= C_4 \left( Tb_{\theta_2,T} (u) \right)^{-1/2} \left( \frac{3H \left( K_{2,\text{opt}} \right) D_1 (u)}{4F \left( K_{2,\text{opt}} \right)} \right)^{1/5} T^{1/5} O_P (1/O_P (1/Tb_{\theta_2,T})).
Therefore,

\[
T^{8q/5(2q+1)} \left( \hat{J}_T (b_{\theta_1,T}, \tilde{b}_{2,T}) - \hat{J}_T (b_{\theta_1,T}, \tilde{b}_{\theta_2,T}) \right) = T^{8q/5(2q+1)} \sum_{k=-T+1}^{T-1} K_1 (b_{\theta_1,T} k) \frac{n_T}{T} \sum_{r=0}^{[T/n_T]} \left( \hat{c} (rn_T/T, k) - \hat{c}_{\theta_2,T} (rn_T/T, k) \right) - T^{8q/5(2q+1)} b_{\theta_1,T} b_{\theta_1,T} \sum_{k=-T+1}^{T-1} K_1 (b_{\theta_1,T} k) \times \frac{n_T}{T} \sum_{r=0}^{[T/n_T]} C_4 T^{-1} \left( \frac{3H (K_2^{opt}) D_1 (u)}{4F (K_2^{opt})} \right) \frac{1}{5} \left( \frac{1}{5} \frac{O_P (1) + O_P (1/Tb_{2,\theta_2,T})}{O_P (1) + o_P (1)} \right) \rightarrow 0,
\]

since \( q \leq 2 \). This completes the proof of part (ii).

We now move to part (i). For some arbitrary \( \phi, \theta \in (0, \infty) \), \( \hat{J}_T (b_{\theta_1,T}, b_{\theta_2,T}) - J_T = o_P (1) \) by Theorem 1.3.2-(i) since \( \tilde{b}_{\theta_2,T} = O \left( T^{-1/5} \right) \) and \( q > 1/2 \) imply that \( \sqrt{Tb_{2,T}b_{1,T}} \rightarrow \infty \) holds. Hence, it remains to show \( \hat{J}_T (b_{\theta_1,T}, b_{\theta_2,T}) - \hat{J}_T (\tilde{b}_{1,T}, \tilde{b}_{2,T}) = o_P (1) \). Note that this result differs from the result of part (ii) only because the scale factor \( T^{8q/5(2q+1)} \) does not appear, Assumption 1.6-(ii) is replaced by part (i) of the same assumption, Assumption 1.6-(iii) is not imposed, and \( q > 1/2 \). Let

\[
\frac{17}{16} + \frac{5q}{8}, \frac{3 + 2q}{4}, 1
\]

and \( S_T \) be defined as in part (ii). We will use the decomposition in (A.2.18), and \( N_1 \) and \( N_2 \) as defined after (A.2.18). Let \( A_{1,T}, A_{2,T} \) and \( A_{3,T} \) be as in (A.2.19) without
the scale factor $T^{8q/10(2q+1)}$. Proceeding as in (A.2.20),

$$|A_{1,1,T}| \leq \sum_{k=1}^{S_T} C_2 \left| \tilde{b}_{1,T} - b_{\tilde{y}_1,T} \right| k \left| \tilde{\Gamma} (k) \right| \leq C \left| \phi \left( q \right) \right|^{1/(2q+1)} - \phi_{\theta^*}^{1/(2q+1)} \right| \times \left( \phi \left( q \right) \phi_{\theta^*}^{-1/(2q+1)} \right) \left( T \tilde{b}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \tilde{\Gamma} (k) \right| ,$$

for some $C < \infty$. By Assumption 1.6-(i),

$$\left| \phi \left( q \right) \right|^{1/(2q+1)} - \phi_{\theta^*}^{1/(2q+1)} \left( \phi \left( q \right) \phi_{\theta^*}^{-1/(2q+1)} \right) = O_p (1) .$$

Then, it suffices to show that $B_{1,T} + B_{2,T} + B_{3,T} \overset{p}{\rightarrow} 0$, where

$$B_{1,T} = \left( T \tilde{b}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \tilde{\Gamma} (k) - \tilde{\Gamma} (k) \right|$$

$$B_{2,T} = \left( T \tilde{b}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \tilde{\Gamma} (k) - \Gamma_{T} (k) \right|$$

$$B_{3,T} = \left( T \tilde{b}_{2,T} \right)^{-1/(2q+1)} \sum_{k=1}^{S_T} k \left| \Gamma_{T} (k) \right| .$$

By a mean-value expansion, we have

$$B_{1,T} \leq \left( T \tilde{b}_{2,T} \right)^{-1/(2q+1)} T^{-1/2} \sum_{k=1}^{S_T} \left| \frac{\partial}{\partial \beta} \tilde{\Gamma} (k) \big|_{\beta = \beta_0} \right| \sqrt{T} \left( \tilde{\beta} - \beta_0 \right)$$

$$\leq C \left( T \tilde{b}_{2,T} \right)^{-1/(2q+1)} \left( T b_{\theta^*} \right)^{2r/(2q+1)} T^{-1/2} \sup_{k \geq 1} \left| \frac{\partial}{\partial \beta} \tilde{\Gamma} (k) \big|_{\beta = \beta_0} \right| \sqrt{T} \left( \tilde{\beta} - \beta_0 \right) ,$$

since $r < 13/16 + 5q/8$, and $\sup_{k \geq 1} \left| \frac{\partial}{\partial \beta} \tilde{\Gamma} (k) \big|_{\beta = \beta_0} \right| = O_p (1)$ using (A.2.10) and
Assumption 1.3-(ii,iii). In addition,

\[
\mathbb{E} \left( B_{2,T}^2 \right) \leq \mathbb{E} \left( \left( T \hat{b}_{2,T} \right)^{-2/(2q+1)} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} k j \left| \Gamma(k) - \Gamma_T(k) \right| \left| \Gamma(j) - \Gamma_T(j) \right| \right) \tag{A.2.33}
\]

\[
\leq \mathbb{E} \left( \left( T \hat{b}_{2,T} \right)^{-2/(2q+1)} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} k j \left| \Gamma(k) - \Gamma_T(k) \right| \left| \Gamma(j) - \Gamma_T(j) \right| \right)
\]

\[
\leq \left( T \hat{b}_{2,T} \right)^{-2/(2q+1)} \sum_{k=1}^{S_T} \sum_{j=1}^{S_T} k j \left| \Gamma(k) - \Gamma_T(k) \right| \left| \Gamma(j) - \Gamma_T(j) \right| \]

\[
\leq \left( T \hat{b}_{2,T} \right)^{-2/(2q+1)} \sup_{k \geq 1} T \sigma_T \text{Var} \left( \Gamma(k) \right)
\]

\[
\leq \left( T \hat{b}_{2,T} \right)^{-2/(2q+1)} \left( T \sigma_T \right)^{4r/(2q+1)} \sup_{k \geq 1} T \sigma_T \text{Var} \left( \Gamma(k) \right)
\]

\[
\leq \left( \hat{T}_{2,T} \right)^{-2/(2q+1)} \left( T \sigma_T \right)^{4r/(2q+1)} \sup_{k \geq 1} T \sigma_T \text{Var} \left( \Gamma(k) \right) \to 0,
\]

given that \( \sup_{k \geq 1} T \sigma_T \text{Var} \left( \Gamma(k) \right) = O(1) \) by Proposition 1.3.1 and \( r < (3 + 2q)/4 \).

The bound in equation (A.2.24) is replaced by,

\[
B_{3,T} \leq \left( T \hat{b}_{2,T} \right)^{-1/(2q+1)} S_T \sum_{k=1}^{\infty} \left| \Gamma_T(k) \right| \tag{A.2.34}
\]

\[
\leq \left( T \hat{b}_{2,T} \right)^{\left( r-1 \right)/(2q+1)} O_{\mathbb{P}}(1) \to 0,
\]

since \( r < 1 \). This gives \( A_{1,T} \overset{\mathbb{P}}{\to} 0 \). Next, we show that \( A_{2,T} \overset{\mathbb{P}}{\to} 0 \). As above, let \( A_{2.1,T} = L_{1,T} + L_{2,T} + L_{3,T} \) where each summand is defined as in (A.2.25) without the factor \( T^{8q/10(2q+1)} \). Equation (A.2.26) is replaced by

\[
\left| L_{1,T} \right| = T^{-1/2} \sum_{k=S_T+1}^{T-1} C_1 \left( \hat{b}_{1,T} k \right)^{-b} \left| \frac{\partial}{\partial \beta'} \hat{\Gamma}(k) \right|_{\beta=\beta_0} \sqrt{T} \left( \hat{\beta} - \beta_0 \right) \tag{A.2.35}
\]

\[
= T^{-1/2+4b/5(2q+1)} \sum_{k=S_T+1}^{T-1} C_1 k^{-b} \left| \frac{\partial}{\partial \beta'} \hat{\Gamma}(k) \right|_{\beta=\beta_0} \sqrt{T} \left( \hat{\beta} - \beta_0 \right)
\]

\[
= T^{-1/2+4b/5(2q+1)+4r(1-b)/5(2q+1)} \left| \frac{\partial}{\partial \beta'} \hat{\Gamma}(k) \right|_{\beta=\beta_0} \sqrt{T} \left( \hat{\beta} - \beta_0 \right)
\]

\[
= T^{-1/2+4b/5(2q+1)+4r(1-b)/5(2q+1)} O(1) O_{\mathbb{P}}(1),
\]

which converges to zero since \( r > (8b - 10q - 5)/8(b - 1) \). Equation (A.2.27) is
replaced by

\[ |L_{2,T}| = \sum_{k=S_T+1}^{T-1} C_1 \left( \hat{b}_{1,T} k \right)^{-b} \left| \hat{\Gamma}(k) - \Gamma_T(k) \right| \]  \hspace{1cm} \text{(A.2.36)}

\[ = C_1 \left( q K_{1,q}^2 \phi(q) \right)^{b/(2q+1)} T^{b/(2q+1) - 1/2} \tilde{b}_{2,T}^{b/(2q+1) - 1/2} \times \left( \sum_{k=S_T+1}^{T-1} k^{-b} \sqrt{T \tilde{b}_{2,T}} \left| \hat{\Gamma}(k) - \Gamma_T(k) \right| \right) \]

and the bound in (A.2.28) is replaced by,

\[ \mathbb{E} \left( T^{b/(2q+1) - 1/2} \tilde{b}_{2,T}^{b/(2q+1) - 1/2} \sum_{k=r(T)}^{T-1} k^{-b} \sqrt{T \tilde{b}_{2,T}} \left| \hat{\Gamma}(k) - \Gamma_T(k) \right| \right)^2 \]  \hspace{1cm} \text{(A.2.37)}

\[ \leq T^{2b/(2q+1) - 1} \tilde{b}_{2,T}^{2b/(2q+1) - 1} \left( \sum_{k=r(T)}^{T-1} k^{-b} \sqrt{T \tilde{b}_{2,T}} \left( \text{Var} \left( \hat{\Gamma}(k) \right) \right)^{1/2} \right)^2 \]

\[ = T^{2b/(2q+1) - 1} \tilde{b}_{2,T}^{2b/(2q+1) - 1} \sum_{k=r(T)}^{T-1} k^{-b} O(1) \]

\[ = T^{2b/(2q+1) - 1} \tilde{b}_{2,T}^{2b/(2q+1) - 1} S_T^{2(1-b)} O(1) \to 0 \]

since \( r > (b - 1/2 - q) / (b - 1) \) and \( T \tilde{b}_{2,T} \text{Var} \left( \hat{\Gamma}(k) \right) = O(1) \) as above. Equations (A.2.36) and (A.2.37) combine to yield \( L_{2,T} \overset{p}{\to} 0 \), since \( \hat{\phi}(q) = O_p(1) \). Let us turn to \( L_{3,T} \). Equation (A.2.29) is replaced by,

\[ \left| \sum_{k=S_T+1}^{T-1} K_1 \left( \hat{b}_{1,T} k \right) \Gamma_T(k) \right| \leq \sum_{k=S_T+1}^{T-1} \frac{n_T}{T} \sum_{r=0}^{[T/n_T]} \left| c \left( r n_T / T, k \right) \right| \]  \hspace{1cm} \text{(A.2.38)}

\[ \leq \sum_{k=S_T+1}^{T-1} \sup_{u \in [0,1]} \left| c(u, k) \right| \to 0. \]

Equations (A.2.35)-(A.2.38) imply \( A_{2,1,T} \overset{p}{\to} 0 \). Thus, as in the proof of part (ii), we have \( A_{2,T} \overset{p}{\to} 0 \) and \( A_{3,T} \overset{p}{\to} 0 \). It remains to show that

\[ \left( \hat{J}_T \left( b_{\theta_1,T}, \hat{b}_{2,T} \right) - \hat{J}_T \left( b_{\theta_1,T}, \tilde{b}_{\theta_2,T} \right) \right) \overset{p}{\to} 0. \]
Let $\hat{c}_{\theta_2,T}(rn_T/T, k)$ be defined as in part (ii). We have for $k \geq 0$,

$$
\hat{c}_T(rn_T/T, k) - \hat{c}_{\theta_2,T}(rn_T/T, k)
$$

$$
= (Tb_{\theta_2,T})^{-1} \sum_{s=k+1}^{T} \left( K_2 \left( \frac{(r+1)n_T - (s + k/2))}{T} \right) - K_2 \left( \frac{(r+1)n_T - (s + k/2))}{b_{2,R}(u)} \right) 
\times \hat{V}_s \hat{V}_s' + O_P \left( 1/Tb_{\theta_2,T} \right).
$$

Given Assumption 1.6-(v,vi) and the delta method, we have for $u = (r+1)n_T/T$,

$$
K_2 \left( \frac{(r+1)n_T - (s + k/2))}{\hat{b}_{2,T}(u)} \right) - K_2 \left( \frac{(r+1)n_T - (s + k/2))}{b_{\theta_2,T}(u)} \right)
\leq C_4 \left( \frac{1}{Tb_{2,T}(u)} - \frac{1}{Tb_{\theta_2,T}(u)} \right)
\leq C_4 \left( \frac{3H(K_{2}^{opt})D_1(u)}{4F(K_{2}^{opt})} \right)^{1/5} T^{1/5} \left( \frac{1}{D_2^{1/5}(u)} - \frac{1}{D_2^{1/5}(u)} \right) o_p \left( 1/Tb_{\theta_2,T} \right)
\leq C_4 \left( \frac{3H(K_{2}^{opt})D_1(u)}{4F(K_{2}^{opt})} \right)^{1/5} T^{1/5} \left( \frac{D_2^{1/5}(u)}{D_2^{1/5}(u)} - \frac{D_2^{1/5}(u)}{D_2^{1/5}(u)} \right) o_p \left( 1/Tb_{\theta_2,T} \right)
\leq C_4 \left( \frac{3H(K_{2}^{opt})D_1(u)}{4F(K_{2}^{opt})} \right)^{1/5} O_P \left( 1 \right) o_p \left( 1/Tb_{\theta_2,T} \right).
Therefore,

\[
\begin{align*}
&\left( \hat{J}_T (b_{\theta_1, T}, \hat{b}_{2,T}) - \hat{J}_T (b_{\theta_1, T}, \bar{b}_{\theta_2, T}) \right) \\
&= \sum_{k=-T+1}^{T-1} K_1 (b_{\theta_1, T} - \bar{b}_{\theta_2, T}) \frac{n_T}{T} \sum_{r=0}^{[T/n_T]} \left( \hat{c} (rn_T/T, k) - \hat{c}_{\theta_2, T} (rn_T/T, k) \right) \\
&= \sum_{k=-T+1}^{T-1} K_1 (b_{\theta_1, T} - \bar{b}_{\theta_2, T}) \frac{n_T}{T} \sum_{r=0}^{[T/n_T]} \left( \hat{c} (rn_T/T, k) - \hat{c}_{\theta_2, T} (rn_T/T, k) \right) \\
&\times \frac{n_T}{T} \sum_{r=0}^{[T/n_T]} \left( \frac{3H (K_{2}^{opt}) D_1 (u)}{4F (K_{2}^{opt})} \right)^{1/5} T^{-3/5} O_{\mathbb{P}} (1) + O_{\mathbb{P}} \left( 1/Tb_{2,b_{\theta_2, T}} \right) \\
&= b_{\theta_1, T} b_{\theta_1, T} \sum_{k=-T+1}^{T-1} K_1 (b_{\theta_1, T} - \bar{b}_{\theta_2, T}) \frac{n_T}{T} \sum_{r=0}^{[T/n_T]} \left( \frac{3H (K_{2}^{opt}) D_1 (u)}{4F (K_{2}^{opt})} \right)^{1/5} O_{\mathbb{P}} (1) + o_{\mathbb{P}} (1) \\
&\overset{\mathbb{P}}{\rightarrow} 0,
\end{align*}
\]

which concludes the proof.

The result of part (iii) follows from the same argument as in Theorem 1.3.2-(iii) with references to Theorem 1.3.2-(i,ii) changed to Theorem 1.5.1-(i,ii). □

**A.2.5 Proofs of the Results in Section 1.7**

**A.2.5.1 Proof of Theorem 1.7.1**

We assume without loss of generality that \( m = 1 \) and provide the proof only for the single break case. Hence, the break date is \( T_2^0 \) (i.e., \( T_1^0 = 0 \) and \( T_3^0 = T \)). Note that by standard properties of approximation to Riemann sums, \( \Gamma (k) \overset{\mathbb{P}}{\rightarrow} \int_0^1 (c (u, k)) \, du \)
even when \( c(\cdot, k) \) has a finite number of discontinuities in \( u \), where

\[
\bar{\Gamma}(k) \triangleq \frac{n_T}{T - n_T} \sum_{r=0}^{[(T-n_T)/n_T]} c(rn_T/T, k).
\]

Let \( \tilde{J}_T \) denote the estimator \( \tilde{J}_T \) that uses the estimated break dates \( \hat{T}_2 \) in place of \( T_2^0 \). We need to show that \( \sqrt{Tb_{1,T}b_{2,T}} (\tilde{J}_T - \hat{J}_T) = o_P(1) \). Given the above-mentioned property of Riemann sums, this holds because \( \sqrt{Tb_{1,T}b_{2,T}n_T/\sqrt{T}} \to 0 \).

We now state a counterpart to Proposition 1.3.1 when \( m = 1 \).

**Lemma A.2.2.** Suppose \( V_{i,T} \) is Segmented Locally Stationary with zero mean where \( A(u, \omega) \) is twice continuously differentiable in \( u \) for all \( Tu \notin \mathcal{T} \) and twice continuously left-differentiable for \( u \in \mathcal{T} \). Suppose \( b_{2,T} \to 0 \) as \( T \to \infty \). For all \( u_0 \in \mathcal{T} \),

\[
E[\tilde{c}_T(u_0, k)] = c(u_0, k) + \frac{1}{2} b_{2,T}^2 \int_0^1 x^2 K_2(x) dx,
\]

\[
\times \int_{-\pi}^{\pi} \exp(i\omega k) (C_1(u_0, \omega) + C_2(u_0, \omega) + C_3(u_0, \omega)) d\omega
\]

\[
+ O\left(\frac{1}{Tb_{2,T}}\right) + o\left(b_{2,T}^2\right),
\]

where

\[
C_1(u_0, \omega) = 2 \frac{\partial A_1(u_0, -\omega)}{\partial_u} \frac{\partial A_2(v, \omega)}{\partial_v}
\]

\[
C_2(u_0, \omega) = \frac{\partial^2 A_2(v, \omega)}{\partial v^2} A_1(u_0, -\omega)
\]

\[
C_3(u_0, \omega) = \frac{\partial^2 A_1(u_0, \omega)}{\partial u^2} A_2(v, \omega),
\]

and \( v = u_0 + k/2T \). For all \( u_0 \notin \mathcal{T} \),
\[
\mathbb{E} [\hat{c}_T (u_0, k)] = c(u_0, k) + \frac{1}{2} b_{2,T}^2 \int_{-1/2}^{1/2} x^2 K_2 (x) \, dx \left[ \frac{\partial^2}{\partial^2 u} c(u_0, k) \right] \\
+ o \left( b_{2,T}^2 \right) + O \left( 1 \left/ (b_{2,T} T) \right\) ,
\]

and

\[
\text{Var} [\hat{c}_T (u_0, k)] = \frac{1}{T b_{2,T}} \int_{-1/2}^{1/2} K_2 (x) \, dx \times \sum_{l=-\infty}^{\infty} c(u_0, l) \left[ c(u_0, l) + c(u_0, l + 2k) \right] + o \left( 1 \left/ (b_{2,T} T) \right\)
\]

For all \( u_0 \in (0, 1) \),

\[
\lim_{T \to \infty} b_{2,T}^{-2} \mathbb{E} [\hat{c}_T (u_0, k) - c(u_0, k)] < \infty,
\]

and if further it holds that \( b_{2,T}^{5/2} T^{1/2} \to \infty \), then

\[
\lim_{T \to \infty} T b_{2,T} \text{Var} [\hat{c}_T (u_0, k)] < \infty.
\]

Furthermore, we have \( \hat{c}_T (u_0, k) - c(u_0, k) = O_p \left( \sqrt{T b_{2,T}} \right) \) for all \( u_0 \in (0, 1) \).

**Proof of Lemma A.2.2.** If \( T u_0 \notin \mathcal{T} \) then the result follows from Proposition 1.3.1.

Suppose \( T u_0 \in \mathcal{T} \) and \( k \geq 0 \). We omit the subscript \( j \) from \( A_{j,s-k,T}^0 (\omega) \) and from \( A_{j} ((s - k) / T, \omega) \) since the value \( j \) is determined by \( s - k \) and can thus be omitted.

Using (1.2.2) we have,

\[
\mathbb{E} [\hat{c}_T (u_0, k)] = \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right) \\
\times \int_{-\pi}^{\pi} \exp (i \omega k) A_{s-k,T}^0 (\omega) A_{s,T}^0 (-\omega) \, d\omega.
\]

Since \( K_2 (x) = 0 \) for \( x < 0 \), the above sum runs up to \( s = u_0 - k/2 \). Hence, the
behavior of $A^0_{s,T}(\omega)$ only matters on a left neighborhood of $u_0$. Using (1.2.4) we have,

$$
E [\tilde{c}_T(u_0, k)] = \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right) \\
\times \int_{-\pi}^{\pi} \exp(i\omega k) A \left( \frac{s - k}{T}, \omega \right) A \left( \frac{s}{T}, -\omega \right) d\omega \\
+ O \left( T^{-1} \right).
$$

By the definition of $f(\cdot, \cdot)$, it follows that,

$$
E [\tilde{c}_T(u_0, k)] = \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right) \\
\times \int_{-\pi}^{\pi} \exp(i\omega k) f \left( \frac{s - k/2}{T}, \omega \right) d\omega + O \left( T^{-1} \right).
$$

By the definition of $f(\cdot, \cdot)$, it follows that,

$$
E [\tilde{c}_T(u_0, k)] = \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right) \\
\times \int_{-\pi}^{\pi} \exp(i\omega k) f \left( \frac{s - k/2}{T}, \omega \right) d\omega + O \left( T^{-1} \right).
$$

Let $u_{\epsilon,T} \triangleq u_0 - \epsilon_T$, where $\epsilon_T > 0$. Since $f(u_{\epsilon,T}, \omega)$ is twice differentiable, by taking a
second-order Taylor’s expansion of $f$ around $u_{\epsilon,T}$ we have

$$
\mathbb{E} \left[ \tilde{c}_T (u_0, k) \right] = \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right) \int_{-\pi}^{\pi} \exp(i \omega k) f \left( u_{\epsilon,T}, \omega \right) d\omega \\
+ \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right) \\
+ \int_{-\pi}^{\pi} \exp(i \omega k) \frac{\partial f \left( u_{\epsilon,T}, \omega \right)}{\partial u} \left( \frac{s - k/2}{T} - u_{\epsilon,T} \right) d\omega \\
+ \frac{1}{2} \frac{1}{b_{2,T} T} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right) \\
+ \int_{-\pi}^{\pi} \exp(i \omega k) \frac{\partial^2 f \left( u_{\epsilon,T}, \omega \right)}{\partial u^2} \left( \frac{s - k/2}{T} - u_{\epsilon,T} \right)^2 d\omega \\
+ o \left( b_{2,T}^2 \right) + O \left( \frac{1}{T b_{2,T}} \right).
$$

Using $\int_0^1 K_2 (x) \, dx = 1$, $K_2 (x) = K_2 (1 - x)$ and the definition of $c \left( u_{\epsilon,T}, k \right)$, the right-hand side above is equal to

$$
c \left( u_{\epsilon,T}, k \right) = \frac{1}{2} b_{2,T} \int_0^1 x^2 K_2 (x) \, dx \int_{-\pi}^{\pi} \exp(i \omega k) \frac{\partial^2 f \left( u_{\epsilon,T}, \omega \right)}{\partial u^2} d\omega \\
+ O \left( \frac{1}{T b_{2,T}} \right) + o \left( b_{2,T}^2 \right).
$$

Since $c \left( u_{\epsilon,T}, k \right)$ and $\partial^2 f \left( u_{\epsilon,T}, \omega \right) / \partial u^2$ are left-Lipschitz continuous,

$$
c \left( u_{\epsilon,T}, k \right) - c \left( u_0, k \right) = O_{\mathbb{P}} \left( \epsilon_T \right), \quad \frac{\partial^2 f \left( u_{\epsilon,T}, \omega \right)}{\partial u^2} - \frac{\partial^2 f \left( u_0, \omega \right)}{\partial u^2} = O_{\mathbb{P}} \left( \epsilon_T \right),
$$

where $\partial^2 f \left( u_0, \omega \right) / \partial u^2$ denote the second left derivative of $f$ in $u$ at $u_0$. Choose $\epsilon_T$ such that $\epsilon_T = o_{\mathbb{P}} \left( b_{2,T}^2 \right)$. Then,

$$
\mathbb{E} \left[ \tilde{c}_T \left( u_0, k \right) - c \left( u_0, k \right) \right] = \frac{1}{2} b_{2,T} \int_0^1 x^2 K_2 (x) \, dx \int_{-\pi}^{\pi} \exp(i \omega k) \frac{\partial^2 f \left( u_0, \omega \right)}{\partial u^2} d\omega \\
+ O \left( \frac{1}{T b_{2,T}} \right) + o \left( b_{2,T}^2 \right).
$$
Next, let us consider $\text{Var} \left[ \tilde{c}_T (u_0, k) \right]$. We begin with

$$\text{Var} \left[ \tilde{c}_T (u_0, k) \right]$$

\[
= \frac{1}{Tb_2 T} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_2 T} \right) \sum_{t=k+1}^{T} K_2 \left( \frac{u_0 - (t + k/2) / T}{b_2 T} \right) \\
\times \left( \int_{-\pi}^{\pi} \exp (i\omega k) A_{s-k,T}^0 (\omega) A_{s,T}^0 (-\omega) d\omega \right) \\
\times \left( \int_{-\pi}^{\pi} \exp (i\omega k) A_{t-k,T}^0 (\omega) A_{t,T}^0 (-\omega) d\omega \right) \\
= \frac{1}{Tb_2 T} \sum_{s=k+1}^{T} K_2 (x) \sum_{t=k+1}^{T} K_2 \left( \frac{u_0 - (t - s + s + k/2) / T}{b_2 T} x + (s - t) / T b_2 T \right) \\
\times \left( \int_{-\pi}^{\pi} \exp (i\omega k) A_{s-k,T}^0 (\omega) A_{s,T}^0 (-\omega) d\omega \right) \\
\times \left( \int_{-\pi}^{\pi} \exp (i\nu k) A_{t-s+s-k,T}^0 (\nu) A_{t-s+s,T}^0 (-\nu) d\nu \right) \epsilon
\]

Proceeding as in (A.2.1),

\[
\text{Var} \left[ \tilde{c}_T (u_0, k) \right] = \int_{0}^{1} K_2 (x)^2 dx \left\{ \sum_{l=0}^{\infty} [c (u_0, l) c (u_0, l) + c (u_0, l) c (u_0, l + 2k)] \\
+ \sum_{l=-\infty}^{-1} [c (u_0, l) c (u_0, l) + c (u_0, l) c (u_0, l + 2k)] \right\}, \quad \text{(A.2.39)}
\]

where $c (u_0, \cdot)$ in the second line above takes the form [cf. the definition of $c (u_0, l)$ for $l < 0$ at the end of Section 1.2.1],

\[
c (u_0, l) = \int_{-\pi}^{\pi} \exp (i\omega l) A_2 (u_0, \omega) A_1 (u_0 + l/T, \omega) d\omega.
\]
The latter results from applying the approximation (1.2.2) to
\[
\int_{-\pi}^{\pi} \exp (i\omega l) A_{T_{u_0}-T_{v},T}^{0} (\omega) A_{T_{u_0}-T_{v},T}^{0} (-\omega) \, d\omega, \quad k < 0, \quad s = T_{u_0}, \quad t = T_{v}, \quad -T_{v} > k,
\]
with the changes in variables \(h = t - s\) and \(l = h - k\).

It remains to consider the case \(T_{u_0} \in \mathcal{T}\) and \(k < 0\). The derivations for \(\text{Var} [\tilde{c}_T (u_0, k)]\) follow the same logic although the arguments used for the summation in the second line of (A.2.39) now should be applied to the first sum. The derivations for the bias expression are different. Again, using (1.2.2) we have,
\[
\mathbb{E} [\tilde{c}_T (u_0, k)] = \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right)
\times \int_{-\pi}^{\pi} \exp (i\omega k) A_{s-k,T}^{0} (\omega) A_{s,T}^{0} (-\omega) \, d\omega.
\]
The symmetry of the kernel and (1.2.4) yield,
\[
\mathbb{E} [\tilde{c}_T (u_0, k)] = \frac{1}{T b_{2,T}} \sum_{s=k+1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} \right)
\int_{-\pi}^{\pi} \exp (i\omega k) A_2 \left( \frac{s - k}{T}, \omega \right) A_1 \left( \frac{s}{T}, -\omega \right) \, d\omega
\quad + O \left( T^{-1} \right).
\]
We cannot use the property \(f_j (u, \omega) = |A_j (u, \omega)|^2\) for \(T_{j-1} / T < u = t / T \leq T_{j} / T\) \((j = 1, 2)\) because now \(s - k > s\) (i.e., \(A_2 ((s - k) / T, \omega) A_1 (s/T, -\omega)\) cannot be approximated by \(f (s - k/2, \omega)\)). However, by taking a second-order Taylor’s expansion
of $A_1$ about $u_0 - \epsilon_{1,T}$ and of $A_2$ about $v + \epsilon_{2,T}$ where $\epsilon_{1,T}, \epsilon_{2,T} > 0$ we have

$$
\mathbb{E} \left[ \hat{c}_T (u_0, k) \right]
= \frac{1}{T b_{2,T}^2} \sum_{s = k + 1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} b_{2,T} \right)
\times \int_{-\pi}^{\pi} \exp (i \omega k) A_2 (v + \epsilon_{2,T}, \omega) A_1 (u_0 - \epsilon_{1,T}, -\omega) \, d\omega
+ \frac{1}{T b_{2,T}^2} \sum_{s = k + 1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} b_{2,T} \right)
\times \int_{-\pi}^{\pi} \exp (i \omega k) \left[ \frac{\partial A_2 (v + \epsilon_{2,T}, \omega)}{\partial v} A_1 (u_0 - \epsilon_{1,T}, -\omega) \left( \frac{s - k}{T} - v - \epsilon_{2,T} \right) + \frac{\partial A_1 (u_0 - \epsilon_{1,T}, -\omega) A_2 (v, \omega)}{\partial u} \left( \frac{s}{T} - u_0 + \epsilon_{1,T} \right) \right] d\omega
+ \frac{1}{2 b_{2,T}^2 T} \sum_{s = k + 1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} b_{2,T} \right)
\int_{-\pi}^{\pi} \exp (i \omega k) \left[ \frac{\partial^2 A_2 (v + \epsilon_{2,T}, \omega)}{\partial v^2} A_1 (u_0, -\omega) \left( \frac{s - k}{T} - v \right)^2 + \frac{\partial^2 A_1 (u_0 - \epsilon_{1,T}, -\omega) A_2 (v, \omega)}{\partial u^2} \left( \frac{s}{T} - u_0 + \epsilon_{1,T} \right)^2 \right] d\omega
+ \frac{1}{b_{2,T}^2 T} \sum_{s = k + 1}^{T} K_2 \left( \frac{u_0 - (s + k/2) / T}{b_{2,T}} b_{2,T} \right)
\times \int_{-\pi}^{\pi} \exp (i \omega k)
\left[ \frac{\partial A_2 (v + \epsilon_{2,T}, \omega)}{\partial v} \frac{\partial A_1 (u_0 - \epsilon_{1,T}, -\omega)}{\partial u} \left( \frac{s - k}{T} - v - \epsilon_{2,T} \right) \left( \frac{s}{T} - u_0 + \epsilon_{1,T} \right) \right] d\omega
+ o \left( b_{2,T}^2 \right).
$$

Since $A_1 (u_0 - \epsilon_{1,T}, -\omega)$ and $\partial^2 A_1 (u_0 - \epsilon_{1,T}, -\omega) / \partial u^2$ are left-continuous at $u_0$ and
$A_2(v + \epsilon_{2,T}, \omega)$ and $\partial^2 A_2(v + \epsilon_{2,T}, \omega) / \partial v^2$ are right-continuous at $v$,

\[
A_1(u_0 - \epsilon_{1,T}, -\omega) - A_1(v, -\omega) = O_P(\epsilon_{1,T}),
\]

\[
\frac{\partial A_1(u_0 - \epsilon_{1,T}, -\omega)}{\partial u} - \frac{\partial A_1(u_0, -\omega)}{\partial u} = O_P(\epsilon_{1,T}),
\]

\[
\frac{\partial^2 A_1(u_0 - \epsilon_{1,T}, -\omega)}{\partial u^2} - \frac{\partial^2 A_1(u_0, -\omega)}{\partial u^2} = O_P(\epsilon_{1,T}),
\]

\[
A_2(v + \epsilon_{2,T}, \omega) - A_2(v, \omega) = O_P(\epsilon_{2,T}),
\]

\[
\frac{\partial A_2(v + \epsilon_{2,T}, \omega)}{\partial v} - \frac{\partial A_2(v, \omega)}{\partial v} = O_P(\epsilon_{2,T}),
\]

\[
\frac{\partial^2 A_2(v + \epsilon_{2,T}, \omega)}{\partial v^2} - \frac{\partial^2 A_2(v, \omega)}{\partial v^2} = O_P(\epsilon_{2,T}),
\]

where $\partial A_2(v, \omega) / \partial v$ (resp., $\partial^2 A_2(v, \omega) / \partial v^2$) denote the first (resp., second) right derivative of $f$ in the first argument at $v$. Choose $\epsilon_{1,T}$ and $\epsilon_{2,T}$ such that $\epsilon_{1,T} = \sigma_{\omega} \left( b_{2,T}^2 \right)$ and $\epsilon_{2,T} = \sigma_{\omega} \left( b_{2,T}^2 \right)$. Using the definition of $c(u, k)$ for $k < 0$, the right-hand side of (A.2.40) is equal to

\[
c(u_0, k) + b_{2,T}^2 \int_0^1 x^2 K_2(x) dx \int_{-\pi}^{\pi} \exp(i\omega k) \times (C_1(u_0, \omega) + C_2(u_0, \omega) + C_3(u_0, \omega)) d\omega + O \left( \frac{1}{T b_{2,T}} \right) + o \left( b_{2,T}^2 \right).
\]

As in the proof of Proposition 1.3.1, basic manipulations lead to bound for the MSE. Then, consistency and the rate of convergence follow from the same arguments. □

Since the results of Lemma A.2.2 about the order of the bias and variance of $\tilde{c}_T(u_0, k)$ are very similar to their counterpart results in Lemma 1.3.1, the proof of Lemma A.2.1 can be repeated with the following changes. When $l = T_0^0$, $k > 0$ and $\tau_2 < 0$, the relationship

\[
\Gamma_{(l - \tau_2)/T}(k) - \Gamma_{l/T}(k) = O_P \left( \left| l/T - (l - \tau_2) / T \right| + T^{-1} \right) = O_P(\tau_2/T),
\]
does not hold because of the discontinuity in the spectrum of \( \{V_{t,T}\} \) at time \( t = T_2^0 \). The same applies to the relationship \( \Gamma_{(s-\tau_1)/T} (k) = \Gamma_{s/T} (k) + \mathcal{O}_T (\tau_1/T) \) when \( s = T_2^0, k \geq 0 \) and \( \tau_1 < 0 \) or when e.g., \( s = T_2^0, k < 0 \), \( \tau_1 \geq 0 \). Thus, one has to carry \( \Gamma_{(t-\tau_2)/T} (k) \) along the proof. The approximations of the form of (A.2.6) still go through and the proofs thus follow with minor changes. This leads to the asymptotic MSE results corresponding to Theorem 1.3.1. Next, \( \sqrt{Tb_{1,T}b_{2,T}} (\hat{J}_T - \tilde{J}_T) = \mathcal{O}_T (1) \) follows using similar arguments as in the proof of \( \sqrt{Tb_{1,T}b_{2,T}} (\tilde{J}_T - \tilde{J}_T) = \mathcal{O}_T (1) \) with references to Proposition 1.3.1 replaced by references to Lemma A.2.2. Thus, we have \( \sqrt{Tb_{1,T}b_{2,T}} (\hat{J}_T - \tilde{J}_T) = \mathcal{O}_T (1) \). With the results of Lemma A.2.2, the proofs of \( \sqrt{Tb_{1,T}b_{2,T}} (\hat{J}_T - J_T) = \mathcal{O}_T (1) \) and of \( \sqrt{Tb_{1,T}} (\hat{J}_T - \tilde{J}_T) = \mathcal{O}_T (1) \) follow the same steps. Thus, Theorem 1.3.2 continues to hold when \( m = 1 \). This implies that the asymptotic MSE of \( \hat{J}_T \) is asymptotically equivalent to that of \( \tilde{J}_T \) also when \( m = 1 \). □

### A.3 Additional Monte Carlo Results

This section presents additional Monte Carlo results about the size and power of HAR inference tests. We consider \( t \)-tests as well as \( F \)-tests for \( m_F \) restrictions \( \beta = \beta_0 \), given by

\[
F = (T-p)\hat{\nabla}^T J_T \hat{\nabla} / m_F \quad \text{where} \quad \hat{\nabla} = T^{-1} \sum_{t=1}^T V_t \left( \hat{\beta} \right) \quad \text{with} \quad V_t \left( \hat{\beta} \right) = z_t \left( y_t - z_t^T \hat{\beta} \right) \quad \text{and} \quad z_t = \left( 1 \ x_t \right)'.
\]

Model M1 is given by \( e_t = 0.4e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N} (0, 1) \) for \( t < 4T/5 \) and \( e_t = 0.6e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N} (0, 1) \) for \( t \geq 4T/5 \), and \( x_t = 0.2x_{t-1} + u_{X,t} \) with \( u_{X,t} \sim \text{i.i.d.} \mathcal{N} (0, 1) \). Model M2 involves locally stationary errors \( e_t = \rho_t e_{t-1} - 0.5e_{t-2} + u_t, u_t \sim \text{i.i.d.} \mathcal{N} (0, 1) \) with \( \rho_t = -0.7 \cos (1.5 - \cos (4\pi t/T)) \) and \( x_t \sim \text{i.i.d.} \mathcal{N} (1, 1) \) for \( t \leq T/8 \) and \( x_t = 0.7x_{t-1} + 2u_{X,t}, u_{X,t} \sim \text{i.i.d.} \mathcal{N} (0, 1) \) for \( t > T/8 \). Model M3 is given by \( e_t = 0.1e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N} (0, 1) \) for \( t < 4T/5 \) and \( e_t = 0.7e_{t-1} + 2u_t, u_t \sim \text{i.i.d.} \mathcal{N} (0, 1) \) for \( t \geq 4T/5 \), and \( x_t = \mu_X + \rho x_{t-1} + u_{X,t} \)
with \( \mu_X = 1, \rho = 0.2 \) and \( u_{X,t} \sim \text{i.i.d.} \mathcal{N}(0, 1) \). Model M3 is given by

\[
y_t = \beta_0^{(1)} + \sum_0 \beta_0^{(2)} x_t + w_t 1 \{ t \in (T/2 - T/20, T/2) \} + e_t, \quad t = 1, \ldots, T,
\]

where \( e_t = 0.2e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \), \( x_t = 5 + 0.2x_{t-1} + u_{X,t} \) with \( u_{X,t} \sim \text{i.i.d.} \mathcal{N}(0, 1) \), and \( w_t \) has the same distribution as \( x_t \) but it is independent from \( x_t \). Model M4 is given by \( e_t = 0.1e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \) for \( t < 4T/5 \) and \( e_t = 0.7e_{t-1} + 2u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \) for \( t \geq 4T/5 \), and \( x_t = \mu_X + \rho x_{t-1} + u_{X,t} \) with \( \mu_X = 1, \rho = 0.2 \) and \( u_{X,t} \sim \text{i.i.d.} \mathcal{N}(0, 1) \). We run a \( t_1 \)-test in M1-M2 and M4, and a \( t_2 \)-test for M3.

Let us turn to the models for the \( F \)-test. Model M5 follows (1.8.1) where \( x_t \sim \text{i.i.d.} \mathcal{N}(1, 1) \) for \( t \leq T/4 \) and \( x_t = 2 + 0.2x_{t-1} + u_{X,t} \) for \( t > T/4 \) with \( u_{X,t} \sim \text{i.i.d.} \mathcal{N}(0, 1) \), and \( e_t = u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \) for all \( t \) except \( T/5 \leq t \leq 3T/5 \) where \( e_t = 0.6e_{t-1} + u_t \) with \( u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \). Model M6 involves segmented locally stationary processes \( e_t = \rho_t e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1), \rho_t = -0.8(\cos (1.5 - \cos (10t/T))) \) for \( t \leq 4T/5 - 1, e_t = 0.4e_{t-1} + 2v_t, v_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \) for \( t \geq 4T/5 \) and \( x_t = 1 + 0.2x_{t-1} + u_{X,t} \) with \( u_{X,t} \sim \text{i.i.d.} \mathcal{N}(0, 1) \).

We report the power function for model M1 and M3. We do not discussed the power of M2 (because it is similar to the ones discussed in the text (cf. Model S1-S2)) or M4 (because the classical HAR inference tests are oversized as in model S5 in the Chapter 1). Model M7 (for \( t_1 \)-test) and M8 (for \( F \)-test) are similar to model P5. M7 is given by

\[
y_t = \beta_0^{(1)} + \delta + \sum_0 \beta_0^{(2)} x_t + \rho_t 1 \{ t \geq 4.5T/5 \} + e_t, \quad t = 1, \ldots, T,
\]

where \( \rho_t = 2\delta (t - 4.5T/5) / T, e_t = 0.5e_{t-1} + u_t, u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \), and \( x_t = 1 + 0.2x_{t-1} + u_{X,t}, u_{X,t} \sim \text{i.i.d.} \mathcal{N}(0, 1) \). For \( t > 4.5T/5 \) the intercept increases
slowly in small increments of magnitude $\rho_t$. Model M8 is similar to M7 but involves $e_t = 0.6e_{t-1} + 2v_t$, $v_t \sim \text{i.i.d. } \mathcal{N}(0, 1)$ for $T/5 \leq t \leq 3T/5$, $e_t = 0.2e_{t-1} + 2v_t$, $v_t \sim \text{i.i.d. } \mathcal{N}(0, 1)$ for all other $t$, and $x_t \sim \text{i.i.d. } \mathcal{N}(1, 1)$ for $1 \leq t \leq T/4 - 1$ and $x_t = 2 + 0.2x_{t-1} + 2u_{X,t}$, $u_{X,t} \sim \text{i.i.d. } \mathcal{N}(0, 1)$ for all other $t$. For M7-M8 we plot the power functions in Figure A.1-A.2. Model M9 is the same as M7 but we compute an $F$-test.

In models in which we run a $F$-test we set $\beta_0^{(1)} = 0$ and $\beta_0^{(2)} = 0$. The results are similar as the ones discussed in the main text. In particular, also the $F$-test suffers of the problems mentioned in the main text.

Table A.1: Empirical small-sample size of the $t_1$-test for model M1-M2

<table>
<thead>
<tr>
<th>$\alpha = 0.05$</th>
<th>Model M1</th>
<th>Model M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 125$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.028</td>
<td>0.048</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel</td>
<td>0.042</td>
<td>0.061</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel, auto, no breaks</td>
<td>0.058</td>
<td>0.058</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel, auto</td>
<td>0.067</td>
<td>0.063</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.079</td>
<td>0.077</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.058</td>
<td>0.055</td>
</tr>
<tr>
<td>Newey-West (1987), &quot;rule&quot;</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.137</td>
<td>0.111</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.088</td>
<td>0.041</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-b</td>
<td>0.031</td>
<td>0.023</td>
</tr>
<tr>
<td>EWP</td>
<td>0.035</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Table A.2: Empirical small-sample size of $t$-test for model M3-M4

<table>
<thead>
<tr>
<th>$\alpha = 0.05$</th>
<th>Model M3, $t_2$</th>
<th>Model M4, $t_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 125$</td>
<td>$T = 200$</td>
</tr>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.060</td>
<td>0.082</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel</td>
<td>0.054</td>
<td>0.079</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel, auto, no breaks</td>
<td>0.048</td>
<td>0.080</td>
</tr>
<tr>
<td>$\tilde{J}_T$, QS kernel, auto</td>
<td>0.600</td>
<td>0.068</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.040</td>
<td>0.043</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.022</td>
<td>0.032</td>
</tr>
<tr>
<td>Newey-West (1987), &quot;rule&quot;</td>
<td>0.042</td>
<td>0.050</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.038</td>
<td>0.049</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.030</td>
<td>0.036</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-b</td>
<td>0.016</td>
<td>0.023</td>
</tr>
<tr>
<td>EWP</td>
<td>0.022</td>
<td>0.020</td>
</tr>
</tbody>
</table>
### Table A.3: Empirical small-sample size of $F$-test for model M5-M6

<table>
<thead>
<tr>
<th>$\alpha = 0.05$</th>
<th>Model M5</th>
<th></th>
<th>Model M6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 125$</td>
<td>$T = 200$</td>
<td>$T = 400$</td>
<td>$T = 125$</td>
</tr>
<tr>
<td>$J_T$, Bartlett</td>
<td>0.048</td>
<td>0.048</td>
<td>0.042</td>
<td>0.020</td>
</tr>
<tr>
<td>$J_T$, QS</td>
<td>0.048</td>
<td>0.049</td>
<td>0.043</td>
<td>0.021</td>
</tr>
<tr>
<td>$J_T$, QS, auto, no breaks</td>
<td>0.046</td>
<td>0.050</td>
<td>0.038</td>
<td>0.067</td>
</tr>
<tr>
<td>$J_T$, QS, auto</td>
<td>0.038</td>
<td>0.041</td>
<td>0.041</td>
<td>0.058</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.109</td>
<td>0.092</td>
<td>0.081</td>
<td>0.029</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.087</td>
<td>0.081</td>
<td>0.078</td>
<td>0.000</td>
</tr>
<tr>
<td>Newey-West (1987), “rule”</td>
<td>0.118</td>
<td>0.105</td>
<td>0.089</td>
<td>0.032</td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.144</td>
<td>0.125</td>
<td>0.106</td>
<td>0.054</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.114</td>
<td>0.098</td>
<td>0.086</td>
<td>0.000</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.053</td>
<td>0.044</td>
<td>0.047</td>
<td>0.002</td>
</tr>
<tr>
<td>EWP</td>
<td>0.052</td>
<td>0.040</td>
<td>0.047</td>
<td>0.004</td>
</tr>
</tbody>
</table>

### Table A.4: Empirical small-sample rejection rates of $t_1$-test for model M1

<table>
<thead>
<tr>
<th>$\alpha = 0.05$, $T = 200$</th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
<th>$\delta = 1.6$</th>
<th>$\delta = 2.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.126</td>
<td>0.324</td>
<td>0.799</td>
<td>0.980</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel</td>
<td>0.160</td>
<td>0.375</td>
<td>0.825</td>
<td>0.985</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no breaks</td>
<td>0.169</td>
<td>0.394</td>
<td>0.826</td>
<td>0.982</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.155</td>
<td>0.360</td>
<td>0.815</td>
<td>0.982</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.085</td>
<td>0.302</td>
<td>0.780</td>
<td>0.953</td>
<td>0.998</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.062</td>
<td>0.227</td>
<td>0.717</td>
<td>0.944</td>
<td>0.997</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.132</td>
<td>0.380</td>
<td>0.795</td>
<td>0.963</td>
<td>0.998</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.097</td>
<td>0.321</td>
<td>0.781</td>
<td>0.945</td>
<td>0.998</td>
</tr>
<tr>
<td>EWP</td>
<td>0.094</td>
<td>0.320</td>
<td>0.771</td>
<td>0.945</td>
<td>0.997</td>
</tr>
</tbody>
</table>

### Table A.5: Empirical small-sample rejection rates of $t_2$-test for model M3

<table>
<thead>
<tr>
<th>$\alpha = 0.05$, $T = 200$</th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.274</td>
<td>0.654</td>
<td>0.984</td>
</tr>
<tr>
<td>$J_T$, QS kernel</td>
<td>0.258</td>
<td>0.640</td>
<td>0.991</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no breaks</td>
<td>0.270</td>
<td>0.652</td>
<td>0.986</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.218</td>
<td>0.582</td>
<td>0.982</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.286</td>
<td>0.664</td>
<td>0.962</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.192</td>
<td>0.580</td>
<td>0.964</td>
</tr>
<tr>
<td>Newey-West (1987), auto, prewhite</td>
<td>0.220</td>
<td>0.594</td>
<td>0.970</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-$b$</td>
<td>0.210</td>
<td>0.604</td>
<td>0.938</td>
</tr>
<tr>
<td>EWP</td>
<td>0.191</td>
<td>0.581</td>
<td>0.922</td>
</tr>
</tbody>
</table>
Table A.6: Empirical small-sample rejection rates of $F$-test for model M9

<table>
<thead>
<tr>
<th>$\alpha = 0.05$, $T = 200$</th>
<th>$\delta = 0$ (size)</th>
<th>$\delta = 0.2$</th>
<th>$\delta = 0.4$</th>
<th>$\delta = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, Bartlett kernel</td>
<td>0.063</td>
<td>0.795</td>
<td>0.940</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel</td>
<td>0.073</td>
<td>0.825</td>
<td>0.960</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto, no breaks</td>
<td>0.059</td>
<td>0.740</td>
<td>0.923</td>
<td>1.000</td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>0.058</td>
<td>0.785</td>
<td>0.950</td>
<td>1.000</td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.103</td>
<td>0.315</td>
<td>0.515</td>
<td>0.954</td>
</tr>
<tr>
<td>Andrews (1991), auto, prewhite</td>
<td>0.060</td>
<td>0.301</td>
<td>0.530</td>
<td>0.956</td>
</tr>
<tr>
<td>Newey-West (1987), prewhite</td>
<td>0.189</td>
<td>0.330</td>
<td>0.530</td>
<td>0.935</td>
</tr>
<tr>
<td>Newey-West (1987), fixed-b</td>
<td>0.065</td>
<td>0.070</td>
<td>0.160</td>
<td>0.665</td>
</tr>
<tr>
<td>EWP</td>
<td>0.053</td>
<td>0.100</td>
<td>0.120</td>
<td>0.652</td>
</tr>
</tbody>
</table>

Figure A.1: Power functions of $t$-test for Model M7 with $T = 200$. 
Figure A.2: Power functions of $F$-test for Model M8 with $T = 200$.

A.4 Empirical Application

We consider the stability of the predictive ability of the Phillips curve when used as a forecast model for inflation. We consider the $t$-test for forecast failure of Giacomini and Rossi (2009) normalized by different HAC estimators. Forecasting inflation via the Phillips curve has been common in applied work and the predicative ability of the the Phillips curve for inflation has always attracted a great deal of attention. We consider two versions of this forecast model. First, similar to Perron and Yamamoto (2018), we apply the Phillips curve model to inflation measured in levels. In addition, we apply the same model to first-differences of inflation, as used also by Giacomini and Rossi (2009). Let $\pi_t^\tau = (1200/\tau) \ln (P_t/P_{t-\tau})$ denote the $\tau$-period inflation in the price level $P_t$ reported at an annual rate, and $u_t$ denote the unemployment gap (i.e., the difference between the unemployment rate and a measure of the NAIRU). The Phillips curve relates changes in inflation to past values of the unemployment gap.
and to past changes in inflation:

$$\pi_{t+\tau} - \pi_t = \theta_0 + \theta_1 (L) u_t + \theta_2 (L) (\pi_t - \pi_{t-1}) + e_{t+\tau}, \quad (A.4.1)$$

where $\pi_t \overset{\Delta}{=} \pi^1_t = (1200) \ln \left( \frac{P_t}{P_{t-1}} \right)$, where $\theta_1 (L)$ and $\theta_2 (L)$ are lag polynomials with $q_u$ and $q_\pi$ lags, respectively. An alternative specification based on inflation level is given by [cf. Perron and Yamamoto (2018)],

$$\pi_{t+\tau} = \theta_0 + \theta_1 (L) u_t + \theta_2 (L) \pi_t + e_{t+\tau}. \quad (A.4.2)$$

The literature suggests that the forecasting ability of the Phillips curve is unstable. In particular, Fisher et al. (2002) documented that the Phillips curve appeared to forecast well 12-month ahead during the 1977-1984 period but not during the period 1993-2000. The same concerns about changes in the performance of Phillips curve for forecasting inflation were expressed by Giacomini and Rossi (2009) and Perron and Yamamoto (2018).

Let us consider model (A.4.2). We assume that the researcher generate a sequence of $\tau$-step-ahead forecasts of $Y_{t+\tau} = \pi_{t+\tau}$ using an out-of-sample procedure. That is, we divide the sample size $T$ into an in-sample window of size $m$ and an out-of-sample window of size $n = T - m - \tau + 1$. Which data constitute the in-sample window depends on the forecasting scheme. We consider the usual forecasting schemes: (1) a fixed forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, m$; (2) a rolling forecasting scheme, where the in-sample window at time $t$ contains observations indexed $t - m + 1, \ldots, t$; and (3) a recursive forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, t$. Let $\beta^* \overset{\Delta}{=} (\theta_0, \theta'_1, \theta'_2)'$ and $f_t \left( \hat{\beta}_t \right)$ be the time-$t$ forecast produced by estimating a model over the in-sample window at time $t$, with $\hat{\beta}_t$ indicating the least-squares estimate of
Each time-$t$ forecast corresponds to a sequence of in-sample fitted values $y_j(\hat{\beta}_t)$, with $j$ varying over the in-sample window.

We evaluate the forecasts by the quadratic loss $L(\cdot)$. Each out-of-sample loss $L_{t+\tau}(\hat{\beta}_t) \triangleq L(Y_{t+\tau}, f_t(\hat{\beta}_t))$ corresponds to in-sample losses $L_j(\hat{\beta}_t) \triangleq L(Y_j, y_j(\hat{\beta}_t))$.

Let $X_t$ collect the set of regressors at time $t$ of model (A.4.2). We have $\hat{\beta}_t = (\sum_{s=1}^{m-\tau} X_s X_s')^{-1} \sum_{s=1}^{m-\tau} X_s Y_{s+\tau}$ for the fixed scheme; $\hat{\beta}_t = (\sum_{s=1}^{t-\tau} X_s X_s')^{-1} \sum_{s=1}^{t-\tau} X_s Y_{s+\tau}$ for the rolling scheme; and $\hat{\beta}_t = (\sum_{s=1}^{t-\tau} X_s X_s')^{-1} \sum_{s=1}^{t-\tau} X_s Y_{s+\tau}$ for the recursive scheme.

The out-of-sample loss corresponding to the forecast at time $t$ is $L_{t+\tau}(\hat{\beta}_t) \triangleq L(Y_{t+\tau}, X_{t+\tau}'\hat{\beta}_t)$ and the corresponding in-sample losses are $L_j(\hat{\beta}_t) \triangleq L(Y_j, X_j'\hat{\beta}_t)$, where $j = \tau + 1, \ldots, m$ for the fixed scheme; $j = \tau + 1, \ldots, m$ for the rolling scheme; and $j = t - m + \tau + 1, \ldots, t$ for the recursive scheme. The same procedure is also applied to (A.4.1).

We verify the presence of forecast failure for the Phillips curve by using the forecast breakdown test of Giacomini and Rossi (2009). This relies on the sequence of so-called surprise losses. The surprise loss at time $t + \tau$ is defined as the difference between the out-of-sample loss at time $t + \tau$ and the average in-sample loss:

$$SL_{t+\tau}(\hat{\beta}_t) = L_{t+\tau}(\hat{\beta}_t) - \bar{T}_t(\hat{\beta}_t), \quad \text{for } t = m, \ldots, T - \tau,$$

(A.4.3)

where $\bar{T}_t(\hat{\beta}_t)$ is the average in-sample loss computed over the in-sample window implied by the forecasting scheme. The null hypotheses is

$$H_0 : \mathbb{E}\left(n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta^*)\right) = 0.$$

The forecast breakdown test statistic of Giacomini and Rossi (2009) is given by $i^{GR}_{m,n,\tau} = n^{1/2} SL_{m,n}/\hat{\sigma}_{m,n}$, where $\hat{\sigma}^2_{m,n} = \lambda \hat{\Sigma}_T$ and (1) $\lambda = 1 + n/m$ for the fixed scheme, (2) $\lambda = 1 - 1/3 (n/m)^2$ for the rolling scheme with $n < m$, (3) $\lambda = 2m/3n$.
for the rolling scheme with \( n \geq m \), (4) \( \lambda = 1 \) for the recursive scheme, and \( \hat{\Sigma}_T \) is the sample variance of the squared losses if the sequence of squared losses are i.i.d. or an HAC estimator otherwise. A level \( \alpha \) test rejects the null hypothesis whenever 
\[
\left| t_{m,n,\tau}^{GR} \right| > z_{\alpha/2},
\]
where \( z_{\alpha/2} \) is the \((1 - \alpha/2)\)-th quantile of a standard normal distribution.

The Breusch-Godfrey test for serial correlation in the squared forecast losses suggests the presence of serial dependence. Thus, we use the HAC estimators in place of \( \hat{\Sigma}_T \). Here we report the results only for the \( \hat{J}_T \) HAC estimator with automatic bandwidth, the Newey-West’s (1987) and Andrews’s (1991) HAC estimator both with automatic bandwidths. Although it is likely that the sequence of squared forecast losses exhibit some kind of nonstationarity in this setting, in order to use the latter two HAC estimators we are implicitly pretending that the data is covariance-stationary. This has been the common practice in economics so far.

We use the same data as in Perron and Yamamoto (2018). We use monthly CPI (consumer price index; revised version), and the unemployment gap for the period 1959:01 to 2004:06. We choose \( q_u = 3 \) and \( q_\pi = 3 \).\(^1\) We consider several sizes for the in-sample windows ranging from \( m = 156 \) (1971:12) to 240 (1978:12). The choice \( m = 240 \) was used also by Perron and Yamamoto (2018) and it implies that the Volker’s Chairmanship period of high inflation enters the out-of-sample period. We consider \( \tau = 1 \) and 12.

For the model in level (A.4.2), Table A.7 shows strong rejections of no change in forecasting accuracy for \( \tau = 1 \) when \( t_{m,n,\tau}^{GR} \) uses the \( \hat{J}_T \) HAC correction. The \( t_{m,n,\tau}^{GR} \) tests that use either Newey and West’s (1987) or Andrews’s (1991) HAC estimator essentially display no evidence for rejection of the hypotheses of no forecast breakdown with exception of the rolling window method when in one case (cf. \( m + 1: 1972 : \)

\(^1\)The results for \( q_\pi = 0 \) and the other combinations of \( q_u \) and \( q_\pi \) are similar and not reported.
01). Indeed, there are cases in which the latter tests are not able to reject the null hypotheses at 10% significance even when the $t_{m,n,\tau}^{GR}$ that uses the $\tilde{J}_T$ HAC correction rejects at the 1% significance level. More specifically, the value of the $t_{m,n,\tau}^{GR}$ when implemented with classical HAC estimators is often less than an half the value of $t_{m,n,\tau}^{GR}$ that uses the $\tilde{J}_T$ HAC correction. Similar comments apply to the case of the “static model” quit $q_x = 0$ (not reported). A similar pattern holds for the case $\tau = 12$ and for the model that uses first-differences in inflation with $\tau = 12$ (bottom panel).

We do not report the test statistics associated to the other classical HAC estimators (e.g., the ones that use prewhitening, long bandwidths with fixed-$b$ critical values, etc.) because they are even smaller than the ones associated to the classical HAC estimators reported here. Hence, the classical HAC standard errors are shown to be unreliable in the sense that a researcher would misleadingly conclude that the forecasting performance of the Phillips curve is stable which, however, contrasts the empirical findings in the literature. When $\tilde{J}_T$ HAC correction is used, inference based on the $t_{m,n,\tau}^{GR}$ confirms the evidence of changes in the forecasting performance of the Phillips curve over time as suggested by the literature.

Overall, this simulation study confirms the concerns about the power issues of test statistics standardized by traditional HAC estimators in context where the data maybe nonstationary and models maybe misspecified. Traditional HAC standard errors become too large and the values of the test statistics become very small so that tests do not have power to reject the null hypotheses.
### A.5 Tables

**Table A.7: Giacomini and Rossi (2009) t-test**

<table>
<thead>
<tr>
<th></th>
<th>Level, dynamic, $\tau = 1$</th>
<th>Fixed</th>
<th>Rolling</th>
<th>Recursive</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>1.34</td>
<td>3.57***</td>
<td>1.84*</td>
<td></td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.96</td>
<td>2.02**</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>1.11</td>
<td>2.42**</td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>1.58*</td>
<td>3.88***</td>
<td>2.50**</td>
<td></td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.50</td>
<td>1.27</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.60</td>
<td>1.54</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>2.74***</td>
<td>5.03***</td>
<td>4.03***</td>
<td></td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.50</td>
<td>1.03</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>0.62</td>
<td>1.29</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>$J_T$, QS kernel, auto</td>
<td>1.97**</td>
<td>4.12***</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>Andrews (1991), auto</td>
<td>0.98</td>
<td>3.21***</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Newey-West (1987), auto</td>
<td>1.82*</td>
<td>1.92*</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td>FD, dynamic, $\tau = 12$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table reports the $t_{m,n,\tau}^{GR}$ statistics proposed by Giacomini and Rossi (2009) for model (A.4.2) with $q_{\pi} = 3$ ("Dynamic model"). "$J_T$, QS kernel, auto" refers to $t_{m,n,\tau}^{GR}$ implemented with the $J_T$ HAC estimator with automatic bandwidth, "Andrews (1991), auto" refers to $t_{m,n,\tau}^{GR}$ with Andrews’s (1991) HAC estimator with automatic bandwidth and asymptotic critical value, and "Newey-West (1987), auto" refers to $t_{m,n,\tau}^{GR}$ with Newey and West’s (1987) HAC estimator with automatic bandwidth. $m + 1$ refers to the start date of the out-of-sample period.
Appendix B

Supplement to Chapter 2: Continuous Record Asymptotics for Structural Change Models

B.1 Mathematical Proofs

B.1.1 Additional Notations

For a matrix $A$, the orthogonal projection matrices $P_A$, $M_A$ are defined as $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$, respectively. For a matrix $A$ we use the vector-induced norm, i.e., $\|A\| = \sup_{x \neq 0} \|Ax\|/\|x\|$. Also, for a projection matrix $P$, $\|PA\| \leq \|A\|$. We denote the $d$-dimensional identity matrix by $I_d$. When the context is clear we omit the subscript notation in the projection matrices. We denote the $(i, j)$-th element of the outer product matrix $A'A$ as $(A'A)_{i,j}$ and the $i \times j$ upper-left (resp., lower-right) sub-block of $A'A$ as $[A'A]_{\{i \times j,\}}$ (resp., $[A'A]_{\{,i \times j\}}$). For a random variable $\xi$ and a number $r \geq 1$, we write $\|\xi\|_r = (\mathbb{E}\|\xi\|^r)^{1/r}$. $B$ and $C$ are generic constants that may vary from line to line; we may sometime write $C_r$ to emphasize the dependence of $C$ on a number $r$. For two scalars $a$ and $b$ the symbol $a \wedge b$ means the infimum of $\{a, b\}$. The symbol $\overset{u,c,p.}{\Rightarrow}$ signifies uniform locally in time convergence under the Skorokhod topology and recall that it implies convergence in probability. The symbol $\overset{d}{\equiv}$ signifies equivalence in distribution. We further use the same notations as explained in Section 2.2.
B.1.2 Preliminary Lemmas

Lemma B.1.1 is Lemma A.1 in Bai (1997). Let $X_\Delta$ be defined as in the display equation after (B.1.11).

**Lemma B.1.1.** The following inequalities hold $P$-a.s.:

\[
\begin{align*}
(Z_0'MZ_0) - (Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0) & \geq R'(X'_\Delta X_\Delta)(X'_2X_2)^{-1}(X'_0X_0) R, \quad T_b < T_0^0 \\
(Z_0'MZ_0) - (Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0) & \geq R'(X'_\Delta X_\Delta)(X'X - X'_2X_2)^{-1}(X'X - X'_0X_0) R, \quad T_b \geq T_0^0.
\end{align*}
\]

The following lemma presents the uniform approximation to the instantaneous covariation between continuous semimartingales. This will be useful in the proof of the convergence rate of our estimator. Below, the time window in which we study certain estimates is shrinking at a rate no faster than $h^{1-\epsilon}$ for some $0 < \epsilon < 1/2$.

**Lemma B.1.2.** Let $X_t$ (resp., $\tilde{X}_t$) be a $q$ (resp., $p$)-dimensional Itô continuous semimartingale defined on $[0, N]$. Let $\Sigma_t$ denote the time $t$ instantaneous covariation between $X_t$ and $\tilde{X}_t$. Choose a fixed number $\epsilon > 0$ and $\varpi$ satisfying $1/2 - \epsilon \geq \varpi \geq \epsilon > 0$. Further, let $B_T \triangleq \lfloor N/h - T^{\varpi} \rfloor$. Define the moving average of $\Sigma_t$ as

\[
\Sigma_{kh} \triangleq (T^{\varpi}h)^{-1} \int_{kh}^{kh+T^{\varpi}h} \Sigma_s ds, \text{ and let } \hat{\Sigma}_{kh} \triangleq (T^{\varpi}h)^{-1} \sum_{i=1}^{\lfloor T^{\varpi} \rfloor} \Delta_h X_{k+i} \Delta_h \tilde{X}^t_{k+i}.
\]

Then, \[ \sup_{1 \leq k \leq B_T} \left\| \hat{\Sigma}_{kh} - \Sigma_{kh} \right\| = o_p(1). \] Furthermore, for each $k$ and some $K > 0$ with $N - K > kh > K$, \[ \sup_{T^\epsilon \leq T \leq T^{1-\epsilon}} \left\| \hat{\Sigma}_{kh} - \Sigma_{kh} \right\| = o_p(1). \]

**Proof.** By a polarization argument, we can assume that $X_t$ and $\tilde{X}_t$ are univariate without loss of generality, and by standard localization arguments, we can assume that the drift and diffusion coefficients of $X_t$ and $\tilde{X}_t$ are bounded. Then, by Itô
Lemma,

\[ \hat{\Sigma}_{kh} - \Sigma_{kh} \equiv \frac{1}{T^\omega h} \sum_{i=1}^{T^\omega} \int_{(k+i-1)h}^{(k+i)h} (X_s - X_{(k+i-1)h}) \, d\bar{X}_s \]

\[ + \frac{1}{T^\omega h} \sum_{i=1}^{T^\omega} \int_{(k+i-1)h}^{(k+i)h} (\bar{X}_s - \bar{X}_{(k+i-1)h}) \, dX_s. \]

For any \( l \geq 1 \), \( \| \hat{\Sigma}_{kh} - \Sigma_{kh} \|_l \leq K_l T^{-\varpi/2} \), which follows from standard estimates for continuous Itô semimartingales. By a maximal inequality, \( \| \sup_{1 \leq k \leq B} |\hat{\Sigma}_{kh} - \Sigma_{kh}|\|_l \leq K_l T^{1/2} T^{-\varpi/2} \), which goes to zero choosing \( l > 2/\varpi \). This proves the first claim.

For the second, note that for \( l \geq 1 \),

\[ \left\| \sup_{T^\omega \leq T^\omega \leq T^{1-\epsilon}} |\hat{\Sigma}_{kh} - \Sigma_{kh}| \right\|_l = \left\| \sup_{1 \leq T^\omega \leq T^{1-2\epsilon}} |\hat{\Sigma}_{kh} - \Sigma_{kh}| \right\|_l \leq K_l T^{(1-2\epsilon)/T - \epsilon/2} \]

Choose \( l > (2 - 4\epsilon)/\epsilon \) to verify the claim. \( \square \)

**B.1.3 Preliminary Results**

As it is customary in related contexts, we use a standard localization argument as explained in Section 1.d in Jacod and Shiryaev (2003), and thus we can replace Assumption 3.1-2.2 with the following stronger assumption.

**Assumption B.1.** Let Assumption 3.1-2.2 hold. The process \( \{Y_t, D_t, Z_t\}_{t \geq 0} \) takes value in some compact set, \( \{\sigma_{\cdot, t}\}_{t \geq 0} \) is bounded càdlàg and the process \( \{\mu_{\cdot, t}\} \) is bounded càdlàg or càglàg.

The localization technique basically translates all the local conditions into global ones. We introduce the following notation which will be useful in some of the proofs below.
B.1.3.1 Approximate Variation, LLNs and CLTs

We review some basic definitions about approximate covariation and more general high-frequency statistics. Given a continuous-time semimartingale $X = (X^i)_{1 \leq i \leq d} \in \mathbb{R}^d$ with zero initial value over the time horizon $[0, N]$, with $P$-a.s. continuous paths, the covariation of $X$ over $[0, t]$ is denoted $[X, X]_t$. The $(i, j)$-element of the quadratic covariation process $[X, X]_t$ is defined as

$$[X^i, X^j]_t = \text{plim}_{T \to \infty} \sum_{k=1}^{T} \left( X^i_{kT} - X^i_{(k-1)T} \right) \left( X^j_{kT} - X^j_{(k-1)T} \right),$$

where plim denotes the probability limit of the sum. $[X, X]_t$ takes values in the cone of all positive semidefinite symmetric $d \times d$ matrices and is continuous in $t$, adapted and of locally finite variation. Associated with this, we can define the $(i, j)$-element of the approximate covariation matrix as

$$\sum_{k \geq 1} \left( h^i X^i_{kT} - h^i X^i_{(k-1)T} \right) \left( h^j X^j_{kT} - h^j X^j_{(k-1)T} \right),$$

which consistently estimates the increments of the quadratic covariation $[X^i, X^j]$. It is an ex-post estimator of the covariability between the components of $X$ over the time interval $[0, t]$. More precisely, as $h \downarrow 0$:

$$\sum_{k \geq 1} \left( X^i_{kT} - X^i_{(k-1)T} \right) \left( X^j_{kT} - X^j_{(k-1)T} \right) \overset{P}{\to} \int_0^t \Sigma^{(i,j)}_{XX,s} ds,$$

where $\Sigma^{(i,j)}_{XX,s}$ is referred to as the spot (not integrated) volatility.

After this brief review, we turn to the statement of the asymptotic results for some statistics to be encountered in the proofs below. We simply refer to Jacod

---

1The reader may refer to Jacod and Protter (2012) or Jacod and Shiryaev (2003) for a complete introduction to the material of this section.
and Protter (2012). More specifically, Lemma B.1.3-B.1.4 follow from their Theorem 3.3.1-(b), while Lemma B.1.5 follows from their Theorem 5.4.2.

**Lemma B.1.3.** Under Assumption B.1, we have as $h \downarrow 0$, $T \to \infty$ with $N$ fixed and for any $1 \leq i, j \leq p$,

(i) $\left|(Z'_2 e)_{i,1}\right| \overset{P}{\to} 0$ where $(Z'_2 e)_{i,1} = \sum_{k=T^0_b+1}^T z_k^{(i)} e_{kh}$;

(ii) $\left|(Z'_0 e)_{i,1}\right| \overset{P}{\to} 0$ where $(Z'_0 e)_{i,1} = \sum_{k=T^0_b+1}^T z_k^{(i)} e_{kh}$;

(iii) $\left|(Z'_2 Z_2)_{i,j} - \int_{(T^0_b+1)h}^N \sum_{Z Z_s} ds\right| \overset{P}{\to} 0$ where $(Z'_2 Z_2)_{i,j} = \sum_{k=T^0_b+1}^T z_k^{(i)} z_k^{(j)}$;

(iv) $\left|(Z'_0 Z_0)_{i,j} - \int_{(T^0_b+1)h}^N \sum_{Z Z_s} ds\right| \overset{P}{\to} 0$ where $(Z'_0 Z_0)_{i,j} = \sum_{k=T^0_b+1}^T z_k^{(i)} z_k^{(j)}$.

For the following estimates involving $X$, we have, for any $1 \leq r \leq p$ and $1 \leq l \leq q + p$,

(v) $\left|(X e)_{i,1}\right| \overset{P}{\to} 0$ where $(X e)_{i,1} = \sum_{k=1}^T x_k^{(i)} e_{kh}$;

(vi) $\left|(Z'_2 X)_{r,l} - \int_{(T^0_b+1)h}^N \sum_{Z X_s} ds\right| \overset{P}{\to} 0$ where $(Z'_2 X)_{r,l} = \sum_{k=T^0_b+1}^T z_k^{(r)} x_k^{(l)}$;

(vii) $\left|(Z'_0 X)_{r,l} - \int_{(T^0_b+1)h}^N \sum_{Z X_s} ds\right| \overset{P}{\to} 0$ where $(Z'_0 X)_{r,l} = \sum_{k=T^0_b+1}^T z_k^{(r)} x_k^{(l)}$.

Further, for $1 \leq u, d \leq q + p$,

(viii) $\left|(X' X)_{u,d} - \int_0^N \sum_{X X_s} ds\right| \overset{P}{\to} 0$ where $(X' X)_{u,d} = \sum_{k=1}^T x_k^{(u)} x_k^{(d)}$.

**Lemma B.1.4.** Under Assumption B.1, we have as $h \downarrow 0$, $T \to \infty$ with $N$ fixed, $|N^0_b - N_b| > \gamma > 0$ and for any $1 \leq i, j \leq p$,

(i) with $(Z'_\Delta Z_\Delta)_{i,j} = \sum_{k=T^0_b}^{T_b} z_k^{(i)} z_k^{(j)}$ we have

\[
\begin{cases}
\left|(Z'_\Delta Z_\Delta)_{i,j} - \int_{(T^0_b+1)h}^{T^0_b} \sum_{Z Z_s} ds\right| \overset{P}{\to} 0, & \text{if } T^0_b < T_b \\
\left|(Z'_\Delta Z_\Delta)_{i,j} - \int_{(T^0_b+1)h}^{T_b} \sum_{Z Z_s} ds\right| \overset{P}{\to} 0, & \text{if } T_b > T^0_b
\end{cases}
\]

and for $1 \leq r \leq p + q$
(ii) with $(Z'_\Delta X_\Delta)_{i,r} = \sum_{k=T^0_0}^{T^0_0+1} \tau_{kh} x_{kh}^{(r)}$ we have

$$
\begin{cases}
| (Z'_\Delta X_\Delta)_{i,r} - \int_{T^0_0}^{T^0_0+1} \sum_{s} x^{(r)}_{s} ds | \xrightarrow{P} 0, & \text{if } T^0_0 < T^0_0 \\
| (Z'_\Delta X_\Delta)_{i,r} - \int_{T^0_0}^{T^0_0+1} \sum_{s} x^{(r)}_{s} ds | \xrightarrow{P} 0, & \text{if } T^0_0 > T^0_0.
\end{cases}
$$

Next, we turn to the central limit theorems, they all feature a limiting process defined on an extension of the original probability space $(\Omega, \mathcal{F}, P)$. In order to avoid non-useful repetitions, we present a general framework valid for all statistics considered in the chapter. The first step is to carry out an extension of the original probability space $(\Omega, \mathcal{F}, P)$. We accomplish this in the usual way. We first fix the original probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$. Consider an additional measurable space $(\Omega^*, \mathcal{F}^*)$ and a transition probability $Q(\omega, d\omega^*)$ from $(\Omega, \mathcal{F})$ into $(\Omega^*, \mathcal{F}^*)$. Next, we can define the products $\tilde{\Omega} = \Omega \times \Omega^*$, $\tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}^*$ and $\tilde{P}(d\omega, d\omega^*) = P(d\omega) Q(\omega, d\omega^*)$. This defines the extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ of the original space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$. Any variable or process defined on either $\Omega$ or $\Omega^*$ is extended in the usual way to $\tilde{\Omega}$ as follows: for example, let $Y_t$ be defined on $\Omega$. Then we say that $Y_t$ is extended in the usual way to $\tilde{\Omega}$ by writing $Y_t(\omega, \omega^*) = Y_t(\omega)$. Further, we identify $\mathcal{F}_t$ with $\mathcal{F}_t \otimes \{\emptyset, \Omega^*\}$, so that we have a filtered space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\mathcal{F}_t\}_{t \geq 0}, \tilde{P})$. Finally, as for the filtration, we can consider another filtration $\{\tilde{\mathcal{F}}_t\}_{t \geq 0}$ taking the product form $\tilde{\mathcal{F}}_t = \cap_{s \geq t} \mathcal{F}_s \otimes \mathcal{F}^*_s$, where $\{\mathcal{F}^*_t\}_{t \geq 0}$ is a filtration on $(\Omega^*, \mathcal{F}^*)$. As for the transition probability $Q$ we can consider the simple form $Q(\omega, d\omega^*) = P^*(d\omega^*)$ for some probability measure on $(\Omega^*, \mathcal{F}^*)$. This defines the way a product filtered extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t \geq 0}, \tilde{P})$ of the original filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ is constructed in this chapter. Assume that the auxiliary probability space $(\Omega^*, \mathcal{F}^*, \{\mathcal{F}^*_t\}_{t \geq 0}, P^*)$ supports a $p^2$-dimensional standard Wiener process $W_s^+$ which is adapted to $\{\mathcal{F}_t\}$. We need some additional ingredients in order to describe the limiting process. We
choose a progressively measurable “square-root” process $\sigma^*_Z$ of the $\mathcal M_{p^2 \times p^2}^+$-valued process $\hat \Sigma_{Z,s}$, whose elements are given by $\hat \Sigma_{Z,s}^{(ij,kl)} = \Sigma_{Z,s}^{(ik)} \Sigma_{Z,s}^{(jl)}$. Due to the symmetry of $\Sigma_{Z,s}$, the matrix with entries $\Sigma_{Z,s}^{(ij,kl)} + \Sigma_{Z,s}^{(ji,kl)} / \sqrt{2}$ is a square-root of the matrix with entries $\Sigma_{Z,s}^{(ij,kl)} + \Sigma_{Z,s}^{(il,jk)}$. Then the process $\mathcal U_t$ with components $\mathcal U_t^{(r,j)} = 2^{-1/2} \sum_{k,l=1}^p \int_0^t \left( \sigma_{Z,s}^{(r,j,kl)} + \sigma_{Z,s}^{(jr,kl)} \right) dW^{(kl)}_s$ is, conditionally on $\mathcal F$, a continuous Gaussian process with independent increments and (conditional) covariance

$$
\mathbb E \left( \mathcal U_t^{(r,j)} (v) \mathcal U_t^{(k,l)} (v) \mid \mathcal F \right) = \int_{T_b^0 h}^{T_b^0 h+1} \left( \Sigma_{Z,s}^{(rk)} \Sigma_{Z,s}^{(il)} + \Sigma_{Z,s}^{(rl)} \Sigma_{Z,s}^{(jk)} \right) ds,
$$

where $v \leq 0$. The CLT of interest is as follows.

**Lemma B.1.5.** Let $Z$ be a continuous Itô semimartingale satisfying Assumption B.1. Then, $(Nh)^{-1/2} \left( Z_T^2 - \left( [Z, Z]_{Th} + [Z, Z]_{(Tb+1)h} \right) \right) \overset{L^2}{\to} \mathcal W$.

**B.1.4 Proofs of Sections 2.3 and 2.4**

**B.1.4.1 Additional Notation**

In some of the proofs we face a setting in which $N_b$ is allowed to vary within a shrinking neighborhood of $N_0^b$. Some estimates only depend on observations in this window. For example, assume $T_b < T_b^0$ and consider $\sum_{k=T_b+1}^{T_b^0} x_{kh}x'_{kh}$. When $N_b$ is allowed to vary within a shrinking neighborhood of $N_0^b$, this sum approximates a local window of asymptotically shrinking size. Introduce a sequence of integers $\{l_T\}$ that satisfies $l_T \to \infty$ and $l_T h \to 0$. Below when we shall establish a $T^{1-\kappa}$-rate of convergence of $\hat \lambda_b$ toward $\lambda_0$, we will consider the case where $N_b - N_0^b = T^{-\gamma}$ for some $\gamma \in (0, 1/2)$. Hence, it is convenient to define

$$
\hat \Sigma_X (T_b, T_b^0) \triangleq \sum_{k=T_b+1}^{T_b^0} x_{kh}x'_{kh} = \sum_{k=T_b^0+1-l_T}^{T_b^0} x_{kh}x'_{kh},
$$

(B.1.3)
where now \( l_T = \lfloor T^\kappa \rfloor \to \infty \) and \( l_T h = h^{1-\kappa} \to 0 \). Note that \( 1/h^{1-\kappa} \) is the rate of convergence and the interpretation for \( \hat{\Sigma}_X(T_b, T_b^0) \) is that it involves asymptotically an infinite number of observations falling in the shrinking (at rate \( h^{1-\kappa} \)) block \(((T_b - 1) h, T_b^0 h] \). Other statistics involving the regressors and errors are defined similarly:

\[
\hat{\Sigma}_{X_e}(T_b, T_b^0) \triangleq \sum_{k=T_b+1}^{T_b^0} x_{kh} e_{kh} = \sum_{k=T_b^0+1-l_T}^{T_b^0} x_{kh} e_{kh}, \tag{B.1.4}
\]

and

\[
\hat{\Sigma}_{Z_e}(T_b, T_b^0) \triangleq \sum_{k=T_b^0+1-l_T}^{T_b^0} z_{kh} e_{kh}. \tag{B.1.5}
\]

Further, we let \( \hat{\Sigma}_{Xe}(T_b, T_b^0) \triangleq h^{-(1-\kappa)} \int_{N_b}^{T_b^0} \Sigma_{X_{e,s}} ds \) and analogously when \( Z \) replaces \( X \). We also define

\[
\hat{\Sigma}_{h,X}(T_b, T_b^0) \triangleq h^{-(1-\kappa)} \sum_{k=T_b^0+1-l_T}^{T_b^0} x_{kh} x'_{kh}. \tag{B.1.6}
\]

The proofs of Section 2.4 are first given for the case where \( \mu_{.,t} \) from equation (2.2.3) are identically zero. In the last step, this is relaxed. Furthermore, throughout the proofs we reason conditionally on the processes \( \mu_{.,t} \) and \( \Sigma_t^0 \) (defined in Assumption 2.2) so that they are treated as if they were deterministic. This is a natural strategy since the processes \( \mu_{.,t} \) are of higher order in \( h \) and they do not play any role for the asymptotic results [cf. Barndorff-Nielsen and Shephard (2004)].
B.1.4.2 Proof of Proposition 2.3.1

Proof. The concentrated sample objective function evaluated at \( \hat{T_b} \) is

\[
Q_T (\hat{T_b}) = \delta_{\hat{T_b}} (Z_2'MZ_2) \delta_{\hat{T_b}}.
\]

We have

\[
\delta_{\hat{T_b}} = (Z_2'MZ_2)^{-1} (Z_2'MY) = (Z_2'MZ_2)^{-1} (Z_2'MZ_0) \delta^0 + (Z_2'MZ_2)^{-1} Z_2Me,
\]

and \( \delta_{T_0} = (Z_0'MZ_0)^{-1} (Z_0'MY) = \delta^0 + (Z_0'MZ_0)^{-1} (Z_0'Me) \) and, therefore,

\[
Q_T (T_b) - Q_T (T_0) = \hat{\delta}_{\hat{T_b}} (Z_2'MZ_2) \hat{\delta}_{T_b} - \hat{\delta}_{T_0} (Z_0'MZ_0) \hat{\delta}_{T_0} \quad \text{(B.1.7)}
\]

\[
= (\delta^0)' \left\{ (Z_0'MZ_2) (Z_2'MZ_2)^{-1} (Z_2'MZ_0) - Z_0'MZ_0 \right\} \delta^0 \quad \text{(B.1.8)}
\]

\[
+ g_e (T_b), \quad \text{(B.1.9)}
\]

where

\[
g_e (T_b) = 2 (\delta^0)' (Z_0'MZ_2) (Z_2'MZ_2)^{-1} Z_2Me - 2 (\delta^0)' (Z_0'Me) \quad \text{(B.1.10)}
\]

\[
+ e'MZ_2 (Z_2'MZ_2)^{-1} Z_2Me - e'MZ_0 (Z_0'MZ_0)^{-1} Z_0'Me. \quad \text{(B.1.11)}
\]

Denote

\[
X_\Delta \triangleq X_2 - X_0 = \left( 0, \ldots, 0, x_{(T_b+1)h}, \ldots, x_{T_b'h}, 0, \ldots, \right)', \quad \text{for } T_b < T_0
\]

\[
X_\Delta \triangleq -(X_2 - X_0) = \left( 0, \ldots, 0, x_{(T_0h+1)}, \ldots, x_{T_0'h}, 0, \ldots, \right)', \quad \text{for } T_b > T_0
\]

\[
X_\Delta \triangleq 0, \quad \text{for } T_b = T_0.
\]
Observe that when $T^0_b \neq T_b$ we have $X_2 = X_0 + X_\Delta \text{sign}(T^0_b - T_b)$. When the sign is immaterial, we simply write $X_2 = X_0 + X\Delta$. Next, let $Z_\Delta = X\Delta R$, and define

$$r(T_b) \triangleq \frac{(\delta^0)' \{ (Z'_0 M Z_0) - (Z'_0 M Z_2) (Z'_2 M Z_2)^{-1} (Z'_2 M Z_0) \} \delta^0}{|T_b - T^0_b|}.$$  \hspace{1cm} (B.1.12)

We arbitrarily define $r(T_b) = (\delta^0)' \delta^0$ when $T_b = T^0_b$. We write (B.1.7) as

$$Q_T(T_b) - Q_T(T_0) = - |T_b - T^0_b| r(T_b) + g_e(T_b), \quad \text{for all } T_b.$$  \hspace{1cm} (B.1.13)

By definition, $\hat{T}_b$ is an extremum estimator and thus it must satisfy $g_e(\hat{T}_b) \geq |\hat{T}_b - T^0_b| r(\hat{T}_b)$. Therefore,

$$P\left( |\hat{\lambda}_b - \lambda_0| > K \right) = P\left( |\hat{T}_b - T^0_b| > TK \right)$$

$$\leq P \left( \sup_{|T_b - T^0_b| > TK} |g_e(T_b)| \geq \inf_{|T_b - T^0_b| > TK} |T_b - T^0_b| r(T_b) \right)$$

$$\leq P \left( \sup_{p \leq T_b \leq T - p} |g_e(T_b)| \geq TK \inf_{|T_b - T^0_b| > TK} r(T_b) \right)$$  \hspace{1cm} (B.1.14)

$$= P \left( r^{-1}_T \sup_{p \leq T_b \leq T - p} |g_e(T_b)| \geq K \right),$$

where recall $p \leq T_b \leq T - p$ is needed for identification, and $r_T \triangleq T \inf_{|T_b - T^0_b| > TK} r(T_b)$. Lemma B.1.6 below shows that $r_T$ is positive and bounded away from zero. Thus, it is sufficient to verify that the stochastic component is negligible as $h \downarrow 0$, i.e.,

$$\sup_{p \leq T_b \leq T - p} |g_e(T_b)| = o_p(1).$$  \hspace{1cm} (B.1.15)
The first term of \( g_e(T_b) \) is

\[
2 \left( \delta^0 \right)' (Z_0' M Z_2) (Z_2' M Z_2)^{-1/2} (Z_2' M Z_2)^{-1/2} Z_2 M e. \tag{B.1.16}
\]

Lemma B.1.5 implies that for any \( 1 \leq j \leq p \), \((Z_2 e)_{j,1} / \sqrt{h} = O_p(1)\) and for any \( 1 \leq i \leq q + p \), \((X e)_{i,1} / \sqrt{h} = O_p(1)\). These hold because they both involve a positive fraction of the data. Furthermore, from Lemma B.1.3, we also have that \( Z_2' M Z_2 \) and \( Z_0' M Z_2 \) are \( O_p(1) \). Therefore, the supremum of \((Z_0' M Z_2) (Z_2' M Z_2)^{-1/2} \) over all \( T_b \) is \( \sup_{T_b} (Z_0' M Z_2) (Z_2' M Z_2)^{-1} \leq Z_0' M Z_0 = O_p(1) \) by Lemma B.1.3. By Assumption (3.1)-(iii) \((Z_2' M Z_2)^{-1/2} Z_2 M e \) is \( O_p(1) \) \( \sqrt{h} \) uniformly, which implies that (B.1.16) is \( O_p(1) \) uniformly over \( p \leq T_b \leq T - p \). As for the second term of (B.1.10), \( Z_0' M e = O_p(1) \). The first term in (B.1.11) is uniformly \( o_p(1) \) and the same holds for the last term. Therefore, combining these results, \( \sup_{T_b} |g_e(T_b)| = O_p(1) \) uniformly when \( |\lambda_b - \lambda_0| > K \). Therefore for some \( B > 0 \), these arguments combined with Lemma B.1.6 below result in

\[
P \left( r_B^{-1} \sup_{p \leq T_b \leq T - p} |g_e(T_b)| \geq K \right) \leq \varepsilon,
\]

from which it follows that the right-hand side of (B.1.14) is weakly smaller than \( \varepsilon \). This concludes the proof since \( \varepsilon > 0 \) was arbitrarily chosen. \( \square \)

**Lemma B.1.6.** For \( B > 0 \), let \( r_B = \inf |T_b - T_0^0| > TB Tr(T_b) \). There exists a \( \kappa > 0 \) such that for every \( \varepsilon > 0 \), there exists a \( B < \infty \) such that \( P(r_B \geq \kappa) \leq 1 - \varepsilon \), i.e., \( r_B \) is positive and bounded away from zero with high probability.

**Proof.** Assume \( T_b \leq T_0^0 \) and observe that \( r_T \geq r_B \) for an appropriately chosen \( B \). From the first inequality result in Lemma B.1.1,

\[
r(T_b) \geq \left( \delta^0 \right)' R' \left( X_2' X_\Delta / (T_b^0 - T_b) \right) (X_2' X_2)^{-1} (X_0' X_0) R \delta^0.
\]
When multiplied by $T$, we have

$$
Tr(T_b) \geq T \left( \delta^0 \right)' R' \frac{X'_{\Delta}X_{\Delta}}{T^0_b - T_b} (X'_{2}X_2)^{-1} (X'_0X_0) R \delta^0
$$

$$
= \left( \delta^0 \right)' R' \frac{X'_{\Delta}X_{\Delta}}{N^0_b - N^0_b} (X'_{2}X_2)^{-1} (X'_0X_0) R \delta^0.
$$

Note that $0 < K < B < h (T^0_b - T_b) < N$. Then,

$$
Tr(T_b) \geq \left( \delta^0 \right)' R' (X'_{\Delta}X_{\Delta}/N) (X'_{2}X_2)^{-1} (X'_0X_0) R \delta^0,
$$

and by standard estimates for Itô semimartingales, $X'_{\Delta}X_{\Delta} = O_p(1)$ (i.e., use the Burkhölder-Davis-Gundy inequality and recall that $|\hat{N}_b - N^0_b| > BN$). Hence, we conclude $Tr(T_b) \geq (\delta^0)' R' O_p(1/N) O_p(1) R \delta^0 \geq \kappa > 0$, where $\kappa$ is some positive constant. The last inequality follows whenever $X'_{\Delta}X_{\Delta}$ is positive definite since $R'X'_{\Delta}X_{\Delta} (X'_{2}X_2)^{-1} (X'_0X_0) R$ can be rewritten as $R' \left[ (X'_0X_0)^{-1} + (X'_{\Delta}X_{\Delta})^{-1} \right] R$. According to Lemma B.1.3, $X'_{2}X_2$ is $O_p(1)$. The same argument applies to $X'_0X_0$, which together with the the fact that $R$ has full common rank in turn implies that we can choose a $B > 0$ such that $r_B = \inf_{|T_b - T^0_b| > TB} Tr(T_b)$ satisfies $P(r_B \geq \kappa) \leq 1 - \varepsilon$. The case with $T_b > T^0_b$ is similar and is omitted. □

B.1.4.3 Proof of Proposition 2.3.2

Proof. Given the consistency result, one can restrict attention to the local behavior of the objective function for those values of $T_b$ in

$$
B_T \triangleq \{ T_b : T \eta \leq T_b \leq T (1 - \eta) \},
$$

where \(\eta > 0\) satisfies \(\eta \leq \lambda_0 \leq 1 - \eta\). By Proposition 2.3.1, the estimator \(\hat{T}_b\) will visit the set \(B_T\) with large probability as \(T \to \infty\). That is, for any \(\varepsilon > 0\), \(P(\hat{T}_b \notin B_T) < \varepsilon\) for sufficiently large \(T\). We show that for large \(T\), \(\hat{T}_b\) eventually falls in the set \(B_{K,T} \equiv \{T_b : |N_b - N_0| \leq KT^{-1}\}\), for some \(K > 0\). For any \(K > 0\), define the intersection of \(B_T\) and the complement of \(B_{K,T}\) by \(D_{K,T} \equiv \{T_b : N_0 \leq N_b \leq N(1 - \eta), |N_b - N_0| > KT^{-1}\}\). Notice that

\[
\left\{ |\hat{\lambda}_b - \lambda_0| > KT^{-1} \right\} \\
= \{ |\hat{\lambda}_b - \lambda_0| > KT^{-1} \cap \hat{\lambda}_b \in (\eta, 1 - \eta) \} \\
\cup \{ |\hat{\lambda}_b - \lambda_0| > KT^{-1} \cap \hat{\lambda}_b \notin (\eta, 1 - \eta) \} \\
\subseteq \{ |\hat{\lambda}_b - \lambda_0| > K(T^{-1}) \cap \hat{\lambda}_b \in (\eta, 1 - \eta) \} \cup \{ \hat{\lambda}_b \notin (\eta, 1 - \eta) \},
\]

and so

\[
P \left( |\hat{\lambda}_b - \lambda_0| > KT^{-1} \right) \leq P \left( \hat{\lambda}_b \notin (\eta, 1 - \eta) \right) \\
+ P \left( |\hat{\lambda}_b - \lambda_0| > KT^{-1} \cap \hat{\lambda}_b \notin (\eta, 1 - \eta) \right),
\]

and for large \(T\),

\[
P \left( |\hat{\lambda}_b - \lambda_0| > KT^{-1} \right) \leq \varepsilon + P \left( |\hat{\lambda}_b - \lambda_0| > KT^{-1} \cap \hat{\lambda}_b \in (\eta, 1 - \eta) \right) \\
\leq \varepsilon + P \left( \sup_{T_b \in D_{K,T}} Q_T(T_b) \geq Q_T(T_0) \right).
\]

Therefore it is enough to show that the second term above is negligible as \(h \downarrow 0\). Suppose \(T_b < T_0\). Since \(\hat{T}_b = \arg\max Q_T(T_b)\), it is enough to show that \(P \left( \sup_{T_b \in D_{K,T}} Q_T(T_b) \geq Q_T(T_0) \right) < \varepsilon\). Note that this implies \(T_b - T_0 > KN^{-1}\). Therefore, we have to deal with a setting where the time span in \(D_{K,T}\) between \(N_b\) and...
\(N_b^0\) is actually shrinking. The difficulty arises from the quantities depending on the difference \(|N_b - N_b^0|\). We can rewrite \(Q_T(T_b) \geq Q_T(T_b^0)\) as \(g_e(T_b) / |T_b - T_b^0| \geq r(T_b)\), where \(g_e(T_b)\) and \(r(T_b)\) were defined above. Thus, we need to show,

\[
P \left( \sup_{T_b \in \mathcal{D}_{K,T}} h^{-1} (T_b^0 - T_b)^{-1} g_e(T_b) \geq B/N \right) < \varepsilon, \tag{B.1.17}
\]

By Lemma B.1.1,

\[
\inf_{T_b \in \mathcal{D}_{K,T}} r(T_b) \geq \inf_{T_b \in \mathcal{D}_{K,T}} \left( \delta^0 \right)^t R' \frac{X_\Delta'X_\Delta}{|T_b - T_b^0|} (X_2'X_2)^{-1} (X_0'X_0) R \delta^0.
\]

The asymptotic results used so far rely on statistics involving integrated covariation between continuous semimartingales. However, since \(|T_b - T_b^0| > K/N\) the context becomes different and the same results do not apply because the time horizon is decreasing as the sample size increases for quantities depending on \(|N_b - N_b^0|\). Thus, we shall apply asymptotic results for the local approximation of the covariation between processes. Moreover, when \(|T_b - T_b^0| > K/N\), there are at least \(K\) terms in this sum with asymptotically vanishing moments. That is, for any \(1 \leq i, j \leq q + p\), we have

\[
E \left[ x_{kh}^{(i)}, x_{kh}^{(j)} \mid \mathcal{F}_{(k-1)h} \right] = \Sigma^{(i,j)}_{X, (k-1)h} h, \quad \text{and note that } x_{kh} / \sqrt{h} \text{ is i.n.d. with finite variance and thus by Assumption 2.4 we can always choose a } K \text{ large enough such that}
\]

\[
(h |T_b - T_b^0|)^{-1} X_\Delta'X_\Delta = (h |T_b - T_b^0|)^{-1} \sum_{k=T_b+1}^{T_b^0} x_{kh} x_{kh}' = A > 0 \text{ for all } T_b \in \mathcal{D}_{K,T}.
\]

This shows that \(\inf_{T_b \in \mathcal{D}_{K,T}} h^{-1} r(T_b)\) is bounded away from zero. Note that for the other terms in \(r(T_b)\) we can use the same arguments since they do not depend on \(|N_b - N_b^0|\). Hence,

\[
h^{-1} (T_b^0 - T_b)^{-1} e'MZ_2 (Z_2'MZ_2)^{-1} Z_2 Me = (T_b^0 - T_b)^{-1} h^{-1} O_p \left( h^{1/2} \right) O_p \left( 1 \right) O_p \left( h^{1/2} \right) = \frac{O_p \left( 1 \right)}{T_b^0 - T_b},
\]
for some $B > 0$. Consider the terms of $g_e(T_b)$ in (B.1.11). When $T_b \in D_{K,T}$, $Z_2$ involves at least a positive fraction $N\eta$ of the data. From Lemma B.1.3, as $h \downarrow 0$, it follows that

$$h^{-1} \left( T_0 - T_b \right)^{-1} e' M Z_2 (Z'_2 M Z_2)^{-1} Z_2 M e = \left( T_0 - T_b \right)^{-1} h^{-1} O_p \left( h^{1/2} \right) O_p \left( 1 \right) O_p \left( h^{1/2} \right) = \frac{O_p \left( 1 \right)}{T_0 - T_b},$$

uniformly in $T_b$. Choose $K$ large enough so that the probability that the right-hand size is larger than $B/N$ is less than $\varepsilon/4$. A similar argument holds for the second term in (B.1.11). Next consider the first term of $g_e(T_b)$. Using $Z_2 = Z_0 \pm Z_\Delta$ we can deduce that

$$\left( \delta^0 \right)' (Z'_0 M Z_2) (Z'_2 M Z_2)^{-1} Z_2 M e = \left( \delta^0 \right)' ( (Z'_2 \pm Z'_\Delta) M Z_2 ) (Z'_2 M Z_2)^{-1} Z_2 M e$$

$$= \left( \delta^0 \right)' Z'_0 M e \pm \left( \delta^0 \right)' Z'_\Delta M e \pm \left( \delta^0 \right)' (Z'_\Delta M Z_2) (Z'_2 M Z_2)^{-1} Z_2 M e,$$

from which it follows that

$$\left| 2 \left( \delta^0 \right)' (Z'_0 M Z_2) (Z'_2 M Z_2)^{-1} Z_2 M e - 2 \left( \delta^0 \right)' (Z'_0 M e) \right|$$

$$= \left| \left( \delta^0 \right)' Z'_\Delta M e \right| + \left| \left( \delta^0 \right)' (Z'_\Delta M Z_2) (Z'_2 M Z_2)^{-1} (Z_2 M e) \right|. \ (B.1.19)$$

First, we can apply Lemma B.1.3 [(vi) and (viii)], and Lemma B.1.4 [(i)-(ii)], together with Assumption 3.1-(iii), to terms that do not involve $|N_b - N_0^b|$,}

$$h^{-1} \left( Z'_\Delta M Z_2 \right) = h^{-1} \left( \delta^0 \right)' (Z'_\Delta Z_2) - h^{-1} \left( \delta^0 \right)' \left( Z'_\Delta X_\Delta (X'X)^{-1} X'Z_2 \right)$$

$$= \left( \delta^0 \right)' \left( \frac{Z'_\Delta X_\Delta}{h} \right) - \left( \delta^0 \right)' \left( \frac{Z'_\Delta X_\Delta}{h} (X'X)^{-1} X'Z_2 \right).$$
Consider $Z'_\Delta Z_\Delta$. By the same reasoning as above, whenever $T_b \in D_{K,T}$,

$$(Z'_\Delta Z_\Delta) / h \left( T^0_b - T_b \right) = O_p(1),$$

for $K$ large enough. The term $Z'_\Delta X_\Delta / h \left( T^0_b - T_b \right)$ is also $O_p(1)$ uniformly. Thus, it follows from Lemma B.1.5 that the second term of (B.1.19) is $O_p\left(h^{1/2}\right)$. Next, note that $Z'_\Delta Me = Z'_\Delta e - Z'_\Delta X \left( X'X \right)^{-1} X'e$. We can write

$$\frac{Z'_\Delta Me}{(T^0_b - T_b) h} = \frac{1}{(T^0_b - T_b) h} \sum_{k=T_b+1}^{T^0_b} z_{kh} e_{kh}$$

$$- \frac{1}{(T^0_b - T_b) h} \left( \sum_{k=T_b+1}^{T^0_b} z_{kh} x'_{kh} \right) (X'X)^{-1} (X'e).$$

Note that the sequence $\{h^{-1/2}z_{kh}h^{-1/2}x_{kh}\}$ is i.n.d. with finite mean identically in $k$. There is at least $K$ terms in this sum, so $\left( \sum_{k=T_b+1}^{T^0_b} z_{kh} x'_{kh} \right) / (T^0_b - T_b) h$ is $O_p(1)$ for a large enough $K$ in view of Assumption 2.4. Then,

$$\frac{1}{(T^0_b - T_b) h} \left( \sum_{k=T_b+1}^{T^0_b} z_{kh} x'_{kh} \right) (X'X)^{-1} (X'e) = O_p(1) O_p(1) O_p\left(h^{1/2}\right), \quad (B.1.20)$$

when $K$ is large. Thus,

$$\frac{1}{(T^0_b - T_b) h} g_e(T_b) = \frac{1}{(T^0_b - T_b) h} \left(\delta^0\right)' 2Z'_\Delta e + \frac{O_p(1)}{T^0_b - T_b} + O_p\left(h^{1/2}\right). \quad (B.1.21)$$
We can now prove (B.1.17) using (B.1.21). To this end, we need a $K > 0$, such that

$$P\left( \sup_{T_b \in \mathcal{D}_{K,T}} \left\| \left( \delta^0 \right)' \frac{1}{h} \frac{1}{T_b - T_{b_k}} \sum_{k=k+1}^{T_b} z_{kh} e_{kh} \right\| > \frac{B}{4N} \right) \leq P\left( \sup_{T_b \leq T_b^0 - KN^{-1}} \left\| \frac{1}{h} \frac{1}{T_b - T_{b_k}} \sum_{k=k+1}^{T_b} z_{kh} e_{kh} \right\| > \frac{B}{8N \| \delta^0 \|} \right) < \varepsilon. \quad (B.1.22)$$

Note that $|T_b - T_b^0|$ is bounded away from zero in $\mathcal{D}_{K,T}$. Observe $(z_{kh}/\sqrt{h}) (e_{kh}/\sqrt{h})$ are independent in $k$ and have zero mean and finite second moments. Hence, by the Hájek-Rényi inequality [see Lemma A.6 in Bai and Perron (1998)],

$$P\left( \sup_{T_b \leq T_b^0 - KN^{-1}} \left\| \frac{1}{h} \frac{1}{T_b - T_{b_k}} \sum_{k=k+1}^{T_b} z_{kh} e_{kh} \right\| > \frac{B}{8N \| \delta^0 \|} \right) \leq A \frac{4N^2}{B^2} \frac{1}{KN^{-1}}$$

where $A > 0$. We can choose $K$ large enough such that the right-hand side is less than $\varepsilon/4$. Combining the above arguments, we deduce the claim in (B.1.17) which then concludes the proof of Proposition 2.3.2. □

**B.1.4.4 Proof of Proposition 2.3.3**

We focus on the case with $T_b \leq T_0$. The arguments for the other case are similar and omitted. From Proposition 2.3.1 the distance $|\hat{\lambda}_b - \lambda_0|$ can be made arbitrary small. Proposition 2.3.2 gives the associated rate of convergence: $T_0 (\hat{\lambda}_b - \lambda_0) = O_p(1)$. Given the consistency result for $\hat{\lambda}_b$, we can apply a restricted search. In particular, by Proposition 2.3.2, for large $T > T$, we know that $\{T_b \notin \mathcal{D}_{K,T} \}$, or equivalently $|T_b - T_b^0| \leq K$, with high probability for some $K$. Essentially, what we shall show is that from the results of Proposition 2.3.1-2.3.2 the error in replacing $T_b^0$ with $\hat{T}_b$ is stochastically small and thus it does not affect the estimation of the parameters $\beta^0$, $\delta_1^0$ and $\delta_2^0$. Toward this end, we first find a lower bound on the convergence rate
for \( \hat{\lambda}_b \) that guarantees its estimation problem to be asymptotically independent from that of the regression parameters. This result will also be used in later proofs. We shall see that the rate of convergence established in Proposition 2.3.2 is strictly faster than the lower bound. Below, we use \( \hat{T}_b \) in order to construct \( Z_2 \) and define \( \hat{Z}_0 \triangleq Z_2 \).

**Lemma B.1.7.** Fix \( \gamma \in (0, 1/2) \) and some constant \( A > 0 \). For all large \( T > T \), if \( |\hat{N}_b - N^0_b| \leq AO_p(h^{1-\gamma}) \), then \( X' (Z_0 - \hat{Z}_0) = O_p(h^{1-\gamma}) \) and \( Z^0_0 (Z_0 - \hat{Z}_0) = O_p(h^{1-\gamma}) \).

**Proof.** Note that the setting of Proposition 2.3.2 satisfies the conditions of this lemma because \( \hat{N}_b - N^0_b = O_p(h) \leq AO_p(h^{1-\gamma}) \) as \( h \downarrow 0 \). By assumption, there exists some constant \( C > 0 \) such that \( P \left( h^\gamma \left| \hat{T}_b - T^0_b \right| > C \right) < \varepsilon \). We have to show that although we only know \( |\hat{T}_b - T^0_b| \leq Ch^{-\gamma} \), the error when replacing \( T^0_b \) by \( \hat{T}_b \) in the construction of \( Z_2 \) goes to zero fast enough. This is achieved because \( |\hat{N}_b - N^0_b| \to 0 \) at rate at least \( h^{1-\gamma} \) which is faster than the standard convergence rate for regression parameters (i.e., \( \sqrt{T} \)-rate). Without loss of generality we take \( C = 1 \). We have

\[
h^{-1/2}X' \left( Z_0 - \hat{Z}_0 \right) = h^{1/2-\gamma} \frac{1}{h^{1-\gamma}} \sum_{T^0_b - [T^\gamma]} T^0_b x_k h \hat{z}_k.
\]

Notice that, as \( h \downarrow 0 \), the number of terms in the sum on the right-hand side, for all \( T > T \), increases to infinity at rate \( 1/h^\gamma \). Since \( \hat{N}_b \) approaches \( N^0_b \) at rate \( T^{-(1-\gamma)} \), the quantity \( X' \left( Z_0 - \hat{Z}_0 \right) /h^{1-\gamma} \) is a consistent estimate of the so-called instantaneous or spot covariation between \( X \) and \( Z \) at time \( N^0_b \). Theorem 9.3.2 part (i) in Jacod and Protter (2012) can be applied since the “window” is decreasing at rate \( h^{1-\gamma} \) and the same factor \( h^{1-\gamma} \) is in the denominator. Thus, we have as \( h \downarrow 0 \),

\[
X'_\Delta Z_\Delta /h^{1-\gamma} \xrightarrow{P} \Sigma_{XX,N^0_b}, \tag{B.1.24}
\]
which implies that \( h^{-1/2}X' \left( Z_0 - \hat{Z}_0 \right) = O_p \left( h^{1/2-\gamma} \right) \). This shows that the order of the error in replacing \( Z_0 \) by \( Z_2 = \hat{Z}_0 \) goes to zero at a enough fast rate. That is, by definition we can write \( Y = X\beta^0 + \tilde{Z}_0 \delta^0 + \left( Z_0 - \hat{Z}_0 \right) \delta^0 + \epsilon \), from which it follows that \( X'\tilde{Z}_0 = X'Z_0 + o_p(1) \), \( X' \left( Z_0 - \tilde{Z}_0 \right) \delta^0 = o_p(1) \) and \( Z_0' \left( Z_0 - \tilde{Z}_0 \right) \delta^0 = o_p(1) \). To see this, consider for example

\[
X' \left( \tilde{Z}_0 - Z_0 \right) = \sum_{T_0^0 \in [T^\gamma]} T_0^0 x_{kh} z_{kh} = h^{1-\gamma} \sum_{T_0^0 \in [T^\gamma]} x_{kh} z_{kh} = h^{1-\gamma} O_p(1),
\]

which clearly implies that \( X'\tilde{Z}_0 = X'Z_0 + o_p(1) \). The other case can be proven similarly. This concludes the proof of the Lemma. □

Using Lemma B.1.7, the proof of the proposition becomes simple.

**Proof of Proposition 2.3.3.** By standard arguments,

\[
\sqrt{T} \begin{bmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta} - \delta^0 \end{bmatrix} = \begin{bmatrix} X'X & X'\tilde{Z}_0 \\ \tilde{Z}_0'X & \tilde{Z}_0'\tilde{Z}_0 \end{bmatrix}^{-1} \sqrt{T} \begin{bmatrix} X'\epsilon + X' \left( Z_0 - \tilde{Z}_0 \right) \delta^0 \\ \tilde{Z}_0'\epsilon + \tilde{Z}_0' \left( Z_0 - \tilde{Z}_0 \right) \delta^0 \end{bmatrix},
\]

from which it follows that

\[
\begin{bmatrix} X'X & X'\tilde{Z}_0 \\ \tilde{Z}_0'X & \tilde{Z}_0'\tilde{Z}_0 \end{bmatrix}^{-1} \frac{1}{h^{1/2}} X' \left( Z_0 - \hat{Z}_0 \right) \delta^0 = O_p(1) o_p(1) = o_p(1),
\]

and a similar reasoning applies to \( \tilde{Z}_0' \left( Z_0 - \tilde{Z}_0 \right) \delta^0 \). All other terms involving \( \tilde{Z}_0 \) can be treated in analogous fashion. In particular, the \( O_p(1) \) result above follows from Lemma B.1.3-B.1.4. The rest of the arguments (including mixed normality) follows from Barndorff-Nielsen and Shephard (2004) and are omitted. □
B.1.4.5 Proof of Proposition 2.4.1

Proof of part (i) of Proposition 2.4.1. Below $C$ is a generic positive constant which may change from line to line. Let $\tilde{e}$ denote the vector of normalized residuals $\tilde{e}_t$ defined by (2.4.1). Recall that $\hat{T}_b = \arg \max_{T_b} Q_T (T_b)$, $Q_T (\hat{T}_b) = \hat{\delta}_{T_b}^t (Z'_2 M Z_2) \hat{\delta}_{T_b}$, and the decomposition

\[
Q_T (T_b) - Q_T (T^0_b) = \hat{\delta}_{T_b}^t (Z'_2 M Z_2) \hat{\delta}_{T_b} - \hat{\delta}_{T^0_b}^t (Z'_0 M Z_0) \hat{\delta}_{T^0_b} \tag{B.1.25}
\]

\[
= \delta'_h \left\{ (Z'_0 M Z_2) (Z'_2 M Z_2)^{-1} (Z'_2 M Z_0) - Z'_0 M Z_0 \right\} \delta_h \tag{B.1.26}
\]

\[
+ g_e (T_b), \tag{B.1.27}
\]

where

\[
g_e (T_b) = 2 \delta'_h (Z'_0 M Z_2) (Z'_2 M Z_2)^{-1} Z_2 M e - 2 \delta'_h (Z'_0 M e) \tag{B.1.28}
\]

\[
+ e' M Z_2 (Z'_2 M Z_2)^{-1} Z_2 M e - e' M Z_0 (Z'_0 M Z_0)^{-1} Z'_0 M e. \tag{B.1.29}
\]

Since $g_e (\hat{T}_b) \geq |\hat{T}_b - T^0_b| r (\hat{T}_b)$, we have

\[
P \left( |\hat{\lambda}_b - \lambda_0| > K \right) = P \left( |\hat{T}_b - T^0_b| > TK \right)
\]

\[
\leq P \left( \sup_{|T_b - T^0_b| > TK} h^{-1/2} |g_e (T_b)| \geq \inf_{|T_b - T^0_b| > TK} h^{-1/2} |T_b - T^0_b| r (T_b) \right)
\]

\[
\leq P \left( \sup_{p \leq T_b \leq T - p} h^{-1/2} |g_e (T_b)| \geq TK \inf_{|T_b - T^0_b| > TK} h^{-1/2} r (T_b) \right)
\]

\[
= P \left( r^{-1}_{T} \sup_{p \leq T_b \leq T - p} h^{-1/2} |g_e (T_b)| \geq K \right), \tag{B.1.30}
\]
where \( r_T = T \inf_{[T_0 - T']^T} h^{-1/2} r (T_b) \), which is positive and bounded away from zero by Lemma B.1.8. Thus, it is sufficient to verify that

\[
\sup_{p \leq T_b \leq T - p} h^{-1/2} |g_e (T_b)| = o_p (1).
\]

Consider the first term of \( g_e (T_b) \):

\[
2 \delta^h (Z_0' M Z_2) (Z_2' M Z_2)^{-1/2} (Z_2' M Z_2)^{-1/2} Z_2 M e
\]

\[
\leq 2 h^{1/4} (\delta^0)' (Z_0' M Z_2) (Z_2' M Z_2)^{-1/2} (Z_2' M Z_2)^{-1/2} Z_2 M e.
\]

For any \( 1 \leq j \leq p \), \( (Z_2 e)_{j,1} / \sqrt{h} = O_p (1) \) by Theorem B.1.5, and similarly, for any \( 1 \leq i \leq q + p \), \( (X e)_{i} / \sqrt{h} = O_p (1) \). Furthermore, from Lemma B.1.3 we also have that \( Z_2' M Z_2 \) and \( Z_0' M Z_2 \) are \( O_p (1) \). Therefore, the supremum of \( (Z_0' M Z_2) (Z_2' M Z_2)^{-1} \) over all \( T_b \) is such that

\[
\sup_{T_b} \left( Z_0' M Z_2 \right) (Z_2' M Z_2)^{-1} \left( Z_2' M Z_0 \right) \leq Z_0' M Z_0 = O_p (1),
\]

by Lemma B.1.3. By Assumption 3.1-(iii) \( (Z_2' M Z_2)^{-1/2} Z_2 M e \) is \( O_p (1) \) \( \left( \sqrt{h} \right) \) uniformly, which implies that (B.1.32) is \( O_p \left( \sqrt{h} \right) \) uniformly over \( p \leq T_b \leq T - p \). In view of Assumption 2.6 [recall (2.4.1)], it is crucial to study the behavior of \( (X e)_{j,1} \) for \( 1 \leq j \leq p + q \). Note first that \( \left| \hat{\lambda}_b - \lambda_0 \right| > K \) or \( N > \left| \hat{N}_b - N_0^b \right| > KN \). Then, by Itô formula proceeding as in the proof of Lemma B.1.2, we have a standard result for the local volatility of a continuous Itô semimartingale; namely that for some \( A > 0 \) (recall the condition \( T^{1-\kappa} \epsilon \rightarrow B > 0 \)),

\[
\left\| \mathbb{E} \left( \frac{1}{\epsilon} \sum_{T_0^b - [T^\kappa]} x_{kh} e_{kh} - \frac{1}{\epsilon} \int_{N_0^b - \epsilon}^{N_0^b} \sum_{X_{k+1},s} d s \mid \mathcal{F} (T_0^b - 1)_{h} \right) \right\| \leq Ah^{1/2}.
\]
From Assumption 3.1-(iv) since $\Sigma_{Xe,t} = 0$ for all $t \geq 0$, we have

$$X'e = \sum_{k=1}^{T_0} x_{kh} \tilde{e}_{kh} + h^{-1/4} \sum_{k=T_b}^{T_0} x_{kh} \tilde{e}_{kh} + \sum_{k=T_b}^{T} x_{kh} \tilde{e}_{kh}$$

$$= O_p \left( h^{1/2} \right) + h^{-1/4} O_p \left( h^{1-k+1/2} \right) + O_p \left( h^{1/2} \right) = O_p \left( h^{1/2} \right). \quad \text{(B.1.33)}$$

The same bound applies to $Z'e$ and $Z_0'e$. Thus, equation (B.1.32) is such that

$$2h^{-1/2}h^{1/4} \left( \delta_0' \right)' \left( Z_0'MZ_2 \right) \left( Z_2'MZ_2 \right)^{-1/2} \left( Z_2'MZ_2 \right)^{-1/2} Z_2Me$$

$$= 2h^{-1/2}h^{1/4} \left\| \delta_0 \right\| O_p \left( 1 \right) O_p \left( h^{1/2} \right) = O_p \left( 1 \right) O_p \left( h^{1/4} \right).$$

As for the second term of (B.1.28),

$$h^{-1/2} \delta_h' \left( Z_0'Me \right) = 2h^{-1/4} \left( \delta_0' \right)' \left( Z_0'Me \right) = Ch^{-1/4}O_p \left( h^{1/2} \right) = CO_p \left( h^{1/4} \right),$$

using (B.1.33). Again using (B.1.33), the first term in (B.1.29) is, uniformly in $T_b$,

$$h^{-1/2}e'MZ_2 \left( Z_2'MZ_2 \right)^{-1} Z_2Me = h^{-1/2}BO_p \left( h^{1/2} \right) O_p \left( 1 \right) O_p \left( h^{1/2} \right) = O_p \left( h^{1/2} \right). \quad \text{(B.1.34)}$$

Similarly, the last term in (B.1.29) is $O_p \left( h^{1/2} \right)$. Therefore, combining these results we have $h^{-1/2} \sup_{T_b} |g_e (T_b)| = BO_p \left( h^{1/4} \right)$, from which it follows that the right-hand side of (B.1.30) is weakly smaller than $\varepsilon$.

**Lemma B.1.8.** For $B > 0$, let $r_{B,h} = \inf_{|T_b-T_0|>TB} Th^{-1/2-r} \left( T_b \right)$. There exists an $A > 0$ such that for every $\varepsilon > 0$, there exists a $B < \infty$ such that $P \left( r_{B,h} \geq A \right) \leq 1 - \varepsilon$.

**Proof.** Assume $N_b \leq N_0^h$, and observe that $r_T \geq r_{B,h}$ for an appropriately chosen $B$. 
From the first inequality result in Lemma B.1.1,

\[ Th^{-1/2}r(T_b) \geq Th^{-1/2}h^{1/2} (\delta^0)' \frac{X_\Delta'X_\Delta}{T^0_b - T_b} (X^0_2X_0)^{-1} (X^0_0X_0) R \delta^0 \]

\[ = (\delta^0)' R' \left( X_\Delta'X_\Delta / \left( N^0_b - N_b \right) \right) (X^0_2X_0)^{-1} (X^0_0X_0) R \delta^0. \]

Note that \( B < h(T^0_b - T_b) < N \). Then

\[ Th^{-1/2}r(T_b) \geq (\delta^0)' R' \left( X_\Delta'X_\Delta / N \right) (X^0_2X_0)^{-1} (X^0_0X_0) R \delta^0 > A \]

by the same argument as in Lemma B.1.6. Following the same reasoning as in the proof of Lemma B.1.6 we can choose a \( B > 0 \) such that

\[ r_{B,h} = \inf_{|T_b - T^0_b| > TB} Th^{-1/2}r(T_b) \]

satisfies \( P(r_{B,h} \geq A) \leq 1 - \varepsilon. \)

**Proof of part (ii) of Proposition 2.4.1.** Suppose \( T_b < T^0_b \). Let

\[ D_{K,T} = \left\{ T_b : N \eta \leq N_b \leq N (1 - \eta), \left| N_b - N^0_b \right| > K (T^{1-\kappa})^{-1} \right\}. \]

It is enough to show \( P \left( \sup_{T_b \in D_{K,T}} Q_T(T_b) \geq Q_T(T^0_b) \right) < \varepsilon \). The difficulty is again to control the estimates that depend on \( |N_b - N^0_b| \). We need to show

\[ P \left( \sup_{T_b \in D_{K,T}} h^{-3/2} g_e(T_b, \delta_h) \geq \inf_{T_b \in D_{K,T}} h^{-3/2}r(T_b) \right) < \varepsilon. \]

By Lemma B.1.1,

\[ \inf_{T_b \in D_{K,T}} r(T_b) \geq \inf_{T_b \in D_{K,T}} \delta_h' R' \frac{X_\Delta'X_\Delta}{T^0_b - T_b} (X^0_2X_0)^{-1} (X^0_0X_0) R \delta_h \]
and since $|T_b - T^0_b| > KT^\kappa$, it is important to consider $X'_\Delta X_\Delta = \sum_{k=T^0_b+1}^{T^0_b} x_{kh}x'_{kh}$. We shall apply asymptotic results for the local approximation of the covariation between processes. Consider
\[
\frac{X'_\Delta X_\Delta}{h (T^0_b - T_b)} = \frac{1}{h (T^0_b - T_b)} \sum_{k=T^0_b+1}^{T^0_b} x_{kh}x'_{kh}.
\]
By Theorem 9.3.2-(i) in Jacod and Protter (2012), as $h \downarrow 0$
\[
\frac{1}{h (T^0_b - T_b)} \sum_{k=T^0_b+1}^{T^0_b} x_{kh}x'_{kh} \xrightarrow{P} \Sigma_{XX,N^0_b}, \tag{B.1.35}
\]
since $|N_b - N^0_b|$ shrinks at a rate no faster than $Kh^{1-\kappa}$ and $1/Kh^{1-\kappa} \to \infty$. By Lemma B.1.2 this approximation is uniform, establishing that
\[
h^{-3/2} \inf_{T_b \in D_{K,T}} (\delta_h)' R' \frac{X'_\Delta X_\Delta}{T^0_b - T_b} (X'_0X_2)^{-1} (X'_0X_0) R\delta_h
\]
\[
= \inf_{T_b \in D_{K,T}} (\delta^0)' R' \frac{X'_\Delta X_\Delta}{h (T^0_b - T_b)} (X'_0X_2)^{-1} (X'_0X_0) R\delta^0,
\]
is bounded away from zero. Thus, it is sufficient to show
\[
P \left( \sup_{T_b \in D_{K,T}} h^{-3/2} g_e (T_b, \delta_h) \left| \frac{T^0_b - T_b}{T^0_b} \right| \geq B \right) < \varepsilon, \tag{B.1.36}
\]
for some $B > 0$. Consider the terms of $g_e (T_b)$ in (B.1.29). Using $Z_2 = Z_0 \pm Z_\Delta$, we can deduce for the first term,
\[
\delta'_h (Z'_0MZ_2) (Z'_2MZ_2)^{-1} Z_2Me
\]
\[
= \delta'_h (\pm Z_\Delta) (Z'_0MZ_2) (Z'_2MZ_2)^{-1} Z_2Me
\]
\[
= \delta'_h Z'_0Me + \delta'_h Z'_0Me + \delta'_h (Z'_\DeltaMZ_2) (Z'_2MZ_2)^{-1} Z_2Me. \tag{B.1.37}
\]
First, we can apply Lemma B.1.3 [(vi)-(viii)], together with Assumption 3.1-(iii), to the terms that do not involve $|N_b - N_0^b|$. Let us focus on the third term,

$$K^{-1}h^{-(1-\kappa)} (Z'_\Delta M Z_2) = \frac{Z'_\Delta Z_2}{Kh^{1-\kappa}} - \frac{Z'_\Delta X_\Delta}{Kh^{1-\kappa}} (X'X)^{-1} X'Z_2. \quad (B.1.38)$$

Consider $Z'_\Delta Z_\Delta$ (the argument for $Z'_\Delta X_\Delta$ is analogous). By Lemma B.1.2, $Z'_\Delta Z_\Delta /Kh^{1-\kappa}$ uniformly approximates the moving average of $\Sigma_{ZZ,t}$ over $(N_0^b - KT^\kappa h, N_0^b]$. Hence, as $h \downarrow 0$,

$$Z'_\Delta Z_\Delta /Kh^{1-\kappa} = BO_p(1), \quad (B.1.39)$$

for some $B > 0$, uniformly in $T_b$. The second term in (B.1.38) is thus also $O_p(1)$ uniformly using Lemma B.1.3. Then, using (B.1.33) and (B.1.38) into the third term of (B.1.37), we have

$$\frac{1}{K}h^{-(1-\kappa)-1/2} (\delta_h)' (Z'_\Delta M Z_2) (Z'_2 M Z_2)^{-1} Z_2 M e \quad (B.1.40)$$

$$\leq \frac{1}{K}h^{-1/4} (\delta^0)' \left( \frac{Z'_\Delta M Z_2}{h^{1-\kappa}} \right) (Z'_2 M Z_2)^{-1} Z_2 M e$$

$$\leq h^{-1/4} \frac{Z'_\Delta M Z_2}{Kh^{1-\kappa}} O_p(1) O_p(h^{1/2}) \leq O_p\left(h^{1/4}\right),$$

where $(Z'_2 M Z_2)^{-1} = O_p(1)$. So the right-hand side of (B.1.40) is less than $\varepsilon/4$ in
probability. Therefore, for the second term of (B.1.37),

\[ K^{-1}h^{-(1-\kappa)-1/2}\delta_h'Z_\Delta'Me \]

\[ = \frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h' \sum_{k=T_b+1}^{T_0} z_{kh}e_{kh} - \frac{h^{-1/2}}{h^{1-\kappa}}\delta_h' \left( \sum_{k=T_b+1}^{T_0} z_{kh}c_{kh}' \right) (X'X)^{-1} (X'e) \]

\[ \leq \frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h' \sum_{k=T_b+1}^{T_0} z_{kh}e_{kh} - B\frac{1}{K}h^{-1/4} \left( \delta^0 \right) \left( \sum_{k=T_b+1}^{T_0} z_{kh}c_{kh}' \right) (X'X)^{-1} (X'e) \]

\[ \leq \frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h' \sum_{k=T_b+1}^{T_0} z_{kh}e_{kh} - h^{-1/4}O_p \left( 1 \right) O_p \left( h^{1/2} \right). \quad (B.1.41) \]

Thus, using (B.1.37), (B.1.28) is such that

\[ 2\delta_h'Z_0'Me + 2\delta_h'Z_\Delta'Me \pm 2\delta_h' \left( Z_\Delta'MZ_2 \right) \left( Z_2'MZ_2 \right)^{-1} Z_2'Me - 2\delta_h' \left( Z_0'Me \right) \]

\[ = 2\delta_h'Z_\Delta'Me + 2\delta_h' \left( Z_\Delta'MZ_2 \right) \left( Z_2'MZ_2 \right)^{-1} Z_2'Me \]

\[ \leq \frac{h^{-1/2}}{Kh^{1-\kappa}} \left( \delta^0 \right)' \sum_{k=T_b+1}^{T_0} z_{kh}c_{kh} - h^{-1/4}O_p \left( 1 \right) O_p \left( h^{1/2} \right) + O_p \left( h^{-1/4} \right), \]

in view of (B.1.40) and (B.1.41). Next, consider equation (B.1.29). We can use the decomposition \( Z_2 = Z_0 \pm Z_\Delta \) and show that all terms involving the matrix \( Z_\Delta \) are negligible. To see this, consider the first term when multiplied by \( K^{-1}h^{-(3/2-\kappa)} \),

\[ K^{-1}h^{-(3/2-\kappa)}c'MZ_2 \left( Z_2'MZ_2 \right)^{-1} Z_2'Me = K^{-1}h^{-(3/2-\kappa)}c'MZ_0 \left( Z_2'MZ_2 \right)^{-1} Z_2'Me \]

\[ \pm K^{-1}h^{-(3/2-\kappa)}c'MZ_\Delta \left( Z_2'MZ_2 \right)^{-1} Z_2'Me. \quad (B.1.42) \]

By the same argument as in (B.1.33), \( Z_2'Me = O_p \left( h^{1/2} \right) \). Then, using the Burkhölder-Davis-Gundy inequality, estimates for the local volatility of continuous Itô semimar-
tingales yield

\[
\tilde{e}' M Z_\Delta = \tilde{e}' Z_\Delta - \tilde{e}' X (X'X)^{-1} X' Z_\Delta
\]

\[
= O_p \left( Kh^{1/2+1-\kappa} \right) - O_p \left( h^{1/2} \right) O_p (1) O_p \left( Kh^{1-\kappa} \right).
\]

Thus, the second term in (B.1.42) is such that

\[
K^{-1}h^{-(3/2-\kappa)} \tilde{e}' M Z_\Delta (Z_2' M Z_2)^{-1} Z_2 M e
\]

\[
= B \left( K^{-1}h^{-(3/2-\kappa)} \right) O_p \left( Kh^{1-\kappa+1/2} \right) O_p (1) O_p \left( h^{1/2} \right) = BO_p \left( h^{1/2} \right).
\]

Next, let us consider (B.1.29). The key here is to recognize that on, \( D_{K,T} \), \( T_b \) and \( T_b^0 \) lies on the same window with right-hand point \( N_b^0 \). Thus the difference between the two terms in (B.1.29) is asymptotically negligible. First, note that using (B.1.33),

\[
\tilde{e}' M Z_0 (Z_0' M Z_0)^{-1} Z_0 M e = O_p \left( h^{1/2} \right) O_p (1) O_p \left( h^{1/2} \right) = O_p \left( h \right).
\]

By the fact that \( Z_0 = Z_2 \pm Z_\Delta \) applied repeatedly in (B.1.42), and noting that the cross-product terms involving \( Z_\Delta \) are \( o_p (1) \) by the same reasoning as in (B.1.43), we obtain that the difference between the first and second term of (B.1.29) is negligible. The more intricate step is the one arising from

\[
e' M Z_0 (Z_0' M Z_2 \pm Z_\Delta' M Z_2)^{-1} Z_0' M e - e' M Z_0 (Z_0' M Z_0)^{-1} Z_0' M e
\]

\[
= e' M Z_0 \left[ (Z_0' M Z_2 \pm Z_\Delta' M Z_2)^{-1} - (Z_0' M Z_0)^{-1} \right] Z_0' M e.
\]

On \( D_{K,T} \), \( |N_b - N_b^0| = O_p (Kh^{1-\kappa}) \), and so each term involving \( Z_\Delta \) is of higher order. By using the continuity of probability limits the matrix in square brackets goes to
zero at rate $h^{1-\kappa}$. Then, this expression when multiplied by $h^{-(3/2-\kappa)}K^{-1}$ and after using the same rearrangements as above, can be shown to satisfy [recall also (B.1.33)]

$$
\begin{align*}
&h^{-(3/2-\kappa)}K^{-1}e'MZ_0 \left[ (Z'_0MZ_0 \pm Z'_\Delta MZ_0)^{-1} - (Z'_0MZ_0)^{-1} \right] Z'_0Me \\
&= h^{-(3/2-\kappa)}O_p(h) \left[ (Z'_0MZ_0 \pm Z'_\Delta MZ_0)^{-1} - (Z'_0MZ_0)^{-1} \right] \\
&= h^{-(3/2-\kappa)}O_p(h) \left[ (Z'_0MZ_0 \pm Z'_0MZ'_0 \pm Z'_\Delta MZ_0)^{-1} - (Z'_0MZ_0)^{-1} \right] \\
&= h^{-(3/2-\kappa)}O_p(h) O_p(h^{1-\kappa}) = O_p(h^{1/2}) o_p(1).
\end{align*}
$$

Therefore, (B.1.29) is stochastically small uniformly in $T_b \in D_{K,T}$ when $T$ is large.

Altogether, we have

$$
\begin{align*}
&h^{-1/2} \exp \left( T_b \right) \leq 2 \frac{h^{-1/2}}{K^{h^{1-\kappa}}} \sum_{k=T_b+1}^{T_0} z_k e_k - h^{-1/4} O_p\left(h^{1/2}\right) + O_p(h^{-1/4}).
\end{align*}
$$

Thus, it remains to find a bound for the first term above. By Itô’s formula, standard estimates for the local volatility of continuous Itô semimartingales yield for every $T_b$,

$$
\begin{align*}
\mathbb{E} \left( \left\| \sum_{k=T_b}^{T_0} e_k \left( T_2, T_b \right) - \sum_{k=T_b}^{T_0} e_k \left( T_2, T_b \right) \right\| | \mathcal{F}_{T_b,h} \right) \leq B h^{1/2}, \quad \text{(B.1.44)}
\end{align*}
$$

for some $B > 0$. Let $R_{1,h} = \sum_{k=T_b}^{T_0} z_k e_k$, $R_{2,h} (T_b) = \sum_{k=T_b+1}^{T_0} z_k e_k$.
Consider first the second probability. By Markov’s inequality,

\[ P \left( \sup_{T_b < T_b^0 - KT^\kappa} \frac{1}{K h^{1-\kappa}} \| R_{2,h} (T_b) \| > 4^{-1} C \| \delta^0 \|^{-1} h^{1/4} \right) \]

\[ \leq (K/B) T^\kappa P \left( \left\| \frac{1}{K h^{1-\kappa}} R_{2,h} (T_b) \right\| > 4^{-1} C \| \delta^0 \|^{-1} h^{1/4} \right) \]

\[ \leq \frac{4 (B + 1) \| \delta^0 \| r}{C} r^{-r/4} K B T^\kappa \mathbb{E} \left( \left\| R_{2,h} (T_b) \right\|^r \right) \]

\[ \leq C_r (B + 1) B^{-1} \| \delta^0 \|^r h^{-r/4} T^\kappa h^{1/2} \leq C_r \| \delta^0 \|^r h^{r/2 - \kappa - r/4} \rightarrow 0, \]

for a sufficiently large \( r > 0 \). We now turn to \( R_{1,h} \). We have,

\[ P \left( \frac{1}{K h^{1-\kappa}} \| R_{1,h} \| > 2^{-1} C \| \delta^0 \|^{-1} h^{1/2} \right) \]

\[ \leq P \left( \frac{(B + 1)}{K} \left\| (B + 1)^{-1} h^{-(1-\kappa)} \sum_{k=T_b^0 - (B+1)[T^\kappa] + 1}^{T_b^0} z_{kh} \right\| > C \| \delta^0 \|^{-1} h^{1/2} \right) \]

\[ \leq P \left( (B + 1) K^{-1} O_p (1) > 4^{-1} C \| \delta^0 \|^{-1} \right) \rightarrow 0, \]

by choosing \( K \) large enough where we have used (B.1.44). Altogether, the right-hand
side of (B.1.45) is less than $\varepsilon$, which concludes the proof. □

**Proof of part (iii) of Proposition 2.4.1.** Observe that Lemma B.1.7 applies under this setting. Then, we have,

$$
\sqrt{T} \left[ \begin{array}{c}
\bar{\beta} - \beta_0 \\
\delta - \delta_h
\end{array} \right] = \left[ \begin{array}{cc}
X'X & X'\tilde{Z}_0 \\
\tilde{Z}_0'X & \tilde{Z}_0'\tilde{Z}_0
\end{array} \right]^{-1} \sqrt{T} \left[ \begin{array}{c}
X'e + X'(Z_0 - \tilde{Z}_0)\delta_h \\
\tilde{Z}_0'e + \tilde{Z}_0'(Z_0 - \tilde{Z}_0)\delta_h
\end{array} \right],
$$

so that we have to show

$$
\left[ \begin{array}{cc}
X'X & X'\tilde{Z}_0 \\
\tilde{Z}_0'X & \tilde{Z}_0'\tilde{Z}_0
\end{array} \right]^{-1} \frac{1}{h^{1/2}} X' (Z_0 - \tilde{Z}_0) \delta_h \overset{p}{\to} 0,
$$

and that the limiting distribution of $X'e/h^{1/2}$ is Gaussian. The first claim can be proven in a manner analogous to that in the proof of Proposition 2.3.3. For the second claim, we have the following decomposition from (B.1.33),

$$
X'e = \sum_{k=1}^{T_b^{0} - |T^\kappa|} x_{kh}\tilde{e}_{kh} + h^{-1/4} \sum_{T_b^{0} - |T^\kappa| + 1}^{T_b^{0} + |T^\kappa|} x_{kh}\tilde{e}_{kh} + \sum_{k=T_b^{0} + |T^\kappa| + 1}^{T} x_{kh}\tilde{e}_{kh}
$$

$$
\triangleq R_{1,h} + R_{2,h} + R_{3,h}.
$$

By Theorem B.1.5, $h^{-1/2} R_{1,h} \overset{L^2}{\to} \mathcal{N}(0, V_1)$, where $V_1 \triangleq \lim_{T \to \infty} T \sum_{k=1}^{T_b^{0} - |T^\kappa|} \mathbb{E}(x_{kh}x_{kh}' \tilde{e}_{kh}^2)$. Similarly, $h^{-1/2} R_{3,h} \overset{L^2}{\to} \mathcal{N}(0, V_3)$, where $V_3 \triangleq \lim_{T \to \infty} T \sum_{k=T_b^{0} + |T^\kappa| + 1}^{T} \mathbb{E}(x_{kh}x_{kh}' \tilde{e}_{kh}^2)$. If $\kappa \in (0, 1/4)$, $h^{-(1-\kappa)} \sum_{T_b^{0} - |T^\kappa| + 1}^{T_b^{0} + |T^\kappa|} x_{kh}\tilde{e}_{kh} \overset{p}{\to} \Sigma_{Xe,N^0_b} \sigma_{Xe}$ by Theorem 9.3.2 in Jacod and Protter (2012) and so $h^{-1/2} R_{2,h} = h^{-3/4} \sum_{T_b^{0} - |T^\kappa|}^{T_b^{0} + |T^\kappa|} x_{kh}\tilde{e}_{kh} \overset{p}{\to} 0$. If $\kappa = 1/4$, then $h^{-1/2} R_{2,h} \overset{p}{\to} \Sigma_{Xe,N^0_b}$ in probability again by Theorem 9.3.2 in Jacod and Protter (2012). Since by Assumption 3.1-(iv) $\Sigma_{Xe,t} = 0$ for all $t \geq 0$, whenever $\kappa \in (0, 1/4]$, $X'e/h^{1/2}$ is asymptotically normally distributed. The rest of the proof is simple and follows the same steps as in Proposition 2.3.3. □
B.1.4.6 Proof of Lemma 2.4.1

First, we begin with the following simple identity. Throughout the proof, $B$ is a generic constant which may change from line to line.

**Lemma B.1.9.** The following identity holds

\[
(\delta_h)' \left\{ Z_0' M Z_0 - (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) \right\} \delta_h
= (\delta_h)' \left\{ Z_0' M Z_0 - (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) \right\} \delta_h.
\]

*Proof.* The proof follows simply from the fact that $Z_0' M Z_2 = Z_0' M Z_2 \pm Z_0' M Z_2$ and so

\[
(\delta_h)' \left\{ Z_0' M Z_0 - (Z_0' M Z_2 \pm Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) \right\} \delta_h
= (\delta_h)' \left\{ Z_0' M Z_0 - (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_2) \right. \\
- (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) \left\} \delta_h
= (\delta_h)' \left\{ Z_0' M Z_0 - (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) \right\} \delta_h. \quad \Box
\]

*Proof of Lemma 2.4.1.* By the definition of $Q_T (T_b) - Q_T (T_0)$ and Lemma B.1.9,

\[
Q_T (T_b) - Q_T (T_0)
= -\delta_h' \left\{ Z_0' M Z_0 - (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) \right\} \delta_h + g_e (T_b, \delta_h), \quad (B.1.46)
\]

where

\[
g_e (T_b, \delta_h) = 2\delta_h' (Z_0' M Z_2) (Z_2' M Z_2)^{-1} Z_2 M e - 2\delta_h' (Z_0' M e) \quad (B.1.47)
+ e' M Z_2 (Z_2' M Z_2)^{-1} Z_2 M e - e' M Z_0 (Z_0' M Z_0)^{-1} Z_0' M e. \quad (B.1.48)
\]
Recall that $N_b(u) \in D(C)$ implies $T_b(u) = T^0_b + uT^\kappa$, $u \in [-C, C]$. We consider the case $u \leq 0$. By Theorem 9.3.2-(i) in Jacod and Protter (2012) combined with Lemma B.1.2, we have uniformly in $u$ as $h \downarrow 0$

\[
\frac{1}{h^{1-\kappa}} \sum_{k=T_b^0 + uT^\kappa}^{T_b^0} x_{kh}x'_{kh} \xrightarrow{P} \Sigma_{XX_N^0}.
\] (B.1.49)

Since $Z'_\Delta X = Z'_\Delta X_\Delta$, we will use this result also for $Z'_\Delta X/h^{1-\kappa}$. With the notation of Section B.1.4.1 [recall (B.1.6)], by the Burkhölder-Davis-Gundy inequality, we have that standard estimates for the local volatility yield,

\[
\left\| \mathbb{E}\left( \tilde{\Sigma}_{ZX}(T_b, T^0_b) - \Sigma_{ZX}(T^0_b - 1)_h | \mathcal{F}_{(T^0_b - 1)_h} \right) \right\| \leq Bh^{1/2}. \tag{B.1.50}
\]

Equation (B.1.49)-(B.1.50) can be used to yield, uniformly in $T_b$,

\[
\psi_h^{-1} Z'_\Delta X (X'X)^{-1} X'Z_\Delta = O_p(1) X'Z_\Delta, \tag{B.1.51}
\]

and

\[
Z'_\Delta MZ_2 = Z'_\Delta Z_\Delta - Z'_\Delta X (X'X)^{-1} X'Z_2 = O_p(\psi_h) - O_p(\psi_h) O_p(1) O_p(1). \tag{B.1.52}
\]

Now, expand the first term of (B.1.46),

\[
\delta'_h Z'_\Delta MZ_\Delta \delta_h = \delta'_h Z'_\Delta Z_\Delta \delta_h - \delta'_h Z'_\Delta X (X'X)^{-1} X'Z_\Delta \delta_h. \tag{B.1.53}
\]
By Lemma B.1.3, \((X'X)^{-1} = O_{p}(1)\) and recall \(\delta_h = h^{1/4}\delta^0\). Then,

\[
\psi_h^{-1}\delta_h' Z'_\Delta M Z_\Delta \delta_h = \psi_h^{-1}\delta_h' Z'_\Delta Z_\Delta \delta_h - \psi_h^{-1}\delta_h' Z'_\Delta X (X'X)^{-1} X'Z_\Delta \delta_h. \tag{B.1.54}
\]

By (B.1.51), the second term above is such that

\[
\|\delta^0\|^2 h^{3/2} \frac{Z'_\Delta X}{\psi_h} (X'X)^{-1} X'Z_\Delta = \|\delta^0\|^2 h^{1/2} O_{p}(1) X'Z_\Delta, \tag{B.1.55}
\]

uniformly in \(T_b(u)\). Therefore,

\[
\psi_h^{-1}\delta_h' Z'_\Delta M Z_\Delta \delta_h = \psi_h^{-1}\delta_h' Z'_\Delta Z_\Delta \delta_h - \|\delta^0\|^2 h^{1/2} O_{p}(1) O_{p}(\psi_h). \tag{B.1.56}
\]

The last equality shows that the second term of \(\delta' Z'_\Delta M Z_\Delta \delta\) is always of higher order. This suggests that the term involving regressors whose parameters are allowed to shift plays a primary role in the asymptotic analysis. The second term is a complicated function of cross products of all regressors around the time of the change. Because of the fast rate of convergence, these high order product estimates around the break date will be negligible. We use this result repeatedly in the derivations that follow.

The second term of (B.1.46) when multiplied by \(\psi_h^{-1}\) is, uniformly in \(T_b(u)\),

\[
\psi_h^{-1}\delta_h (Z'_\Delta M Z_2) (Z'_2 M Z_2)^{-1} (Z'_2 M Z_\Delta) \delta_h' = \|\delta^0\|^2 h^{1/2} O_{p}(1) O_{p}(1) O_{p}(\psi_h),
\]

where we have used the fact that \(Z'_\Delta M Z_2/\psi_h = O_{p}(1)\) [cf. (B.1.52)]. Hence, the second term of (B.1.46), when multiplied by \(\psi_h^{-1}\), is \(O_{p}\left(h^{3/2-\kappa}\right)\) uniformly in \(T_b\).

Finally, let us consider \(g_{e}(T_b, \delta_h)\). Recall that \(\tilde{e}_{kh}\) defined in (2.4.1) is i.n.d. with zero mean and conditional variance \(\sigma^2_{e,k-1} h\). Upon applying the continuity of probability limits repeatedly one first obtains that the difference between the two terms in
(B.1.48) goes to zero at a fast enough rate as in the last step of the proof of Proposition 2.4.1-(ii). That is, for $T$ large enough, we can find a $c_T$ sufficiently small such that,

$$
\psi_h^{-1} \left[ e' M Z_2 (Z'_2 M Z_2)^{-1} Z_2 M e - e' M Z_0 (Z'_0 M Z_0)^{-1} Z'_0 M e \right] = o_p (c_T h).
$$

Next, consider the first two terms of $g_e (T_b, \delta_h)$. Using $Z'_0 M Z_2 = Z'_2 M Z_2 \pm Z'_\Delta M Z_2$, it is easy to show that

$$
2h^{1/4} (\delta^0)' (Z'_0 M Z_2) (Z'_2 M Z_2)^{-1} Z_2 M e - 2h^{1/4} (\delta^0)' (Z'_0 M e) = 2h^{1/4} (\delta^0)' Z'_\Delta M e \pm 2h^{1/4} (\delta^0)' Z'_\Delta M Z_2 (Z'_2 M Z_2)^{-1} Z'_2 M e. \quad (B.1.57)
$$

Note that, uniformly in $T_b (u)$,

$$
\psi_h^{-1} h^{1/4} (\delta^0)' Z'_\Delta M Z_2 = h^{1/4} (\delta^0)' Z'_\Delta Z_\Delta + (\delta^0)' h^{1/4} Z'_\Delta X \psi_h (X'X)^{-1} X' Z_2
$$

$$
= h^{1/4} (\delta^0)' Z'_\Delta Z_\Delta + (\delta^0)' h^{1/4} h O_p (1) = h^{1/4} \| \delta^0 \| O_p (1)
$$

$$
+ \| \delta^0 \| h^{1/4} O_p (1),
$$

where we have used (B.1.49) and the fact that $(X'X)^{-1}$ and $X' Z_2$ are each $O_p (1)$. Recall the decomposition in (B.1.33):

$$
X' e = O_p \left( h^{1-\kappa+1/4} \right) + O_p \left( h^{1/2} \right). \quad (B.1.58)
$$
Thus, the last term in (B.1.57) multiplied by $\psi^{-1}_h$ is

$$
\psi^{-1}_h 2 h^{1/4} \left( \delta^0 \right)' \dagger \left( \dagger \right)' Z' \Delta M Z_2 (Z'_2 M Z_2)^{-1} Z'_2 M e
$$

$$
= h^{1/4} \left\| \delta^0 \right\| O_p (1) O_p (1) \left[ O_p \left( h^{1-\kappa+1/4} \right) + O_p \left( h^{1/2} \right) \right]
$$

$$
= \left\| \delta^0 \right\| h^{1/4} O_p (1) O_p \left( h^{1/2} \right) = \left\| \delta^0 \right\| O_p \left( h^{3/4} \right).
$$

The first term of (B.1.57) can be decomposed further as follows

$$
2 h^{1/4} \left( \delta^0 \right)' Z' \Delta e = 2 h^{1/4} \left( \delta^0 \right)' Z' \Delta e - 2 h^{1/4} \left( \delta^0 \right)' Z' \Delta X (X'X)^{-1} X' e.
$$

Then, when multiplied by $\psi^{-1}_h$, the second term above is, uniformly in $T_b$,

$$
h^{1/4} \left( \delta^0 \right)' \left( Z'_\Delta X / \psi_h \right) (X'X)^{-1} X' e
$$

$$
= h^{1/4} \left( \delta^0 \right)' O_p (1) O_p (1) \left[ O_p \left( h^{1-\kappa+1/4} \right) + O_p \left( h^{1/2} \right) \right] = O_p \left( h^{3/4} \right),
$$

where we have used (B.1.49) and (B.1.58). Combining the last results, we have uniformly in $T_b$,

$$
\psi^{-1}_h g_c (T_b, \delta_h) = 2 h^{1/4} \left( \delta^0 \right)' \left( Z'_\Delta e / \psi_h \right) + O_p \left( h^{3/4} \right) + \left\| \delta^0 \right\| O_p \left( h^{3/4} \right) + o_p (c_T h),
$$

when $T$ is large and $c_T$ is a sufficiently small number. Then,

$$
\psi^{-1}_h \left( Q_T (T_b) - Q_T (T_b^0) \right) = -\delta_h \left( Z'_\Delta Z / \psi_h \right) \delta_h - 2 \delta_h \left( Z'_\Delta e / \psi_h \right)
$$

$$
+ O_p \left( h^{3/2-\kappa} \right) + O_p \left( h^{3/4} \right) + \left\| \delta^0 \right\| O_p \left( h^{3/4} \right) + o_p (c_T h).
$$
Therefore, for $T$ large enough,

$$\psi_h^{-1}\left(Q_T(T_b) - Q_T(T_b^0)\right) = -\delta_h (Z^T \mathfrak{D}Z_\Delta / \psi_h) \delta_h \pm 2\delta_h (Z^T \mathfrak{D}e / \psi_h) + o_p\left(h^{1/2}\right).$$

This concludes the proof of Lemma 2.4.1. □

### B.1.4.7 Proof of Proposition 2.5.1

**Proof.** Replace $\xi_1, \xi_2, \rho$ and $\vartheta$ in (2.4.5) by their corresponding estimates $\hat{\xi}_1, \hat{\xi}_2, \hat{\rho}$ and $\hat{\vartheta}$, respectively. Multiply both sides of (2.4.5) by $h^{-1}$ and apply a change in variable $v = s/h$. Consider the case $s < 0$. On the “fast time scale” $W^*$ is replaced by $\hat{W}_{1,h}(s) = W^*_{1,h}(sh)$ ($s < 0$) where $W^*_{1,h}(s)$ is a sample-size dependent Wiener process. It follows that

$$-h^{-1} \frac{|s|}{2} + h^{-1} W^*_{1,h} (hs) = -\frac{|v|}{2} + W^*_1(v).$$

A similar argument can be applied for $s \geq 0$. Let $\hat{\mathcal{V}}(s)$ denote our estimate of $\mathcal{V}(s)$ constructed with the proposed estimates in place of the population parameters. Then,

$$h^{-1} \argmax_{s \in [-\lambda_0 \vartheta, (1-\lambda_0) \vartheta]} \hat{\mathcal{V}}(s) = \argmax_{v \in [-\lambda_0 \vartheta/h, (1-\lambda_0) \vartheta/h]} \hat{\mathcal{V}}(v) \Rightarrow \argmax_{v \in [-\lambda_0 \vartheta, (1-\lambda_0) \vartheta]} \mathcal{V}(v),$$

which is equal to the right-hand side of (2.4.5). Recall that

$$\vartheta = \|\delta_0\|^2 \sigma^{-2} \left((\delta_0) \left((Z, Z) \| \delta_0\|^2 / \Omega_{\mathcal{W},1} (\delta_0)\right).$$

Therefore, equation (2.4.5) holds when we use the proposed plug-in estimates. □
B.1.5 Proofs of Section 2.9.2

The steps are similar to those used for the case when the model does not include predictable processes. However, we need to rely occasionally on different asymptotic results since the latter processes have distinct statistical properties. Recall that the dependent variable $\Delta_h Y_k$ in model (2.2.6) is the increment of a discretized process which cannot be identified as an ordinary diffusion. However, its normalized version, $\tilde{Y}_{(k-1)h} \triangleq h^{1/2} Y_{(k-1)h}$, is well-defined and we exploit this property in the proof. $\Delta_h Y_k$ has first conditional moment on the order $O(h^{-1/2})$, it has unbounded variation and does not belong to the usual class of semimartingales.\footnote{For an introduction to the terminology used in this sub-section, we refer the reader to first chapters in Jacod and Shiryaev (2003).} The predictable process $\{\tilde{Y}_{(k-1)h}\}_{k=1}^T$ derived from it has different properties. Its “quadratic variation” exists, and thus it is finite in any fixed time interval. That is, the integrated second moments of the regressor $Y_{(k-1)h}$ are finite:

$$\sum_{k=1}^T \left( Y_{(k-1)h} h \right)^2 = \sum_{k=1}^T \left( h^{1/2} Y_{(k-1)h} h^{1/2} \right)^2 = h \sum_{k=1}^T \left( \tilde{Y}_{(k-1)h} \right)^2 = O_p(1),$$

by a standard approximation for Riemann sums and recalling that $\tilde{Y}_{(k-1)h}$ is scaled to be $O_p(1)$. Then it is easy to see that $\{\tilde{Y}_{(k-1)h}\}_{k=1}^T$ has nice properties. It is left-continuous, adapted, and of finite variation in any finite time interval. When used as the integrand of a stochastic integral, the integral itself makes sense. Importantly, its quadratic variation is null and the process is orthogonal to any continuous local martingale. These properties will be used in the sequel. In analogy to the previous section we use a localization procedure and thus we have a corresponding assumption to Assumption B.1.

Assumption B.2. Assumption 2.8 holds, the process $\{\tilde{Y}_t, D_t, Z_t\}_{t \geq 0}$ takes value in
some compact set and the processes \( \{ \mu_{-t}, \sigma_{-t} \}_{t \geq 0} \) (except \( \{ \mu_{-t}^h \}_{t \geq 0} \)) are bounded.

Recall the notation \( M = I - X (X'X)^{-1} X' \), where now

\[
X = \begin{bmatrix}
    h^{1/2} & Y_0 h & \Delta_h D'_1 & \Delta_h Z'_1 \\
    h^{1/2} & Y_1 h & \Delta_h D'_2 & \Delta_h Z'_2 \\
    \vdots & \vdots & \vdots & \vdots \\
    h^{1/2} & Y_T h & \Delta_h D'_T & \Delta_h Z'_T \\
\end{bmatrix}_{T \times (q+p+2)}.
\]  

(B.1.59)

Thus, \( X'X \) is a \((q + p + 2) \times (q + p + 2)\) matrix given by \([XX_1, XX_2]\) where

\[
XX_1 \doteq \begin{bmatrix}
    \sum_{k=1}^T h & h^{1/2} \sum_{k=1}^T \left( Y_{(k-1)h} h \right) \\
    h^{1/2} \sum_{k=1}^T \left( Y_{(k-1)h} h \right) & \sum_{k=1}^T \left( Y_{(k-1)h}^2 h^2 \right) \\
    \sum_{k=1}^T h^{1/2} \left( \Delta_h D_k' \right) & \sum_{k=1}^T \left( \Delta_h D_k \right) \left( Y_{(k-1)h} \right) \\
    \sum_{k=1}^T h^{1/2} \left( \Delta_h Z_k' \right) & \sum_{k=1}^T \left( \Delta_h Z_k \right) \left( Y_{(k-1)h} \right) \\
\end{bmatrix},
\]

and

\[
XX_2 \doteq \begin{bmatrix}
    \sum_{k=1}^T h^{1/2} \left( \Delta_h D_k' \right) & \sum_{k=1}^T \left( \Delta_h D_k \right) \left( Y_{(k-1)h} \right) & X'_D X_D \\
    \sum_{k=1}^T \left( \Delta_h D_k \right) \left( Y_{(k-1)h} \right) & \sum_{k=1}^T \left( \Delta_h Z_k \right) \left( Y_{(k-1)h} \right) & X'_D X_Z \\
    X'_Z X_D & X'_Z X_Z & \\
\end{bmatrix},
\]

where \( X'_D X_D \) is a \( q \times q \) matrix whose \((j, r)\)-th component is the approximate covariance between the \( j \)-th and \( r \)-th element of \( D \), with \( X'_D X_Z \) defined similarly. In view of the properties of \( Y_{(k-1)h} \) outlined above and Assumption B.2, \( X'X \) is \( O_p(1) \) as \( h \downarrow 0 \).

The limit matrix is symmetric positive definite where the only zero elements are in the \( 2 \times (q + p) \) upper right sub-block, and by symmetry in the \((q + p) \times 2\) lower left sub-block. Furthermore, we have

\[
X' e = \begin{bmatrix}
    \sum_{k=1}^T h^{1/2} e_{kh} \\
    \sum_{k=1}^T \left( Y_{(k-1)h} \right) e_{kh} \\
    \sum_{k=1}^T \left( \Delta_h D_k e_{kh} \right) \\
    \sum_{k=1}^T \left( \Delta_h Z_k e_{kh} \right) \\
\end{bmatrix}.
\]  

(B.1.60)

The other statistics are omitted in order to save space. Again the proofs are first given
for the case where the drift processes $\mu_{Z,t}$, $\mu_{D,t}$ of the semimartingale regressors $Z$ and $D$ are identically zero. In the last step we extend the results to nonzero $\mu_{Z,t}$, $\mu_{D,t}$. We also reason conditionally on the processes $\mu_{Z,t}$, $\mu_{D,t}$ and on all the volatility processes so that they are treated as if they were deterministic. We begin with a preliminary lemma.

**Lemma B.1.10.** For $1 \leq i \leq 2$, $3 \leq j \leq p + 2$ and $\gamma > 0$, the following estimates are asymptotically negligible: $\sum_{k=[s/h]}^{[t/h]} z_{kh}^{(i)} z_{kh}^{(j)} \overset{u.c.p.}{\Rightarrow} 0$, for all $N > t > s + \gamma > s > 0$.

**Proof.** Without loss of generality consider any $3 \leq j \leq p + 2$ and $N > t > s > 0$. We have $\sum_{k=[s/h]}^{[t/h]} z_{kh}^{(1)} z_{kh}^{(j)} = \sum_{k=[s/h]}^{[t/h]} \sqrt{h} (\Delta_h M_{Z,k}^{(j)})$, with further $E[z_{kh}^{(1)} z_{kh}^{(j)} | \mathcal{F}_{(k-1)h}] = 0$, $|z_{kh}^{(1)} z_{kh}^{(j)}| \leq K$ for some $K$ by Assumption B.2. Thus $\left\{z_{kh}^{(i)} z_{kh}^{(j)}, \mathcal{F}_{kh}\right\}$ is a martingale difference array. Then, for any $\eta > 0$,

$$P \left( \sum_{k=[s/h]}^{[t/h]} |z_{kh}^{(1)} z_{kh}^{(j)}|^2 > \eta \right) \leq \frac{K}{\eta} \mathbb{E} \left( \sum_{k=[s/h]}^{[t/h]} h^2 (\Delta_h M_{Z,k}^{(j)})^2 \right) \leq \frac{K}{\eta} h O_p (t - s) \to 0,$$

where the second inequality follows from the Burkhölder-Davis-Gundy inequality with parameter $r = 2$. This shows that the array $\left\{|z_{kh}^{(i)} z_{kh}^{(j)}|^2\right\}$ is asymptotically negligible. By Lemma 2.2.11 in the Appendix of Jacod and Protter (2012), we verify the claim for $i = 1$. For the case $i = 2$ note that $z_{kh}^{(2)} z_{kh}^{(j)} = (Y_{(k-1)h}) (\Delta_h M_{Z,k}^{(j)})$, and recall that $\tilde{Y}_{(k-1)h} = h^{1/2} Y_{(k-1)h} = O_p (1)$. Thus, the same proof remains valid for the case $i = 2$. The assertion of the lemma follows. □
B.1.5.1 Proof of Proposition 2.9.1

Proof of part (i) of Proposition 2.9.1. Following the same steps that led to (B.1.12), we can write

\[ Q_T (T_b) - Q_T (T_0) = - |T_b - T_0^0| d(T_b) + g_e(T_b), \quad \text{for all } T_b, \quad (B.1.61) \]

where

\[ d(T_b) \triangleq \frac{(\delta^0)' \left\{ (Z_0'MZ_0) - (Z_0'MZ_2) (Z_2'MZ_2)^{-1} (Z_2'MZ_0) \right\} \delta^0}{|T_b - T_0^0|}, \quad (B.1.62) \]

and we arbitrarily define \( d(T_b) = (\delta^0)' \delta^0 \) when \( T_b = T_0^0 \). Let

\[ d_T = T \inf_{|T_b - T_0^0| > TK} d(T_b); \]

it is positive and bounded away from zero by Lemma B.1.11 below. Then

\[ P \left( |\tilde{\lambda}_b - \lambda_0| > K \right) = P \left( |\tilde{T}_b - T_0^0| > TK \right) \]

\[ \leq P \left( \sup_{|T_b - T_0^0| > TK} |g_e(T_b)| \geq \inf_{|T_b - T_0^0| > TK} |T_b - T_0^0| d(T_b) \right) \]

\[ \leq P \left( \sup_{p+2 \leq T_b \leq T-2} |g_e(T_b)| \geq TK \inf_{|T_b - T_0^0| > TK} d(T_b) \right) \]

\[ = P \left( d_T^{-1} \sup_{p+2 \leq T_b \leq T-2} |g_e(T_b)| \geq K \right). \quad (B.1.63) \]

We can write the first term of \( g_e(T_b) \) as

\[ 2 (\delta^0)' (Z_0'MZ_2) (Z_2'MZ_2)^{-1/2} (Z_2'MZ_2)^{-1/2} Z_2 M e. \quad (B.1.64) \]
For the stochastic regressors, Theorem B.1.5 implies that for any $3 \leq j \leq p + 2$, 
$(Z_2e)_{j,1}/\sqrt{h} = O_p(1)$ and for any $3 \leq i \leq q + p + 2$, 
$(Xe)_{i,1}/\sqrt{h} = O_p(1)$, since these estimates include a positive fraction of the data. We can use the above expression for 
$X'X$ to verify that 
$Z'_2MZ_2$ and $Z'_0MZ_2$ are $O_p(1)$.

Then, 
\[
\sup_{T_b} \left( Z'_0MZ_2 \right) \left( Z'_2MZ_2 \right)^{-1} \left( Z'_2MZ_0 \right) \leq Z'_0MZ_0 = O_p(1),
\]

by Lemma B.1.3. Next, note that the first two elements of the vector $X'e$ and $Z'_2e$ 
are $O_p\left(h^{1/2}\right)$ [recall (B.1.60)]. By Assumption 3.1-(iii) and the inequality 
\[
\sup_{T_b} \left\| (Z'_2MZ_2)^{-1/2} Z_2Me \right\| \leq \sup_{T_b} \left( Z'_2MZ_2 \right)^{-1/2} \sup_{T_b} \left\| Z_2Me \right\|,
\]

we have that $(Z'_2MZ_2)^{-1/2} Z_2Me$ is $O_p\left(h^{1/2}\right)$ uniformly in $T_b$ since the last $q + p$ 
(resp., $p$) elements of $X'e$ (resp., $Z'_2e$) are $o_p(1)$ locally uniformly in time. Therefore, uniformly over $p + 2 \leq T_b \leq T - p - 2$, the overall expression in (B.1.64) is 
$O_p\left(h^{1/2}\right)$. As for the second term of (B.1.10), $Z'_0Me = O_p\left(h^{1/2}\right)$. The first term in 
(B.1.11) is uniformly negligible and so is the last. Therefore, combining these results we can show that 
$\sup_{T_b} |g_e(T_b)| = O_p\left(\sqrt{h}\right)$. Using Lemma B.1.11 below, we have 
$P \left( d_T^{-1} \sup_{p+2 \leq T_b \leq T - p - 2} |g_e(T_b)| \geq K \right) \leq \varepsilon$, which shows that $\hat{\lambda}_b \xrightarrow{P} \lambda_0$. □

**Lemma B.1.11.** Let $d_B = \inf_{|T_b - T_b^0| > TB} Td(T_b)$. There exists a $\kappa > 0$ and for every 
$\varepsilon > 0$, there exists a $B < \infty$ such that $P \left( d_B \geq \kappa \right) \leq 1 - \varepsilon$.

**Proof.** Assuming $N_b \leq N^0_b$ and following the same steps as in Lemma B.1.6 (but
replacing $R$ by $\bar{R}$)

$$Td(T_b) \geq T \left(\delta^0\right)' \bar{R} \frac{X'_\Delta X_\Delta}{T_b - T_b} (X'_2 X_2)^{-1} (X'_0 X_0) \bar{R} \left(\delta^0\right)$$

$$= \left(\delta^0\right)' \bar{R} \frac{X'_\Delta X_\Delta}{B} (X'_2 X_2)^{-1} (X'_0 X_0) \bar{R} \delta^0 \geq \kappa > 0.$$

Under Assumption 3.1-(iii) and in view of (B.1.59), it can be seen that $X'_\Delta X_\Delta$ is positive definite: for the $p \times p$ lower-right sub-block apply Lemma B.1.3 as in the proof of Lemma B.1.6, whereas for the remaining elements of $X'_\Delta X_\Delta$ the result follows from the convergence of approximations to Riemann sums. Note that $X'_2 X_2$ and $X'_0 X_0$ are $O_p(1)$. It follows that

$$Td(T_b) \geq \left(\delta^0\right)' \bar{R} \frac{X'_\Delta X_\Delta}{N} (X'_2 X_2)^{-1} (X'_0 X_0) \bar{R} \delta^0 \geq \kappa > 0.$$

The result follows choosing $B > 0$ such that $P(d_B \geq \kappa)$ is larger than $1 - \varepsilon$. □

*Proof of part (ii) of Proposition 2.9.1.* We introduce again

$$D_{K,T} = \{T_b : N \eta \leq N_b \leq N (1 - \eta), \ |N'_0 - N_0| > KT^{-1}\},$$

and observe that it is enough to show that $P \left(\sup_{T_b \in D_{K,T}} Q_T(T_b) \geq Q_T(T'^0_b)\right) < \varepsilon$, which is equivalently to

$$P \left( \sup_{T_b \in D_{K,T}} h^{-1} g_e(T_b) \geq \inf_{T_b \in D_{K,T}} h^{-1} |T_b - T'^0_b| d(T_b) \right) < \varepsilon. \quad (B.1.65)$$

By Lemma B.1.1,

$$\inf_{T_b \in D_{K,T}} d(T_b) \geq \inf_{T_b \in D_{K,T}} \left(\delta^0\right)' \bar{R} \frac{X'_\Delta X_\Delta}{T'^0_b - T_b} (X'_2 X_2)^{-1} (X'_0 X_0) \bar{R} \delta^0.$$
For the $(q + p) \times (q + p)$ lower right sub-block of $X'_\Delta X_\Delta$ the arguments of Proposition 2.3.2 apply: $(h (T^0_b - T_b))^{-1} [X'_\Delta X_\Delta]_{(q+p)\times(q+p)}$ is bounded away from zero for all $T_b \in D_{K,T}$ by choosing $K$ large enough (recall $|T^0_b - T_b| > K$), where $[A]_{i,i\times j}$ is the $i \times j$ lower right sub-block of $A$. Furthermore, this approximation is uniform in $T_b$ by Assumption 2.4. It remains to deal with the upper left sub-block of $X'_\Delta X_\Delta$. Consider its $(1, 1)$-th element. It is given by 
\[ \sum_{k=T_b}^{T_b+1} \left( h \left( T^0_b - T_b \right) \right)^2. \]
Thus
\[ (1/h (T^0_b - T_b)) \sum_{k=T_b+1}^{T_b} \left( h^{1/2} \right)^2 > 0. \]
The same argument applies to the $(2, 2)$-th element of the upper left sub-block of $X'_\Delta X_\Delta$. The latter results imply that $\inf_{T_b \in D_{K,T}} Td (T_b)$ is bounded away from zero. It remains to show that $\sup_{T_b \in D_{K,T}} (h |T_b - T^0_b|)^{-1} g_e (T_b)$ is small when $T$ is large. Recall that the terms $Z_2$ and $Z_0$ involve a positive fraction $N\eta$ of the data. We can apply Lemma B.1.3 to those elements which involve the stochastic regressors only, whereas the other terms are treated directly using the definition of $X'\epsilon$ in (B.1.60). Consider the first term of $g_e (T_b)$. Using the same steps which led to (B.1.19), we have
\[ \begin{align*}
2 \left( \left( \delta^0 \right)' \left( Z'_0 M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} Z_2 M \epsilon - 2 \left( \delta^0 \right)' \left( Z'_0 M \epsilon \right) \right) \\
= \left| \left( \delta^0 \right)' Z'_\Delta M \epsilon \right| + \left| \left( \delta^0 \right)' \left( Z'_\Delta M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} (Z_2 M \epsilon) \right|. \quad \text{(B.1.66)}
\end{align*} \]

We can apply Lemma B.1.3 to the terms that do not involve $|N_b - N^0_b|$ but only stochastic regressors. Next consider the first term of
\[ \begin{align*}
(h (T^0_b - T_b))^{-1} \left( \delta^0 \right)' \left( Z'_\Delta M Z_2 \right) \\
= \frac{\left( \delta^0 \right)' \left( Z'_\Delta Z_\Delta \right)}{h (T^0_b - T_b)} - \left( \delta^0 \right)' \left( \frac{Z'_\Delta X_\Delta}{h (T^0_b - T_b)} (X'X)^{-1} X'Z_2 \right). \end{align*} \]
Applying the same manipulations as those used above for the $p \times p$ lower right sub-block of $Z'_\Delta Z_\Delta$, we have $(h (T^0_b - T_b))^{-1} [Z'_\Delta Z_\Delta]_{p \times p} = O_p (1)$, since there are $T^0_b - T_b$
summands whose conditional first moments are each $O(h)$. The $O_p(1)$ result is uniform by Assumption 2.4. The same argument holds for the corresponding sub-block of $Z'_\Delta X_\Delta/(h(T_0^b - T_b))$. Hence, as $h \downarrow 0$ the second term above is $O_p(1)$. Next, consider the upper left $2 \times 2$ block of $Z'_\Delta Z_\Delta$ (the same argument holds true for $Z'_\Delta X_\Delta$). Note that the predictable variable $Y_{(k-1)h}$ in the $(2,2)$-th element can be treated as locally constant after multiplying by $h^{1/2}$ (recall $h^{1/2}Y_{(k-1)h} = \tilde{Y}_{(k-1)h} = O_p(1)$ by Assumption B.2),

$$\sum_{k=T_b+1}^{T_0^b} \left(\tilde{Y}_{(k-1)h}h\right)^2 \leq C \sum_{k=T_b+1}^{T_0^b} h,$$

where $C = \sup_k |\tilde{Y}_{(k-1)h}|$ is a fixed constant given the localization in Assumption B.2. Thus, when multiplied by $(h(T_0^b - T_b))^{-1}$, the $(2,2)$-th element of $Z'_\Delta Z_\Delta$ is $O_p(1)$. The same reasoning can be applied to the corresponding $(1,1)$-th element. Next, let us consider the cross-products between the semimartingale regressors and the predictable regressors. Consider any $3 \leq j \leq p + 2$,

$$\frac{1}{h(T_0^b - T_b)} \sum_{k=T_b+1}^{T_0^b} z^{(2)j}_{kh} \tilde{z}^{(j)}_{kh} = \frac{1}{h(T_0^b - T_b)} \sum_{k=T_b+1}^{T_0^b} \left(\tilde{Y}_{(k-1)h}h^{1/2}\right) z^{(j)}_{kh} = \frac{1}{T_0^b - T_b} \sum_{k=T_b+1}^{T_0^b} \tilde{Y}_{(k-1)h} \frac{z^{(j)}_{kh}}{\sqrt{h}}.$$

Since $z_{kh}/\sqrt{h}$ is i.n.d. with zero mean and finite variance and $\tilde{Y}_{(k-1)h}$ is $O_p(1)$ by Assumption B.2, Assumption 2.4 implies that we can find a $K$ large enough such that the right hand side is $O_p(1)$ uniformly in $T_b$. The same argument applies to $(Z'_\Delta Z_\Delta)_{1,j}, 3 \leq j \leq p + 2$. This shows that the term $Z'_\Delta X_\Delta/(h(T_0^b - T_b))$ is bounded
and so is $Z'_\Delta X_\Delta / \left( h \left( T_b^0 - T_b \right) \right)$ using the same reasoning. Thus,

$$
\left( h \left( T_b^0 - T_b \right) \right)^{-1} \left( \delta^0 \right)' \left( Z'_\Delta M Z_2 \right).
$$

is $O_p(1)$. By the same arguments as before, we can use Theorem B.1.5 to show that the second term of (B.1.66) is $O_p\left( h^{1/2} \right)$ when multiplied by $\left( h \left( T_b^0 - T_b \right) \right)^{-1}$ since the last term involves a positive fraction of the data. Now, expand the $(p + 2)$-dimensional vector $Z'_\Delta M e$ as

$$
\frac{Z'_\Delta M e}{h \left( T_b^0 - T_b \right)} = \frac{1}{h \left( T_b^0 - T_b \right)} \sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh}
$$

$$
- \frac{1}{h \left( T_b^0 - T_b \right)} \left( \sum_{k=T_b+1}^{T_b^0} z_{kh} x_{kh}' \right) \left( X' X \right)^{-1} \left( X' e \right).
$$

The arguments for the last $p$ elements are the same as above and yield [recall (B.1.20)]

$$
\frac{\left[ Z'_\Delta M e \right]_{(p,p)}}{h \left( T_b^0 - T_b \right)} = o_p \left( K^{-1} \right) - O_p \left( 1 \right) O_p \left( h^{1/2} \right),
$$

where we recall that by Assumption 3.1-(iv) $\Sigma_{e,e,N_b^0} = 0$. Note that the convergence is uniform over $T_b$ by Lemma B.1.2. We now consider the first two elements of $Z'_\Delta e$:

$$
\left| \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(2)} e_{kh} \right| = \left| \sum_{k=T_b+1}^{T_b^0} h^{1/2} Y_{(k-1)h} h^{1/2} e_{kh} \right| \leq A \sum_{k=T_b+1}^{T_b^0} \left| \tilde{Y}_{(k-1)h} h^{1/2} e_{kh} \right|,
$$

for some positive $A < \infty$. Noting that $e_{kh}/\sqrt{h} \sim \text{i.n.d.} N \left( 0, \sigma_{e,k-1}^2 \right)$, we have

$$
\left( h \left( T_b^0 - T_b \right) \right)^{-1} \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(2)} e_{kh} \leq C \left( h \left( T_b^0 - T_b \right) \right)^{-1} \sum_{k=T_b+1}^{T_b^0} \left| e_{kh}/h^{1/2} \right|
$$

where $C = \sup_k \left| \tilde{Y}_{(k-1)h} \right|$ is finite by Assumption B.2. Choose $K$ large enough such
that the probability that the right-hand side is larger than $B/3N$ is less than $\varepsilon$. For the first element of $Z'_{\Delta e}$ the argument is the same and thus

$$P \left( \left( h \left( T^0_b - T_b \right) \right)^{-1} \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(1)} e_{kh} > \frac{B}{3N} \right) \leq \varepsilon,$$

when $K$ is large. For the last product in the second term of $Z'_{\Delta Me/h}$ the argument is easier. This includes a positive fraction of data and thus

$$\sum_{k=1}^{T} x_{kh}^{(1)} e_{kh} = \sum_{k=1}^{T} h^{1/2} e_{kh} = h^{1/2} o_p \left( 1 \right), \quad (B.1.67)$$

using the basic result $\sum_{k=1}^{[t/h]} e_{kh} \Rightarrow \int_0^t \sigma_{e,s} dW_{e,s}$. A similar argument applies to $x_{kh}^{(2)} e_{kh}$ by using in addition the localization Assumption B.2. Combining the above derivations, we have

$$\frac{1}{h \left( T^0_b - T_b \right)} g_e \left( T_b \right) = \frac{1}{h \left( T^0_b - T_b \right)} \left( \delta^0 \right)' 2Z'_{\Delta e} + o_p \left( 1 \right). \quad (B.1.68)$$

In order to prove

$$P \left( \sup_{T_b \in D_{K,T}} \left( h \left( T^0_b - T_b \right) \right)^{-1} g_e \left( T_b \right) \geq \inf_{T_b \in D_{K,T}} h^{-1} d \left( T_b \right) \right) < \varepsilon,$$

we can use (B.1.68). To this end, we shall find a $K > 0$, such that

$$P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left| \mu^0 b \frac{2}{h} \left( T^0_b - T_b \right)^{-1} \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(1)} e_{kh} \right| > \frac{B}{3N} \right) \leq P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left( T^0_b - T_b \right)^{-1} \left| \sum_{k=T_b+1}^{T_b^0} e_{kh} \right| > \frac{B}{6 \left| \mu^0 b \right| N} \right) < \frac{\varepsilon}{3}. \quad (B.1.69)$$
Recalling that \( e_{kh}/h^{1/2} \sim \mathcal{N} \left( 0, \sigma^2_{e,k-1} \right) \), the Hájek-Rényi inequality yields

\[
P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left( T^0_b - T_b \right)^{-1} \left| \sum_{k=T_b+1}^{T^0_b} e_{kh} \right| > \frac{B}{6 \mu^0_b} \right) \leq A \frac{36 (\mu^0_b)^2 N^2}{B^2} \frac{1}{KN-1}.
\]

We can choose \( K \) sufficiently large such that the right-hand side is less than \( \varepsilon/3 \). The same bound holds for the second element of \( Z'_e \). Next, by equation (B.1.22),

\[
P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left( T^0_b - T_b \right)^{-1} \left\| 2 \left( \delta^0_Z \right)' \sum_{k=T_b+1}^{T^0_b} [Z'_e]_{(\cdot,p)} \right\| > \frac{B}{3N} \right) < \frac{\varepsilon}{3},
\]

since for each \( j = 3, \ldots, p \), \( \{z_{kh} e_{kh}/h\} \) is i.n.d. with finite variance, and thus the result is implied by the Hájek-Rényi inequality for large \( K \). Using the latter results into (B.1.68), we have

\[
P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left( T^0_b - T_b \right)^{-1} \left\| 2 \left( \delta^0_Z \right)' \sum_{k=T_b+1}^{T^0_b} z_{kh} e_{kh} \right\| > \frac{B}{N} \right) < \varepsilon,
\]

which verifies (B.1.65) and thus proves our claim. \( \square \)

### B.1.5.2 Proof of Theorem 2.9.1

Part (i)-(ii) follows the same steps as in the proof of Proposition 2.4.1 part (i)-(ii) but using the results developed throughout the proof of part (i)-(ii) of Proposition 2.9.1. As for part (iii), we begin with the following lemma, where again \( \psi_h = h^{1-\kappa} \). Without loss of generality we set \( B = 1 \) in Assumption 2.6.

**Lemma B.1.12.** Under Assumption B.2, uniformly in \( T_b \),

\[
\left( Q_T (T_b) - Q_T (T^0_b) \right) / \psi_h = -\delta_h (Z'_\Delta Z_\Delta / \psi_h) \delta_h + 2 \delta'_h (Z'_\Delta \tilde{e} / \psi_h) + O_p \left( h^{3/4N1-\kappa/2} \right).
\]
Proof. By the definition of \( Q_T(T_b) - Q_T(T^0_b) \) and Lemma B.1.9,

\[
Q_T(T_b) - Q_T(T^0_b) = -\delta_h \left\{ Z'_\Delta M Z_\Delta + \left( Z'_\Delta M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} \left( Z'_2 M Z_\Delta \right) \right\} \delta_h 
\]

\[\text{(B.1.70)}\]

\[+ g_c(T_b, \delta_h).\]

We can expand the first term of (B.1.70) as

\[
\delta'_h Z'_\Delta M Z_\Delta \delta_h = \delta'_h Z'_\Delta Z_\Delta \delta_h - \delta'_h A \delta_h, \quad \text{(B.1.71)}
\]

where \( A = Z'_\Delta X (X'X)^{-1} X' Z_\Delta \). We show that \( \delta'_h A \delta_h \) is uniformly of higher order than \( \delta'_h Z'_\Delta Z_\Delta \delta_h \). The cross-products between the semimartingale and the predictable regressors (i.e., the \( p \times 2 \) lower-left sub-block of \( Z'_\Delta X \)) are \( o_p(1) \), as can be easily verified. Lemma B.1.10 provides the formal statement of the result for \( Z'_\Delta Z_\Delta \). Hence, the result carries over to \( Z'_\Delta X \) with no changes. By symmetry so is the \( 2 \times p \) upper-right block. This allows us to treat the \( 2 \times 2 \) upper-left block and the \( p \times p \) lower-right block of statistics such as \( A \) separately. By Lemma B.1.3, \((X'X)^{-1} = O_p(1)\). Using Proposition 2.4.1-(ii), we let \( N_b - N^0_b = K \psi_h \). By the Burkholder-Davis-Gundy inequality, we have standard estimates for local volatility so that

\[
\left\| \mathbb{E} \left( \hat{\Sigma}^{(i,j)}_{Z_X}(T_b, T^0_b) - \Sigma^{(i,j)}_{Z_X,T^0_b}(T^0_b-1)h | \mathcal{F}(T^0_b-1)h \right) \right\| \leq K h^{1/2},
\]

with \( 3 \leq i \leq p + 2 \) and \( 3 \leq j \leq q + p + 2 \) which in turn implies \( [Z'_\Delta X_\Delta]_{p \times p} = O_p(1/ (h(T^0_b - T_b))) \). The same bound applies to the corresponding blocks of \( Z'_\Delta Z_\Delta \).
and $X'_\Delta Z_\Delta$. Now let us focus on the $(2, 2)$-th element of $A$. First notice that

$$(Z'_\Delta X)_{2,2} = \sum_{k=T_b+1}^{T^0_b} z^{(2)}_{kh} x^{(2)}_{kh} = \sum_{k=T_b+1}^{T^0_b} \left( \tilde{Y}_{(k-1)h} \right)^2 h.$$ 

By a localization argument (cf. Assumption B.2), $\tilde{Y}_{(k-1)h}$ is bounded. Then, since the number of summands grows at a rate $T^\kappa$, we have $(Z'_\Delta X)_{2,2} = O_p(K h^{-\kappa})$. The same proof can be used for $(Z'_\Delta X)_{1,1}$, which gives $(Z'_\Delta X)_{1,1} = O_p(K h^{-\kappa})$. Thus, in view of (B.1.72), we conclude that (B.1.71) when divided by $\psi_h$ is such that

$$\delta'_h Z'_\Delta M Z_\Delta \delta_h / \psi_h = \delta'_h Z'_\Delta Z_\Delta \delta_h / \psi_h - \delta'_h Z'_\Delta X (X'X)^{-1} X'Z_\Delta \delta_h / \psi_h$$

$$= \psi^{-1}_h \left( \delta^0 \right)' Z'_\Delta Z_\Delta \delta^0 - \psi^{-1}_h h^{1/2} O_p (h^{2(1-\kappa)}). \quad \text{(B.1.72)}$$

For the second term of (B.1.70), we have

$$\psi^{-1}_h h^{1/2} \left( \delta^0 \right)' \left\{ (Z'_\Delta M Z_2) (Z'_2 M Z_2)^{-1} (Z'_2 M Z_\Delta) \right\} \delta^0 \quad \text{(B.1.73)}$$

$$= \psi^{-1}_h h^{1/2} \|\delta_0\|^2 O_p (\psi_h) O_p (1) O_p (\psi_h) \leq K \psi^{-1}_h h^{1/2} O_p (h^{2(1-\kappa)})$$

uniformly in $T_b$, which follows from applying the same reasoning used for $Z'_\Delta (I - M) Z_\Delta$ above to each of these three elements. Finally, consider the stochastic term $g_e(T_b, \delta_h)$. We have

$$g_e(T_b, \delta_h) = 2\delta'_h (Z'_0 M Z_2) (Z'_2 M Z_2)^{-1} Z_2 M e - 2\delta'_h (Z'_0 M e)$$

$$+ e' M Z_2 (Z'_2 M Z_2)^{-1} Z_2 M e - e' M Z_0 (Z'_0 M Z_0)^{-1} Z'_0 M e. \quad \text{(B.1.74)}$$

Recall (B.1.60), and $\sum_{k=T_h+1}^{T^0_b} x_{kh} e_{kh} = h^{-1/4} \sum_{k=T_h+1}^{T^0_b} x_{kh} \tilde{e}_{kh}$. Introduce the following
decomposition,

\[(X'e)_{2,1} = \sum_{k=1}^{T_b^0 - \lfloor T^\kappa \rfloor} x_{kh}^{(2)} e_{kh} + h^{-1/4} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor + 1}^{T_b^0 + \lceil T^\kappa \rceil} x_{kh}^{(2)} e_{kh} + \sum_{k=T_b^0 + \lceil T^\kappa \rceil + 1}^{T} x_{kh}^{(2)} e_{kh},\]

where \( e_{kh} \sim \text{i.n.d. } N\left(0, \sigma^2 e,k - h^{-1/4}\right). \) The first and third terms are \( O_p\left(h^{1/2}\right) \) in view of (B.1.67). The term in the middle is \( h^{3/4} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor + 1}^{T_b^0 + \lceil T^\kappa \rceil} \tilde{Y}_{(k-1)h} h^{-1/2} e_{kh} \), which involves approximately \( 2T^\kappa \) summands. Since \( \tilde{Y}_{(k-1)h} \) is bounded by the localization procedure,

\[h^{3/4} \frac{T^\kappa}{T^{\kappa/2}} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor}^{T_b^0 + \lceil T^\kappa \rceil} \tilde{Y}_{(k-1)h} \frac{e_{kh}}{\sqrt{h}} = h^{3/4} T^\kappa / 2 O_p(1),\]

or \( h^{-1/4} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor}^{T_b^0 + \lceil T^\kappa \rceil} x_{kh}^{(2)} e_{kh} = h^{3/4 - \kappa/2} O_p(1) \). This implies that

\[(X'e)_{2,1} = O_p\left(h^{1/2} h^{3/4 - \kappa/2}\right).\]

The same observation holds for \( (X'e)_{1,1} \). Therefore, one follows the same steps as in the concluding part of the proof of Lemma 2.4.1 [cf. equation (B.1.55) and the derivations thereafter]. That is, for the first two terms of \( g_e(T_b, \delta h) \), using \( Z_0'MZ_2 = Z_2'MZ_2 \pm Z_2'MZ_2 \), we have

\[2h^{1/4} \left(\delta^0\right)' (Z_0'MZ_2) (Z_2'MZ_2)^{-1} Z_2'Me - 2h^{1/4} \left(\delta^0\right)' (Z_0'Me) = 2h^{1/4} \left(\delta^0\right)' Z_2'Me \pm 2h^{1/4} \left(\delta^0\right)' Z_2'MZ_2 (Z_2'MZ_2)^{-1} Z_2'Me.\]  

(B.1.75)

The last term above when multiplied by \( \psi_h^{-1} \) is such that

\[\psi_h^{-1} 2h^{1/4} \left(\delta^0\right)' Z_2'MZ_2 (Z_2'MZ_2)^{-1} Z_2'Me = \|\delta^0\| O_p(1) O_p\left(h^{1.5/4 - \kappa/2}\right),\]
where we have used the fact that \( Z'_\Delta M Z_2 / \psi_h = O_p(1) \). For the first term of (B.1.75),

\[
2h^{1/4} \left( \delta^0 \right)' Z'_\Delta M e / \psi_h = 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta e / \psi_h - 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta X (X'X)^{-1} X' e / \psi_h
\]

\[
= 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta e - 2 \left( \delta^0 \right)' O_p(1) O_p \left(h^{1.5/4 - \kappa/2} \right).
\]

As in the proof of Lemma 2.4.1, we can now use part (i) of the theorem so that the difference between the terms on the second line of \( g_e (T_b, \delta_h) \) is negligible. That is, we can find a \( c_T \) sufficiently small such that,

\[
\psi_h^{-1} \left[ e' M Z_2 (Z'_2 M Z_2)^{-1} Z_2 Me - e' M Z_0 (Z'_0 M Z_0)^{-1} Z'_0 Me \right] = o_p (c_T h).
\]

This leads to

\[
g_e (T_b, \delta_h) / \psi_h = 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta e / \psi_h + O_p \left(h^{3/4 - \kappa/2} \right)
\]

\[
+ \left\| \delta^0 \right\| O_p \left(h^{3/4 - \kappa/2} \right) + o_p (c_T h),
\]

for sufficiently small \( c_T \). This together with (B.1.72) and (B.1.73) yields,

\[
\psi_h^{-1} \left( Q_T (T_b) - Q_T (T_b^0) \right) = -\delta_h (Z'_\Delta Z_\Delta / \psi_h) \delta_h \pm 2\delta'_h (Z'_e / \psi_h)
\]

\[
+ O_p \left(h^{3/4 - \kappa/2} \right) + o_p \left(h^{1/2} \right),
\]

when \( T \) is large, where \( c_T \) is a sufficiently small number. This concludes the proof. □

**Proof of part (iii) of Theorem 2.9.1.** We proceed as in the proof of Theorem 2.4.1 and, hence, some details are omitted. We again change the time scale \( s \mapsto t \equiv \psi_h^{-1} s \) on \( D(C) \) and observe that the re-parameterization \( \theta_h, \sigma_{h,t} \) does not alter the result.
of Lemma B.1.12. In addition, we have now,

\[ dZ_{\psi,s}^{(1)} = \psi_h^{-1/2} (ds)^{1/2} = (ds)^{1/2}, \]

and

\[ dZ_{\psi,s}^{(2)} = \psi_h^{-1/2} Y_s ds = \psi_h^{-1/2} \tilde{Y}_s ds = (ds)^{1/2}, \]

where the first equality in the second term above follows from \( \tilde{Y}_{(k-1)h} = h^{1/2}Y_{(k-1)h} \) on the old time scale. \( N^0_b(v) \) varies on the time horizon \([N^0_b - |v|, N^0_b + |v|]\) as implied by \( D^*(C) \), as defined in Section 2.4. Again, in order to avoid clutter, we suppress the subscript \( \psi_h \). We then have equation (2.9.10)-(2.9.11). Consider \( T_b \leq T^0_b \) (i.e., \( v \leq 0 \)). By Lemma B.1.12, there exists a \( T \) such that for all \( T > T \), \( h^{-1/2} (Q_T (T_b) - Q_T (T^0_b)) \) is

\[
\bar{Q}_T (\theta^*) = -h^{-1/2} \delta_h' Z'_\Delta \Delta \delta_h + h^{-1/2} 2 \delta_h' Z'_\Delta e + o_p (1)
\]

\[
= - (\delta^0)' \left( \sum_{k=T_b+1}^{T^0_b} z_{kh} z_{kh}' \right) \delta^0 + 2 (\delta^0)' \left( h^{-1/2} \sum_{k=T_b+1}^{T^0_b} z_{kh} \tilde{e}_{kh} \right) + o_p (1),
\]

and note that this relationship corresponds to (2.9.12). As in the proof of Theorem 2.4.1 it is convenient to associate to the continuous time index \( t \) in \( D^* \), a corresponding \( D^* \)-specific index \( t_v \). We then define the following functions which belong to \( D (D^*, \mathbb{R}) \),

\[ J_{Z,h} (v) \triangleq \sum_{k=T_b(v)+1}^{T^0_b} z_{kh} z_{kh}', \quad J_{e,h} (v) \triangleq \sum_{k=T_b(v)+1}^{T^0_b} z_{kh} \tilde{e}_{kh}, \]

for \( (T_b (v) + 1) h \leq t_v < (T_b (v) + 2) h \). Recall that the lower limit of the summation is \( T_b (v) + 1 = T^0_b + \lfloor v/h \rfloor (v \leq 0) \) and thus the number of observations in each sum increases at rate \( 1/h \). We first note that the partial sums of cross-products
between the predictable and stochastic semimartingale regressors is null because the
drift processes are of higher order (recall Lemma B.1.10). Given the previous lemma
we can decompose \( \overline{Q}_T(\theta, v) \) as follows,

\[
\overline{Q}_T(\theta, v) = \left( \delta_0^0 \right)' R_{1,h}(v) \delta_0^0 + \left( \delta_Z^0 \right)' R_{2,h}(v) \delta_Z^0 + 2 \left( \delta^0 \right)' \left( \frac{1}{\sqrt{h}} \sum_{k=T_b+1}^{T_b} z_{kh} \tilde{e}_{kh} \right),
\]

(B.1.76)

where

\[
R_{1,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b} \begin{bmatrix} h & Y_{(k-1)h}h^{3/2} \\ Y_{(k-1)h}h^{3/2} & (Y_{(k-1)h})^2 \end{bmatrix}, \quad R_{2,h}(v) \triangleq [Z'_\Delta Z_\Delta]_{\{i,p\times p\}},
\]

and \( \delta^0 \) has been partitioned accordingly; that is, \( \delta_p^0 = (\mu_0^0, \alpha_0^0)' \) is the vector of
parameters associated with the predictable regressors whereas \( \delta_Z^0 \) is the vector of
parameters associated with the stochastic martingale regressors in \( Z \). By ordinary
results for convergence of Riemann sums,

\[
\left( \delta_p^0 \right)' R_{1,h}(v) \delta_p^0 \overset{u.c.p.}{\Rightarrow} \left( \delta_p^0 \right)' \begin{bmatrix} N_b^0 - N_b & \int_{N_b^0+v}^{N_b} \tilde{Y}_s ds \\ \int_{N_b^0+v}^{N_b} \tilde{Y}_s ds & \int_{N_b^0+v}^{N_b} \tilde{Y}_s^2 ds \end{bmatrix} \delta_p^0.
\]

(B.1.77)

Next, since \( Z^{(j)}_t \) (\( j = 3, \ldots, p + 2 \)) is a continuous Itô semimartingale, we have by
Theorem 3.3.1 in Jacod and Protter (2012),

\[
R_{2,h}(v) \overset{u.c.p.}{\Rightarrow} \langle Z_\Delta, Z_\Delta \rangle(v).
\]

(B.1.78)

We now turn to examine the asymptotic behavior of the second term in (B.1.76) on
\( \mathcal{D}^* \). We follow the following steps. First, we present a stable central limit theorem
for each component of \( Z'_\Delta e \). Second, we show the joint convergence stably in law.
to a continuous Gaussian process and finally we verify tightness of the sequence of processes which in turn yields the stable convergence under the uniform metric. We begin with the second element of $Z'_{\Delta e}$,

$$
\frac{1}{\sqrt{h}} \sum_{k = T_b(v) + 1}^{T_b^0} \alpha^0 \delta_{z_{kh}} \tilde{e}_{kh} = \frac{1}{\sqrt{h}} \sum_{k = T_b(v) + 1}^{T_b^0} \alpha^0 \delta \left( Y_{(k-1)h} \right) \tilde{e}_{kh},
$$

and using $\bar{Y}_{(k-1)h} = h^{1/2} Y_{(k-1)h}$ [recall that $\bar{Y}_{(k-1)h}$ is bounded by the localization Assumption B.2] we then have

$$
\frac{1}{\sqrt{h}} \sum_{k = T_b(v) + 1}^{T_b^0} \alpha^0 \delta \left( Y_{(k-1)h} \right) \tilde{e}_{kh} \Rightarrow \int_{N_0^0 + v}^{N_0^0} \alpha^0 \bar{Y}_s dW_{e,s},
$$

which follows from the convergence of Riemann approximations for stochastic integrals [cf. Proposition 2.2.8 in Jacod and Protter (2012)]. For the first component, the argument is similar:

$$
\frac{1}{\sqrt{h}} \sum_{k = T_b(v) + 1}^{T_b^0} \mu^0 \delta_{z_{kh}} \tilde{e}_{kh} \Rightarrow \int_{N_0^0 + v}^{N_0^0} \mu^0 \bar{Y}_s dW_{e,s}.
$$

(B.1.79)

Next, we consider the $p$-dimensional lower subvector of $Z'_{\Delta e}$, which can be written as

$$
2 \left( \delta^0_Z \right)' \left( \frac{1}{\sqrt{h}} \sum_{k = T_b(v) + 1}^{T_b^0} \bar{z}_{kh} \tilde{e}_{kh} \right),
$$

(B.1.80)

where we have partitioned $z_{kh}$ as $z_{kh} = \left[ h^{1/2} \ Y_{(k-1)h} \ \bar{z}_{kh} \right]'$. Then, note that the small-dispersion asymptotic re-parametrization implies that $\bar{z}_{kh} \tilde{e}_{kh}$ corresponds to $z_{kh} \tilde{e}_{kh}$ from Theorem 2.4.1. Hence, we shall apply the same arguments as in the proof of Theorem 2.4.1 since (B.1.80) is simply $2 \left( \delta^0_Z \right)'$ times $H_b(v) = h^{-1/2} J_{e,h}(v)$, where $J_{e,h}(v) \triangleq \sum_{k = T_b(v) + 1}^{T_b^0} \bar{z}_{kh} \tilde{e}$ with $(T_b(v) + 1) h < t_v < (T_b(v) + 2) h$. By The-
orem 5.4.2 in Jacod and Protter (2012), \( W_h(v) \xrightarrow{\mathcal{L}} W_{\mathcal{Z}_v}(v) \). Since the convergence of the drift processes \( R_{1,h}(v) \) and \( R_{2,h}(v) \) occur in probability locally uniformly in time while \( W_h(v) \) converges stably in law to a continuous limit process, we have for each \((\theta, \cdot)\) a stable convergence in law under the uniform metric. This is a consequence of the property of stable convergence in law [cf. section VIII.5c in Jacod and Shiryaev (2003)]. Since the case \( v > 0 \) is analogous, this proves the finite-dimensional convergence of the process \( \overline{Q}_T(\theta, \cdot) \), for each \( \theta \). It remains to verify stochastic equicontinuity. As for the terms in \( R_{1,h}(v) \), we can decompose \( (\alpha_3)^2 \sum_{k=1}^{T_h(v)+1} (\xi_{kh}^{(2)})^2 - \left( \sum_{k=1}^{T_h(v)+1} \xi_{kh}^{(1)} \right)^2 \) as \( \overline{Q}_{6,T}(\theta, v) + \overline{Q}_{7,T}(\theta, v) \), where \( \overline{Q}_{6,T}(\theta, v) \Deltaq (\alpha_3)^2 \sum_{k} \xi_{2,h,k}^{*} \) and \( \overline{Q}_{7,T}(\theta, v) \Deltaq (\alpha_3)^2 \sum_{k} \xi_{2,h,k}^{**} \), with

\[
\zeta_{2,h,k}^{*} = \left( \zeta_{kh}^{(2)} \right)^2 - \left( \int_{(k-1)h}^{kh} \tilde{Y}_s^2 ds \right) - 2\tilde{Y}_{(k-1)h} \int_{(k-1)h}^{kh} (\tilde{Y}_{(k-1)h} - \tilde{Y}_s) ds + 2\mathbb{E} \left[ \tilde{Y}_{(k-1)h} \left( \tilde{Y}_{(k-1)h} \cdot h - \int_{(k-1)h}^{kh} \tilde{Y}_s ds \right) \mid \mathcal{F}_{(k-1)h} \right] \Deltaq L_{1,h,k} + L_{3,h,k},
\]

and

\[
\zeta_{2,h,k}^{**} = 2\tilde{Y}_{(k-1)h} \left( \tilde{Y}_{(k-1)h} h - \int_{(k-1)h}^{kh} \tilde{Y}_s ds - \mathbb{E} \left[ \left( \tilde{Y}_{(k-1)h} h - \int_{(k-1)h}^{kh} \tilde{Y}_s ds \right) \mid \mathcal{F}_{(k-1)h} \right] \right).
\]

Then, we have the following decomposition for \( \overline{Q}_T^c(\theta^*) \Deltaq \overline{Q}_T(\theta^*) + (\delta^0)' \Lambda(v) \delta^0 \) (if \( v \leq 0 \) and defined analogously for \( v > 0 \)):

\[
\overline{Q}_T^c(\theta^*) = \sum_{r=1}^{4} \overline{Q}_{r,T}(\theta, v), \quad \overline{Q}_{r,T}(\theta, v) \Deltaq (\mu_3)^2 \sum_{k} \zeta_{1,h,k},
\]

\( r = 1, \ldots, 4 \) are defined in (2.9.13) and \( \overline{Q}_{5,T}(\theta, v) \Deltaq (\mu_3)^2 \sum_{k} \zeta_{1,h,k}, \overline{Q}_{6,T}(\theta, v) \Deltaq (\mu_3)^2 \left( h^{-1/2} \sum_{k} \xi_{1,h,k} \right), \overline{Q}_{7,T}(\theta, v) \Deltaq (\mu_3)^2 \left( h^{-1/2} \sum_{k} \xi_{2,h,k} \right) \) where \( \zeta_{1,h,k} \Deltaq (\zeta_{kh}^{(1)})^2 - h, \xi_{1,h,k} \Deltaq h^{1/2} \tilde{c}_{kh} \) and \( \xi_{2,h,k} \Deltaq (\tilde{Y}_{(k-1)h} h^{1/2}) \tilde{c}_{kh} \). Moreover, recall that \( \sum_{k} \) replaces \( \sum_{T_h(v)+1} \) for \( N_b(v) \in \mathcal{D}^*(C) \). Let us consider \( \overline{Q}_{6,T}(\theta, v) \) first. For \( s \in [(k-1)h, kh] \),

\[
\Lambda(v) \delta^0 \]

\[
\sum_{k} \zeta_{2,h,k}^{**}
\]

\[
\left( h^{-1/2} \sum_{k} \xi_{1,h,k} \right)
\]

\[
\left( h^{-1/2} \sum_{k} \xi_{2,h,k} \right)
\]

\[
\left( \tilde{Y}_{(k-1)h} h^{1/2} \right) \tilde{c}_{kh}
\]
by the Burkholder-Davis-Gundy inequality

\[ |E \left[ \tilde{Y}_{(k-1)h} \left( \tilde{Y}_{(k-1)h} - \tilde{Y}_s \right) \mid \mathcal{F}_{(k-1)h} \right] | \leq Kh, \]

from which we can deduce that, by using a maximal inequality for any \( r > 1 \),

\[ \left[ E \left( \sup_{(\theta, v)} \left| (\alpha \delta)^2 \sum_k L_{2,h,k} \right| \right) \right]^{1/r} \leq K_r \left( \sup_{(\theta, v)} (\alpha \delta)^2 \sum_k h^r \right)^{1/r} = K_r h^{\frac{r-1}{r}}. \]  

(B.1.81)

By a Taylor series expansion for the mapping \( f : y \to y^2 \), and \( s \in [(k-1)h, kh] \),

\[ E \left| \tilde{Y}_{(k-1)h}^2 - \tilde{Y}_s^2 - 2 \tilde{Y}_{(k-1)h} \left( \tilde{Y}_{(k-1)h} - \tilde{Y}_s \right) \right| \leq K E \left[ (\tilde{Y}_{(k-1)h} - \tilde{Y}_s)^2 \right] \leq Kh, \]

where the second inequality follows from the Burkholder-Davis-Gundy inequality. Thus, using a maximal inequality as in (B.1.81), we have for \( r > 1 \)

\[ \left[ E \left( \sup_{(\theta, v)} \left| (\alpha \delta)^2 \sum_k L_{1,h,k} \right| \right) \right]^{1/r} = K_r h^{\frac{r-1}{r}}. \]  

(B.1.82)

(B.1.81) and (B.1.82) imply that \( \overline{Q}_{6,T} (\cdot, \cdot) \) is stochastically equicontinuous. Next, note that \( \overline{Q}_{7,T} (\theta, v) \) is a sum of martingale differences times \( h^{1/2} \) (recall the definition of \( \Delta_h \tilde{V}_k = h^{1/2} \Delta_h V_k (\pi, \delta_{Z,1}, \delta_{Z,2}) \)). Therefore by Assumption B.2, for any \( 0 \leq s < t \leq N, V_t - V_s = O_p (1) \) uniformly and therefore,

\[ \sup_{(\theta, v)} \left| \overline{Q}_{7,T} (\theta, v) \right| \leq K O_p \left( h^{1/2} \right). \]  

(B.1.83)

Given (B.1.77) and (B.1.81)-(B.1.83), we deduce that

\[ \sup_{(\theta, v)} \left\{ \left| \overline{Q}_{6,T} (\theta, v) \right| + \left| \overline{Q}_{7,T} (\theta, v) \right| \right\} = o_p (1). \]
As for the term involving \( R_{1,h} (v) \), it is easy to see that \( \sup_{(\theta, v)} \left| \mathcal{Q}_{5,T} (\theta, v) \right| \to 0 \). Next, we can use some of the results proved in the proof of Theorem 2.4.1. In particular, the asymptotic stochastic equicontinuity of the sequence of processes \( \{ 2 (\delta Z) \}' \mathcal{W}_h (v) \} \) follows from the same property for \( \{ \mathcal{Q}_{3,T} (\theta, v) \} \) and \( \{ \mathcal{Q}_{4,T} (\theta, v) \} \) proved in that proof. The stochastic equicontinuity of

\[
(\delta Z)’ (R_{2,h} (\theta, v) - \langle Z_\Delta, Z_\Delta \rangle (v)) \delta Z
\]

also follows from the same proof. Recall \( \mathcal{Q}_{1,T} (\theta, v) + \mathcal{Q}_{2,T} (\theta, v) \) as defined in (2.9.13). Thus, stochastic equicontinuity follows from (2.9.15) and the equation right before that. Next, let us consider \( \mathcal{Q}_{9,T} (\theta, v) \). We use the alternative definition (ii) of stochastic equicontinuity in Andrews (1994). Consider any sequence \( \{(\theta, v)\} \) and \( \{ (\bar{\theta}, \bar{v}) \} \) (we omit the dependence on \( h \) for simplicity). Assume \( N_b \leq N^0_b \leq \bar{N}_b \) (the other cases can be proven similarly) and let \( N d_h \triangleq \bar{N}_b - N_b \). Then,

\[
| \mathcal{Q}_{9,T} (\theta, v) - \mathcal{Q}_{9,T} (\bar{\theta}, \bar{v}) | = | \alpha_\delta \sum_{k = T_b (v) + 1}^{T_b (\bar{v})} \tilde{Y}_{(k-1)h} e_{kh} - \tilde{\alpha}_\delta \sum_{k = T_b (\bar{v})}^{T_b (\bar{v})} \tilde{Y}_{(k-1)h} e_{kh} | \\
\leq | \alpha_\delta | \sum_{k = T_b (v) + 1}^{T_b (\bar{v})} \tilde{Y}_{(k-1)h} e_{kh} + | \tilde{\alpha}_\delta | \sum_{k = T_b (\bar{v})}^{T_b (\bar{v})} \tilde{Y}_{(k-1)h} e_{kh} .
\]

For the second term, by the Burkholder-Davis-Gundy inequality for any \( r \geq 1 \),

\[
\mathbb{E} \left[ \sup_{0 \leq u \leq d_h} \left| \sum_{k = T_b (v) + 1}^{T_b (\bar{v})} \tilde{Y}_{(k-1)h} e_{kh} \right|^r | \mathcal{F}_{N^0_b} \right] \\
\leq K_r \left( N d_h \right)^{r/2} \mathbb{E} \left[ \frac{1}{N d_h} \left( \sum_{k = T_b (v) + 1}^{T_b (\bar{v})} \int_{(k-1)h}^{kh} \tilde{Y}_s^2 ds \right)^{r/2} | \mathcal{F}_{N^0_b} \right] \leq K_r d_h^{r/2} .
\]
By the law of iterated expectations, and using the property that $d_h \downarrow 0$ in probability, we can find a $T$ large enough such that for any $B > 0$

\[
\left( \mathbb{E} \left[ \sup_{0 \leq u \leq d_h} \left| T_0^u + [Nu/h] \right| \sum_{k=T_0^u}^{\lfloor Nu/h \rfloor} \tilde{Y}_{(k-1)h}, \tilde{e}_{kh} \right|^r \right)^{1/r} \leq K_r d_h^{1/2} P \left( Nd_h > B \right) \to 0.
\]

The argument for the first term in (B.1.84) is analogous. By Markov’s inequality and combining the above steps we have that for any $\varepsilon > 0$ and $\eta > 0$ there exists some $T$ such that for $T > T$,

\[ P \left( \left| \mathcal{Q}_{9,T} \left( \theta, v \right) - \mathcal{Q}_{9,T} \left( \bar{\theta}, \bar{v} \right) \right| > \eta \right) < \varepsilon. \]

Thus, the sequence $\{ \mathcal{Q}_{9,T} (\cdot, \cdot) \}$ is stochastically equicontinuous. Noting that the same proof can be repeated for $\mathcal{Q}_{8,T} (\cdot, \cdot)$, we conclude that the sequence of processes $\{ \mathcal{Q}_{T} (\theta^*) \}, T \geq 1$ in (B.1.76) is stochastically equicontinuous. Furthermore, by (B.1.77) and (B.1.78) we obtain,

\[ \left( \delta_p^0 \right)' R_{1,h} \left( \theta, v \right) \delta_p^0 + \left( \delta_Z^0 \right)' \left( R_{2,h} \left( \left( \theta, v \right) \right) \right) \delta_Z^0 \overset{\text{u.c.p.}}{\Rightarrow} \left( \delta^0 \right)' \Lambda \left( v \right) \delta^0. \]

This suffices to guarantee the $\mathcal{G}$-stable convergence in law of the process $\{ \mathcal{Q}_T (\cdot, \cdot), T \geq 1 \}$ towards a process $\mathcal{W} (\cdot)$ with drift $\Lambda (\cdot)$ which, conditional on $\mathcal{G}$, is a two-sided Gaussian martingale process with covariance matrix given in (2.9.6). By definition, $\mathcal{D}^* (C)$ is compact and $Th \left( \lambda - \lambda_0 \right) = O_p (1)$, which together with the fact that the limit process is a continuous Gaussian process enable one to deduce the main assertion from the continuous mapping theorem for the argmax functional. □
B.1.5.3 Proof of Proposition 2.9.2

We begin with a few lemmas. Let \( \tilde{Y}_t^* \triangleq \tilde{Y}_{[t/h]}^h \). The first result states that the observed process \( \{\tilde{Y}_t^*\} \) converges to the non-stochastic process \( \{\tilde{Y}_t^0\} \) defined in (2.9.4) as \( h \downarrow 0 \). Assumption B.2 is maintained throughout and the constant \( K > 0 \) may vary from line to line.

**Lemma B.1.13.** As \( h \downarrow 0 \), \( \sup_{0 \leq t \leq N} |\tilde{Y}_t^* - \tilde{Y}_t^0| = o_p(1) \).

**Proof.** Let us introduce a parameter \( \gamma_h \) with the property \( \gamma_h \downarrow 0 \) and \( h^{1/2}/\gamma_h \to B \) where \( B < \infty \). By construction, for \( t < N^0_b \),

\[
\tilde{Y}_t - \tilde{Y}_t^0 = \int_0^t \alpha_1^0 (\tilde{Y}_s - \tilde{Y}_s^0) \, ds + B\gamma_h (\pi^0)' D_t + B\gamma_h (\delta_{Z,1}^0)' \int_0^t dZ_s + B\gamma_h \int_0^t \sigma_{e,s} dW_{e,s}.
\]

We can use Cauchy-Schwarz’s inequality,

\[
|\tilde{Y}_t - \tilde{Y}_t^0|^2 \leq 2K \left[ \int_0^t \alpha_1^0 |\tilde{Y}_s - \tilde{Y}_s^0| \, ds \right]^2 \\
+ \left( |\pi^0 D_t|^2 + \left| \int_0^t dZ_s \right|^2 + \left| \int_0^t \sigma_{e,s} dW_{e,s} \right|^2 \right) (B\gamma_h)^2 \\
\leq 2Kt \left[ \alpha_1^0 \int_0^t |\tilde{Y}_s - \tilde{Y}_s^0|^2 \, ds + \sup_{0 \leq s \leq t} |\pi^0 D_s|^2 + \sup_{0 \leq s \leq t} \left| \delta_{Z,1}^0 \int_0^t dZ_s \right|^2 \\
+ \sup_{0 \leq s \leq t} \left| \int_0^s \sigma_{e,u} dW_{e,u} \right|^2 \right] (B\gamma_h)^2.
\]

By Gronwall’s inequality,

\[
|\tilde{Y}_t - \tilde{Y}_t^0|^2 \leq 2 (B\gamma_h)^2 C \exp \left( \int_0^t 2K^2 t ds \right) \\
\leq 2 (B\gamma_h)^2 C \exp \left( 2K^2 t^2 \right),
\]

implying the result.
where $C < \infty$ is a bound on the sum of the supremum terms in the last equation above. The bound follows from Assumption B.2. Then, $\sup_{0 \leq t \leq N} |\tilde{Y}_t - \tilde{Y}_t^0| \leq K\sqrt{2B\gamma h} \exp(K^2N^2) \to 0$, as $h \downarrow 0$ (and so $\gamma h \downarrow 0$). The assertion then follows from $\lfloor t/h \rfloor h \to t$ as $h \downarrow 0$. For $t \geq N_0^0$, one follows the same steps. □

Lemma B.1.14. As $h \downarrow 0$, uniformly in $(\mu_1, \alpha_1)$, $(N/T) \sum_{k=1}^{T_0} (\mu_1 + \alpha_1 \tilde{Y}_{(k-1)h}) \overset{P}{\to} \int_0^{N_0^0} (\mu_1 + \alpha_1 \tilde{Y}_s^0) \, ds$.

Proof. Note that

$$\sup_{\mu_1, \alpha_1} \left| \frac{N}{T} \sum_{k=1}^{T_0} (\mu_1 + \alpha_1 \tilde{Y}_{(k-1)h}) - \int_0^{N_0^0} (\mu_1 + \alpha_1 \tilde{Y}_s^0) \, ds \right|$$

$$= \sup_{\mu_1, \alpha_1} \left| \int_0^{N_0^0} (\mu_1 + \alpha_1 \tilde{Y}_s^0) \, ds - \int_0^{N_0^0} (\mu_1 + \alpha_1 \tilde{Y}_s^0) \, ds \right|$$

$$\leq \sup_{\alpha_1} \int_0^{N_0^0} |\alpha_1| |\tilde{Y}_s^* - \tilde{Y}_s^0| \, ds \leq KOP(\gamma h) \sup_{\alpha_1} \alpha_1,$$

which goes to zero as $h \downarrow 0$ by Lemma B.1.13 (recall $h^{1/2}/\gamma h \to B$) and by Assumption B.2. □

Lemma B.1.15. For each $3 \leq j \leq p + 2$ and each $\theta$, as $h \downarrow 0$,

$$\sum_{k=1}^{\lfloor N_0^0/h \rfloor} (\mu_1 + \alpha_1 \tilde{Y}_{(k-1)h}) \delta^{(j)}_{Z_1,1} \Delta_h Z_k^{(j)} \overset{P}{\to} \int_0^{N_0^0} (\mu_1 + \alpha_1 \tilde{Y}_{(k-1)h}) \, dZ_s^{(j)}.$$

Proof. Note that

$$\sum_{k=1}^{\lfloor N_0^0/h \rfloor} (\mu_1 + \alpha_1 \tilde{Y}_{(k-1)h}) \delta^{(j)}_{Z_1,1} \Delta_h Z_k^{(j)} = \int_0^{N_0^0} (\mu_1 + \alpha_1 \tilde{Y}_s^*) \, dZ_s^{(j)}.$$

By Markov’s inequality and the dominated convergence theorem, for every $\varepsilon > 0$ and
every $\eta > 0$

$$P \left( \left\| \int_0^{N_0 b} \alpha_1 \left( \tilde{Y}_s^* - \tilde{Y}_s^0 \right) \delta_{Z,1}^{(j)} dZ_s^{(j)} \right\| > \eta \right) \leq \frac{\left( \sup_{0 \leq s \leq N} \left( \sum_{r=1}^p \sigma_{Z,s}^{(j,r)} \right)^2 \right)^{1/2}}{\eta} \left| \alpha_1 \right| \left\| \delta_{Z,1}^{(j)} \right\| \left( \int_0^{N_0 b} \mathbb{E} \left[ (\tilde{Y}_s^* - \tilde{Y}_s^0)^2 \right] ds \right)^{1/2},$$

which goes to zero as $h \downarrow 0$ in view of Lemma B.1.13 and Assumption B.2. \qed

**Lemma B.1.16.** As $h \downarrow 0$, uniformly in $\mu_1, \alpha_1$,

$$\sum_{k=1}^{T_0} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \tilde{Y}_{kh} - \tilde{Y}_{(k-1)h} - \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_{(k-1)h} \right) h \right) \overset{P}{\to} 0.$$

**Proof.** By definition [recall the notation in (2.9.3)],

$$\tilde{Y}_{kh} - \tilde{Y}_{(k-1)h} = \int_{(k-1)h}^{kh} \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_s \right) ds + \Delta_h \tilde{V}_k \left( \pi_0, \delta_{Z,1}, \delta_{Z,2} \right).$$

Then,

$$\sum_{k=1}^{T_0} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \tilde{Y}_{kh} - \tilde{Y}_{(k-1)h} - \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_{(k-1)h} \right) h \right) = \sum_{k=1}^{T_0} \int_{(k-1)h}^{kh} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_s - \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_{(k-1)h} \right) \right) ds$$

$$+ \sum_{k=1}^{T_0} \int_{(k-1)h}^{kh} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \Delta_h \tilde{V}_k \left( \pi_0, \delta_{Z,1}, \delta_{Z,2} \right)$$

$$= \int_0^{N_0 b} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \mu_1^0 \left( \tilde{Y}_s - \tilde{Y}_{(k-1)h} \right) \right) ds + B\gamma h \int_0^{N_0 b} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) dV_s.$$
For the first term on the right-hand side,

\[
\sup_{\mu_1, \alpha_1} \left| \int_0^{N_0} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) \left( \alpha_0^0 \left( \tilde{Y}_s - \tilde{Y}_s^* \right) \right) \, ds \right|
\]

\[
\leq |\alpha_1^0| \int_0^{N_0} \sup_{\mu_1, \alpha_1} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) \left( \tilde{Y}_s - \tilde{Y}_0 + \tilde{Y}_s^0 - \tilde{Y}_s^* \right) \, ds
\]

\[
\leq \alpha_0^0 K \left( \int_0^{N_0} \sup_{0 \leq s \leq N_0^0} \left| \tilde{Y}_s - \tilde{Y}_s^0 \right| + \sup_{0 \leq s \leq N_0^0} \left| \tilde{Y}_s^0 - \tilde{Y}_s^* \right| \, ds \right),
\]

which is \(o_p(1)\) as \(h \downarrow 0\) from Lemma B.1.13 and Assumption B.2. Next, consider the vector of regressors \(Z\), and note that for any \(3 \leq j \leq p + 2\),

\[
B_{\gamma h} \sup_{\mu_1, \alpha_1} \int_0^{N_0} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) dZ^{(j)} \leq B_{\gamma h} \sup_{\mu_1, \alpha_1} \int_0^{N_0} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) \sum_{r=1}^{p} \sigma_{Z,s}^{(j,r)} dW^{(r)}.
\]

Let \(R_{j,h} = R_{j,h} (\mu_1, \alpha_1) \triangleq \int_0^{N_0} B_{\gamma h} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) \sum_{r=1}^{p} \sigma_{Z,s}^{(j,r)} dW^{(r)}\) (we index \(R_j\) by \(h\) because \(\tilde{Y}_s^*\) depends on \(h\)). Then, we want to show that, for every \(\varepsilon > 0\) and \(K > 0\),

\[
P \left( \sup_{\mu_1, \alpha_1} |R_{j,h} (\mu_1, \alpha_1)| > K \right) \leq \varepsilon. \quad \text{(B.1.85)}
\]

In view of Chebyshev’s inequality and the Itô’s isometry,

\[
P \left( |R_{j,h}| > K \right) \leq \left( \frac{B_{\gamma h}}{K} \right)^2 \mathbb{E} \left[ \int_0^{N_0} \left( R_{j,h} / (B_{\gamma h}) \right)^2 \right],
\]

\[
\leq \left[ \sup_{0 \leq s \leq N_0} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right] \left( \frac{B_{\gamma h}}{K} \right)^2 \mathbb{E} \left[ \int_0^{N_0} \left| \mu_1 + \alpha_1 \tilde{Y}_s^* \right|^2 \, ds \right],
\]

so that by the boundness of the processes (cf. Assumption B.2) and the compactness
of $\Theta_0$, we have for some $A < \infty$,

$$P(|R_{j,h}| > K) \leq A \left[ \sup_{0 \leq s \leq T} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right] \left( \frac{B \gamma_h}{K} \right)^2 \rightarrow 0, \quad \text{(B.1.86)}$$

since $\gamma_h \downarrow 0$. This demonstrates pointwise convergence. It remains to show the stochastic equicontinuity of the sequence of processes $\{R_{j,h}(\cdot)\}$. Choose $2m > p$ and note that standard estimates for continuous Itô semimartingales result in $E \left[ |R_{j,h}|^{2m} \right] \leq K$ which follows using the same steps that led to (B.1.86) with the Burkhölder-Davis-Gundy inequality in place of the Itô’s isometry. Let $g \left( \tilde{Y}_s^*, \tilde{\theta} \right) \triangleq \mu_{1,1} + \alpha_{1,1} \tilde{Y}_s^*$, $\tilde{\theta}_1 \triangleq (\mu_{1,1}, \alpha_{1,1})'$ and $\tilde{\theta}_1 \triangleq (\mu_{2,1}, \alpha_{2,1})'$. For any $\tilde{\theta}_1$, $\tilde{\theta}_2$, first use the Burkhölder-
Davis-Gundy inequality to yield,

\[
\mathbb{E} \left[ \left| R_{j,h} (\tilde{\theta}_2) - R_{j,h} (\tilde{\theta}_1) \right|^{2m} \right]
\]

\[
\leq (B \gamma_h)^{2m} K_m \left[ \sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(r)} \right)^2 \right]^{m} \mathbb{E} \left[ \left( \int_0^{N_0^b} \left( g \left( \bar{Y}_s, \tilde{\theta}_2 \right) - g \left( \bar{Y}_s, \tilde{\theta}_1 \right) \right)^2 ds \right)^m \right]
\]

\[
\leq (B \gamma_h)^{2m} K_m \left[ \sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(r)} \right)^2 \right]^{m} \times \mathbb{E} \left[ \left( \int_0^{N_0^b} \left( (\mu_{1,2} - \mu_{1,1}) + (\alpha_{1,2} - \alpha_{1,1}) \bar{Y}_s \right)^2 ds \right)^m \right]
\]

\[
\leq (B \gamma_h)^{2m} K_m \left[ \sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(r)} \right)^2 \right]^{m} \times \mathbb{E} \left[ \left( \int_0^{N_0^b} \left( (\mu_{1,2} - \mu_{1,1}) + (\alpha_{1,2} - \alpha_{1,1}) C \right)^2 ds \right)^m \right]
\]

\[
\leq (B \gamma_h)^{2m} K_m \mathbb{E} \left[ \left( \int_0^{N_0^b} \left( 2 (\mu_{1,2} - \mu_{1,1})^2 + 2C (\alpha_{1,2} - \alpha_{1,1})^2 \right) ds \right)^m \right]
\]

\[
\leq (B \gamma_h)^{2m} K_m \left[ \int_0^{N_0^b} \left( 2 (\mu_{1,2} - \mu_{1,1})^2 + 2 (\alpha_{1,2} - \alpha_{1,1})^2 \right. \right.
\]

\[
- 2 (\alpha_{1,2} - \alpha_{1,1})^2 + 2C (\alpha_{1,2} - \alpha_{1,1})^2 ds \right)^m \right]
\]

\[
\leq 2^m (B \gamma_h)^{2m} K_m \left\| \bar{Y}_s \right\|^{2m} \left( \int_0^{N_0^b} ds \right)^m + 2^m (B \gamma_h)^{2m} K \left( \tilde{\theta}_1, \tilde{\theta}_2, m, C \right)
\]

(B.1.87)

where \( C = \sup_{s \geq 0} |\bar{Y}_s| \), \( K \left( \tilde{\theta}_1, \tilde{\theta}_2, m, C \right) \) is some constant that depends on its arguments and we have used that \((a + b)^2 \leq 2a^2 + 2b^2\). Thus, since \( \gamma_h \downarrow 0 \), the mapping \( R_{j,h} (\cdot) \) satisfies a Lipschitz-type condition [cf. Section 2 in Andrews (1992)]. This is sufficient for the asymptotic stochastic equicontiuity of \( \{R_{j,h} (\cdot)\} \). Therefore, using Theorem 20 in Appendix I of Ibragimov and Has’minskii (1981), (B.1.86) and (B.1.87)
yield (B.1.85). Since the same result can be shown to remain valid for each term in
the stochastic element $\Delta_hV_k(\pi, \delta_{Z,1}, \delta_{Z,2})$, this establishes the claim. □

**Proof of Proposition 2.9.2.** To avoid clutter, we prove the case for which the true
parameters are $(\mu^0, \alpha^0)'$. The extension to parameters being local-to-zero is straight-
forward. The least-squares estimates of $(\mu^0, \alpha^0)'$ are given by,

$$
\hat{\mu}_1N_0 = \bar{Y}_{N_0} - \bar{Y}_0 - \hat{\alpha}_1 h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}
$$

(B.1.88)

Then, assuming $\hat{T}_b < T^0_b$,

$$
\hat{\alpha}_1 = \frac{\sum_{k=1}^{T_b} \left( \bar{Y}_{kh} - \bar{Y}_{(k-1)h} \right) \bar{Y}_{(k-1)h} - \left( \hat{N}_b^{-1} \left( \bar{Y}_{N_0} - \bar{Y}_0 \right) \right) h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}}{h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}^2 - \hat{N}_b^{-1} \left( h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h} \right)^2}.
$$

(B.1.89)

$$
\hat{\alpha}_1 = \frac{\sum_{k=1}^{T_b} \left( \mu^0 + \alpha^0 \bar{Y}_{(k-1)h} + \Delta_h \bar{V}_{h,k} \right) \bar{Y}_{(k-1)h}}{h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}^2 - \hat{N}_b^{-1} \left( h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h} \right)^2} - \left( \mu^0 + \alpha^0 \hat{N}_b^{-1} \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}h + \hat{N}_b^{-1} B\gamma_h \left( \bar{V}_{N_0} - \bar{V}_0 \right) \right) h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h} + o_p(1),
$$
and thus
\[ \hat{\alpha}_1 \]
\[ = \frac{\sum_{k=1}^{T_b^0} \left( \mu_1^0 h + \alpha_1^0 \tilde{Y}_{(k-1)h} h + \Delta_h \tilde{V}_k \right) \tilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h}^2 - \tilde{N}_b^{-1} \left( h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h} \right)^2} \]
\[ - \frac{\sum_{k=1}^{T_b^0} \left( \mu_1^0 h + \alpha_1^0 \tilde{Y}_{(k-1)h} h + \Delta_h \tilde{V}_k \right) \tilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h}^2 - \tilde{N}_b^{-1} \left( h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h} \right)^2} + \]
\[ \tilde{N}_b^{-1} \left( \sum_{k=1}^{\tilde{T}_b+1} \mu_1^0 h + \alpha_1^0 \sum_{k=1}^{\tilde{T}_b+1} \tilde{Y}_{(k-1)h} h + B \gamma_h \left( V_{N_b^0} - V_{\tilde{N}_b} \right) \right) \frac{\sum_{k=1}^{T_b^0} \tilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h}^2 - \tilde{N}_b^{-1} \left( h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h} \right)^2} \]
\[ \times \tilde{N}_b^{-1} \frac{h \sum_{k=1}^{\tilde{T}_b+1} \tilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h}^2 - \tilde{N}_b^{-1} \left( h \sum_{k=1}^{\tilde{T}_b} \tilde{Y}_{(k-1)h} \right)^2}. \]

By part (ii) of Theorem 2.9.1, \( N_b^0 - \tilde{N}_b = O_p(h^{1-\kappa}) \), and thus it is easy to see that the third and fourth terms go to zero in probability at a slower rate than \( h^{1-\kappa} \). As for the first and second terms, recalling that \( \Delta_h \tilde{V}_{h,k} = h^{1/2} \Delta V_{h,k} \) from (2.9.3), we have by ordinary convergence of approximations to Riemann sums, Lemma B.1.14 and the continuity of probability limits,
\[ \alpha_1^0 \sum_{k=1}^{T_b^0} \tilde{Y}_{(k-1)h} h \xrightarrow{P} \alpha_1^0 \int_0^{N_b^0} \tilde{Y}_s ds, \quad \sum_{k=1}^{T_b^0} \mu_1^0 h \xrightarrow{P} \mu_1^0 \int_0^{N_b^0} ds, \]
and by Lemma B.1.15, \( \sum_{k=1}^{T_b^0} \tilde{Y}_{(k-1)h} \Delta_h \tilde{V}_k \xrightarrow{P} 0 \). Thus, we deduce that
\[ \hat{\alpha}_1 = \alpha_1^0 + O_p(B \gamma_h). \] (B.1.90)
Using (B.1.90) into (B.1.88),
\[
\tilde{\mu}_1 \tilde{N}_b = \tilde{Y}_{N_b} - \tilde{Y}_0 - \alpha_1^0 h \sum_{k=1}^{T_h} \tilde{Y}_{(k-1)h} - O_p (B \gamma_h),
\]
\[
= \tilde{Y}_{N_b} - \tilde{Y}_0 - \alpha_1^0 h \sum_{k=1}^{T_h} \tilde{Y}_{(k-1)h} - \alpha_1^0 h \sum_{k=T_h+1}^{T_b} \tilde{Y}_{(k-1)h} - o_p (1).
\]

By part (ii) of Theorem 2.9.1, the number of terms in the second sum above increases at rate \( T^\kappa \) and thus, \( \alpha_0^0 h \sum_{k=T_h+1}^{T_b} \tilde{Y}_{(k-1)h} = KO_p (h^{1-\kappa}) \), where we have also used standard estimates for the drift arising from the Burkhölder-Davis-Gundy inequality. This gives
\[
\tilde{\mu}_1 \tilde{N}_b = \tilde{Y}_{N_b} - \tilde{Y}_0 - \alpha_1^0 h \int_0^{N_b^0} \tilde{Y}_s ds - \alpha_1^0 O_p \left( h^{1-\kappa} \right) - o_p (1).
\]

Noting that
\[
\tilde{Y}_{N_b^0} - \tilde{Y}_0 = \mu_1^0 N_b^0 + \alpha_1^0 \int_0^{N_b^0} \tilde{Y}_s ds + O_p (B \gamma_h) \left( V_{N_b^0} - V_0 \right),
\]
we have \( \tilde{\mu}_1 N_b^0 = \mu_1^0 N_b^0 + O_p (B \gamma_h) \left( V_{N_b^0} - V_0 \right) \), which yields
\[
\tilde{\mu}_1 = \mu_1^0 + O_p (B \gamma_h).
\]

Thus, as \( h \downarrow 0 \), \( \tilde{\mu}_1 \) is consistent for \( \mu_1^0 \). The case where \( T_h > T_b^0 \) can be treated in the same fashion and is omitted. Further, the consistency proof for \( (\tilde{\mu}_2, \tilde{\alpha}_2)' \) is analogous and also omitted. The second step is to construct the least-squares residuals and
scaling them up. The residuals are constructed as follows,

\[
\hat{u}_{kh} = \begin{cases} 
  h^{-1/2} \left( \Delta_h \tilde{Y}_k - \hat{\mu}_1 \tilde{x}^{(1)}_{kh} - \hat{\alpha}_1 \tilde{x}^{(2)}_{kh} \right), & k \leq \hat{T}_b \\
  h^{-1/2} \left( \Delta_h \tilde{Y}_k - \hat{\mu}_2 \tilde{x}^{(1)}_{kh} - \hat{\alpha}_2 \tilde{x}^{(2)}_{kh} \right), & k > \hat{T}_b,
\end{cases}
\]

where \( \tilde{x}^{(1)}_{kh} = h \) and \( \tilde{x}^{(2)}_{kh} = \tilde{Y}_{(k-1)h}h \). This yields, for \( k \leq T^0_b \leq \hat{T}_b \),

\[
\hat{u}_{kh} = h^{-1/2} \left( \mu_1^0 h + \alpha_1^0 \tilde{Y}_{(k-1)h}h + B \gamma h \Delta_h V_k - \hat{\mu}_1 h - \hat{\alpha}_1 \tilde{Y}_{(k-1)h}h \right),
\]

and using (B.1.90) and (B.1.91),

\[
\hat{u}_{kh} = h^{-1/2} \left( \mu_1^0 h + \alpha_1^0 \tilde{Y}_{(k-1)h}h + B \gamma h \Delta_h V_k - \mu_1^0 h - O_p \left( h^{3/2} \right) \right) \\
- \alpha_1^0 \tilde{Y}_{(k-1)h}h - O_p \left( h^{3/2} \right) \\
= h^{-1/2} B \gamma h \Delta_h V_k - O_p \left( h \right). \quad (B.1.92)
\]

Similarly, for \( T^0_b \leq \hat{T}_b \leq k \),

\[
\hat{u}_{kh} = h^{-1/2} B \gamma h \Delta_h V_k - O_p \left( h \right), \quad (B.1.93)
\]

whereas for \( \hat{T}_b < k \leq T^0_b \),

\[
\hat{u}_{kh} = h^{-1/2} \left( \mu_1^0 h + \alpha_1^0 \tilde{Y}_{(k-1)h}h + B \gamma h \Delta_h V_k - \mu_2^0 h - O_p \left( h^{3/2} \right) \right) \\
- \alpha_2^0 \tilde{Y}_{(k-1)h}h - O_p \left( h^{3/2} \right) \\
= h^{-1/2} \left( -\mu_2^0 h - \alpha_2^0 \tilde{Y}_{(k-1)h}h + B \gamma h \Delta_h V_k - O_p \left( h^{3/2} \right) \right) \\
= -\mu_2^0 h^{1/2} - \alpha_2^0 \tilde{Y}_{(k-1)h}h^{1/2} + h^{-1/2} B \gamma h \Delta_h V_k - O_p \left( h \right). \quad (B.1.94)
\]
Next, note that \( \sum_{k=\hat{T}_b+1}^{T_b} \mu_k^bh^{1/2} \leq Kh^{1/2-\kappa} \) and \( \sum_{k=\hat{T}_b+1}^{T_b} \alpha_k^0 \hat{y}_{(k-1)}h^{1/2} \leq Kh^{1/2-\kappa} \) since by Theorem 2.9.1-(ii) there are \( T^* \) terms in each sum. Moreover, recall that \( \epsilon_{kh} = \Delta_h e_k^* \sim \mathcal{N}(0, \sigma_{\epsilon_{k-1}}^2) \) and thus \( \sum_{k=\hat{T}_b+1}^{T_b} \epsilon_{kh} = \sqrt{h} \sum_{k=\hat{T}_b+1}^{T_b} h^{-1/2} \epsilon_{kh} = h^{1/2-\kappa} o_p(1) \). Therefore, \( \sum_{k=\hat{T}_b+1}^{T_b} \tilde{u}_{kh} = K o_p \left( h^{1/2-\kappa} \right) \). Since \( \kappa \in (0, 1/2) \), this shows that the residuals \( \tilde{u}_{kh} \) from equation (B.1.94) are asymptotically negligible. That is, asymptotically the estimator of \( (\beta_0^S)' \), \( (\delta_0^0)' \), \( (\delta_0^1)' \), \( (\delta_0^2)' \) minimizes (assuming \( \hat{T}_b \leq T_b^0 \)),

\[
\sum_{k=1}^{\hat{T}_b} (\tilde{u}_{kh} - \tilde{x}_{kh}'\beta^S)^2 + \sum_{k=\hat{T}_b+1}^{T_b} (\tilde{u}_{kh} - \tilde{x}_{kh}'\beta^S - \tilde{z}_{0,kh}'\delta^S)^2 + o_p(1),
\]

where \( X = [\tilde{X}^{(1)} \ \tilde{X}^{(2)} \ \tilde{X}] \), \( \beta^0 = [\mu_0^0 \ \alpha_0^0 \ \beta^0_S]' \), and \( Z_0 \) and \( \delta_S^0 \) are partitioned accordingly. The subscript \( S \) indicates that these are the parameters of the stochastic semimartingale regressors. But this is exactly the same regression model as in Proposition 2.3.3. Hence, the consistency result for the slope coefficients of the semimartingale regressors follows from the same proof. The following regression model estimated by least-squares provides consistent estimates for \( \beta_S^0 \) and \( \delta_S^0 \): \( \hat{U} = \tilde{X} \beta_S^0 + \tilde{Z}_0 \delta_S^0 + \text{residuals} \), where

\[
\hat{Z}_0 = \begin{bmatrix}
\tilde{z}_1^{(1)} & \ldots & \tilde{z}_1^{(p)} \\
\vdots & \ddots & \vdots \\
\tilde{z}_{T_b}^{(1)} & \ldots & \tilde{z}_{T_b}^{(p)} \\
\tilde{z}_{(T_b^0+1)}^{(1)} & \ldots & \tilde{z}_{(T_b^0+1)}^{(p)} \\
\vdots & \ddots & \vdots \\
\tilde{z}_N^{(1)} & \ldots & \tilde{z}_N^{(p)}
\end{bmatrix},
\]

and \( \hat{U} = (\tilde{u}_{kh}; \ k = 1, \ldots, \hat{T}_b, \ T_b^0 + 1, \ldots, N) \). Therefore, using (B.1.92) and (B.1.93),

\[\text{The same bound holds for the corresponding sum involving the other terms in } \Delta_h V_k.\]
we have

\[ h^{-1/2} \begin{bmatrix} \hat{\beta}_S - \beta^0 \\ \hat{\delta}_S - \delta^0 \end{bmatrix} = \begin{bmatrix} \tilde{X}'\tilde{X} & \tilde{X}'\tilde{Z}_0 \\ \tilde{Z}_0'\tilde{X} & \tilde{Z}_0'\tilde{Z}_0 \end{bmatrix}^{-1} h^{-1/2} \begin{bmatrix} \tilde{X}'e & \tilde{X}' (Z_0 - \tilde{Z}_0) \delta^0 + \tilde{X}'AO_p(h) \\ \tilde{Z}_0'e & \tilde{Z}_0' (Z_0 - \tilde{Z}_0) \delta^0 + \tilde{Z}_0'AO_p(h) \end{bmatrix}, \]

for some matrix \( A = O_p(1) \). It then follows by the same proof as in Proposition 2.3.3 that

\[ \begin{bmatrix} \tilde{X}'\tilde{X} & \tilde{X}'\tilde{Z}_0 \\ \tilde{Z}_0'\tilde{X} & \tilde{Z}_0'\tilde{Z}_0 \end{bmatrix}^{-1} \tilde{X}'AO_p(h^{1/2}) = o_p(1), \] (B.1.95)

and

\[ \begin{bmatrix} \tilde{X}'\tilde{X} & \tilde{X}'\tilde{Z}_0 \\ \tilde{Z}_0'\tilde{X} & \tilde{Z}_0'\tilde{Z}_0 \end{bmatrix}^{-1} \frac{1}{h^{1/2}} \tilde{X}' (Z_0 - \tilde{Z}_0) \delta^0 = O_p(1) o_p(1) = o_p(1). \] (B.1.96)

The same arguments can be used for \( \tilde{Z}_0' (Z_0 - \tilde{Z}_0) \delta^0 \) and \( \tilde{Z}_0'AO_p(h) \). Therefore, in view of (B.1.90) and (B.1.91), we obtain \( \hat{\mu}_1 = \mu_1^0 + o_p(1) \) and \( \hat{\alpha}_1 = \alpha_1^0 + o_p(1) \), respectively, whereas (B.1.95) and (B.1.96) imply \( \hat{\beta}_S = \beta_S^0 + o_p(1) \) and \( \hat{\delta}_S = \delta_S^0 + o_p(1) \), respectively. Under the setting where the magnitude of the shifts is local to zero, we observe that by Proposition 2.4.1, \( \hat{N}_b - \tilde{N}_b^0 = O_p(h^{1-\kappa}) \) and one can follow the same steps that led to (B.1.90) and (B.1.91) and proceed as above. The final result is \( \hat{\theta} = \theta^0 + o_p(1) \), which is what we wanted to show. □

**B.1.5.4 Negligibility of the Drift Term**

Recall Lemma B.1.10 and apply the same proof as in Section 2.9.3.3. Of course, the negligibility only applies to the drift processes \( \mu_{.,t} \) from (2.2.3) (i.e., only the drift processes of the semimartingale regressors) and not to \( \mu_1^0, \mu_2^0, \alpha_1^0 \) or \( \alpha_2^0 \). The steps are omitted since they are the same.
B.2 Additional Discussion about the Continuous Record Asymptotic Density Function

B.2.1 Further Discussion from Section 2.4.2

In this section, we continue our discussion about the properties of the continuous record asymptotic distribution from Section 2.4. It is useful to plot the probability densities for a fractional break date $\lambda_0$ close to the endpoints. Figure B.1 presents the densities of $\rho\left(\hat{T}_b - T_0^b\right)$ given in equation (2.4.5) for $\rho^2 = 0.2, 0.3, 0.5$ (the left, middle and right panel, respectively) and a true break point $\lambda_0 = 0.2$. The figure also reports the density of the shrinkage large-$N$ asymptotic distribution. We report corresponding plots for $\lambda_0 = 0.35, 0.5, 0.75$ in Figure B.2-B.4. The shape of the density of the shrinkage large-$N$ asymptotic distribution is seen to remain unchanged as we raise the signal-to-noise ratio. It is always symmetric, uni-modal and centered at the true value $\lambda_0$. This contrasts with the density derived under a continuous record. From Figure B.1 it is easily seen that when the break size is small, the density from Theorem 2.4.2 is always asymmetric suggesting that the location of the break date indeed plays a key role in shaping the asymptotic distribution even if the regressors and errors have the same distribution across adjacent regimes. As we raise the signal-to-noise ratio (from left to right panel) the distribution becomes less asymmetric and accordingly less positively skewed but both features are still evident. An additional feature arises from this plot. There are only two modes when $\lambda_0 = 0.2$ (cf. Figure B.1, left and middle panels), the mode at the true value being no longer present. When the date of the break is not in the middle 80% of the sample, the density shows bi-modality rather than tri-modality as we discussed in Section 2.4. This constitutes the only exception to the otherwise similar comments that can be
made when \( \lambda_0 = 0.35 \) and 0.5 (cf. Figure B.2-B.3). Figure B.4 displays the case for \( \lambda_0 = 0.75 \). Once again, the distribution is asymmetric. Since \( \lambda_0 \) is located in the second half of the sample, the density is negatively skewed.

When we consider nearly stationary regimes, that is, we allow for low heterogeneity across regimes according to the restrictions in (2.4.6), the results are not affected. However, observe that when the heterogeneity is higher, there are few notable distinctions as explained in the main text.

The features of the density under a continuous record arise from the properties of the limiting process. Consider the process \( \mathcal{V}(s) \) as defined before Theorem 2.4.2. The limiting distribution is related to the extremum of \( \mathcal{V}(s) \) over a fixed time interval with boundary points \(-\rho N_b^0 / \| \delta^0 \|^{-2} \sigma^2 \) and \( \rho(N - N_b^0) / \| \delta^0 \|^{-2} \sigma^2 \). \( \mathcal{V}(s) \) has a continuous sample path and it is the sum of a deterministic component or drift and a stochastic Gaussian component. The deterministic part is given by the second moments of the regressors and thus it is always negative because of the minus in front of it. The term \((|s|/2)(\delta^0)'(Z,Z)(\delta^0)\) is of \(|s|\) whereas the stochastic term is of order \(|s|^{1/2}\). This means that for small \(|s|\), the highly-fluctuating Gaussian part is more influential. However, when the signal-to-noise ratio is large (\( \rho \) is high), the deterministic part dominates the stochastic one. Thus, the maximum of \( \mathcal{V}(s) \) cannot be attained at large values of \(|s|\). This explains why there is only one mode at the origin when the signal is high. In contrast, when the signal-to-noise ratio is low, the interval over which \( \mathcal{V}(s) \) is maximized is short. Hence, the fluctuations in the stochastic part dominates that of the deterministic one as the former is of higher order on that interval. This has at least two consequences. First, there is another mode at each of the endpoints because by the property of the Gaussian part of \( \mathcal{V}(s) \) it is much more likely to attain a maximum close to the boundary points than at any interior point. We refer
to Karatzas and Shreve (1996) for an accessible treatment about the probabilistic aspects of this class of processes. Second, when $\rho$ is low, so is $(|s|/2)(\delta^0)'(Z, Z)_{(\cdot)} \delta^0$, and thus it is more likely that the maximum is achieved at either endpoint than at zero. This explains why when the signal is low the highest mode is not at the origin. When $\lambda_0 \neq 0.5$, the interval over which $V(s)$ is maximized is asymmetric and as a consequence the density is also asymmetric. If $\rho$ is not very large, when $\lambda_0$ is less (larger) than 0.5 there is a higher probability for $V(s)$ to attain a maximum closer to the left (right) boundary point since the deterministic component takes a less negative value at $-\rho N^0_b / \|\delta^0\|^{-2} \sigma^2$ and $\rho (N - N^0_b) / \|\delta^0\|^{-2} \sigma^2$. When the size of the break is sufficiently high, the density is always symmetric and has unique mode at a value corresponding to $\lambda_b$ being close to $\lambda_0$ because the deterministic component $(|s|/2)(\delta^0)'(Z, Z)_{(\cdot)} (\delta^0)$ is large enough that $V(s)$ decreases as it moves away from the origin. Thus, with very high probability, the maximum is located at the origin.

When considering nonstationary regimes, the heterogeneity across regimes determines the stochastic order of the process $V(s)$. If the post-break regime has higher volatility, there is a high probability that the limiting process attains a maximum on the interval $[0, \rho (N - N^0_b) / \|\delta^0\|^{-2} \sigma^2]$ since it fluctuates more in that region. This explains why the density is clearly negatively skewed and the mode near the right boundary point is always higher than the mode near the left boundary point (Figure B.9-B.11, right panel).

Consider now the extreme cases $\lambda_0 = 0.1, 0.45, 0.55$ and $\lambda_0 = 0.9$. The characteristics discussed in Section 2.5 remain valid as can be seen from Figure B.5-B.8. The features of skewness, asymmetry, tri-modaility (only when $\rho$ is low) and peakedness (when $\rho$ is high) are all more pronounced for those relatively more extreme cases. For example, in Figure B.5 we plot the densities of $\tilde{T}_b$ for a true break fraction $\lambda_0 = 0.1$. 

(near the beginning of the sample). When $\rho$ is low there are now only two modes because the mode associated with the middle point has disappeared. This bi-modaility vanishes as we increase $\rho$, and the density is positively skewed for all values of the signal-to-noise ratio. Similar comments apply to the other cases.

That the density is symmetric only if the break date is at half sample ($\lambda_0 = 0.5$) and that this property is sharp, can be seen from Figure B.6-B.7, left panel. When the true break date is not exactly at 0.5 but, e.g. as close as 0.45, the density is visibly asymmetric. Further, in such a case the density is positively skewed and the highest mode is towards beginning of the sample.

**B.2.2 Further Discussion from Section 2.5**

We continue with the analysis of cases allowing differences between the distribution of the errors and regressors in the pre- and post-break regimes (i.e., nonstationary regimes). We consider a scenario where the second regime is twice as volatile as the first. Here the signal-to-noise ratio is given by $\delta^0/\sigma_{e,1}$, where $\sigma_{e,1}^2$ is the variance of the error term in the first regime. We notice substantial similarities with the cases considered above but there is one notable exception. In Figure B.9-B.12, the shrinkage asymptotic density of Bai (1997) is asymmetric and unimodal for all pairs $(\rho^2, \lambda_0)$ considered. The density is negatively skewed and the right tail much fatter then the left tail. Turning to the density from (2.4.6), we can make the following observations. Even if the signal-to-noise ratio is moderately high, the asymptotic distribution deviates from being symmetric when the break occurs at exactly middle sample ($\lambda_0 = 0.5$, Figure B.11, right panel). This is in contrast to the nearly stationary framework since the density was shown to be always symmetric no matter the value taken by $\rho$ if $\lambda_0 = 0.5$. This suggests that when the statistical properties of the errors and regressors display significant differences across the two regimes, the
probability densities derived under a continuous record is not symmetric even with $\lambda_0 = 0.5$. This means that the asymptotic distribution attributes different weights to the informational content of the two regimes since they possess highly heterogeneous statistical characteristics. Figure B.11 makes this point clear. It reports plots for the case with $\lambda_0 = 0.5$ and $\rho^2 = 0.3, 0.8, 1.5$ (from left to right panels). The density is no longer symmetric and the right tail is much fatter than the left one. This follows simply because there is more variability in the post-break region. In such cases, there is a tendency to overestimate the break point which leads to an upward bias if the post-break regime displays larger variability. There are important differences with respect to Bai’s (1997) density. First, although the density under a continuous record asymptotics is also asymmetric for all $\lambda_0$, the degree of asymmetry varies across different break dates. Second, there is multi-modality when the size of the break is small which is not shared with Bai’s (1997) density since the latter is always unimodal. Finally, one should expect the density to be symmetric when the magnitude of the break is large as the distribution should collapse at $\lambda_0$ for large breaks. The continuous record asymptotics reproduces this property whereas the large-$N$ asymptotic distribution of Bai (1997) remains asymmetric even for large break sizes (Figure B.12).
B.2.3 Figures

Figure B.1: Distributions with \( \lambda_0 = 0.2 \)

The asymptotic probability density of \( \rho \left( \hat{T}_b - T_b^\theta \right) \) derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) for \( \lambda_0 = 0.2 \) and \( \rho^2 = 0.2, 0.3 \) and 0.5 (the left, middle and right panel, respectively).
Figure B.2: Distributions with $\lambda_0 = 0.35$

The asymptotic probability density of $\rho\left(\hat{T}_b - T^0_b\right)$ derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.35$ and $\rho^2 = 0.2, 0.3$ and 0.5 (the left, middle and right panel, respectively).

Figure B.3: Distributions with $\lambda_0 = 0.5$

The asymptotic probability density of $\rho\left(\hat{T}_b - T^0_b\right)$ derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) for a true fractional break date $\lambda_0 = 0.5$ and $\rho^2 = 0.2, 0.3$ and 0.5 (the left, middle and right panel, respectively).
Figure B.4: Distributions with $\lambda_0 = 0.75$

The asymptotic probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.75$ and $\rho^2 = 0.2, 0.3$ and 0.5 (the left, middle and right panel, respectively).

Figure B.5: Distributions with $\lambda_0 = 0.1$

The asymptotic probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ derived under a continuous record (blue solid line) and the density of of Bai’s (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.1$ and $\rho^2 = 0.3, 0.8$ and 1.5 (the left, middle and right panel, respectively).
Figure B.6: Distributions with $\lambda_0 = 0.45$

The asymptotic probability density of $\rho \left( \hat{T}_b - T_0^b \right)$ derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.45$ and $\rho^2 = 0.3, 0.8$ and 1.5 (the left, middle and right panel, respectively).

Figure B.7: Distributions with $\lambda_0 = 0.55$

The asymptotic probability density of $\rho \left( \hat{T}_b - T_0^b \right)$ derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) for a true break point $\lambda_0 = 0.55$ and $\rho^2 = 0.3, 0.8$ and 1.5 (the left, middle and right panel, respectively).
Figure B.8: Distributions with \( \lambda_0 = 0.9 \)

The asymptotic probability density of \( \rho \left( \hat{T}_b - T_0^b \right) \) derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) for \( \lambda_0 = 0.9 \) and \( \rho^2 = 0.3, 0.8 \) and 1.5 (the left, middle and right panel, respectively).

Figure B.9: Distributions with \( \lambda_0 = 0.2 \) for nonstationary regimes

The asymptotic probability density of \( \rho \left( \hat{T}_b - T_0^b \right) \) derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) under nonstationary regimes for \( \lambda_0 = 0.2 \) and \( \rho^2 = 0.3, 0.5 \) and 1 (the left, middle and right panel, respectively).
Figure B.10: Distributions with $\lambda_0 = 0.35$ for nonstationary regimes
The asymptotic probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ derived under a continuous record (blue solid line) and the density of Bai’s (1997) asymptotic distribution (black broken line) under nonstationary regimes for $\lambda_0 = 0.35$, and $\rho^2 = 0.3, 0.5$ and 1 (the left, middle and right panel, respectively).

Figure B.11: Distributions with $\lambda_0 = 0.5$ for nonstationary regimes
The asymptotic probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ derived under a continuous record (solid line) and the density of Bai’s (1997) asymptotic distribution (broken line) under nonstationary regimes for $\lambda_0 = 0.5$, and $\rho^2 = 0.3, 0.5$ and 1 (the left, middle and right panel, respectively).
Figure B.12: Distributions with $\lambda_0 = 0.5$ for nonstationary regimes
The asymptotic probability density of $\rho \left( \hat{T}_b - T_0^b \right)$ derived under a continuous record (solid line) and the density of Bai’s (1997) asymptotic distribution (broken line) under nonstationary regimes for $\lambda_0 = 0.5$, and $\rho^2 = 1.2$, 1.5 and 2 (the left, middle and right panel, respectively).

Figure B.13: Distributions with $\lambda_0 = 0.7$ for nonstationary regimes
The asymptotic probability density of $\rho \left( \hat{T}_b - T_0^b \right)$ derived under a continuous record (solid line) and the density of Bai’s (1997) asymptotic distribution (broken line) under nonstationary regimes for $\lambda_0 = 0.7$, and $\rho^2 = 0.3$, 0.5 and 1 (the left, middle and right panel, respectively).
Figure B.14: Distributions for model (2.5.1) with $\delta^0 = 0.3$ for non-stationary regimes

The probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ for model (6.1) with break magnitude $\delta^0 = 0.3$ and $\lambda_0 = 0.3, 0.5$ and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0 / \sigma_{e,1} = \delta^0$ since $\sigma_{e,1}^2 = 1$ where $\sigma_{e,1}^2$ is the variance of the errors in the first regime. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution of Bai (1997) and the red broken line break is the density of the finite-sample distribution.
Figure B.15: Distributions for model (2.5.1) with $\delta^0 = 1$

The probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ for model (6.1) with $\delta^0 = 0.5$ and $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0 / \sigma_{e,1} = \delta^0$ since $\sigma_{e,1}^2 = 1$ where $\sigma_{e,1}^2$ is the variance of the errors in the first regime. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution of Bai (1997) and the red broken line break is the density of the finite-sample distribution.
Figure B.16: Distributions for model (2.5.1) with $\delta^0 = 1.5$

The probability density of $\rho \left( \hat{T}_b - T^0_b \right)$ for model (6.1) with $\delta^0 = 1.5$ and $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_{e,1} = \delta^0$ since $\sigma_{e,1}^2 = 1$ where $\sigma_{e,1}^2$ is the variance of the errors in the first regime. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution of Bai (1997) and the red broken line break is the density of the finite-sample distribution.
Appendix C

Supplement to Chapter 3: Tests for Forecast Instability and Forecast Failure under a Continuous Record Asymptotic Framework

C.1 Mathematical Proofs

The Mathematical Appendix is structured as follows. The proofs of the results in Section 3.3 and 3.4 are collected in Section C.1.4 and C.1.5, respectively. The results of Section 3.5 are covered in Section C.1.6.

C.1.1 Additional Notation

Throughout the proofs, $C$ is a generic constant that may vary from line to line; we may sometime write $C_r$ to emphasize the dependence of $C$ on a scalar $r$. For brevity, we indicate that a sequence $\{U_k\}$ is formed by independent and non-identically distributed random variables by labeling it as i.n.d. For the variables $\Delta_h e_k$ and $\Delta_h X_k$ we use a tilde notation to denote their normalized version: $\Delta_h \tilde{e}_k = h^{-1/2} \Delta_h e_k$ and $\Delta_h \tilde{X}_k = h^{-1/2} \Delta_h X_k$. We use a star superscript ($\ast$) on $\Delta_h e_k$ to indicate the residuals obtained when $\beta = \beta^\ast$: $\Delta_h \tilde{e}_k^\ast = h^{-1/2} \left( \Delta_h Y_k - (\beta^\ast)' \Delta_h X_{k-\tau} \right)$. We sometime omit
the index from $\hat{\beta}_k$ and simply use $\hat{\beta}$ when it is clear from the context.

\section*{C.1.2 Localization}

As it is typical in the high-frequency statistics literature, we use a localization argument [cf. Section I.1.d in Jacod and Shiryaev (2003)]. Thus, we replace Assumption 3.1 and Assumption 3.2 by the following stronger assumption which basically turns the local restrictions into global.

\textbf{Assumption C.1.} Let Assumption 3.1-3.2, Assumption 3.3-3.8 and Condition 2 hold. When $\mu_{e,t} = 0$ for all $t \geq 0$ the process $\{Z_t\}_{t \geq 0}$ takes value in some compact set; the processes $\{\sigma_{X,t}, \sigma_{e,t}\}_{t \geq 0}$ are bounded càdlàg and $\{\mu_{X,t}, \mu_{e,t}\}_{t \geq 0}$ are bounded càdlàg. Furthermore, $\phi_{\sigma,\eta,N} \leq C\eta$ for some $C < \infty$.

\section*{C.1.3 Preliminary Lemmas}

\textbf{Lemma C.1.1.} For any $1 \leq r, l \leq q$, and $1 \leq i \leq n_T$, we have

(i) $\sup_{b=0,...,\lfloor T_n/n_T \rfloor-2} \sum_{j=1}^{T_n+b_nT+i-1} \Delta_h X_k^{(r)} \Delta_h e_k^* \overset{p}{\to} 0$;

(ii) $\sup_{b=0,...,\lfloor T_n/n_T \rfloor-2} \left\| \sum_{j=1}^{T_n+b_nT+i-1} \Delta_h X_k^{(r)} \Delta_h X_{k'}^{(l')'} - \int_0^{T_n+b_nT} \sum_{j=1}^{T_n+b_nT} X_{k,j}^{(r)} X_{k',j}^{(l')} ds \right\| \overset{p}{\to} 0$;

(iii) the central limit theorem in Lemma S.A.5 in Casini and Perron (2017a) holds for $X_t$.

\textit{Proof.} Part (i)-(ii) are a consequence of the law of large numbers for quadratic variation; see Section S.A.3 in Casini and Perron (2017a). For part (iii) see the above referenced theorem. $\square$

\section*{C.1.4 Proofs of Section 3.3}

Throughout this section we maintain Assumption C.1.
C.1.4.1 Proof of Theorem 3.3.1

The idea behind the proof of both Theorem 3.3.1-3.3.2 is the same. Thus, the quadratic loss case serves as a guide and we then use some of these derivations for the general loss case. All the results in this section are proved under $H_0$.

C.1.4.1.1 Proof of part (i) of Theorem 3.3.1

The theorem is proved through several lemmas. The first step involves showing that the error in replacing $\hat{\beta}$ by $\beta^*$ is asymptotically negligible. We provide the proof of this first step by assuming that $\mu_{e,t} = 0$ in (3.2.2). That is, in Lemma C.1.2-C.1.3 we have $\mu_{e,t} = 0$ and we show how these results continue to hold without this restriction in Section C.1.4.1.3. We focus for simplicity on the recursive scheme only; the proofs for the other cases are similar and omitted. Let $U_{h,b} \triangleq n^{-1}_T \sum_{j=1}^{n_T} SL_{\psi,T_m+\tau+bn_T+j-1}(\beta^*)$, $U_{h,b} \triangleq n^{-1}_T \sum_{j=1}^{n_T} L_{\psi,T_m+\tau+bn_T+j-1}(\beta^*)$ and

$$U_{\max,h}(T_n, \tau) \triangleq \max_{b=0,\ldots,\lfloor T_n/n_T \rfloor-2} \left| \frac{(U_{h,b+1} - U_{h,b})}{U_{h,b+1}} \right|.$$ 

In some steps of the proof, we will use the following simple result. For any integer $m \geq 1$, let $c_{1,b}$ and $c_{2,b}$ ($b = 1, \ldots, m$) be arbitrary real numbers, then

$$|c_{1,b}| \leq |c_{1,b} - c_{2,b}| + |c_{2,b}| \leq \max_{b=1,\ldots,m} |c_{1,b} - c_{2,b}| + \max_{b=1,\ldots,m} |c_{2,b}|.$$  \hspace{1cm} (C.1.1)

Lemma C.1.2. As $h \downarrow 0$, $(\log (T_n) n_T)^{1/2} (U_{\max,h}(T_n, \tau) - B_{\max,h}(T_n, \tau)) \xrightarrow{p} 0$.

Proof. By the reverse triangle inequality, inequality (C.1.1) and Lemma C.1.3 below,
for some $C_1, C_2 < \infty$,

$$| U_{\max,h}(T_n, \tau) - B_{\max,h}(T_n, \tau) |$$

$$\leq \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( SL_{\psi,T_m+\tau+(b+1)n_T+j-1} (\beta^*) / B_{h,b+1} - SL_{\psi,T_m+\tau+(b+1)n_T+j-1} (\hat{\beta}) / U_{h,b+1} \right) \right|$$

$$+ \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \left( \sum_{j=1}^{n_T} SL_{\psi,T_m+\tau+bn_T+j-1} (\beta^*) / U_{h,b+1} - SL_{\psi,T_m+\tau+bn_T+j-1} (\hat{\beta}) / U_{h,b+1} \right) \right|$$

$$\leq C_1 \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \left( \sum_{j=1}^{n_T} SL_{\psi,T_m+\tau+(b+1)n_T+j-1} (\beta^*) - SL_{\psi,T_m+\tau+(b+1)n_T+j-1} (\hat{\beta}) \right) / U_{h,b+1}^2 \right|$$

$$+ C_2 \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \left( \sum_{j=1}^{n_T} SL_{\psi,T_m+\tau+bn_T+j-1} (\beta^*) - SL_{\psi,T_m+\tau+bn_T+j-1} (\hat{\beta}) \right) / U_{h,b+1}^2 \right|.$$  

(C.1.2)

Note that for any $j = 1, \ldots, n_T$,

$$SL_{\psi,T_m+\tau+bn_T+j-1} (\beta^*) - SL_{\psi,T_m+\tau+bn_T+j-1} (\hat{\beta})$$

$$= L_{\psi,T_m+\tau+bn_T+j-1} (\beta^*) - L_{\psi,T_m+\tau+bn_T+j-1} (\hat{\beta}) + o_P \left( T^{-1/2} \right),$$

where the $o_P \left( T^{-1/2} \right)$ term arises from the proof of Lemma C.1.3. Recall that for $1 \leq j \leq n_T$,

$$\Delta_h \tilde{e}_{T_m+\tau+bn_T+j} = \sigma_{e,(T_m+\tau+bn_T)h} \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+bn_T+j} \right),$$
so that

\[ L_{\psi,T_{m+\tau+bn_T+j}}(\beta^*) - L_{\psi,T_{m+\tau+bn_T+j}}(\hat{\beta}) \]

\[ = - \left( \hat{\beta} - \beta^* \right)' \Delta_h \tilde{X}_{T_{m+bn_T+j}} \Delta_h \tilde{X}'_{T_{m+bn_T+j}} \left( \Delta_h \tilde{X}_{T_{m+bn_T+j}} \right)' \Delta_h \tilde{X}_{T_{m+bn_T+j}}. \]

\[ (C.1.3) \]

Recall that \( \Delta_h \tilde{X}_k = h^{-1/2} \Delta_h X_k \) and thus

\[ n_T^{-1} \sum_{j=1}^{n_T} \Delta_h \tilde{X}_{T_{m+bn_T+j}} \Delta_h \tilde{X}'_{T_{m+bn_T+j}} - \Sigma X_{(T_{m+bn_T})}h = o_p(1), \]

which follows from Theorem 9.3.2 part (i) in Jacod and Protter (2012). This implies that

\[ n_T^{-1} \sum_{j=1}^{n_T} \Delta_h \tilde{X}_{T_{m+bn_T+j-1}} \Delta_h \tilde{X}'_{T_{m+bn_T+j-1}} = O_p(1), \]

by Assumption 3.1-(iv). By Assumption 3.1-(v) and the aforementioned theorem,

\[ n_T^{-1} \sum_{j=1}^{n_T} \left( h^{-1/2} \Delta_h W_{\epsilon,T_{m+\tau+bn_T+j-1}} \right) \Delta_h \tilde{X}_{T_{m+bn_T+j-1}} \xrightarrow{p} 0. \]

Note that by Assumption 3.8, \( \hat{\beta}_k - \beta^* = O_p \left( 1/\sqrt{T} \right) \) uniformly in \( k \geq T_m \). Therefore, from these arguments we deduce that

\[ n_T^{-1} \sum_{j=1}^{n_T} \left( \Delta L_{\psi,T_{m+\tau+bn_T+j-1}}(\beta^*) - \Delta L_{\psi,T_{m+\tau+bn_T+j-1}}(\hat{\beta}) \right) = o_p \left( 1/\sqrt{T} \right). \]

\[ (C.1.4) \]

Then, for any \( \varepsilon > 0 \) and any constant \( K > 0 \), the first term on the right-hand side of
\((C.1.2)\) is

\[
\mathbb{P}\left( \max_{b=0,\ldots,\lfloor T_n/n_T \rfloor - 2} \left| \frac{(\log (T_n) n_T)^{1/2} (U_{h,b+1} - B_{h,b+1})}{U_{h,b+1}^2} \right| > \varepsilon \right) \\
\leq \mathbb{P}\left( \max_{b=0,\ldots,\lfloor T_n/n_T \rfloor - 2} \left| (\log (T_n) n_T)^{1/2} (U_{h,b+1} - B_{h,b+1}) \right| > \varepsilon/K \right) \\
+ \mathbb{P}\left( \max_{b=0,\ldots,\lfloor T_n/n_T \rfloor - 2} 1/\|U_{h,b+1}\| > K \right). \tag{C.1.5}
\]

Given the result on the negligibility of the drift term from Section C.1.4.1.3, we can apply Lemma C.1.4 to \(U_{h,b}\). Then, the second probability term above is equal to \(\mathbb{P}\left( \min_{b=0,\ldots,\lfloor T_n/n_T \rfloor - 2} \|U_{h,b+1}\| < 1/K \right)\) which converges to zero by letting \(K = 4/\sigma^4\).

As for the first probability term, we use (C.1.4) and choose \(r > 0\) sufficiently large to deduce that,

\[
\mathbb{P}\left( \max_{b=0,\ldots,\lfloor T_n/n_T \rfloor - 2} \left| (\log (T_n) n_T)^{1/2} (U_{h,b+1} - B_{h,b+1}) \right| > \varepsilon/K \right) \\
\leq \left( \frac{K}{\varepsilon} \right)^r \sum_{b=0}^{\lfloor T_n/n_T \rfloor - 2} \mathbb{E} \left[ \left| (\log (T_n) n_T)^{1/2} (U_{h,b+1} - B_{h,b+1}) \right|^r \right] \\
= \left( \frac{K}{\varepsilon} \right)^r (\log (T_n))^{r/2} n_T^{r/2-1} O_{\mathbb{P}} \left( 1/T^{r/2-1} \right) \to 0,
\]

in view of Condition 2 and \(T_n = O(T)\). We can repeat the same argument for the second term of (C.1.2). Altogether, this establishes the claim of the lemma. \(\Box\)

**Lemma C.1.3.** As \(h \downarrow 0\),

\[
\max_{b=0,\ldots,\lfloor T_n/n_T \rfloor - 2} (\log (T_n) n_T)^{1/2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( T_{\psi'(T_m+(b+1)n_T+j-1)h} (\hat{\beta}) - T_{\psi'(T_m+bn_T+j-1)h} (\hat{\beta}) \right) \right|_{\mathbb{P}} \to 0,
\]
and the same result holds with $\beta^*$ in place $\tilde{\beta}$. Furthermore, as $h \downarrow 0$,

$$
\max_{b=0, \ldots, [T_n/n_T] - 2} \left( \log \left( \frac{T_n}{n_T} \right) \right)^{1/2}
$$

$$
\left| n_T^{-1} \sum_{j=1}^{n_T} \left( T_{\psi, (T_m + bn_T + j - 1)h} (\tilde{\beta}) - T_{\psi, (T_m + bn_T + j - 1)h} (\beta^*) \right) \right| \overset{P}{\to} 0.
$$

**Proof.** By definition,

$$
\left| n_T^{-1} \sum_{j=1}^{n_T} \left( T_{\psi, (T_m + (b+1)n_T + j - 1)h} (\beta^*) - T_{\psi, (T_m + bn_T + j - 1)h} (\beta^*) \right) \right|
$$

$$
= \left| \left[ n_T^{-1} \sum_{j=1}^{n_T} \left( \sum_{l=1}^{T_m + (b+1)n_T + j - 1} \frac{(\Delta h \tilde{e}_l^*)^2}{T_m + (b+1)n_T + j - 1} \right) - \sum_{l=1}^{T_m + bn_T + j - 1} \frac{(\Delta h \tilde{e}_l^*)^2}{T_m + bn_T + j - 1} \right] \right|
$$

$$
= n_T^{-1} \sum_{j=1}^{n_T} \left( \sum_{l=1}^{T_m + (b+1)n_T + j - 1} (\Delta h \tilde{e}_l^*)^2 \left( T_m + (b+1)n_T + j - 1 \right) \right)
$$

$$
\times \left( \frac{1}{T_m + (b+1)n_T + j - 1 - \frac{1}{T_m + bn_T + j - 1}} \right)
$$

$$
+ \sum_{l=T_m + bn_T + j}^{T_m + (b+1)n_T + j - 1} \frac{(\Delta h \tilde{e}_l^*)^2}{T_m + bn_T + j - 1} \right) \right|.
$$

By a law of large numbers for a sequence of i.n.d. random variables [see White (1984), Section 3.2] and the boundedness of $\{\sigma_t\}_{t \geq 0}$, we have

$$
(T_m + (b+1)n_T + j - 1)^{-1} \sum_{l=1}^{T_m + (b+1)n_T + j - 1} (\Delta h \tilde{e}_l^*)^2 = O_P(1).
$$

On the other hand, the second term is negligible because there are $n_T - 1$ summands
and $T_m = O(T)$. Altogether,

$$\left| n_T^{-1} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m+(b+1)n_T+j-1)h} (\beta^*) - L_{\psi,(T_m+b n_T+j-1)h} (\beta^*) \right) \right| \leq C \left( \mathbb{O}_P \left( \frac{n_T}{T_m} \right) + \mathbb{O}_P \left( \frac{n_T}{T_m + m_T n_T} \right) \right).$$

Thus, for any $\varepsilon > 0$,

$$\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} (\log (T_n) n_T)^{1/2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m+(b+1)n_T+j-1)h} (\beta^*) - L_{\psi,(T_m+b n_T+j-1)h} (\beta^*) \right) \right| > \varepsilon \right) \leq \sum_{b=0}^{\lfloor T_n/n_T \rfloor - 2} \mathbb{P} \left( (\log (T_n) n_T)^{1/2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m+(b+1)n_T+j-1)h} (\beta^*) - L_{\psi,(T_m+b n_T+j-1)h} (\beta^*) \right) \right| > \varepsilon \right) \leq \varepsilon^{-r} \sum_{b=0}^{\lfloor T_n/n_T \rfloor - 2} \mathbb{E} \left[ (\log (T_n) n_T)^{r/2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m+(b+1)n_T+j-1)h} (\beta^*) - L_{\psi,(T_m+b n_T+j-1)h} (\beta^*) \right) \right| \right] ^r \leq \varepsilon^{-r} C \left( \log (T_n) n_T \right)^{r/2} \mathbb{O}_P \left( \frac{n_T^{r-1} T_n^{1-r}}{T_m} \right) \to 0, \quad \text{(C.1.6)}$$

for $r > 0$ sufficiently large and in view of Condition 2 since $T_m$ is of the same order as $T_n$. By a law of large numbers for a sequence of i.n.d. random variables [see White (1984), Section 3.2] and the boundedness of $\{\sigma_t\}_{t \geq 0}$, we have

$$(T_m + (b + 1)n_T + j - 1)^{-1} \sum_{t=1}^{T_m+(b+1)n_T+j-1} \left( \Delta_n \tilde{e}_t^* \right)^2 = \mathbb{O}_P (1).$$
On the other hand, the second term is negligible because there are $n_T - 1$ summands and $T_m = O(T)$. Altogether,

$$
\left| n_T^{-1} \sum_{j=1}^{n_T} \left( \mathcal{L}_{\psi,T_m+(b+1)n_T+j-1,h}(\beta^*) - \mathcal{L}_{\psi,T_m+bn_T+j-1,h}(\beta^*) \right) \right| \leq C \left( O_p \left( \frac{n_T}{T_m} \right) + O_p \left( \frac{n_T}{T_m + m_T n_T} \right) \right).
$$

Thus, for any $\varepsilon > 0$,

$$
P \left( \max_{b=0,\ldots,\lfloor T_n/n_T \rfloor-2} \left( \log (T_n) n_T \right)^{1/2} 
\left| n_T^{-1} \sum_{j=1}^{n_T} \left( \mathcal{L}_{\psi,T_m+(b+1)n_T+j-1,h}(\beta^*) - \mathcal{L}_{\psi,T_m+bn_T+j-1,h}(\beta^*) \right) \right| > \varepsilon \right) 
\leq \sum_{b=0}^{\lfloor T_n/n_T \rfloor-2} P \left( \left( \log (T_n) n_T \right)^{1/2} 
\left| n_T^{-1} \sum_{j=1}^{n_T} \left( \mathcal{L}_{\psi,T_m+(b+1)n_T+j-1,h}(\beta^*) - \mathcal{L}_{\psi,T_m+bn_T+j-1,h}(\beta^*) \right) \right| > \varepsilon \right) 
\leq \varepsilon^{-r} \sum_{b=0}^{\lfloor T_n/n_T \rfloor-2} 
\mathbb{E} \left[ \left( \log (T_n) n_T \right)^{r/2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( \mathcal{L}_{\psi,T_m+(b+1)n_T+j-1,h}(\beta^*) - \mathcal{L}_{\psi,T_m+bn_T+j-1,h}(\beta^*) \right) \right|^r \right] 
\leq \varepsilon^{-r} C \left( \log (T_n) n_T \right)^{r/2} O_p \left( n_T^{r-1} T_n^{1-r} \right) \rightarrow 0,
$$

for $r > 0$ sufficiently large and in view of Condition 2 since $T_m$ is of the same order
as $T_n$. For the last claim of the lemma, note that

$$
\begin{align*}
L_{\psi,(T_m+bn_T+j-1)h} (\tilde{\beta}) - L_{\psi,(T_m+bn_T+j-1)h} (\beta^*) \\
= \sum_{l=1}^{T_m+bn_T+j-1} \frac{(\Delta_h \tilde{e}_l)^2}{T_m + bn_T + j - 1} - \sum_{l=1}^{T_m+bn_T+j-1} \frac{(\Delta_h \tilde{e}_l)^2}{T_m + bn_T + j - 1} \\
= \frac{1}{T_m + bn_T + j - 1} \sum_{l=1}^{T_m+bn_T+j-1} (\beta - \beta^*)' \Delta_h \tilde{X}_l \Delta_h \tilde{X}_l' (\tilde{\beta} - \beta^*) \
- \frac{2}{T_m + bn_T + j - 1} \sum_{l=1}^{T_m+bn_T+j-1} \Delta_h \tilde{e}_l^2 (\beta - \beta^*)' \Delta_h \tilde{X}_l. \tag{C.1.8}
\end{align*}
$$

By Lemma C.1.1, $(T_m + bn_T + j - 1)^{-1} \sum_{l=1}^{T_m+bn_T+j-1} \Delta_h \tilde{X}_l \Delta_h \tilde{X}_l' = O_p (1)$. Since \(\hat{\beta}_k - \beta^* = O_p \left(1/\sqrt{T}\right)\) uniformly in $k \geq T_m$ by Assumption 3.8, the term in (C.1.8) is $O_p (T^{-1})$ whereas the term (C.1.9) is $o_p \left(T^{-1/2}\right)$ by Lemma C.1.1. Therefore, upon using Condition 2 and the same argument that led to (C.1.7) we show the last claim of the lemma. The proof of the second claim then follows from combining the result of the first and last claim. \(\square\)

**Lemma C.1.4.** Let $B_{h,b}^0 = (n_T h)^{-1} \sum_{j=1}^{n_T} \sigma_{e,(T_m+\tau+bn_T-j-1)h}^2 (\Delta_h W_{e,T_m+\tau+bn_T-j-1})^2$. For any $\varepsilon > 0$ and some constant $K > 0$, $\mathbb{P} \left( \max_{b=0,\ldots, [n_T/\tau]-2} |1/B_{h,b}^0| > K \right) \to 0$.

**Proof.** Note that

$$
\begin{align*}
\mathbb{P} \left( \max_{b=0,\ldots, [n_T/\tau]-2} \left| 1/B_{h,b}^0 \right| > K \right)
= \mathbb{P} \left( \min_{b=0,\ldots, [n_T/\tau]-2} \left| B_{h,b}^0 \right| < K^{-1} \right)
= \mathbb{P} \left( \min_{b=0,\ldots, [n_T/\tau]-2} \frac{1}{n_T h} \sum_{j=1}^{n_T} \left( \sigma_{e,(T_m+\tau+bn_T-j-1)h}^2 (\Delta_h W_{e,T_m+\tau+bn_T-j-1})^2 \right) < K^{-1} \right)
\leq \sum_{b=0}^{[T/n_T]-2} \mathbb{P} \left( \frac{1}{n_T} \sum_{j=1}^{n_T} \left( \sigma_{e,(T_m+\tau+bn_T-j-1)h}^2 (h^{-1/2} \Delta_h W_{e,T_m+\tau+bn_T-j-1})^2 \right) < K^{-1} \right).
\end{align*}
$$

With $K = 2/\sigma^2_\tau$ [with $\sigma_\tau$ defined in Assumption 3.1-(iii)], we can use Markov's
inequality to deduce, for any $r > 0,$

$$\mathbb{P}\left( \frac{1}{n_T} \sum_{j=1}^{n_T} (\sigma_e(T_{m+j+bT-1}) \sigma e(T_{m+j+bT-1}) h_{T_{m+j+bT-1}+1}^2) < \frac{\sigma_e^2}{2} \right)$$

$$\leq \mathbb{P}\left( \frac{1}{n_T} \sum_{j=1}^{n_T} (\sigma_e(T_{m+j+bT-1}) h_{T_{m+j+bT-1}+1}^2) < \frac{\sigma_e^2}{2} \right)$$

$$\leq \left( \frac{2}{\sigma^2} \right)^r n_T^{-r/2}$$

$$\mathbb{E}\left[ \left| \frac{n_T^{-1/2}}{n_T} \sum_{j=1}^{n_T} (\sigma_e(T_{m+j+bT-1}) h_{T_{m+j+bT-1}+1}^2) < K \right|^r \right].$$

From a standard central limit theorem for i.i.d. observations we have

$$\mathbb{E}\left[ \left| \frac{n_T^{-1/2}}{n_T} \sum_{j=1}^{n_T} (\sigma_e(T_{m+j+bT-1}) h_{T_{m+j+bT-1}+1}^2) < K \right|^r \right] < C_{2,r},$$

where $C_{2,r} < \infty.$ Thus, since we can choose $r$ sufficiently large we can deduce,

$$\sum_{b=0}^{T/n_T-2} \mathbb{P}\left( \frac{1}{n_T} \sum_{j=1}^{n_T} (\sigma_e(T_{m+j+bT-1}) h_{T_{m+j+bT-1}+1}^2) < K \right)$$

$$\leq C_r \left( \frac{2}{\sigma^2} \right)^r (T/n_T) n_T^{-r/2} \rightarrow 0,$$

where we have also used Condition 2. This concludes the proof. □

Next, let

$$B_{\max,h}^0 (T_n, \tau) \triangleq \max_{b=0, \ldots, [T_n/n_T]-2} \left| \frac{B_{h+b}^0 - B_{h,b}^0}{B_{h,b+1}^0} \right|$$

$$B_{\max,h}^* (T_n, \tau) \triangleq \max_{b=0, \ldots, [T_n/n_T]-2} \left| \frac{B_{h+b}^* - \mathcal{U}_{h,b}}{\mathcal{U}_{h,b+1}} \right|,$$
where \( B_{h,b}^0 = (n_T h)^{-1} \sum_{j=1}^{n_T} \sigma_e^2 (T_{m+\tau+bn_T+j-1}) h (\Delta_h W_{e,T_{m+\tau+bn_T+j-1}})^2 \) and

\[
B_{h,b}^* = n_T^{-1} \sum_{j=1}^{n_T} (\Delta_h \tilde{e}_{T_{m+\tau+bn_T+j-1}})^2.
\]

The following lemma shows that, under \( H_0 \), the difference in the in-sample losses \( L_{\psi,kh}(\beta_k) \) across adjacent blocks is negligible asymptotically.

**Lemma C.1.5.** As \( h \downarrow 0 \), \( (\log(T_n) n_T)^{1/2} (B_{\max,h}^* (T_n, \tau) - U_{\max,h} (T_n, \tau)) \xrightarrow{P} 0 \).

**Proof.** We begin with the inequality,

\[
| B_{\max,h}^* (T_n, \tau) - U_{\max,h} (T_n, \tau) | \leq \max_{b=0, \ldots, \lfloor T_n/\tau \rfloor - 2} \left( \sum_{j=1}^{n_T} (\mathcal{L}_{\psi,(T_{m+(b+1)n_T+j-1})h (\beta^*) - \mathcal{L}_{\psi,(T_{m+bn_T+j-1})h (\beta^*)} / U_{h,b} | > \varepsilon \right)
\]

For any \( \varepsilon > 0 \) and any \( K > 0 \),

\[
P \left( \max_{b=0, \ldots, \lfloor T_n/\tau \rfloor - 2} \left| \frac{(\log(T_n) n_T)^{1/2} \sum_{j=1}^{n_T} (\mathcal{L}_{\psi,(T_{m+(b+1)n_T+j-1})h (\beta^*) - \mathcal{L}_{\psi,(T_{m+bn_T+j-1})h (\beta^*)} / U_{h,b} | > \varepsilon \right) \right. \right.
\]

\[
\leq P \left( \max_{b=0, \ldots, \lfloor T_n/\tau \rfloor - 2} \left| (\log(T_n) n_T)^{1/2} \sum_{j=1}^{n_T} (\mathcal{L}_{\psi,(T_{m+(b+1)n_T+j-1})h (\beta^*) - \mathcal{L}_{\psi,(T_{m+bn_T+j-1})h (\beta^*)} / U_{h,b} | > \varepsilon / \sqrt{K} \right) \right.
\]

\[
+ P \left( \max_{b=0, \ldots, \lfloor T_n/\tau \rfloor - 2} 1 / | U_{h,b+1} | > \sqrt{K} \right).
\]

By the second result in Lemma C.1.3 the first term converges to zero. As for the second term, it was already treated in (C.1.5) with \( U_{h,b+1}^2 \) in place of \( U_{h,b+1} \), and a
similar argument can be applied to yield the same result. □

Lemma C.1.5 implies that the asymptotic behavior of the test statistics under $H_0$ is determined by the sequence of out-of-sample losses only. Next, let us define the following quantity which has the volatility shifted back by one block of time-length $n_T h$:

$$
\tilde{B}_{h,b}^0 = (n_T h)^{-1} \sum_{j=1}^{n_T} \sigma^2_{e,(T_m+\tau+(b-1)n_T-1)h} \left( \Delta_h W_{e,T_m+\tau+bn_T+j-1} \right)^2,
$$

and use it to define the statistic

$$
\tilde{B}_{\text{max},h}^0 (T_n, \tau) \triangleq \max_{b=0,\ldots, \lfloor T_n/n_T \rfloor-2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\tilde{B}_{h,b+1}^0} \right|.
$$

Our final goal is to show that $(\log (T_n) n_T)^{1/2} \left( V_{\text{max},h} (T_n, \tau) - \tilde{B}_{\text{max},h}^0 (T_n, \tau) \right)$ converges to zero in probability, where

$$
V_{\text{max},h} (T_n, \tau) \triangleq \max_{b=0,\ldots, \lfloor T_n/n_T \rfloor-2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\sigma^2_{e,(T_m+\tau+bn_T-1)h}} \right|.
$$

We deduce this result from several small lemmas. We begin by replacing $B_{\text{max},h}^* (T_n, \tau)$ by $B_{\text{max},h}^0 (T_n, \tau)$.

**Lemma C.1.6.** As $h \downarrow 0$, $(\log (T_n) n_T)^{1/2} \left( B_{\text{max},h}^* (T_n, \tau) - B_{\text{max},h}^0 (T_n, \tau) \right) \xrightarrow{p} 0$. 

Proof. We begin by using inequality (C.1.1),

\[
\left| B_{\max,h}^* (T_n, \tau) - B_{\max,h}^0 (T_n, \tau) \right|
\]

\[
= \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| B_{h,b+1}^* - B_{h,b}^* \right|/U_{h,b+1} + \max_{i=0,\ldots,\lfloor T_n/n \rfloor-2} \left| B_{h,b}^0 - B_{h,b+1}^0 - 1 \right|
\]

\[
\leq \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| B_{h,b}^* /U_{h,b+1} - 1 - \left( B_{h,b}^0 /B_{h,b+1}^0 - 1 \right) \right|
\]

\[
\leq \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| B_{h,b}^* \left( \frac{1}{U_{h,b+1}} - \frac{1}{B_{h,b+1}^0} \right) \right| + \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| B_{h,b}^* - B_{h,b}^0 \right|
\]

(C.1.12)

Consider the second term of (C.1.12). Let \( K > 0 \). For any \( \varepsilon > 0 \),

\[
\mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| \frac{\log (T_n) n_T}{n_T}^{1/2} \left( B_{h,b}^* - B_{h,b}^0 \right) \right| > \varepsilon \right)
\]

\[
\leq \mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| \frac{\log (T_n) n_T}{n_T}^{1/2} \left( B_{h,b}^* - B_{h,b}^0 \right) \right| > \varepsilon/K \right)
\]

\[
+ \mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| 1/B_{h,b+1}^0 \right| > K \right).
\]

(C.1.13)

By Lemma C.1.4, \( \mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n/n \rfloor-2} \left| 1/B_{h,b+1}^0 \right| > K \right) = o_P (1) \) if we set for instance \( K = 2/\sigma^2 \). Let us consider the first term of (C.1.13). By Itô’s formula,
Consider the first term of (C.1.14),

\[ B_{h,b}^* - B_{h,b}^0 = (n_T h)^{-1} \sum_{j=1}^{n_T} \left( \Delta_h e_{T_m + \tau + bn_T + j - 1}^* \right)^2 \]

\[ - (n_T h)^{-1} \sum_{j=1}^{n_T} \sigma_{e,(T_m + \tau + bn_T - 1)h}^2 \left( \Delta_h W_{e,(T_m + \tau + bn_T + j - 1)} \right)^2 \]

\[ = (n_T h)^{-1} \sum_{j=1}^{n_T} 2 \int_{(T_m + \tau + bn_T + j - 1)h}^{(T_m + \tau + bn_T + j - 1)h} \left( \sigma_{e,s}^2 - \sigma_{e,(T_m + \tau + bn_T - 1)h}^2 \right) ds \]

\[ + (n_T h)^{-1} \sum_{j=1}^{n_T} 2 \int_{(T_m + \tau + bn_T + j - 1)h}^{(T_m + \tau + bn_T + j - 1)h} \left( e_s - e_{(T_m + \tau + bn_T + j - 1)h} \right) \sigma_{e,s} ds \]

\[ - (W_{e,s} - W_{e,(T_m + \tau + bn_T + j - 1)h}) \sigma_{e,(T_m + \tau + bn_T - 1)h} ds \]

\[ + (n_T h)^{-1} \sum_{j=1}^{n_T} 2 \int_{(T_m + \tau + bn_T + j - 1)h}^{(T_m + \tau + bn_T + j - 1)h} \left( e_s - e_{(T_m + \tau + bn_T + j - 1)h} \right) \mu_{e,s} h^{-\vartheta} ds. \] (C.1.14)

Consider the first term of (C.1.14),

\[ \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} \right| \]

\[ \sum_{j=1}^{n_T} 2 \int_{(T_m + \tau + bn_T + j - 1)h}^{(T_m + \tau + bn_T + j - 1)h} \left( \sigma_{e,s}^2 - \sigma_{e,(T_m + \tau + bn_T - 1)h}^2 \right) ds \] (C.1.15)

\[ \leq \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} \]

\[ \times \sum_{j=1}^{n_T} 2 \int_{(T_m + \tau + bn_T + j - 1)h}^{(T_m + \tau + bn_T + j - 1)h} \left| \sigma_{e,s}^2 - \sigma_{e,(T_m + \tau + bn_T - 1)h}^2 \right| ds \]

\[ \leq \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} 2 \phi_{\sigma,N_T h,N} \cdot n_T h \]

\[ \leq C \left( \log \left( T_n \right) n_T \right)^{1/2} n_T h \rightarrow 0, \] (C.1.16)
by Condition 2. Let us now turn to the last term. We have for any integer \( r > 0 \),

\[
\mathbb{P} \left( \max_{b=0,\ldots,[T_n/n_T]-2} \left( \log (T_n) n_T \right)^{1/2} (n_T h)^{-1} \right.
\]
\[
\times \left. \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \left( e_s - e(T_m+\tau+bn_T+j-2)h \right) \mu e,s h^{-\vartheta} ds \right) > \varepsilon / (4K) \right) \right.
\]
\[
\leq \sum_{b=0}^{[T_n/n_T]-2} \mathbb{P} \left( \left( \log (T_n) n_T \right)^{1/2} (n_T h)^{-1} \right.
\]
\[
\times \left. \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \left( e_s - e(T_m+\tau+bn_T+j-2)h \right) \mu e,s h^{-\vartheta} ds \right) > \varepsilon / (4K) \right) \right)
\]
\[
\leq \left( \frac{4K}{\varepsilon} \right)^r \sum_{b=0}^{[T_n/n_T]-2}
\]
\[
\times \mathbb{E} \left[ \left( \log (T_n) n_T \right)^{1/2} (n_T h)^{-1} \right.
\]
\[
\times \left. \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \left( e_s - e(T_m+\tau+bn_T+j-2)h \right) \mu e,s h^{-\vartheta} ds \right] \right)^r \right.
\]
\[
\leq C_r \left( \frac{4K}{\varepsilon} \right)^r \sum_{b=0}^{[T_n/n_T]-2}
\]
\[
\times \left[ \left( \log (T_n) n_T \right)^{1/2} (n_T h)^{-1} \right.
\]
\[
\times \left. \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \left( \mathbb{E} \left[ \left| \left( e_s - e(T_m+\tau+bn_T+j-2)h \right) \mu e,s \right|^r h^{-\vartheta} \right] \right)^{1/r} ds \right] \right],
\]

where the last inequality follows from using Jensen’s and Minkowski’s inequalities.

By the Burkhölder-Davis-Gundy inequality, for any

\[
s \in [(T_m + \tau + bn_T + j - 2) h, (T_m + \tau + bn_T + j - 1) h],
\]

we have

\[
\mathbb{E} \left[ \left| \left( e_s - e(T_m+\tau+bn_T+j-2)h \right) \mu e,s \right|^r h^{-\vartheta} \right] \leq C_r h^{r/2-\vartheta r},
\]
and therefore since $\vartheta \in [0, 1/8)$,

$$
\mathbb{P}\left(\max_{b=0,\ldots,\lfloor T_n/n_T\rfloor-2} \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} \left| \sum_{j=1}^{\lfloor T_n/n_T\rfloor} 2 \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \left( e_s - c(T_m+\tau+bn_T+j-2) h \right) \mu_{e,s} h^{-\vartheta} ds \right| > \varepsilon/(4K) \right) 
\leq C_r \left( \frac{8K}{\varepsilon} \right)^r \sum_{b=0}^{\lfloor T_n/n_T\rfloor-2} \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} (n_T h^{1+3/8})^r 
\leq C_r \left( \frac{8K}{\varepsilon} \right)^r \sqrt{\log \left( T_n \right) T_n^{1/3+\epsilon}} \left( h^{1/24+\epsilon} \right)^r \to 0,
$$

for $r > 0$ sufficiently large and where $\epsilon > 0$ is a small real number. Next, consider the second term of \eqref{C.1.14},

\begin{align*}
(e_s - c(T_m+\tau+bn_T+j-2)h)\sigma_{e,s} - \left( W_{e,s} - W_{e,(T_m+\tau+bn_T+j-2)h} \right) \sigma_{e,(T_m+\tau+bn_T-1)h} & \tag{C.1.17} \\
= \sigma_{e,s} \int_{(T_m+\tau+bn_T+j-2)h}^{s} \sigma_{e,v} dW_{e,v} & \\
- \sigma_{e,(T_m+\tau+bn_T-1)h} \int_{(T_m+\tau+bn_T+j-2)h}^{s} dW_{e,v} & \\
+ \sigma_{e,s} \int_{(T_m+\tau+bn_T+j-2)h}^{s} \mu_{e,v} h^{-\vartheta} dv & \\
= \left( \sigma_{e,s} - \sigma_{e,(T_m+\tau+bn_T-1)h} \right) \int_{(T_m+\tau+bn_T+j-2)h}^{s} \sigma_{e,v} dW_{e,v} & \\
+ \sigma_{e,(T_m+\tau+bn_T-1)h} \int_{(T_m+\tau+bn_T+j-2)h}^{s} \left( \sigma_{e,v} - \sigma_{e,(T_m+\tau+bn_T-1)h} \right) dW_{e,v} & \\
+ \sigma_{e,s} \int_{(T_m+\tau+bn_T+j-2)h}^{s} \mu_{e,v} h^{-\vartheta} dv.
\end{align*}
For any integer \( r > 2 \),

\[
\mathbb{P} \left( \max_{b_0, \ldots, b_{\lfloor Tn / nT \rfloor - 2}} \left( \frac{1}{2} \left( \frac{1}{nT} \right)^{1/2} \left( nT - 2 \right)^{-1} \sum_{j=1}^{nT} \int_{(T_m + \tau + bn_T + j - 1)h}^{(T_m + \tau + bn_T + j - 2)h} \right) \sigma_{e,s} - \sigma_{e,(T_m + \tau + bn_T - 1)h} \right) \left. \int_{(T_m + \tau + bn_T + j - 2)h}^{s} \sigma_{e,v} dW_{e,v} dW_{e,s} \right| > \frac{\varepsilon}{12K} \right)
\]

\[
 \leq \left( \frac{\varepsilon}{12K} \right)^{-r} \sum_{b=1}^{\lfloor Tn / nT \rfloor - 2} \left( \frac{1}{2} \left( \frac{1}{nT} \right)^{1/2} \left( nT - 2 \right)^{-1} \sum_{j=1}^{nT} \int_{(T_m + \tau + bn_T + j - 1)h}^{(T_m + \tau + bn_T + j - 2)h} \right) \sigma_{e,s} - \sigma_{e,(T_m + \tau + bn_T - 1)h} \right) \left. \int_{(T_m + \tau + bn_T + j - 2)h}^{s} \sigma_{e,v} dW_{e,v} dW_{e,s} \right| .
\]

Then, by Hölder’s inequality,
\[
\mathbb{E} \left[ \left( \log \left( T_n \right) n_T \right)^{1/2} \left( n_T h \right)^{-1} \times \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-1)h}^{(T_m+\tau+bn_T+j-2)h} \left( \sigma_{e,s} - \sigma_{e,(T_m+\tau+bn_T-1)h} \right)^r \sigma_{e,v} dW_{e,v} dW_{e,s} \right]^{r/2} \leq C_r \left( \frac{\sqrt{\log \left( T_n \right)}}{h \sqrt{n_T}} \right)^r \\
\times \left( \int_{(T_m+\tau+(b+1)n_T-1)h}^{(T_m+\tau+bn_T-1)h} \sum_{j=1}^{n_T} \mathbb{E} \left[ \left( \sigma_{e,s} - \sigma_{e,(T_m+\tau+bn_T-1)h} \right)^r \sigma_{e,v} dW_{e,v} \right] \right) \times 1_{\left[ (T_m+\tau+bn_T+j-2)h, (T_m+\tau+bn_T+j-1)h \right]} (s) \right)^{2/r} ds \right)^{r/2} \\
\leq C_r \left( \frac{\sqrt{\log \left( T_n \right)}}{h \sqrt{n_T}} \right)^r \\
\times \left( \int_{(T_m+\tau+(b+1)n_T-1)h}^{(T_m+\tau+bn_T-1)h} \sum_{j=1}^{n_T} \mathbb{E} \left[ \phi_{\sigma,n_T h,N}^r \left( \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \sigma_{e,v} dW_{e,v} \right) \right] \right) \times 1_{\left[ (T_m+\tau+bn_T+j-2)h, (T_m+\tau+bn_T+j-1)h \right]} (s) \right)^{2/r} ds \right)^{r/2} \\
\leq C_r \left( \frac{\sqrt{\log \left( T_n \right)}}{h \sqrt{n_T}} \right)^r \left( \int_{(T_m+\tau+(b+1)n_T-1)h}^{(T_m+\tau+bn_T-1)h} \left( n_T h \right)^{r/2} \left( h^{r/2} ds \right)^{r/2} \right) \\
\leq C_r \left( \frac{\sqrt{\log \left( T_n \right)}}{h \sqrt{n_T}} \right)^r h^n n_T^{r} \rightarrow 0. \quad (C.1.18)
\]

The same bound holds for the second term in (C.1.17). Finally, the last term of
(C.1.17) is such that

\[
\mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n/n_T \rfloor - 2} \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} \right.
\]
\[
\times \left| \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-1)h}^{(T_m+\tau+bn_T+j-2)h} \sigma_{e,s} \int_{(T_m+\tau+bn_T+j-2)h}^{s} \mu_{e,v} h^{-\theta} \, dv \right| > \varepsilon / (12K) \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1}
\]
\[
\times \left| \sum_{b=0}^{\lfloor T_n/n_T \rfloor - 2} \mathbb{P} \left( \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} \right. \right.
\]
\[
\left. \times \left| \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-1)h}^{(T_m+\tau+bn_T+j-2)h} \sigma_{e,s} \int_{(T_m+\tau+bn_T+j-2)h}^{s} \mu_{e,v} h^{-1/8} \, dv \right| > \varepsilon / (12K) \right) \right)
\]
\[
\leq \left( \frac{12K}{\varepsilon} \right)^r \left[ \sum_{b=0}^{\lfloor T_n/n_T \rfloor - 2} \mathbb{P} \left( \left( \log \left( T_n \right) n_T \right)^{1/2} (n_T h)^{-1} \sum_{j=1}^{n_T} \right.ight.
\]
\[
\left. \times \left| \sum_{j=1}^{n_T} 2 \int_{(T_m+\tau+bn_T+j-1)h}^{(T_m+\tau+bn_T+j-2)h} \sigma_{e,s} \int_{(T_m+\tau+bn_T+j-2)h}^{s} \mu_{e,v} h^{-1/8} \, dv \right| \right)^r \right).
\]

Let

\[
\mathbb{E} \left( f_s^r \right) = 2^r \sum_{j=1}^{n_T} \mathbb{E} \left[ \sigma_{e,s}^r \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-2)h} \mu_{e,v} h^{-\theta} \, dv \right]^r \right]
\]
\[
\times 1_{[(T_m+\tau+bn_T+j-2)h, (T_m+\tau+bn_T+j-1)h]} \left( s \right)
\]
\[
\leq 2^r C_r h^{r(1-\theta)}.
\]

and observe that for any integer \( r > 1 \),

\[
\mathbb{E} \left( f_s^r \right) = 2^r \sum_{j=1}^{n_T} \mathbb{E} \left[ \sigma_{e,s}^r \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-2)h} \mu_{e,v} h^{-\theta} \, dv \right]^r \right]
\]
\[
\times 1_{[(T_m+\tau+bn_T+j-2)h, (T_m+\tau+bn_T+j-1)h]} \left( s \right)
\]
\[
\leq 2^r C_r h^{r(1-\theta)}.
\]
Therefore, the right-hand side of (C.1.19) is less than

\[
\left( \frac{12K}{\varepsilon} \right)^{r} \left( (\log (T_n) n_T)^{1/2} (n_T h)^{-1} \right)^{r} E \left[ \int_{(T_m+\tau+(b+1)n_T-1)h}^{(T_m+\tau+bn_T-1)h} f_s dW_{e,s} \right]^{r/2}
\]

\[
\leq \left( \frac{12K}{\varepsilon} \right)^{r} \sum_{b=0}^{[T_n/n_T]-2} \left( (\log (T_n) n_T)^{1/2} (n_T h)^{-1} \right)^{r} \times E \left[ \int_{(T_m+\tau+bn_T-1)h}^{(T_m+\tau+(b+1)n_T-1)h} (\mathbb{E} (f_s)^{r})^{2/r} ds \right] \times (\log (T_n) n_T)^{1/2} (n_T h)^{-1} \right)^{r} \times \left( \int_{(T_m+\tau+bn_T-1)h}^{(T_m+\tau+(b+1)n_T-1)h} h^{2(1-\vartheta)} ds \right)^{r/2}
\]

\[
\leq C_r \left( \frac{12K}{\varepsilon} \right)^{r} \sum_{b=0}^{[T_n/n_T]-2} \left( (\log (T_n) n_T)^{1/2} (n_T h)^{-1} \right)^{r} \times \left( \int_{(T_m+\tau+bn_T-1)h}^{(T_m+\tau+(b+1)n_T-1)h} h^{2(1-\vartheta)} ds \right)^{r/2}
\]

\[
\leq C_r \left( \frac{12K}{\varepsilon} \right)^{r} \left( (\log (T_n))^{1/2} \right)^{r} \left( h^{r/24-4/3} \right) \to 0,
\]

for \( r \) sufficiently large. This leads to

\[
\mathbb{P} \left( \max_{b=0,\ldots,[T_n/n_T]-2} \left| \log (T_n) n_T \right|^{1/2} \left| B_{h,b}^* - B_{h,b}^0 \right| > \varepsilon / K \right) \to 0. \quad \text{(C.1.20)}
\]

We now turn to the first term on the right hand side of (C.1.12). Choose any \( \varepsilon > 0 \) and positive \( K < \infty \), and note that

\[
\mathbb{P} \left( \max_{b=0,\ldots,[T_n/n_T]-2} \left( (\log (T_n) n_T)^{1/2} \right) \left| B_{h,b}^* - B_{h,b}^0 \right| \geq \varepsilon / K \right) \to 0.
\]

\[
\leq \mathbb{P} \left( \max_{b=0,\ldots,[T_n/n_T]-2} \left( (\log (T_n) n_T)^{1/2} \right) \left| B_{h,b}^* - B_{h,b}^0 \right| > \varepsilon / K \right) + \mathbb{P} \left( \max_{b=0,\ldots,[T_n/n_T]-2} \left| U_{h,b+1} B_{h,b+1}^0 \right| > K \right).
\]
We can manipulate the second term as follows:

\[
P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} 1/ |U_{h,b+1}B_{h,b+1}^0| > K \right)
= P \left( \min_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |U_{h,b+1}B_{h,b+1}^0| < 1/K \right)
\leq P \left( \min_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |U_{h,b+1}| < 1/\sqrt{K} \right)
+ P \left( \min_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b+1}^0| < 1/\sqrt{K} \right)
\leq P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |U_{h,b+1} - B_{h,b+1}^0| > 1/\sqrt{K} \right)
+ P \left( \min_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b+1}^0| < 1/\sqrt{K} \right).\]

The second and third term on the right-hand side of the the last inequality converge to zero in view of Lemma C.1.4. Noting that \( U_{h,b} \) coincides with \( B_{h,b}^* \), we can use the same arguments that led to (C.1.20) which shows a tighter bound since it involves \( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b+1}^* - B_{h,b+1}^0| \) multiplied by \((\log (T_n) n_T)^{1/2}\). Turning to the first term in (C.1.21), note that for any \( K_2 > 0 \),

\[
P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} (\log (T_n) n_T)^{1/2} |B_{h,b}^* (U_{h,b+1} - B_{h,b+1}^0)| > \varepsilon/K \right)
\leq P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b}^*| > K_2 \right)
+ P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} (\log (T_n) n_T)^{1/2} |U_{h,b+1} - B_{h,b+1}^0| > \varepsilon/(K \cdot K_2) \right).
\]

Noting that \( U_{h,b+1} \) coincides with \( B_{h,b+1}^* \), we can use the same arguments as in (C.1.20). The second term converges to zero by the same argument as above. Then,

\[
P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b}^*| > K_2 \right)
\leq P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b}^* - B_{h,b}^0| > K_2/2 \right)
+ P \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b}^0| > K_2/2 \right).\]
The first term was already discussed above whereas the second term converges to zero by invoking again Lemma C.1.4 together with the localization assumption [cf. Assumption 3.1-(iii)] which implies the $\sigma_{e,t}$ is bounded from above for all $t \geq 0$. □

**Lemma C.1.7.** As $h \downarrow 0$, $(\log (T_n) n_T)^{1/2} \left( B_{\text{max},h}^0 (T_n, \tau) - \tilde{B}_{\text{max},h}^0 (T_n, \tau) \right) \xrightarrow{p} 0$.

**Proof.** By simple rearrangements,

$$\left| B_{\text{max},h}^0 (T_n, \tau) - \tilde{B}_{\text{max},h}^0 (T_n, \tau) \right| \leq \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \frac{B_{h,b}^0 (\tilde{B}_{h,b+1}^0 - B_{h,b+1}^0)}{B_{h,b+1}^0 B_{h,b+1}^0}.$$  

We shall show that

$$\max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} (\log (T_n) n_T)^{1/2} \left| B_{h,b}^0 (\tilde{B}_{h,b+1}^0 - B_{h,b+1}^0) \right| = o_P (1). \quad (C.1.22)$$

By Lemma C.1.4, $\mathbb{P} \left( \min_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |\tilde{B}_{h,b+1}^0 B_{h,b+1}^0| < 1/K \right) \to 0$ for some $K > 0$; for example, set $\sqrt{K} = 2/\sigma^2$. Turning to the numerator of (C.1.22), for any $\varepsilon > 0$ and any $K > 0$,

$$\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} (\log (T_n) n_T)^{1/2} \left| B_{h,b}^0 (\tilde{B}_{h,b+1}^0 - B_{h,b+1}^0) \right| > \varepsilon \right)$$

$$\leq \mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} |B_{h,b}^0| > K \right)$$

$$+ \mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} (\log (T_n) n_T)^{1/2} |\tilde{B}_{h,b+1}^0 - B_{h,b+1}^0| > \varepsilon/K \right),$$

where the first term converges to zero by the same argument as in the last part of the proof of Lemma C.1.6. Therefore, it remains to deal with the second term for which
\[
\mathbb{P}\left(\max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} (\log (T_n) nT)^{1/2} \left| \tilde{B}_{h,b+1}^0 - B_{h,b+1}^0 \right| > \varepsilon/K \right)
\]

\[
= \mathbb{P}\left(\max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} \frac{(\log (T_n) nT)^{1/2}}{nT} \right)
\]

\[
\times \left| \sigma_{e,T_m + \tau + bnT - 1}^2 - \sigma_{e,T_m + \tau + (b+1)nT - 1}^2 \right|
\]

\[
\times \sum_{j=1}^{nT} \left( \Delta_h W_{e,T_m + \tau + (b+1)nT + j - 1} \right)^2 > \varepsilon/K \right)
\]

\[
\leq \mathbb{P}\left(\max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} (\log (T_n) nT)^{1/2}
\right)
\]

\[
\left| \sigma_{e,T_m + \tau + bnT - 1}^2 - \sigma_{e,T_m + \tau + (b+1)nT - 1}^2 \right| > \varepsilon/2K \right)
\]

\[
+ \mathbb{P}\left(\max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} \left| (nT)^{-1} \sum_{j=1}^{nT} \left( \Delta_h W_{e,T_m + \tau + (b+1)nT + j - 1} \right)^2 \right| > 2 \right).
\]

By Assumption 3.2, Markov’s inequality and sufficiently large \( r > 0 \),

\[
\mathbb{P}\left(\max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} (\log (T_n) nT)^{1/2}
\right)
\]

\[
\left| \sigma_{e,T_m + \tau + bnT - 1}^2 - \sigma_{e,T_m + \tau + (b+1)nT - 1}^2 \right| > \varepsilon/2K \right)
\]

\[
\leq C_r \left( \frac{2K}{\varepsilon} \right)^r (\log (T_n) nT)^{r/2} (T_n/nT) \phi_{\sigma, nT, h, N} \to 0. \tag{C.1.24}
\]
Finally, for all integers $r > 0$,

\[
\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T \sigma T / n \rfloor - 2} (n_T h)^{-1} \sum_{j=1}^{n_T} \left( \Delta_h W_{e,T_m + \tau + (b+1)\sigma T, j-1} \right)^2 > 2 \right) > 0 \tag{C.1.25}
\]

\[
\leq \sum_{b=0}^{\lfloor T \sigma T / n \rfloor - 2} \mathbb{P} \left( \left| \sum_{j=1}^{n_T} \left( (h^{-1/2} \Delta_h W_{e,T_m + \tau + (b+1)\sigma T, j-1} \right)^2 - 1 \right|^r > 1 \right) \tag{C.1.26}
\]

\[
\leq \sum_{b=0}^{\lfloor T \sigma T / n \rfloor - 2} \mathbb{E} \left( \left| \sum_{j=1}^{n_T} \left( (h^{-1/2} \Delta_h W_{e,T_m + \tau + (b+1)\sigma T, j-1} \right)^2 - 1 \right|^r \right) \tag{C.1.27}
\]

\[
\leq C_r \left( \frac{\sigma T}{n} \right)^{r/2} = C_r \sigma T \sigma T^{-r/2} \rightarrow 0,
\]

in view of Condition 2 by choosing $r$ sufficiently large. Using this together with (C.1.24) into (C.1.23) we deduce (C.1.22). □

**Lemma C.1.8.** As $h \downarrow 0$, $(\log (T \sigma T / n))^{1/2} \left( V_{\max,h} (T \sigma T, \tau) - \tilde{B}_{\max,h} (T \sigma T, \tau) \right) \xrightarrow{\mathbb{P}} 0$.

**Proof.** Note that

\[
V_{\max,h} (T \sigma T, \tau) - \tilde{B}_{\max,h} (T \sigma T, \tau)
\]

\[
= \max_{b=0, \ldots, \lfloor T \sigma T / n \rfloor - 2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\sigma^2 (T_m + \tau + b-n \sigma T - 1) h} \right| - \max_{b=0, \ldots, \lfloor T \sigma T / n \rfloor - 2} \left| \frac{B_{h,b}^0 - \tilde{B}_{h,b+1}^0}{\sigma^2 (T_m + \tau + b-n \sigma T - 1) h} \right|,
\]

and thus we show

\[
(\log (T \sigma T / n))^{1/2} \max_{b=0, \ldots, \lfloor T \sigma T / n \rfloor - 2} \left| \frac{\left( B_{h,b+1}^0 - \tilde{B}_{h,b}^0 \right) \left( \tilde{B}_{h,b+1}^0 - \sigma^2 (T_m + \tau + b-n \sigma T - 1) h \right)}{\sigma^2 (T_m + \tau + b-n \sigma T - 1) h} \right| \xrightarrow{\mathbb{P}} \sigma^2 (1).
\]

By the boundedness of $\sigma_{e,t}$, $t \geq 0$, and upon using the same arguments as in the
previous lemmas for $\tilde{B}_{h,b}^0$, the denominator is $O_p(1)$. Since for any $\varepsilon > 0$,

\[
\mathbb{P} \left( \left( \log \left( T_n \right) n_T \right)^{1/4} \max_{b=0, \ldots, \left[ T_n/n_T \right]-2} \left| \tilde{B}_{h,b+1}^0 - B_{h,b}^0 \right| > \sqrt{\varepsilon} \right)
\]

\[
\leq \mathbb{P} \left( \max_{b=0, \ldots, \left[ T_n/n_T \right]-2} \left( \log \left( T_n \right) n_T \right)^{1/4} \frac{n_T}{n_T h} \sigma_{\tau(T_m+\tau+bn_T-1)h}^2 \right.
\]

\[
\times \left| \sum_{j=1}^{n_T} \left( (\Delta_h W_{e,T_m+\tau+bn_T+j-1})^2 - (\Delta_h W_{e,T_m+\tau+(b+1)nt+j-1})^2 \right) \right| > \sqrt{\varepsilon}/2 \right)
\]

\[
\leq \mathbb{P} \left( \max_{b=0, \ldots, \left[ T_n/n_T \right]-2} \left( \log \left( T_n \right) n_T \right)^{1/4} \frac{n_T}{n_T h} \right.
\]

\[
\sigma_{\tau(T_m+\tau+bn_T-1)h}^2 \left| \sum_{j=1}^{n_T} (\Delta_h W_{e,T_m+\tau+bn_T+j-1})^2 - 1 \right| > \sqrt{\varepsilon}/2 \right)
\]

\[
+ \mathbb{P} \left( \max_{b=0, \ldots, \left[ T_n/n_T \right]-2} \left( \log \left( T_n \right) n_T \right)^{1/4} \frac{n_T}{n_T h} \sigma_{\tau(T_m+\tau+bn_T-1)h}^2 \right.
\]

\[
\left| \sum_{j=1}^{n_T} (\Delta_h W_{e,T_m+\tau+(b+1)nh+j-1})^2 - 1 \right| > \sqrt{\varepsilon}/2 \right)
\]

We consider the first probability term; the argument for the second is analogous. By using a similar argument as in (C.1.25), the first term is less than

\[
\sum_{b=0}^{\left[ T_n/n_T \right]-2} \mathbb{P} \left( \left( \log \left( T_n \right) n_T \right)^{1/4} \sigma_{\tau(T_m+\tau+bn_T-1)h}^2 \right.
\]

\[
\left| \sum_{j=1}^{n_T} \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+(b+1)nt+j-1} \right)^2 - 1 \right| > \sqrt{\varepsilon}/2 \right)
\]

\[
\leq C_r \left( \frac{2}{\sqrt{\varepsilon}} \right)^r (\log \left( T_n \right) n_T)^{r/4}
\]

\[
\sum_{b=0}^{\left[ T_n/n_T \right]-2} \mathbb{E} \left( \left| \sum_{j=1}^{n_T} \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+(b+1)nt+j-1} \right)^2 - 1 \right|^r \right)
\]

\[
= \left( \frac{2}{\sqrt{\varepsilon}} \right)^r (\log \left( T_n \right))^{r/4} (T_n/n_T)^{-r/4}
\]
which goes to zero by choosing $r > 0$ sufficiently large. It remains to show that
\[
\mathbb{P} \left( (\log (T_n) n_T)^{1/4} \max_{b=0, \ldots, [T_n/n_T]-2} \left| \hat{B}_{h,b+1}^0 - \sigma_e^2 (T_{m+\tau+bn_T-1}) h \right| > \varepsilon^{1/2} \right) \to 0. \tag{C.1.27}
\]

Simple manipulations yield for some $C < \infty$, with $\sigma_+ \leq \sqrt{C}$,
\[
\mathbb{P} \left( (\log (T_n) n_T)^{1/4} \max_{b=0, \ldots, [T_n/n_T]-2} \left| \hat{B}_{h,b+1}^0 - \sigma_e^2 (T_{m+\tau+bn_T-1}) h \right| > \varepsilon^{1/2} \right)
\leq C^r \left( \frac{1}{\sqrt{\varepsilon}} \right)^r (\log (T_n) n_T)^{r/4} \sum_{b=0}^{[T_n/n_T]-2} \left| \left( \frac{1}{n_T} \sum_{j=1}^{n_T} \left( h^{-1/2} \Delta_h W_{e,T_{m+\tau+(b+1)n_T+j-1}} \right)^2 - 1 \right) \right| > \sqrt{\varepsilon}
\leq C^r \left( \frac{2}{\sqrt{\varepsilon}} \right)^r (\log (T_n))^r (T_n/n_T)^{-r/4} \to 0.
\]

We have (C.1.27) and thus (C.1.26), which concludes the proof. \(\square\)

From Lemma C.1.2-C.1.8 we deduce
\[
\sqrt{\log (T_n) n_T} \left( B_{\max,h} (T_n, \tau) - V_{\max,h} (T_n, \tau) \right) = o_\mathbb{P} (1),
\]
where $V_{\max,h} (T_n, \tau)$ was defined in (C.1.11). By the properties of the Wiener process, for each block $b$ the variables $(\Delta_h W_{e,T_{m+\tau+bn_T+j-1}})^2$ are a sequence of $\chi^2_1$ random variables which are independent over $j$. After centering these variables, we can apply the results in Lemma 1-2 in Wu and Zhao (2007). This leads us to a limit theorem for the statistic $V_{\max,h} (T_n, \tau)$ which takes a similar form to the statistic in equation (13) of Wu and Zhao (2007). Therefore, in Lemma C.1.10 we provide a limit theorem which
adapted Theorem 1 of Wu and Zhao (2007) to our context. The difference hinges on (i) the dependence structure of the variables \( \{ (\Delta_h W_{e,T_{m+n+\tau+b_n T+j-1}})^2 - 1 \}_{i=1}^{n_T} \) relative to the sequence \( \{ X_k \}_{k \geq 1} \) appearing in Wu and Zhao (2007), and on (ii) the form of our test statistics which allow both for additive and multiplicative structure. For the quadratic loss case, our problem is then similar to that of Bibinger et al. (2017) who also uses Lemma 1-2 in Wu and Zhao (2007); yet even in the quadratic loss case our context differs from that of Bibinger et al. (2017) because we allow for model misspecification via the additional term \( \mu_{e,t} \) in (3.2.3) and estimation of \( \beta_k \).

Assumption C.2. The sequence of rescaled forecasts errors \( \{ \Delta_h \tilde{e}_k^* \}_{k \geq 1} \) satisfies, for some \( p \geq 4 \), \( \mathbb{E} [ |\Delta_h \tilde{e}_k^*|^p ] < \infty \) for all \( k \geq 1 \). Furthermore, the sequence of forecast losses \( \{ L_{\psi,kh} \}_{k \geq 1} \) satisfies the same assumption.

We now explain how to verify Assumption C.2.

Lemma C.1.9. Given the model in (3.2.3), Assumption C.2 holds.

Proof. We know that \( \Delta_h e_k^* = \int_{(k-1)h}^{kh} \mu_{e,s} h^{-\theta} ds + \int_{(k-1)h}^{kh} \sigma_{e,s} dW_{e,s} \). Note that conditional on \( \{ \mu_{e,t} \}_{t \geq 0} \) and \( \{ \sigma_{e,t} \}_{t \geq 0} \),

\[
(\Delta_h e_k^*)^2 = \left( \int_{(k-1)h}^{kh} \mu_{e,s} h^{-\theta} ds \right)^2 + \left( \int_{(k-1)h}^{kh} \sigma_{e,s} dW_{e,s} \right)^2 + 2 \int_{(k-1)h}^{kh} \int_{(k-1)h}^{kh} \mu_{e,s} h^{-\theta} \sigma_{e,v} dS_{e,s} \\
= O\left(h^2(1-\theta)\right) + \left( \int_{(k-1)h}^{kh} \sigma_{e,s} dW_{e,s} \right)^2 + O_P\left(h^{3/2-\theta}\right) \\
= o(h) + \left( \int_{(k-1)h}^{kh} \sigma_{e,s} dW_{e,s} \right)^2 + o_P\left(h^{3/2}\right). \tag{C.1.28}
\]

Hence, \( \mathbb{E} [ |\Delta_h e_k^*|^p | \mathcal{F}_{(k-1)h} ] = \mathbb{E} [ |\int_{(k-1)h}^{kh} \sigma_{e,s} dW_{e,s}|^p | \mathcal{F}_{(k-1)h} ] + C_p \mathbb{O}_P\left(h^{p/2}\right) \) and Assumption C.2 is verified given the properties of the Wiener process and \( \psi_h = h^{1/2} \).
Lemma C.1.10. For \( n = 1, \ldots, T_n \), let \( \mu_n = \mu (n_T/T_n) \) with \( \mu \in \text{Lip} ([0, 1]) \). Let \( \{ U_n \}_{n \geq 1} \) denote a sequence of i.n.d. random variables with \( U_n = \mu_n + \tilde{U}_n \), \( \mathbb{E} (\tilde{U}_n) = 0 \), \( \text{Var} (\tilde{U}_n) = \sigma^2_U \) and \( \mathbb{E} [\mid \tilde{U}_n \mid^p] < \infty \) for some \( p \geq 4 \). Set \( m_T = \lfloor T_n/n_T \rfloor \) and define

\[
B_{\text{max}, T_n} \triangleq \frac{1}{n_T} \max_{0 \leq b \leq T_n/n_T - 2} \left\{ \sum_{j=1}^{n_T} (U_{(b+1)n_T+j} - U_{bn_T+j}) \right\},
\]

and

\[
MB_{\text{max}, T_n} \triangleq \frac{1}{n_T} \max_{n_T \leq i \leq T_n - n_T} \left\{ \sum_{j=i+1}^{n_T+i} U_j - \sum_{j=i-n_T+1}^{i} U_j \right\}.
\]

If the following condition holds,

\[
n_T^{-p/2} T_n = o \left( (\log (T_n))^{-p/2} \right), \tag{C.1.29}
\]

then

\[
\sqrt{\log (m_T)} \left( \frac{\sqrt{n_T}}{\sigma_U} B_{\text{max}, T_n} - \gamma_{m_T} \right) \Rightarrow \mathcal{N}, \tag{C.1.30}
\]

and

\[
\sqrt{\log (m_T)} \left( \frac{\sqrt{n_T}}{\sigma_U} MB_{\text{max}, T_n} - 2 \log (m_T) - \frac{1}{2} \log \log (m_T) - \log 3 \right) \Rightarrow \mathcal{N}, \tag{C.1.31}
\]

where \( \gamma_{m_T} = [4 \log (m_T) - 2 \log (\log (m_T))]^{1/2} \) and

\[
\mathbb{P} (\mathcal{N} \leq v) = \exp \left( -\pi^{-1/2} \exp (-v) \right).
\]
Proof. Without loss of generality, we set $\sigma_{\tilde{U}} = 1$. By the Donsker-Prokhorov invariance principle $T_n^{-1/2} \sum_{j=1}^{\lfloor sT_n \rfloor} \tilde{U}_j \Rightarrow \mathbb{B}(s)$, where $\{\mathbb{B}(s)\}_{s \in [0,1]}$ is a standard Wiener process on $[0, 1]$. Then, we have by definition that

$$Z_{b+1} \triangleq n_T^{-1/2} (\mathbb{B}((b+1)n_T) - \mathbb{B}(bn_T)),$$

$b = 0, \ldots, m_T - 1$, are i.i.d. standard normal random variables. We have the decomposition

$$n_T^{-1} \sum_{j=1}^{n_T} U_{(b+1)n_T+j} = \frac{Z_{b+1}}{\sqrt{n_T}} + \frac{1}{n_T} \sum_{j=1}^{n_T} \mu((b+1)n_T+j) + \frac{R_{b+1,n_T}}{n_T},$$

where $R_{b,n_T} \triangleq \sum_{j=1}^{bn_T} \tilde{U}_j - \mathbb{B}(bn_T) - \left(\sum_{j=1}^{(b-1)n_T} \tilde{U}_j - \mathbb{B}((b-1)n_T)\right)$ and recall $\tilde{U}_j = U_j - \mu_j$. By the strong invariance principle of Komlós et al. (1975),

$$\max_{b \leq m_T-1} |R_{b+1,n_T}| = o_{a.s.} \left(T_n^{1/p}\right),$$

where we have used the independence structure of $\{\tilde{U}_j\}$. Since $\mu \in \textbf{Lip}([0, 1])$, we have uniformly over $b$ and $j$, $n_T^{-1} \sum_{j=1}^{n_T} (\mu((b+1)n_T+j) - \mu_{bn_T+j}) = O(n_T/T_n)$. Altogether,

$$n_T^{-1} \sum_{j=1}^{n_T} (U_{(b+1)n_T+j} - U_{bn_T+j}) = Z_{b+1} - Z_b + O_{a.s.} \left(n_T^{3/2}/T_n + n_T^{-1/2}T_n^{1/p}\right) = Z_{b+1} - Z_b + o_{a.s.} \left((\log (m_T))^{-1/2}\right).$$

The result in equation (C.1.30) then follows from Lemma 1 in Wu and Zhao (2007).

We now turn to the corresponding result for the overlapping case. We redefine
\{Z_j\}_{j \geq 1}$ as being a sequence of standard normal random variables. Then,

$$
\max_{n_T \leq i \leq T_n - n_T} \left| \sum_{j=i+1}^{n_T+i} (U_j - Z_j) - \sum_{j=n_T-i+1}^{i} (U_j - Z_j) \right|
$$

$$
= \max_{n_T \leq i \leq T_n - n_T} \left| \sum_{j=i+1}^{n_T+i} (U_j - \mu_j - Z_j) - \sum_{j=n_T-i+1}^{i} (U_j - \mu_j - Z_j) + \sum_{j=i+1}^{n_T+i} \mu_j - \sum_{j=n_T-i+1}^{i} \mu_j \right|
$$

$$
\leq 4 \max_{n_T \leq i \leq T_n - n_T} \sum_{j=1}^{i} (\bar{U}_j - Z_j) + \max_{n_T \leq i \leq T_n - n_T} \sum_{j=i+1}^{n_T+i} \mu_j - \sum_{j=n_T-i+1}^{i} \mu_j
$$

$$
= 4 \max_{n_T \leq i \leq T_n - n_T} \left| \sum_{j=1}^{i} (\bar{U}_j - Z_j) + O \left( \frac{n^2_T}{T_n} \right) \right|
$$

where the last equality follows from \( \mu \in \text{Lip}([0, 1]) \) and \( O \left( \frac{n^2_T}{T_n} \right) \) being uniform.

Next, we use Theorem 4 of Komlós et al. (1976) to derive a bound on the approximation error for the first term above. Let \( \{a_{T_n}\}_{T_n \in \mathbb{N}} \) be a positive sequence. By Markov's inequality,

$$
P \left( \max_{n_T \leq i \leq T_n} \left| \sum_{j=1}^{i} (\bar{U}_j - Z_j) \right| \geq a_{T_n} \right) \leq C_{1,p} \frac{1}{a_{T_n}} \sum_{j=1}^{T_n} \mathbb{E} \left( |\bar{U}_j|^p \right) \leq C_{2,p} \frac{T_n}{a_{T_n}},
$$

where \( C_{1,p}, C_{2,p} < \infty \). The conditions of Theorem 4 in Komlós et al. (1976) are satisfied if we set \( a_{T_n} = \sqrt{n_T \log(T_n)} \). This leads to

$$
\max_{n_T \leq i \leq T_n - n_T} \left| \sum_{j=i+1}^{n_T+i} (\bar{U}_j - Z_j) - \sum_{j=n_T-i+1}^{i} (\bar{U}_j - Z_j) \right| = o_P \left( \sqrt{n_T \log(T_n)}^{-1/2} \right),
$$

(C.1.32)

where we have used (C.1.29). Let \( B(i) = \sum_{j=1}^{i} Z_j \) and define

$$
H(u) \triangleq (1 (0 \leq u < 1) - 1 (-1 < u < 0)) / \sqrt{2}.
$$
Use (C.1.32) to deduce that,

\[
\sqrt{n_T MB_{\max, T_n}} \leq \sqrt{\frac{\max_{n T \leq i \leq T_n - n_T} \left| \sum_{j=i}^{n_T+i} \tilde{U}_j - \sum_{j=i-n_T+1}^{i} \tilde{U}_j \right| + O \left( \frac{n_T^{3/2}}{T_n} \right)}{\sqrt{2}}}
\]

Therefore, letting \(\varphi_n = \sup \{|B(u) - B(u')|: u, u' \in [0, T_n], |u - u'| \leq 1\}\), we have

\[
\sqrt{n_T MB_{\max, T_n}} \leq \sqrt{\frac{\max_{n T \leq i \leq T_n - n_T} \left| B(i + n_T) - B(i) - (B(i) - B(i - n_T)) \right| + O \left( \frac{\varphi_n}{\sqrt{n_T}} \right)}{\sqrt{2}}}
\]

By the global modulus of continuity of the standard Wiener process [cf. Theorem 2.9.25 in Karatzas and Shreve (1996)], we know that \(\varphi_n = o_P \left( \log (T_n) \right)^{1/2} \). The result for the overlapping case then follows from Lemma 2 in Wu and Zhao (2007) with \(\alpha = 1\), \(D_{H,1} = 3\), bandwidth \(b_n = m_T^{-1}\) and \(n = T_n\); see their Definition 1 as well and note that their lemma can be applied because \((\log (T_n))^6 = o(n_T)\) holds by condition (C.1.29). □

**Proof of Theorem 3.3.1-(i).** From Lemma

\[
\sqrt{\log (T_n)} n_T \left( B_{\max, h} (T_n, \tau) - V_{\max, h} (T_n, \tau) \right) = o_P(1).
\]

Lemma C.1.9 shows that Assumption C.2 holds. Then, under Condition 2, we can apply Lemma C.1.10 to \(V_{\max, h} (T_n, \tau)\) which in turn leads to the result for \(B_{\max, h} (T_n, \tau)\) in part (i) of Theorem 3.3.1. □
C.1.4.1.2 Proof of part (ii) of Theorem 3.3.1  The proof can be simplified considerably by using arguments similar to those of part (i) of Theorem 3.3.1. Let

\[
MB^*_{\text{max}, h}(T_n, \tau) = \max_{i = n\tau, \ldots, T_n - n\tau} \left| \frac{n_T^{-1} \sum_{j=i-n\tau+1}^{i} \left( \Delta_h \tilde{e}_T^* \right)^2}{n_T^{-1} \sum_{j=i+1}^{i+n\tau} \left( \Delta_h \tilde{e}_T^* \right)^2} - 1 \right| , \tag{C.1.33}
\]

and

\[
MB^0_{\text{max}, h}(T_n, \tau) = \max_{i = n\tau, \ldots, T_n - n\tau} \left| \frac{n_T^{-1} \sum_{j=i-n\tau+1}^{i} \sigma_e^2(T_m+\tau+i-n\tau-1)h \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+j-1} \right)^2}{n_T^{-1} \sum_{j=i+1}^{i+n\tau} \sigma_e^2(T_m+\tau+i-1)h \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+j-1} \right)^2} - 1 \right| . \tag{C.1.34}
\]

Lemma C.1.11. \( \sqrt{\log(T_n)} n_T \left( MB_{\text{max}, h}(T_n, \tau) - MB^0_{\text{max}, h}(T_n, \tau) \right) \overset{p}{\to} 0. \)

Proof. Note that the choice of overlapping blocks does not alter the results of Lemma C.1.2-C.1.5, which in turn give

\[
(\log(T_n) n_T)^{1/2} \left( MB_{\text{max}, h}(T_n, \tau) - MB^*_{\text{max}, h}(T_n, \tau) \right) \overset{p}{\to} 0.
\]

Thus, we can begin by proving a result analogous to Lemma C.1.6:

\[
(\log(T_n) n_T)^{1/2} \left( MB^*_{\text{max}, h}(T_n, \tau) - MB^0_{\text{max}, h}(T_n, \tau) \right) \overset{p}{\to} 0.
\]
Note that proceeding as in (C.1.12), we have

\[
\left| MB_{\text{max},h}^* (T_n, \tau) - MB_{\text{max},h}^0 (T_n, \tau) \right|
\leq \max_{i=n_T, \ldots, T_n-n_T} \left| n_T^{-1} \sum_{j=i-n_T+1}^{i} (\Delta h \tilde{e}_{T_m+\tau+j-1}^x)^2 \right|
\times \left( \frac{1}{n_T^{-1} \sum_{j=i+1}^{i+n_T} (\Delta h \tilde{e}_{T_m+\tau+j-1}^r)^2} - \frac{1}{\sigma_{e,(T_m+\tau+i-1)h}^2 (h^{-1/2} \Delta h W_{e,T_m+\tau+j-1})^2} \right)
\]
\]

\[
\left| \frac{n_T^{-1} \sum_{j=i-n_T+1}^{i} (\Delta h \tilde{e}_{T_m+\tau+j-1}^r)^2}{n_T^{-1} \sum_{j=i+1}^{i+n_T} \sigma_{e,(T_m+\tau+i-1)h}^2 (h^{-1/2} \Delta h W_{e,T_m+\tau+j-1})^2} \right|
\]
\[
\left| \frac{n_T^{-1} \sum_{j=i-n_T+1}^{i} \sigma_{e,(T_m+\tau+i-1)h}^2 (h^{-1/2} \Delta h W_{e,T_m+\tau+j-1})^2}{n_T^{-1} \sum_{j=i+1}^{i+n_T} \sigma_{e,(T_m+\tau+i-1)h}^2 (h^{-1/2} \Delta h W_{e,T_m+\tau+j-1})^2} \right|
\].
Then, we can use the same decomposition as in (C.1.13),

\[
\mathbb{P} \left( \max_{i=nT, \ldots, T-nT} (\log (T_n) n_T)^{1/2} \times \left| \frac{n_T^{-1} \sum_{j=i-nT+1}^{i} \left( \left( \Delta_h \hat{e}_{i,T_{m+\tau+j-1}}^* \right)^2 \right)}{n_T^{-1} \sum_{j=i+1}^{i+nT} \sigma^2_{e,T_{m+\tau+i-1} h} \left( h^{-1/2} \Delta_h W_{\nu,T_{m+\tau+j-1}} \right)^2} \right| > \varepsilon \right) > \varepsilon/K \right)
\]

\[
\leq \mathbb{P} \left( \max_{i=nT, \ldots, T-nT} (\log (T_n) n_T)^{1/2} \left| n_T^{-1} \sum_{j=i-nT+1}^{i} \left( \left( \Delta_h \hat{e}_{i,T_{m+\tau+j-1}}^* \right)^2 \right) - \sigma^2_{e,T_{m+\tau+i-nT-1} h} \left( h^{-1/2} \Delta_h W_{\nu,T_{m+\tau+j-1}} \right)^2 \right| > \varepsilon/K \right)
\]

\[
+ \mathbb{P} \left( \min_{i=nT, \ldots, T-nT} \left| n_T^{-1} \sum_{j=i+1}^{i+nT} \sigma^2_{e,T_{m+\tau+i-1} h} \left( h^{-1/2} \Delta_h W_{\nu,T_{m+\tau+j-1}} \right)^2 \right| < 1/K \right),
\]

which holds for any \( \varepsilon > 0 \) and any constant \( K > 0 \). Using the same reasoning as in the proof involving the second term of (C.1.12) and choosing \( K \) appropriately, we have for the second term,

\[
\mathbb{P} \left( \max_{i=nT, \ldots, T-nT} \left| n_T^{-1} \sum_{j=i+1}^{i+nT} \sigma^2_{e,T_{m+\tau+i-1} h} \left( h^{-1/2} \Delta_h W_{\nu,T_{m+\tau+j-1}} \right)^2 \right| > K \right) \to 0.
\]

Thus, it remains to consider the first term on the right-hand side above. For the non-overlapping case it was treated in (C.1.14) and its final bound can be obtained from (C.1.16)-(C.1.20). However, for the overlapping block case, the maximum is over a larger number of arguments. Indeed, the final bound is an order \( O(n_T) \) larger than the one for the non-overlapping case. Nonetheless, the same conclusion holds.
upon choosing $r$ large enough there:

$$\mathbb{P}\left(\max_{i=nT-1}^{nT} (\log (T_n) n_T)^{1/2} \left| n_T^{-1} \sum_{j=i-n_T+1}^{i} (\Delta_h \tilde{e}_{T_m+\tau+j-1})^2 \right. \right. - \sigma^2 \epsilon_{(T_m+\tau+i-n_T-1)h} \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+j-1} \right)^2 > \varepsilon/K) \right) \rightarrow 0.$$  

(C.1.35)

Generalizing the arguments that led to (C.1.35) and noting that the bounds involving the Lipschitz continuity of $\{\sigma_{e,t}\}_{t \geq 0}$ remain the same as in the non-overlapping case, the corresponding results in Lemma C.1.6-C.1.8 can be verified. This together with Lemma C.1.2-C.1.5—which are valid for both cases with minor changes in notation—yield the conclusion of the lemma. □

Proof of Theorem 3.3.1-(ii). From Lemma C.1.11,

$$\sqrt{\log (T_n) n_T} \left( MB_{\max,h} (T_n, \tau) - MB^0_{\max,h} (T_n, \tau) \right) = o_p (1).$$

For the non-overlapping case, Lemma C.1.9 shows that Assumption C.2 is satisfied. Given Condition 2, Lemma C.1.10 [cf. the result pertaining to $MB_{\max,T_n}$ there] applied to $MB^0_{\max,h} (T_n, \tau)$ gives part (ii) of the theorem. □

C.1.4.1.3 Negligibility of the $\mu_{e,t}$ term The negligibility of the drift term can be proven by using similar arguments to those in Section A.3.3 in Casini and Perron
From the decomposition in (C.1.28) we have for any \( b = 0, \ldots, \left\lfloor T_n/n_T \right\rfloor - 2, \)

\[
\sum_{j=1}^{n_T} \left( \Delta_h c_{T_m + \tau + bn_T + j - 1}^e \right)^2
\]

\[
= \sum_{j=1}^{n_T} \left( \int_{(T_m + \tau + bn_T + j - 2)h}^{(T_m + \tau + bn_T + j - 1)h} \mu_{e,s} h^{-\alpha} ds \right)^2 + \sum_{j=1}^{n_T} \left( \int_{(T_m + \tau + bn_T + j - 2)h}^{(T_m + \tau + bn_T + j - 1)h} \sigma_{e,s} dW_{e,s} \right)^2
\]

\[
+ 2 \sum_{j=1}^{n_T} \int_{(T_m + \tau + bn_T + j - 2)h}^{(T_m + \tau + bn_T + j - 1)h} \int_{(T_m + \tau + bn_T + j - 2)h}^{(T_m + \tau + bn_T + j - 1)h} \mu_{e,v} h^{-\alpha} \sigma_{e,s} dW_{e,s}
\]

\[
= o\left( n_T h^{2(1-\theta)} \right) + \sum_{j=1}^{n_T} \left( \int_{(T_m + \tau + bn_T + j - 2)h}^{(T_m + \tau + bn_T + j - 1)h} \sigma_{e,s} dW_{e,s} \right)^2 + o_{\mathbb{P}}\left( n_T h^{3/2-\theta} \right), \quad (C.1.37)
\]

for small \( \epsilon > 0 \). The limit theorems involve normalizing the above sums by the factor \( \sqrt{\log (T_n) n_T / (n_T h)} = h^{-2/3-\epsilon/2} \). Then the first term is \( o\left( h^{5/12+\epsilon} \right) \). The bound can be extended to hold for the maximum over blocks \( b = 0, \ldots, \left\lfloor T_n/n_T \right\rfloor - 2 \) by using the same argument as in (C.1.19). The latter bound also applies to the third term of (C.1.37) which is even of higher order. Therefore, the results of Lemma C.1.2-C.1.5 still holds when \( \mu_{e,t} \) is not restricted to be null for all \( t \geq 0 \).

### C.1.4.2 Proof of Corollary 3.3.1

**Proof.** The proof follows easily from Lemma C.1.10 with \( \sigma_U = \nu_L \). That is, we have now \( R_{b,T_n} \triangleq \sum_{j=1}^{bn_T} \bar{U}_j - \nu_L B (bn_T) - \left( \sum_{j=1}^{(b-1)n_T} \bar{U}_j - \nu_L B ((b - 1) n_T) \right) \) which satisfies the same bound as above. Then, proceeding as above,

\[
\nu_L^{-1} n_T^{-1/2} \sum_{j=1}^{n_T} \left( U_{(b+1)n_T+j} - U_{bn_T+j} \right) = Z_{b+1} - Z_b + o_{\mathbb{P}} \left( (\log (m_T))^{-1/2} \right),
\]

and the final result for \( Q_{\max,h} \) can be deduced again from Lemma 1 in Wu and Zhao (2007). □
C.1.4.3 Proof of Theorem 3.3.2

C.1.4.3.1 Proof of part (i) of Theorem 3.3.2 Recall the notation for the normalized forecast error $\Delta_h \tilde{e}_k \triangleq \Delta_h e_k / \psi_h$ and for the normalized forecast loss $L_{\psi,T_m+\tau+bn+1} (\beta^*) = g(\Delta_h \tilde{e}_{T_m+\tau+bn+1}; \beta^*)$. We use the quantity $U_{\max,h} (T, \tau)$ as defined in the proof for the quadratic case. However, $U_{h,b}$ is now defined as $D_{h,b}$ but with $\beta^*$ in place of $\hat{\beta}$. Let

$$B_{h,b}^* = n_T^{-1} \sum_{j=1}^{n_T} g \left( \Delta_h \tilde{e}_{T_m+\tau+bn+1}; \beta^* \right).$$

We only provide the proof for the recursive forecasting scheme. As in the proof of Theorem 3.3.1, we first assume that $\mu_{e,t} = 0$ in (3.2.2) and relax such restriction in Section C.1.4.3.3. We again omit the index from $\hat{\beta}$ when it is clear from the context.

Lemma C.1.12. For any $L \in L_e$, the results of Lemma C.1.3 hold.
Proof. By definition and upon using basic manipulations,

\[
\left| n_T^{-1} \sum_{j=1}^{n_T} \left( T_{\psi(T_m+(b+1)n_T+j-1)h}^{(\beta^*)} - T_{\psi(T_m+bn_T+j-1)h}^{(\beta^*)} \right) \right|
\]

\[
= \left| n_T^{-1} \sum_{j=1}^{n_T} \left( \sum_{l=1}^{T_m+(b+1)n_T+j-1} \frac{g \left( \Delta_h \tilde{e}_l^*; \beta^* \right)}{T_m + (b + 1) n_T + j - 1} \right) \right|
\]

\[
- \sum_{l=1}^{T_m + bn_T + j-1} \frac{g \left( \Delta_h \tilde{e}_l^*; \beta^* \right)}{T_m + bn_T + j - 1}
\]

\[
= \left| n_T^{-1} \sum_{j=1}^{n_T} \left( \sum_{l=1}^{T_m+(b+1)n_T+j-1} g \left( \Delta_h \tilde{e}_l^*; \beta^* \right) \right) \right|
\]

\[
\times \left( \frac{T_m + (b + 1) n_T + j - 1}{T_m + bn_T + j - 1} - \frac{1}{T_m + bn_T + j - 1} \right)
\]

\[
+ \sum_{l=T_m+bn_T+j}^{T_m+(b+1)n_T+j-1} \frac{g \left( \Delta_h \tilde{e}_l^*; \beta^* \right)}{T_m + bn_T + j - 1}
\]

\[
= O_p \left( \frac{n_T}{T_m} \right) + O_p \left( \frac{n_T}{T_m} \right),
\]

where the latter bounds are implied by basic law of large numbers given Assumption 3.5. Then use the same arguments as in the proof of Lemma C.1.3 to yield a bound similar to (C.1.7). Finally, consider a mean-value expansion of \( g \left( \Delta_h \tilde{e}_i; \beta \right) \) around \( \beta^* \),

\[
g \left( \Delta_h \tilde{e}_i; \beta \right) = g \left( \Delta_h \tilde{e}_i^*; \beta^* \right) + \frac{\partial g \left( \Delta_h \tilde{e}_i^*; \beta^* \right)}{\partial \beta} (\beta - \beta^*)
\]

\[
+ \frac{1}{2} (\beta - \beta^*) \frac{\partial^2 g \left( \Delta_h \tilde{e}_i; \beta \right)}{\partial \beta \partial \beta'} (\beta - \beta^*),
\]
where $\overline{\beta}$ is an intermediate point between $\beta^*$ and $\hat{\beta}$. It follows that

$$L_{\psi,(T_{m_{nT}}+bT_{nT}+j-1)}h\left(\hat{\beta}\right) - L_{\psi,(T_{m_{nT}}+bT_{nT}+j-1)}h\left(\beta^*\right)$$

$$= \sum_{l=1}^{T_{m_{nT}}+bT_{nT}+j-1} \frac{g\left(\Delta_h \overline{\epsilon}_l; \hat{\beta}\right)}{T_{m_{nT}}+bT_{nT}+j-1} - \sum_{l=1}^{T_{m_{nT}}+bT_{nT}+j-1} \frac{g\left(\Delta_h \overline{\epsilon}_l^*; \beta^*\right)}{T_{m_{nT}}+bT_{nT}+j-1}$$

$$= \frac{1}{T_{m_{nT}}+bT_{nT}+j-1} \sum_{l=1}^{T_{m_{nT}}+bT_{nT}+j-1} \left(\frac{\partial g\left(\Delta_h \overline{\epsilon}_l; \beta^*\right)}{\partial \beta} \left(\hat{\beta} - \beta^*\right) + \frac{1}{2} \left(\hat{\beta} - \beta^*\right)' \frac{\partial^2 g\left(\Delta_h \overline{\epsilon}_l; \beta^*\right)}{\partial \beta \partial \beta'} \left(\hat{\beta} - \beta^*\right)\right)$$

By Assumption 3.4, $\left|\frac{\partial^2 g\left(\Delta_h \overline{\epsilon}_l; \beta^*\right)}{\partial \beta \partial \beta'}\right| < C$ and thus the second term is $O_P\left(1/T\right)$ uniformly in $l$. By Assumption 3.5, $E\left(\|\partial g\left(\Delta_h \overline{\epsilon}_l; \beta^*\right) / \partial \beta\|\right)^{2+\infty} < \infty$ uniformly in $l$. Since

$$\{\partial g\left(\Delta_h \overline{\epsilon}_l; \beta^*\right) / \partial \beta - E\left(\partial g\left(\Delta_h \overline{\epsilon}_l; \beta^*\right) / \partial \beta\right)\}_{t \geq T_m},$$

forms a martingale difference sequence we can use classical bounds on averages of m.d.s. By Assumption 3.8, $\hat{\beta} - \beta^* = O_P\left(1/\sqrt{T}\right)$ because $\hat{\beta}_l - \beta^* = O_P\left(1/\sqrt{T}\right)$ uniformly in $l \geq T_m$. Thus,

$$L_{\psi,(T_{m_{nT}}+bT_{nT}+j-1)}h\left(\hat{\beta}\right) - L_{\psi,(T_{m_{nT}}+bT_{nT}+j-1)}h\left(\beta^*\right) = O_P\left(1/\sqrt{T}\right).$$

Proceeding as in (C.1.7)-(C.1.9) one verifies,

$$\max_{b=0,...,\left[T_{nT}/nT\right]-2} \left(\log\left(T_{nT}\right) nT\right)^{1/2}$$

$$\left|nT^{-1} \sum_{j=1}^{nT} \left(\frac{\Psi_{\psi,(T_{m_{nT}}+bT_{nT}+j-1)}h\left(\hat{\beta}\right) - L_{\psi,(T_{m_{nT}}+bT_{nT}+j-1)}h\left(\beta^*\right)}{T_{m_{nT}}+bT_{nT}+j-1}\right)\right|$$

$$\overset{P}{\to} 0. \square$$
We now have a corresponding result to Lemma C.1.2.

**Lemma C.1.13.** As $h \downarrow 0$, $(\log (T_n) n_T)^{1/2} (U_{\max, h} (T_n, \tau) - G_{\max, h} (T_n, \tau)) \overset{p}{\to} 0$.

**Proof.** The same manipulations as in Lemma C.1.2 yield,

$$|U_{\max, h} (T_n, \tau) - G_{\max, h} (T_n, \tau)| \leq \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( SL_{\psi, T_m+\tau+(b+1)n_T+j-1} (\beta^*) - SL_{\psi, T_m+\tau+(b+1)n_T+j-1} (\hat{\beta}) \right) \right|,$$

$$+ \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( SL_{\psi, T_m+\tau+bn_T+j-1} (\beta^*) - SL_{\psi, T_m+\tau+bn_T+j-1} (\hat{\beta}) \right) \right|,$$

$$\leq C_1 \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( SL_{\psi, T_m+\tau+(b+1)n_T+j-1} (\beta^*) - SL_{\psi, T_m+\tau+(b+1)n_T+j-1} (\hat{\beta}) \right) \right|,$$

$$+ C_2 \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{-1} \sum_{j=1}^{n_T} \left( SL_{\psi, T_m+\tau+bn_T+j-1} (\beta^*) - SL_{\psi, T_m+\tau+bn_T+j-1} (\hat{\beta}) \right) \right|.$$

From Lemma C.1.12, for any $j = 1, \ldots, n_T$,

$$SL_{T_m+\tau+bn_T+j-1} (\beta^*) - SL_{T_m+\tau+bn_T+j-1} (\hat{\beta}) = L_{\psi, T_m+\tau+bn_T+j-1} (\beta^*) - L_{\psi, T_m+\tau+bn_T+j-1} (\hat{\beta}) + o_P \left( T^{-1/2} \right).$$
Note that,

\[ L_{\psi,T_m+\tau+bn_T+j-1}(\hat{\beta}) - L_{\psi,T_m+\tau+bn_T+j-1}(\beta^*) \]

\[ = g \left( \Delta_h \bar{e}_{T_m+\tau+bn_T+j-1}(\hat{\beta}) \right) \]

\[ - g \left( \sigma_e,(T_m+\tau+bn_T-1)h \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+bn_T+j-1} ; \beta^* \right) \right), \]

and taking a mean-value expansion of \( g \left( \Delta_h \bar{e}_{T_m+\tau+bn_T+j-1}(\hat{\beta}) \right) \) around \( \beta^* \) we have

\[ g \left( \Delta_h \bar{e}_{T_m+\tau+bn_T+j-1}(\hat{\beta}) \right) \]

\[ = g \left( \sigma_e,(T_m+\tau+bn_T-1)h \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+bn_T+j-1} ; \beta^* \right) \right) (\hat{\beta} - \beta^*) \]

\[ + \frac{1}{2} (\hat{\beta} - \beta^*) \left\| \frac{\partial^2 g \left( \Delta_h \bar{e}_{T_m+\tau+bn_T+j-1}(\hat{\beta}) \right)}{\partial \beta \partial \beta'} \right\| (\hat{\beta} - \beta^*). \]

Therefore, using the last three relationships above, Assumption 3.4-3.5 and Assumption 3.8 we have for the numerator of (C.1.38),

\[
\left| \frac{1}{n_T} \sum_{j=1}^{n_T} \left( SL_{\psi,T_m+\tau+bn_T+j-1}(\beta^*) - SL_{\psi,T_m+\tau+bn_T+j-1}(\hat{\beta}) \right) \right|
\]

\[
\leq \left| \frac{1}{n_T} \sum_{j=1}^{n_T} \left( \frac{\partial g \left( \sigma_e,(T_m+\tau+bn_T-1)h \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+bn_T+j-1} ; \beta^* \right) \right)}{\partial \beta} (\hat{\beta} - \beta^*) \right| \]

\[
+ \frac{1}{2} (\hat{\beta} - \beta^*) \left\| \frac{\partial^2 g \left( \Delta_h \bar{e}_{T_m+\tau+bn_T+j-1}(\hat{\beta}) \right)}{\partial \beta \partial \beta'} \right\| (\hat{\beta} - \beta^*) \]

\[
\leq \left\| \hat{\beta} - \beta^* \right\| \left| \frac{1}{n_T} \sum_{j=1}^{n_T} \left\| \frac{\partial g \left( \sigma_e,(T_m+\tau+bn_T-1)h \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+bn_T+j-1} ; \beta^* \right) \right)}{\partial \beta} \right\| \]

\[
+ \left\| \hat{\beta} - \beta^* \right\|^2 \left| \frac{1}{2n_T} \sum_{j=1}^{n_T} \left\| \frac{\partial^2 g \left( \Delta_h \bar{e}_{T_m+\tau+bn_T+j-1}(\hat{\beta}) \right)}{\partial \beta \partial \beta'} \right\| \right| .
\]
Since $\tilde{\beta}_k - \beta^* = O_P \left( 1/\sqrt{T} \right)$ uniformly, the first term on the right-hand side above is $CO_P \left( T^{-1/2} \right)$ by Assumption 3.5 while the second term is $O_P \left( T^{-1} \right)$ by Assumption 3.4. Both bounds are uniform in $b$. Combining the latter two results we have

$$\left| \frac{1}{nT} \sum_{j=1}^{nT} \left( SL_{\psi, T_m + r + bn_T + j-1} (\beta^*) - SL_{\psi, T_m + r + bn_T + j-1} (\tilde{\beta}) \right) \right| = KO_P \left( T^{-1/2} \right).$$

(C.1.39)

Thus, for any $\varepsilon > 0$ and any constant $K > 0$, the first term on the right-hand side of (C.1.38) is such that

$$\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} \left| \frac{\left( \log \left( T_n \right) n_T \right)^{1/2} \sum_{j=1}^{nT} \left( \bar{L}_{\psi,(T_m+(b+1)n_T+j-1)} h (\beta^*) - \bar{L}_{\psi,(T_m+bn_T+j-1)} h (\beta^*) \right)}{\sqrt{U_{h,b+1}}} \right| > \varepsilon \right) \leq \mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} \left| \frac{\left( \log \left( T_n \right) n_T \right)^{1/2} \sum_{j=1}^{nT} \left( \bar{L}_{\psi,(T_m+(b+1)n_T+j-1)} h (\beta^*) - \bar{L}_{\psi,(T_m+bn_T+j-1)} h (\beta^*) \right)}{\sqrt{U_{h,b+1}}} \right| > \varepsilon / K \right) + \mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} 1/ \sqrt{U_{h,b+1}} > K \right).$$

By Lemma C.1.16 below, $\mathbb{P} \left( \min_{b=0, \ldots, \lfloor T_n/nT \rfloor - 2} \left| U_{b+1,h} \right| < 1/K \right) \rightarrow 0$ by letting, for example, $K = 2/\sigma_{L,-}^2$, where $\sigma_{L,-}$ was introduced in that proof. As for the first probability term, by using Markov’s inequality and the relationship in (C.1.39), we
have for any \( r > 0 \),

\[
\mathbb{P}\left( \max_{b=0,\ldots,[T_n/n_T]-2} \left| (\log (T_n)) n_T \right|^{1/2} \frac{n_T^{-1} \sum_{j=1}^{n_T} (SL_{\psi,T_n+\tau+(b+1)n_T+j-1} (\beta) - SL_{\psi,T_n+\tau+(b+1)n_T+j-1} (\hat{\beta}))}{\sqrt{U_{b+1,h} D_{b+1,h}}} > \varepsilon \right) \\
\leq \mathbb{P}\left( \max_{b=0,\ldots,[T_n/n_T]-2} \left| (\log (T_n)) n_T \right|^{1/2} n_T^{-1} \sum_{j=1}^{n_T} (SL_{\psi,T_n+\tau+(b+1)n_T+j-1} (\beta) - SL_{\psi,T_n+\tau+(b+1)n_T+j-1} (\hat{\beta})) > \varepsilon / K \right) \\
+ \mathbb{P}\left( \max_{b=0,\ldots,[T_n/n_T]-2} \frac{1}{|U_{b+1,h}|} < K \right).
\]

(C.1.40)

using Condition 2. The argument for the second term of (C.1.38) is equivalent. This concludes the proof of the lemma. □

Let \( u_{T_n+\tau+bn_T+j-1} \triangleq g \left( \sigma_{e,(T_n+\tau+bn_T-1)h} h^{-1/2} (\Delta_{h,e,T_n+\tau+bn_T+j-1} \beta^\star) \right) \) and define \( B^{0}_{\max,h} (T_n, \tau) \triangleq \max_{b=0,\ldots,[T_n/n_T]-2} \left| (B^{0}_{h,b+1} - B^{0}_{h,b}) / \sqrt{D^{0}_{h,b+1}} \right| \), where

\[
B^{0}_{h,b} = n_T^{-1} \sum_{j=1}^{n_T} u_{T_n+\tau+bn_T+j-1}
\]

and

\[
D^{0}_{h,b} \triangleq n_T^{-1} \sum_{j=1}^{n_T} (u_{T_n+\tau+bn_T+j-1} - \bar{g}_b)^2,
\]

with \( \bar{g}_b \triangleq n_T^{-1} \sum_{j=1}^{n_T} u_{T_n+\tau+bn_T+j-1} \).

Similarly, define \( B^{\star}_{h,b} = n_T^{-1} \sum_{j=1}^{n_T} u_{T_n+\tau+bn_T+j-1}^\star \), where

\[
u_{T_n+\tau+bn_T+j-1}^\star \triangleq g \left( \Delta_{h,e,T_n+\tau+bn_T+j-1} \beta^\star \right).
\]

The next quantity that we define is similar to \( B^{0}_{h,b} \) but has all the parameters shifted
back by one block of time length \( n_T h \):

\[
\tilde{B}_{\max,h}^0 (T_n, \tau) \triangleq \max_{b=0,...,\lfloor T_n/n_T \rfloor - 2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\sqrt{\tilde{D}_{h,b+1}^0}} \right|
\]

where

\[
\tilde{u}_{T_m+\tau+bn_T+j-1} \triangleq g \left( \sigma_u(T_m+\tau+(b-1)n_T-1)h h^{-1/2} \Delta_h W_{e,T_m+\tau+bn_T+j-1}, \beta^* \right).
\]

With this notation we can define the statistic

\[
\tilde{B}_{\max,h}^0 (T_n, \tau) \triangleq \max_{b=0,...,\lfloor T_n/n_T \rfloor - 2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\sqrt{\tilde{D}_{h,b+1}^0}} \right|
\]

where \( \tilde{D}_{h,b}^0 \triangleq n_T^{-1} \sum_{j=1}^{n_T} \left( \tilde{u}_{T_m+\tau+bn_T+j-1} - \tilde{g}_b \right)^2 \) with \( \tilde{g}_b \triangleq n_T^{-1} \sum_{j=1}^{n_T} \tilde{u}_{T_m+\tau+bn_T+j-1} \). We want to show that

\[
P \left( (\log (T_n) n_T)^{1/2} \left( V_{\max,h}(T_n, \tau) - \tilde{B}_{\max,h}^0 (T_n, \tau) \right) > \varepsilon \right) \to 0,
\]

for any \( \varepsilon > 0 \), where

\[
V_{\max,h}(T_n, \tau) \triangleq \max_{b=0,...,\lfloor T_n/n_T \rfloor - 2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\sigma_u(T_m+\tau+bn_T-1)h} \right|
\]

with \( \sigma^2 u_{(T_m+\tau+bn_T-1)h} \triangleq \text{Var} (u_{T_m+\tau+bn_T}) \). The normalization by \( \sigma_u(T_m+\tau+bn_T-1)h \) ensures that we obtain a distribution-free limit theory. Note that the localization assumption implies that there exist \( 0 < \sigma_{u,-} < \sigma_{u,+} < \infty \) defined by \( \sigma_{u,-} \triangleq \inf_{k \geq 1} \{ \sigma_{u,kh} \} \) and \( \sigma_{u,+} \triangleq \sup_{k \geq 1} \{ \sigma_{u,kh} \} \). Furthermore, under \( H_0 \), \( \sigma_u(T_m+\tau+bn_T-1)h \) is a smooth function of Lipschitz parameters and therefore Condition 3.3.1 applies to
\[ \sigma_{u,(T_m+\tau+bn_T-1)h} \] as well: \( \phi_{u,\eta,N} \leq K\eta. \) Finally, let

\[ B_{\text{max},h}^*(T_n, \tau) \triangleq \max_{b=0,\ldots,\left[\frac{T_n}{n_T}\right]-2} \left| \left( \frac{B_{h,b+1}^* - B_{h,b}^*}{U_{h,b+1}} \right) \right|. \quad (C.1.41) \]

We proceed via small lemmas which parallel Lemma C.1.6-C.1.8. The following lemma shows that, under \( H_0 \), the difference in the in-sample losses \( L_{\psi,kh}(\beta) \) between adjacent blocks is negligible asymptotically.

**Lemma C.1.14.** As \( h \downarrow 0 \), \( (\log(T_n) n_T)^{1/2} \left( B_{\text{max},h}^*(T_n, \tau) - U_{\text{max},h}(T_n, \tau) \right) \xrightarrow{\mathbb{P}} 0. \)

**Proof.** Apply (C.1.1) to yield

\[
\left| B_{\text{max},h}^*(T_n, \tau) - U_{\text{max},h}(T_n, \tau) \right| =
\left| \max_{b=0,\ldots,\left[\frac{T_n}{n_T}\right]-2} \left( \frac{B_{h,b+1}^* - B_{h,b}^*}{\sqrt{U_{h,b+1}}} \right) - \max_{b=0,\ldots,\left[\frac{T_n}{n_T}\right]-2} \left( \frac{U_{h,b+1} - U_{h,b}}{\sqrt{U_{h,b+1}}} \right) \right|
\leq \max_{b=0,\ldots,\left[\frac{T_n}{n_T}\right]-2} \left| \sum_{j=1}^{n_T} \frac{L_{\psi,(T_m+(b+1)n_T+j-1)h}(\beta^*) - L_{\psi,(T_m+bn_T+j-1)h}(\beta^*)}{\sqrt{U_{h,b+1}}} \right|.
\]

For any \( \varepsilon > 0 \) and any \( K > 0 \),

\[
\mathbb{P}\left( \max_{b=0,\ldots,\left[\frac{T_n}{n_T}\right]-2} \left| \left( \log(T_n) n_T \right)^{1/2} \sum_{j=1}^{n_T} \frac{L_{\psi,(T_m+(b+1)n_T+j-1)h}(\beta^*) - L_{\psi,(T_m+bn_T+j-1)h}(\beta^*)}{\sqrt{U_{h,b+1}}} \right| > \varepsilon \right) \]
\[
\leq \mathbb{P}\left( \max_{b=0,\ldots,\left[\frac{T_n}{n_T}\right]-2} \left| \left( \log(T_n) n_T \right)^{1/2} \sum_{j=1}^{n_T} \frac{L_{\psi,(T_m+(b+1)n_T+j-1)h}(\beta^*) - L_{\psi,(T_m+bn_T+j-1)h}(\beta^*)}{\sqrt{U_{h,b+1}}} \right| > \varepsilon/K \right) + \mathbb{P}\left( \max_{b=0,\ldots,\left[\frac{T_n}{n_T}\right]-2} 1/\sqrt{U_{h,b+1}} > K \right).
\]

By Lemma C.1.12 the first term on the right-hand size converges to zero. As for the
second term, use the same argument as in (C.1.40). The result then follows. □

Lemma C.1.15. As $h \downarrow 0$, $(\log (T_n) n_T)^{1/2} \left( B_{\text{max},h}^* (T_n, \tau) - B_{\text{max},h}^0 (T_n, \tau) \right) \xrightarrow{P} 0$.

Proof. Note that

\[
| B_{\text{max},h}^* (T_n, \tau) - B_{\text{max},h}^0 (T_n, \tau) |
\leq \max_{b = 0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| B_{h,b+1}^* - B_{h,b}^* \right| \left( \frac{1}{\sqrt{U_{h,b+1}}} - \frac{1}{\sqrt{D_{h,b+1}^0}} \right)
\]

\[+ \max_{b = 0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| B_{h,b+1}^* - B_{h,b+1}^0 \right| \frac{1}{\sqrt{D_{h,b+1}^0}}
\]

\[+ \max_{b = 0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| B_{h,b}^* - B_{h,b}^0 \right| \frac{1}{\sqrt{D_{h,b+1}^0}}.
\]

(C.1.42)

Consider the first term of (C.1.42). We can write for any $\varepsilon > 0$, any $0 < K, C < \infty$, and some small positive number $\varpi < 1/2$,

\[
P \left( \max_{b = 0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| \left( \log (T_n) n_T \right)^{1/2} \left( B_{h,b+1}^* - B_{h,b}^* \right) \left( \frac{1}{\sqrt{U_{h,b+1}}} - \frac{1}{\sqrt{D_{h,b+1}^0}} \right) \right| > \varepsilon \right)
\]

\[\leq P \left( \max_{b = 0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| n_T^{\varpi} \left( B_{h,b+1}^* - B_{h,b}^* \right) \right| > \varepsilon /K \right)
\]

\[+ P \left( \max_{b = 0, \ldots, \lfloor T_n/n_T \rfloor - 2} \sqrt{\log (T_n) n_T^{1/2-\varpi}} \left( \sqrt{U_{h,b+1}} - \sqrt{D_{h,b+1}^0} \right) > K/C \right)
\]

\[+ P \left( \max_{b = 0, \ldots, \lfloor T_n/n_T \rfloor - 2} \frac{1}{\sqrt{D_{h,b+1}^0 U_{h,b+1}}} > C \right) \triangleq A_{1,h} + A_{2,h} + A_{3,h}. \quad (C.1.43)
\]
We first discuss $A_{1,h}$:

$$
P\left(\max_{b=0,\ldots,\left\lfloor T_n/n_T\right\rfloor-2} \left| n_T^{\infty} \left( B_{h,b+1} - B_{h,b}^* \right) \right| > \varepsilon/K \right)
\leq \sum_{b=0}^{\left\lfloor T_n/n_T\right\rfloor-2} P\left( \left| n_T^{\infty} \left( B_{h,b+1}^* - B_{h,b}^* \right) \right| > \varepsilon/K \right)
\leq \sum_{b=0}^{\left\lfloor T_n/n_T\right\rfloor-2} P\left( \left| n_T^{\infty} \left( B_{h,b+1}^* - \mu_{T_m+\tau+(b+1)n_T-1} \right) \right| > \varepsilon/(3K) \right)
\leq \sum_{b=0}^{\left\lfloor T_n/n_T\right\rfloor-2} P\left( \left| n_T^{\infty} \left( B_{h,b}^* - \mu_{T_m+\tau+bn_T-1} \right) \right| > \varepsilon/(3K) \right)
\leq \sum_{b=0}^{\left\lfloor T_n/n_T\right\rfloor-2} P\left( \left| n_T^{\infty} \left( \mu_{T_m+\tau+(b+1)n_T-1} - \mu_{T_m+\tau+bn_T-1} \right) \right| > \varepsilon/(3K) \right). \tag{C.1.44}
$$

Since $B_{h,b}^* \xrightarrow{\mathbb{P}} \mu_{T_m+\tau+bn_T-1} \triangleq \mathbb{E}(u_{T_m+\tau+bn_T})$ and $\mathbb{E}\left[ \sqrt{n_T} \left( B_{h,b}^* - \mu_{T_m+\tau+bn_T-1} \right) \right] < \infty$ by a standard CLT, we have for $r > 0$ sufficiently large and by choosing $\varpi$ sufficiently small,

$$
\sum_{b=0}^{\left\lfloor T_n/n_T\right\rfloor-2} P\left( \left| n_T^{\infty} \left( B_{h,b}^* - \mu_{T_m+\tau+bn_T-1} \right) \right| > \varepsilon/(3K) \right)
\leq K_r \left( \frac{3K}{\varepsilon} \right)^r \sum_{b=0}^{\left\lfloor T_n/n_T\right\rfloor-2} \mathbb{E}\left( \left| n_T^{\infty} \left( B_{h,b}^* - \mu_{T_m+\tau+bn_T-1} \right) \right|^r \right)
\leq K_r \left( \frac{3K}{\varepsilon} \right)^r T_n n_T^{r(\varpi-1/2)-1} \to 0.
$$

The term involving $B_{h,b+1}^*$ admits a similar bound. For the last term of (C.1.44), we
use the Lipschitz continuity of $\mu$ to yield

\[
\sum_{b=0}^{[T_n/n_T]-2} P \left( |n_T^{\sigma} (\mu_{T_m+\tau+(b+1)n_T}-\mu_{T_m+\tau+bn_T})| > \frac{\varepsilon}{3K} \right) \\
\leq \left( \frac{3K}{\varepsilon} \right)^r \sum_{b=0}^{[T_n/n_T]-2} \mathbb{E} \left( |n_T^{\sigma} (\mu_{T_m+\tau+(b+1)n_T}-\mu_{T_m+\tau+bn_T})| \right) \\
\leq \left( \frac{3K}{\varepsilon} \right)^r (T_n/n_T - 2) n_T^{r(\omega+1)} h^r \to 0,
\]

for $r > 0$ sufficiently large since $\omega$ is chosen to be small. Thus, $A_{1,h} \to 0$ while Lemma C.1.16 implies that $A_{3,h} \to 0$ by setting $\sqrt{C} = 1/(2\sigma_{u,-})$. Further, note that $\sigma^2_{u,(T_m+\tau+(b+1)n_T-1)h}$ is the limit of both $U_{h,b+1}$ and $D^0_{h,b+1}$. Thus, given the i.i.d. structure, we can use a standard CLT to yield

\[
P \left( \max_{b=0, \ldots, [T_n/n_T]-2} \log \left( n_T \right)^{1/2-\omega} \left( \sqrt{U_{h,b+1}} - \sqrt{D^0_{h,b+1}} \right) > K/C \right) \\
\leq \sum_{b=0}^{[T_n/n_T]-2} P \left( \left| \log (T_n)^{1/2-\omega} \left( \sqrt{U_{h,b+1}} - \sigma_{u,(T_m+\tau+(b+1)n_T-1)h} \right) \right|^r > (K/2C)^r \right) \\
+ \sum_{b=0}^{[T_n/n_T]-2} P \left( \left| \log (T_n)^{1/2-\omega} \left( \sqrt{D^0_{h,b+1}} - \sigma_{u,(T_m+\tau+(b+1)n_T-1)h} \right) \right|^r > (K/2C)^r \right) \\
\leq 2 (K/2C)^{-r} (T_n/n_T) O_{\mathbb{P}} \left( \left( \log (T_n)^{-\omega} \right)^r \right) \to 0,
\]

for $r > 1/\omega$ sufficiently large. This shows that $A_{2,h} \to 0$. It remains to discuss the second term of (C.1.42); the argument for the third term is equivalent and omitted. Recall the definition of $u^*_{T_m+\tau+(b+1)n_T-1}$ and $u_{T_m+\tau+(b+1)n_T-1}$. By a mean-value
expansion,

\[
B_{h,b+1}^* - B_{h,b+1}^0 = n_T^{-1} \sum_{j=1}^{n_T} (u_{T_m + \tau + (b+1)n_T + j-1}^* - u_{T_m + \tau + (b+1)n_T + j-1})
= n_T^{-1} \sum_{j=1}^{n_T} (g (\Delta h e_{T_m + \tau + (b+1)n_T + j-1}; \beta^*))
- g (\sigma_{e,(T_m + \tau + (b+1)n_T - 1)h} h^{-1/2} (\Delta_h W_{e,T_m + \tau + (b+1)n_T + j-1}; \beta^*))
= n_T^{-1} \sum_{j=1}^{n_T} \left[ \partial e g (\sigma_{e,(T_m + \tau + (b+1)n_T - 1)h} h^{-1/2} (\Delta_h W_{e,T_m + \tau + (b+1)n_T + j-1}; \beta^*))
\times (\Delta_h e_{T_m + \tau + (b+1)n_T + j-1})
- \sigma_{e,(T_m + \tau + (b+1)n_T - 1)h} h^{-1/2} \Delta_h W_{e,T_m + \tau + (b+1)n_T + j-1}
+ \partial^2 e g (\sigma_{e,(T_m + \tau + (b+1)n_T - 1)h} h^{-1/2} (\Delta_h W_{e,T_m + \tau + (b+1)n_T + j-1}; \beta^*))
\times (\Delta_h e_{T_m + \tau + (b+1)n_T + j-1} - \sigma_{e,(T_m + \tau + (b+1)n_T - 1)h} h^{-1/2} \Delta_h W_{e,T_m + \tau + (b+1)n_T + j-1})^2 \right].
\]

(C.1.46)

Since for \( r = 1, 2 \), \( |\partial^r e g (e; \beta)| < C_r \) for some \( C_r < \infty \) by Assumption 3.7, the right-hand side above is less than

\[
(\log T_n n_T)^{1/2} C_1 \left( n_T \sqrt{h} \right)^{-1} \times \sum_{j=1}^{n_T} (\Delta_h e_{T_m + \tau + (b+1)n_T + j-1}^* - \sigma_{e,(T_m + \tau + (b+1)n_T - 1)h} \Delta_h W_{e,T_m + \tau + (b+1)n_T + j-1})
+ (\log (T_n) n_T)^{1/2} C_2 \left( n_T h \right)^{-1} \times \sum_{j=1}^{n_T} (\Delta_h e_{T_m + \tau + (b+1)n_T + j-1}^* - \sigma_{e,(T_m + \tau + (b+1)n_T - 1)h} \Delta_h W_{e,T_m + \tau + (b+1)n_T + j-1})^2.
\]

(C.1.47)
Let us consider the first term of (C.1.47). By Itô’s formula,

\[
\Delta h e^{T_m+\tau+(b+1)n_T+j-1} - \sigma_e(T_m+\tau+(b+1)n_T-1) \Delta T_m+\tau+(b+1)n_T+j-1 W_e \tag{C.1.48}
\]

Then, for an integer \( r > 2 \), by Jensen’s inequality,

\[
E \left[ \left( \log (T_n) n_T \right)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \sum_{j=1}^{n_T} \int_{(T_m+\tau+(b+1)n_T+j-1)h}^{(T_m+\tau+(b+1)n_T+j)h} \left( \sigma_{e,s} - \sigma_{e,(T_m+\tau+(b+1)n_T-1)h} \right) dW_{e,s} \right] 
\leq K_r \left( (\log (T_n) n_T)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \right)^r 
\leq K_r \left( (\log (T_n) n_T)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \right)^r \left( \int_{(T_m+\tau+(b+1)n_T)h}^{(T_m+\tau+(b+2)n_T)h} \left( \frac{\phi^{2r}_{\sigma,N,h}}{\phi^{2r}_{\sigma,N,h}} \right)^{1/r} ds \right)^{r/2} 
\leq K_r \left( (\log (T_n) n_T)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \right)^r \left( n_T h \right)^{r/2} 
\leq K_r \left( (\log (T_n))^{1/2} \right)^r h^{r/3-\epsilon} \to 0, \tag{C.1.49}
\]

by choosing \( r \) large enough. Next, we consider the second term of (C.1.48),

\[
E \left[ \left( \log (T_n) n_T \right)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \sum_{j=1}^{n_T} \int_{(T_m+\tau+(b+1)n_T+j-1)h}^{(T_m+\tau+(b+1)n_T+j)h} \mu_{e,s} h^{-\theta} ds \right]^r 
\leq K_r \left( (\log (T_n) n_T)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \right)^r \left( n_T h^{1-\theta} \right)^r 
\leq K_r \left( (\log (T_n))^{1/2} \right)^r h^{21r/24-\epsilon} \to 0. \tag{C.1.50}
\]
For the term in the second line of (C.1.47) apply the same arguments as in (C.1.49)-(C.1.50) with $m = r/2$ in place of $r$ above. Choosing $m$ large enough yields the same result. Thus, using the latter results into the second term of (C.1.42) via (C.1.46) we have

$$\mathbb{P}\left(\max_{b=0,...,\lfloor T/n\rfloor-2} \left| (\log (T_n) n)^{1/2}\left( B_{h,b+1}^* - B_{h,b+1}^0 \right) \right| > \varepsilon/K \right) \to 0. \quad (C.1.51)$$

Note that the same result holds for $B_{h,b}^* - B_{h,b}^0$. The first term of (C.1.42) has been treated above and so the claim of the lemma follows. □

**Lemma C.1.16.** Assume $\mu_{e,t} = 0$ for all $t \geq 0$. Then,

$$\mathbb{P}\left(\max_{b=0,...,\lfloor T/n\rfloor-2} \left| 1/U_{h,b} \right| > K \right) \to 0$$

for some constant $K > 0$.

**Proof.** Note that

$$\mathbb{P}\left(\max_{b=0,...,\lfloor T/n\rfloor-2} \left| 1/U_{h,b} \right| > K \right) = \mathbb{P}\left(\min_{b=0,...,\lfloor T/n\rfloor-2} \left| U_{h,b} \right| < K^{-1} \right) = \mathbb{P}\left(\min_{b=0,...,\lfloor T/n\rfloor-2} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m+\tau+bn_T+j-1)h} (\beta^*) - \mathcal{L}_{\psi,b} (\beta^*) \right)^2 < K^{-1} \right) \leq \sum_{b=0}^{\lfloor T/n\rfloor-2} \mathbb{P}\left(\sum_{j=1}^{n_T} \left( L_{\psi,(T_m+\tau+bn_T+j-1)h} (\beta^*) - \mathcal{L}_{\psi,b} (\beta^*) \right)^2 < K^{-1} \right).$$

The rest of the proof continues by setting $K^{-1} = \sigma_{L,-}^2/2$ where $\sigma_{L,-}^2 \triangleq \inf_{k \geq 1} \sigma_{L,kh}^2$ with $\sigma_{L,kh}^2 \triangleq \text{Var}(L_{\psi,kh} (\beta^*))$. We can use Markov’s inequality to deduce for any
r > 0,

\[
P \left( \frac{n_T^{-1}}{T} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m+\tau+bn_T+j-1)h} (\beta^*) - L_{\psi,b} (\beta^*) \right)^2 < \sigma_{L,-}^2 / 2 \right) \leq \frac{2}{\left( \sigma_{L,-}^2 \right)^r} \times 

\mathbb{E} \left[ \left| \frac{n_T^{-1}}{T} \sum_{j=1}^{n_T} \left( \left( L_{\psi,(T_m+\tau+bn_T+j-1)h} (\beta^*) - L_{\psi,b} (\beta^*) \right)^2 - \sigma_{L,(T_m+\tau+bn_T-1)h}^2 \right) \right|^r \right].
\]

Observe that, conditional on \( \{ \sigma_{e,t} \}_{t \geq 0} \), \( \text{Var}_\sigma \left[ L_{\psi,(T_m+\tau+bn_T+j-1)h} (\beta^*) \right] \) is constant across \( j = 1, \ldots, n_T \) for a given \( b \). Then, Assumption C.2 implies that we can rely on a basic CLT for i.i.d. observations to yield,

\[
\mathbb{E} \left[ \left| \frac{n_T^{-1}}{T} \sum_{j=1}^{n_T} \left( \left( L_{\psi,(T_m+\tau+bn_T+j-1)h} (\beta^*) - L_{\psi,b} (\beta^*) \right)^2 - \sigma_{L,(T_m+\tau+bn_T-1)h}^2 \right) \right|^r \right] < C_r,
\]

where \( C_r < \infty \). Thus, choose \( r \) sufficiently large so that

\[
\sum_{b=0}^{[T/n_T] - 2} \mathbb{P} \left( \frac{n_T^{-1}}{T} \sum_{j=1}^{n_T} \left( L_{\psi,(T_m+\tau+bn_T+j-1)h} (\beta^*) - L_{\psi,b} (\beta^*) \right)^2 < K^{-1} \right) \leq C_r \left( \frac{2}{\sigma_{L,-}^2} \right)^r O_P \left( T_n/n_T \right) n_T^{-r/2} \to 0,
\]

and the proof is concluded. \( \square \)

**Lemma C.1.17.** As \( h \downarrow 0 \), \( (\log (T_n) n_T)^{1/2} \left( B_{\max,h}^0 (T_n, \tau) - \tilde{B}_{\max,h}^0 (T_n, \tau) \right) \) \( \xrightarrow{P} 0 \).
Proof. By basic manipulations,

\[
\begin{align*}
| B_{\text{max},h}^0 (T_n, \tau) - \tilde{B}_{\text{max},h}^0 (T_n, \tau) | \\
\leq \max_{b=0, \ldots, \lfloor T_n/n \rfloor - 2} \left| \frac{\sqrt{D_{h,b+1}^0} (B_{h,b+1}^0 - \tilde{B}_{h,b+1}^0)}{\sqrt{D_{h,b+1}^0 D_{h,b+1}^0}} \right|
\end{align*}
\]

\[
\leq \max_{b=0, \ldots, \lfloor T_n/n \rfloor - 2} \left| \frac{\sqrt{D_{h,b+1}^0} (B_{h,b+1}^0 - \tilde{B}_{h,b+1}^0)}{\sqrt{D_{h,b+1}^0 D_{h,b+1}^0}} \right|
\]

\[
+ \max_{b=0, \ldots, \lfloor T_n/n \rfloor - 2} \left| \frac{(\tilde{B}_{h,b+1}^0 - B_{h,b}^0) \left( \sqrt{D_{h,b+1}^0} - \sqrt{D_{h,b+1}^0} \right)}{\sqrt{D_{h,b+1}^0 D_{h,b+1}^0}} \right|
\]

\[\triangleq R_{1,h} + R_{2,h}. \quad (C.1.52)\]

We begin with showing that \((\log (T_n) n_T)^{1/2} R_{1,h} \overset{p}{\to} 0, \) or

\[
\max_{b=0, \ldots, \lfloor T_n/n \rfloor - 2} (\log (T_n) n_T)^{1/2} \left| \frac{\sqrt{D_{h,b+1}^0} (B_{h,b+1}^0 - \tilde{B}_{h,b+1}^0)}{\sqrt{D_{h,b+1}^0 D_{h,b+1}^0}} \right| = o_P (1). \quad (C.1.53)
\]

By Lemma C.1.16, \(\mathbb{P} \left( \min_{b=0, \ldots, \lfloor T_n/n \rfloor - 2} \sqrt{D_{h,b+1}^0} < K^{-1/2} \right) \to 0, \) where, for example, \(\sqrt{K} = 2/\sigma_u.\) A similar argument can be used for \(\tilde{D}_{h,b+1}^0\) and therefore it remains to consider the first term of the following decomposition which is valid for any \(\varepsilon > 0\)
and any $K > 0$,

$$
P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} (\log(T_n) n_T)^{1/2} \left| \frac{\sqrt{D_{h,b+1}^0 (B_{h,b+1}^0 - \bar{B}_{h,b+1}^0)} \right| }{\sqrt{D_{h,b+1}^0 D_{h,b+1}^0}} > \varepsilon/K \right)
\leq P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} (\log(T_n) n_T)^{1/2} \left| \sqrt{D_{h,b+1}^0 (B_{h,b+1}^0 - \bar{B}_{h,b+1}^0)} > \varepsilon/K \right) \right.
+ P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} \sqrt{D_{h,b+1}^0 \bar{D}_{h,b+1}^0} > K \right).
\tag{C.1.54}
$$

We have for any positive $K_2 < \infty$,

$$
P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} (\log(T_n) n_T)^{1/2} \left| \sqrt{D_{h,b+1}^0 (B_{h,b+1}^0 - \bar{B}_{h,b+1}^0)} > \varepsilon/K \right) \right.
\leq P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} (\log(T_n) n_T)^{1/2} \left| B_{h,b+1}^0 - \bar{B}_{h,b+1}^0 > \varepsilon/(K \cdot K_2) \right) \right.
+ P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} \sqrt{D_{h,b+1}^0 \bar{D}_{h,b+1}^0} > K_2 \right).
$$

It is straightforward to see that Lemma C.1.16 can be applied also to the second term on the right-hand side above. Hence, it is sufficient to focus on the first term only. Recall the definition of $\tilde{u}_{T_m+\tau+b n_T−1}$ and $u_{T_m+\tau+b n_T−1}$ introduced before Lemma C.1.14. We write

$$
P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} (\log(T_n) n_T)^{1/2} \left| B_{h,b+1}^0 - \bar{B}_{h,b+1}^0 \right| > \varepsilon/K \right)
= P \left( \frac{\max_{b=0, \ldots, \left[ T_n/n_T \right]-2} (\log(T_n) n_T)^{1/2} n_T^{-1} \right.
\times \sum_{j=1}^{n_T} \left( u_{T_m+\tau+(b+1)n_T+j-1} - \tilde{u}_{T_m+\tau+(b+1)n_T+j-1} \right) > \varepsilon/K \right)
\tag{C.1.55}
$$

By a mean-value expansion (omitting the second argument of $g(\cdot; \cdot)$ which is for both
terms here equal to $\beta^*$,

$$
g \left( \sigma_{e,(T_m+\tau+(b+1)n_T-1)h} h^{-1/2} \left( \Delta_h W_{e,T_m+\tau+(b+1)n_T+j-1} \right) \right)
- g \left( \sigma_{e,(T_m+\tau+bn_T-1)h} h^{-1/2} \left( \Delta_h W_{e,T_m+\tau+(b+1)n_T+j-1} \right) \right)
= g_e \left( \sigma_{e,(T_m+\tau+bn_T-1)h} h^{-1/2} \Delta_h W_{e,T_m+\tau+(b+1)n_T+j-1} \right)
\times \left[ \left( \sigma_{e,(T_m+\tau+bn_T-1)h} - \sigma_{e,(T_m+\tau+(b+1)n_T-1)h} \right) \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+(b+1)n_T+j-1} \right) \right]
+ 2^{-1} g_{ee,b,j}(\tau) \times
\left[ \left( \sigma_{e,(T_m+\tau+bn_T-1)h} - \sigma_{e,(T_m+\tau+(b+1)n_T-1)h} \right) \left( h^{-1/2} \Delta_h W_{e,T_m+\tau+(b+1)n_T+j-1} \right) \right]^2
$$

In view of Assumption 3.2, for $r = 1, 2$,

$$
\left| \sigma_{e,(T_m+\tau+bn_T-1)h} - \sigma_{e,(T_m+\tau+(b+1)n_T-1)h} \right|^r \leq C_r (n_T h)^r , \quad (C.1.56)
$$

uniformly in $b$ where $C_r < \infty$. Let

$$
\overline{C}_1 \triangleq 2 \sup_{k \geq 1} \sup_{t \geq 0} \left| g_e \left( \sigma_{e,t} h^{-1/2} \Delta_h W_{e,k} \right) \right| , \quad \overline{C}_2 \triangleq 2 \sup_{k \geq 1} \sup_{t \geq 0} \left| g_{ee} \left( \sigma_{e,t} h^{-1/2} \Delta_h W_{e,k} \right) \right| .
$$

Then, the right-hand side of (C.1.55) can be decomposed as follows with $K_1 = \sqrt{2 \overline{C}_1}$
and $K_2 = \sqrt{2C_2}$,

$$
\mathbb{P}\left( \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \sum_{j=1}^{n_T} \left( \bar{u}_{T_m+\tau+(b+1)nT-1} - u_{T_m+\tau+(b+1)nT-1} \right) > \varepsilon/K \right) 
\leq \mathbb{P}\left( \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left( \log (T_n) n_T \right)^{1/2} n_T^{-1} \sum_{j=1}^{n_T} \left( \bar{u}_{T_m+\tau+(b+1)nT-1} - u_{T_m+\tau+(b+1)nT-1} \right) > \varepsilon/K \right) 
\leq \mathbb{P}\left( \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left( \log (T_n) n_T \right)^{1/2} \left( \sigma_e, (T_m+\tau+bnT-1)h - \sigma_e, (T_m+\tau+(b+1)nT-1)h \right) > \varepsilon/K_1 \cdot K \right) 
+ \mathbb{P}\left( \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left( \log (T_n) n_T \right)^{1/2} \left( \sigma_e, (T_m+\tau+bnT-1)h \right) h^{-1/2} \Delta_h W_{e,T_m+\tau+(b+1)nT+j-1} \right) > K_1 
+ \mathbb{P}\left( \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left( \log (T_n) n_T \right)^{1/2} \left( \sigma_e, (T_m+\tau+bnT-1)h \right) \right)^2 > \varepsilon/(2K_2 \cdot K) 
+ \mathbb{P}\left( \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left( \log (T_n) n_T \right)^{1/2} \right)^2 > K_2 
\triangleq A_{1,h} + A_{2,h} + A_{3,h} + A_{4,h}.
$$

The relationship in (C.1.56) implies that $A_{1,h}, A_{3,h} \to 0$ using Condition 2 because

$$
\max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left( \log (T_n) n_T \right)^{1/2} \left( \sigma_e, (T_m+\tau+bnT-1)h \right) \leq (\log (T_n) n_T)^{1/2} \phi_{\sigma,nT,h,N} \to 0.
$$
The boundedness of \( g(\cdot, \cdot) \) [cf. Assumption 3.7], implies that for \( r > 0 \) large enough,

\[
\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor-2} n_T^{-1} \sum_{j=1}^{n_T} g e \left( \sigma, (T_m + \tau + bn_T - 1) h^{-1/2} \Delta h W_{e,T_m + \tau + (b+1)n_T + j - 1} \right) \times h^{-1/2} \Delta h W_{e,T_m + \tau + (b+1)n_T + j - 1} > K_1 \right) \leq \mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor-2} n_T^{-2} \sum_{j=1}^{n_T} \left( h^{-1/2} \Delta h W_{e,T_m + \tau + (b+1)n_T + j - 1} \right)^2 - 1 \right)^{r/2} > \left( \frac{K_1^2}{2} \right)^{r/2}
\]

\[
\leq \left( \frac{2}{K_1^2} \right)^{r/2} C_r \sum_{b=0}^{\lfloor T_n/n_T \rfloor-2} \mathbb{E} \left[ n_T^{-2} \sum_{j=1}^{n_T} \left( \left( h^{-1/2} \Delta h W_{e,T_m + \tau + (b+1)n_T + j - 1} \right)^2 - 1 \right) \right]^{r/2} \leq \left( \frac{2}{K_1^2} \right)^{r} C_r T_n n_T^{-1-3r/2} \rightarrow 0.
\]

We can apply the same argument with \( K_2 = \sqrt{2C_2} \) to \( A_{4,h} \) to show that

\[
\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor-2} n_T^{-1} \sum_{i=1}^{n_T} g_{e,b,j} \left( \sigma \right) \left( h^{-1/2} \Delta T_m + \tau + bn_T + i - 1 \right) W_{e,i}^2 > K_2 \right) \leq K_2^{2r} C_r T_n n_T^{-1-r} \rightarrow 0,
\]

and so \( A_{4,h} \rightarrow 0 \). This gives (C.1.54) and thus (C.1.53). Next, we consider \( R_{2,h} \) and want to show \( (\log (T_n) n_T)^{1/2} R_{2,h} \mathbb{P} \rightarrow 0 \), or

\[
\max_{b=0, \ldots, \lfloor T_n/n_T \rfloor-2} (\log (T_n) n_T)^{1/2} \left| \frac{\sqrt{D_{h,b+1}^0} - B_{h,b}^0} {\sqrt{D_{h,b+1}^0 D_{h,b+1}^0}} \left( \sqrt{\frac{D_{h,b+1}^0}{D_{h,b+1}^0}} - \sqrt{\frac{D_{h,b+1}^0}{D_{h,b+1}^0}} \right) \right| = o_p (1).
\]

(C.1.57)
Proceeding as in (C.1.54), it is sufficient show

\[
\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T n / n T \rfloor - 2} (\log (T n) n T)^{1/2} \left| (\bar{B}_{h,b+1}^0 - B_{h,b}^0) \left( \sqrt{\tilde{D}_{h,b+1}^0} - \sqrt{D_{h,b+1}^0} \right) > \varepsilon / K \right| \right) \to 0.
\]

The argument for \((\log (T n) n T)^{1/2} (\bar{B}_{h,b+1}^0 - B_{h,b}^0)\) is similar to the one used above, but now one needs an additional step using a Taylor series expansion of \(g\); we omit the details. Thus, we have to show

\[
\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T n / n T \rfloor - 2} \left| \frac{\sqrt{\tilde{D}_{h,b+1}^0} - \sqrt{D_{h,b+1}^0}}{\sqrt{\tilde{D}_{h,b+1}^0} - \sum_{D_{h,b}^0} (\tilde{D}_{h,b+1}^0)} > C \varepsilon \right| \right) \to 0,
\]

for some finite \(C > 0\). Note that

\[
\sqrt{\tilde{D}_{h,b}^0} - \sqrt{D_{h,b}^0} = \sigma_{u,(T n+\tau+(b-1) n T -1)h} - \sigma_{u,(T n+\tau+b n T -1)h} + O_{\mathbb{P}} (n_T^{-1/2})
\]

\[
= \phi_{\sigma_{u,n_T h,N}} + O_{\mathbb{P}} (n_T^{-1/2}).
\]

Since \(\phi_{\sigma_{u,n_T h,N}} \leq C n_T h\) uniformly over \(h, \ldots, T h = N\), we can show using the same arguments employed above that

\[
\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T n / n T \rfloor - 2} \left| \bar{D}_{h,b+1}^0 - D_{h,b+1}^0 \right| > C \varepsilon \right) \to 0.
\]

Therefore, we have \((\log (T n) n T)^{1/2} R_{2,h} \xrightarrow{\mathbb{P}} 0\). The claim of the lemma follows. □

**Lemma C.1.18.** As \(h \downarrow 0\), \((\log (T n) n T)^{1/2} \left( V_{\max,h} (T n, \tau) - \bar{B}_{\max,h}^0 (T n, \tau) \right) \xrightarrow{\mathbb{P}} 0\).
Proof. We have the inequality,

\[
V_{\text{max},h}(T_n, \tau) - \tilde{B}_{\text{max},h}(T_n, \tau) \\
= \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\sigma_u(T_m + \tau + bnT - 1)h} \right| - \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left| \frac{\tilde{B}_{h,b+1}^0 - B_{h,b}^0}{\sqrt{\tilde{D}_{h,b+1}^0}} \right| \\
\leq \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left| \frac{(\tilde{B}_{h,b+1}^0 - B_{h,b}^0)(\sqrt{\tilde{D}_{h,b+1}^0} - \sigma_u(T_m + \tau + bnT - 1)h)}{\sigma_u(T_m + \tau + bnT - 1)h \sqrt{\tilde{D}_{h,b+1}^0}} \right|.
\]

Thus, we want to show that

\[
(\log(T_n) nT)^{1/2} \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left| \frac{(\tilde{B}_{h,b+1}^0 - B_{h,b}^0)(\sqrt{\tilde{D}_{h,b+1}^0} - \sigma_u(T_m + \tau + bnT - 1)h)}{\sigma_u(T_m + \tau + bnT - 1)h \sqrt{\tilde{D}_{h,b+1}^0}} \right| = o_p(1).
\]

By Assumption 3.5, \(0 < \sigma_u(T_m + \tau + bnT - 1)h < \infty\) for all \(b \geq 0\) while \(\tilde{D}_{h,b+1}^0\) was already shown to bounded from below and above. Thus, basic manipulations as in the previous lemmas show that the denominator is also \(O_p(1)\). Turning to the numerator, we have

\[
P \left( (\log(T_n) nT)^{1/2} \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left| \frac{(\tilde{B}_{h,b+1}^0 - B_{h,b}^0)(\sqrt{\tilde{D}_{h,b+1}^0} - \sigma_u(T_m + \tau + bnT - 1)h)}{\sigma_u(T_m + \tau + bnT - 1)h \sqrt{\tilde{D}_{h,b+1}^0}} \right| > \varepsilon \right) \\
\leq P \left( (\log(T_n) nT)^{1/2} \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left| \tilde{B}_{h,b+1}^0 - B_{h,b}^0 \right| \right) \varepsilon \\
+ P \left( \max_{b=0,\ldots,\lceil T_n/nT \rceil-2} \left| \sqrt{\tilde{D}_{h,b+1}^0} - \sigma_u(T_m + \tau + bnT - 1)h \right| > \sqrt{\varepsilon} \right).
\]
In view of the proof of the last part of Lemma C.1.17,

\[
\mathbb{P} \left( (\log (T_n) n_T)^{1/2} \max_{b=0,\ldots,\lfloor T_n / n_T \rfloor - 2} |\tilde{B}_{h,b+1}^0 - B_{h,b}^0| > \sqrt{\varepsilon} \right) \to 0.
\]

To conclude the proof of the lemma it remains to show that

\[
\mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n / n_T \rfloor - 2} \left| \sqrt{D_{h,b+1}^0} - \sigma_{u,(T_m + \tau + b n_T - 1)h} \right| > \sqrt{\varepsilon} \right) \to 0. \tag{C.1.59}
\]

By the definition of \( \tilde{D}_{h,b+1}^0 \), the summands \( \tilde{u}_{T_m + \tau + b n_T + j - 1}, (j = 1, \ldots, n_T) \) are independent and each satisfies \( \text{Var} \left[ \tilde{u}_{T_m + \tau + b n_T + j - 1} \right] = \sigma_{u,(T_m + \tau + b n_T - 1)h}^2 \). Then,

\[
\mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n / n_T \rfloor - 2} \left| \sqrt{D_{h,b+1}^0} - \sigma_{u,(T_m + \tau + b n_T - 1)h} \right| > \sqrt{\varepsilon} \right)
\]

\[
\leq \mathbb{P} \left( \max_{b=0,\ldots,\lfloor T_n / n_T \rfloor - 2} \left| \sqrt{D_{h,b+1}^0} - \sigma_{u,(T_m + \tau + b n_T - 1)h} \right| > \sqrt{\varepsilon} \right)
\]

\[
\leq \prod \left[ \sqrt{n_T - 1} \sum_{j=1}^{n_T} \left( g \left( \sigma_{e,(T_m + \tau + b n_T - 1)h} h^{-1/2} \left( \Delta_{h} W_{e,T_m + \tau + (b+1) n_T + j - 1} \right) - \tilde{g}_{b+1} \right)^2 - \sigma_{u,(T_m + \tau + b n_T - 1)h} \right) \right]
\]

Note that the variables \( g \left( \sigma_{e,(T_m + \tau + b n_T - 1)h} h^{-1/2} \left( \Delta_{h} W_{e,T_m + \tau + (b+1) n_T + j - 1} \right) - \tilde{g}_{b+1} \right)^2 - \sigma_{u,(T_m + \tau + b n_T - 1)h} \) are independent over \( j \) and their variances are constant and equal to \( \sigma_{u,(T_m + \tau + b n_T - 1)h}^2 \). Due to the i.i.d. structure we can rely on a basic CLT for the sample variance which, given Assumption C.2, yields

\[
\mathbb{E} \left[ \sqrt{n_T - 1/2} \sum_{j=1}^{n_T} \left( g \left( \sigma_{e,(T_m + \tau + b n_T - 1)h} h^{-1/2} \left( \Delta_{h} W_{e,T_m + \tau + (b+1) n_T + j - 1} \right) - \tilde{g}_{b+1} \right)^2 - \sigma_{u,(T_m + \tau + b n_T - 1)h} \right) \right]
\]
and thus for \( r > 0 \) sufficiently large, we have by Condition 2,

\[
\begin{align*}
\mathbb{P} \left( \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| \sqrt{D_{n,b+1}^0} - \sigma_{u,(T_m+\tau+bn_T-1)h} \right| > \sqrt{\varepsilon} \right) \\
\leq C_r \varepsilon^{-r/2} \sum_{b=0}^{\lfloor T_n/n_T \rfloor - 2} \sum_{j=1}^{n_T} \left| \sum_{j=1}^{n_T} \left( g \left( \sigma_{e,(T_m+\tau+bn_T-1)h} \right) - \bar{g}_{b+1} \right) \right| \\
\leq C_r \varepsilon^{-r/2} \mathcal{O}_P \left( T_n/n_T \right) n_T^{-r/2} \to 0.
\end{align*}
\]

Altogether, we have (C.1.59) and thus (C.1.58), which concludes the proof. \( \square \)

**Proof of Theorem 3.3.2-(i).** By Lemma C.1.13-C.1.18,

\[
(\log (T_n) n_T)^{1/2} (G_{\max,h} (T_n, \tau) - V_{\max,h} (T_n, \tau)) \overset{\mathbb{P}}{\to} 0.
\]

We now apply Lemma C.1.10 to \( V_{\max,h} (T_n, \tau) \). Let

\[
\begin{align*}
\bar{U}_{T_m+\tau+bn_T+j-1} &\triangleq \frac{\bar{u}_{T_m+\tau+bn_T+j-1} - \mu_{u,T_m+\tau+(b-1)n_T-1}}{\sigma_{u,(T_m+\tau+(b-1)n_T-1)h}} \\
U_{T_m+\tau+bn_T+j-1} &\triangleq \frac{u_{T_m+\tau+bn_T+j-1} - \mu_{u,T_m+\tau+bn_T-1}}{\sigma_{u,(T_m+\tau+bn_T-1)h}},
\end{align*}
\]

with \( \mu_{u,T_m+\tau+bn_T-1} \triangleq \mathbb{E} (u_{T_m+\tau+bn_T-1}) \). Then, write

\[
V_{\max,h} (T_n, \tau) = \max_{b=0, \ldots, \lfloor T_n/n_T \rfloor - 2} \left| \sum_{j=1}^{n_T} \left( \bar{U}_{T_m+\tau+(b+1)n_T+j-1} - U_{T_m+\tau+bn_T+j-1} \right) \right|.
\]

Observe that the variables \( \bar{U}_{T_m+\tau+(b+1)n_T+j-1} \) and \( U_{T_m+\tau+bn_T+j-1} \) (\( j = 1, \ldots, n_T \))
have both zero mean, unit variance and are independent over $b$ and $j$. Thus, $V_{\text{max}, h}(T_n, \tau)$ corresponds to $B_{\text{max}, T_n}$ from Lemma C.1.10. In addition, under Assumption C.2 and Condition 2 the final result can be deduced from the same lemma. □

C.1.4.3.2 Proof of part (ii) of Theorem 3.3.2 The proof follows similar steps as those used for $\text{MQ}_{\text{max}, h}$. More specifically, we can repeat the same proof as in Lemma C.1.12 so that corresponding results for a general loss function are still valid.

Let $U_{h, i} \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( \frac{L_{\psi,Tm+\tau+j-1} h (\beta^*) - L_{\psi,i} (\beta^*)}{n_T} \right)^2$ and define

$$U_{\text{max}, h}(T_n, \tau) \triangleq \max_{i=n_T, \ldots, T_n-n_T} \left| \frac{n_T^{-1} \sum_{j=i+1}^{i+n_T} SL_{\psi,Tm+\tau+j-1} (\beta^*) - n_T^{-1} \sum_{j=i-n_T+1}^{i} SL_{\psi,Tm+\tau+j-1} (\beta^*)}{\sqrt{U_{h, i}}} \right|.$$ 

Lemma C.1.19. For any $L \in L_e$, we have the results of Lemma C.1.12 and

$$(\log (T_n) n_T)^{1/2} \left( U_{\text{max}, h}(T_n, \tau) - \text{MG}_{\text{max}, h}(T_n, \tau) \right) \xrightarrow{p} 0.$$ 

Proof. The first claim can be proven in the same fashion as in Lemma C.1.12 with
Following the same derivations as in the non-overlapping case we have a result corresponding to equation (C.1.39),

\[
\left| \sum_{j=i+1}^{i+n_T} \left( SL_{\psi,T_m+\tau+j-1} (\beta^*) - SL_{\psi,T_m+\tau+j-1} (\bar{\beta}) \right) \right| = KO_p \left( T^{-1/2} \right). \tag{C.1.61}
\]

For any \( \varepsilon > 0 \) and any constant \( K > 0 \), we then have the decomposition,

\[
P \left( \max_{i=n_T, \ldots, T_n-n_T} \left| \sum_{j=i+1}^{i+n_T} \left( SL_{\psi,T_m+\tau+j-1} (\beta^*) - SL_{\psi,T_m+\tau+j-1} (\bar{\beta}) \right) \right| > \varepsilon \right) \leq P \left( \max_{i=n_T, \ldots, T_n-n_T} \left| \sum_{j=i+1}^{i+n_T} \left( SL_{\psi,T_m+\tau+j-1} (\beta^*) - SL_{\psi,T_m+\tau+j-1} (\bar{\beta}) \right) \right| > \varepsilon/K \right) + P \left( \max_{i=n_T, \ldots, T_n-n_T} 1/|U_{h,i}| > K \right). \tag{C.1.62}
\]
Observe that Lemma C.1.16 remains valid when blocks overlap and so

\[ \mathbb{P} \left( \max_{i=n_T, \ldots, T_n-n_T} \frac{1}{n_T} |U_{h,i}| > K \right) \rightarrow 0, \]

by setting, for example, \( K = 1/\sigma_{L,-}^2 \). Upon using Markov’s inequality, (C.1.61) (which holds uniformly in \( i \)) and Condition 2 we can conclude the proof with

\[ \mathbb{P} \left( \max_{i=n_T, \ldots, T_n-n_T} \times \right. \]

\[ \left. \left( \log (T_n) n_T \right)^{1/2} \right| \left. n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( SL_{\psi,T_m+\tau+j-1} (\beta^*) - SL_{\psi,T_m+\tau+j-1} (\tilde{\beta}) \right) \right| > \varepsilon/K \]

\[ \leq \frac{K}{\varepsilon} \mathbb{E} \left[ \left( \log (T_n) n_T \right)^{1/2} \right| \left. n_T^{-1} \sum_{i=1}^{n_T} \left( SL_{\psi,T_m+\tau+j-1} (\beta^*) - SL_{\psi,T_m+\tau+j-1} (\tilde{\beta}) \right) \right] \]

\[ = \frac{K}{\varepsilon} \left( \log (T_n) n_T \right)^{1/2} \mathcal{O}_P \left( 1/\sqrt{T} \right) \rightarrow 0. \square \]

Define

\[ MB_{\max,h}^0 (T_n, \tau) \]

\[ \triangleq \max_{i=n_T, \ldots, T_n-n_T} \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} g \left( \sigma_{e,(T_m+\tau+i-1)h} h^{-1/2} (\Delta_h W_{e,T_m+\tau+j-1}) \right) \right| \]

\[ - n_T^{-1} \sum_{j=1-n_T+1}^{i} g \left( \sigma_{e,(T_m+\tau+i-n_T-1)h} h^{-1/2} (\Delta_h W_{e,T_m+\tau+j-1}) \right) / \sqrt{D_{h,i}^0} \]

where

\[ D_{h,i}^0 \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( g \left( \sigma_{e,(T_m+\tau+i-1)h} h^{-1/2} (\Delta_h W_{e,T_m+\tau+j-1}) \right) - \bar{g}_i \right)^2 , \]
with \( \bar{g}_i \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} g \left( \sigma_{u,(T_m+\tau+i-1)h} h^{-1/2} (\Delta h W_{e,T_m+\tau+j-1}) \right) \). Next, let

\[
MB_{\text{max},h}^* (T_n, \tau) \triangleq \max_{i=n_T, \ldots, T_n} \left| \frac{n_T^{-1} \sum_{j=i+1}^{i+n_T} u_{T_m+\tau+j-1} - n_T^{-1} \sum_{j=i-n_T+1}^{i} u_{T_m+\tau+j-1} \right| \sqrt{U_{h,i}},
\]

(C.1.63)

where \( u_{T_m+\tau+j-1} \triangleq g (\Delta h e_{T_m+\tau+j-1}; \beta^*) \). Then, define

\[
\overline{MB}_{\text{max},h}^0 (T_n, \tau) \triangleq \max_{i=n_T, \ldots, T_n} \left| \frac{n_T^{-1} \sum_{j=i+1}^{i+n_T} \bar{u}_{T_m+\tau+j-1} - n_T^{-1} \sum_{j=i-n_T+1}^{i} u_{T_m+\tau+j-1} \right| \sqrt{\bar{D}_{h,i}^0},
\]

where \( \bar{D}_{h,i}^0 \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} (\bar{u}_{T_m+\tau+j-1} - \bar{g}_i)^2 \), with \( \bar{g}_i \triangleq n_T^{-1} \sum_{j=i+1}^{i+n_T} \bar{u}_{T_m+\tau+j-1} \) and

\[
\bar{u}_{T_m+\tau+j-1} \triangleq g \left( \sigma_{e,(T_m+\tau+i-n_T-1)h} h^{-1/2} \Delta h W_{e,T_m+\tau+j-1}; \beta^* \right).
\]

In the final step we shall show that

\[
\mathbb{P} \left( \left( \log (T_n) n_T \right)^{1/2} \left( \text{MV}_{\text{max},h} (T_n, \tau) - \overline{MB}_{\text{max},h}^0 (T_n, \tau) \right) > \varepsilon \right) \rightarrow 0,
\]

for any \( \varepsilon > 0 \), where

\[
\text{MV}_{\text{max},h} (T_n, \tau) \triangleq \max_{i=n_T, \ldots, T_n} \left| \frac{n_T^{-1} \sum_{j=i+1}^{i+n_T} \bar{u}_{T_m+\tau+j-1} - n_T^{-1} \sum_{j=i-n_T+1}^{i} u_{T_m+\tau+j-1} \right| \sigma_{u,(T_m+\tau+i-n_T-1)h},
\]

with \( \sigma_{u,(T_m+\tau+i-n_T-1)h} \triangleq \left( \text{Var} \left( u_{T_m+\tau+i-n_T} \mid \mathcal{F}_{(T_m+\tau+i-n_T-1)h} \right) \right)^{1/2} \). By Assumption C.1 there exist \( 0 < \sigma_{u,-} < \sigma_{u,+} < \infty \) defined by \( \sigma_{u,-} \triangleq \inf_{k \geq 1} \{ \sigma_{u,kh} \} \) and \( \sigma_{u,+} \triangleq \sup_{k \geq 1} \{ \sigma_{u,kh} \} \). In parts of the derivations below we shall use some of the results from the non-overlapping case. In particular, the only difference arises from the fact that now the maximum is over a larger set and therefore the bounds should be adjusted.
Lemma C.1.20. As $h \downarrow 0$,

$$\left( \log \left( \frac{T_n}{nT} \right) \right)^{1/2} \left( U_{\max,h} (T_n, \tau) - MB_{\max,h}^0 (T_n, \tau) \right) \xrightarrow{P} 0.$$ 

Proof. First, given Lemma C.1.19 it follows that

$$\left( \log \left( \frac{T_n}{nT} \right) \right)^{1/2} \left( MB_{\max,h}^* (T_n, \tau) - U_{\max,h} (T_n, \tau) \right) \xrightarrow{P} 0.$$ 

Thus, we have to show

$$\left( \log \left( \frac{T_n}{nT} \right) \right)^{1/2} \left( MB_{\max,h}^* (T_n, \tau) - MB_{\max,h}^0 (T_n, \tau) \right) \xrightarrow{P} 0.$$ 

Note that the result of Lemma C.1.16 still holds. Thus, we have decompositions similar to (C.1.42) and (C.1.43) and then one can follow the same steps as above. However, the bounds in (C.1.44) and (C.1.45) are now different because the maximum is over $i = n_T, \ldots, T_n - n_T$. The bound in (C.1.44) is now $(K/\varepsilon)^r n_T^{\varpi r - r} (2h^{r-1})$ which converges to zero by choosing $r$ sufficiently large and $\varpi$ small. The bound corresponding to (C.1.45) also goes to zero for large enough $r > 0$. All the steps leading to (C.1.49) can be repeated with minor changes. Indeed, the bound (C.1.49) also remains the same because it involves using the condition on Lipschitz continuity,
which gives for $r > 0$ large enough,

$$
\mathbb{E} \left[ \left( \log \left( \frac{T_n}{n_T} \right) \right)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \sum_{j=1}^{n_T} \int_{(T_m + \tau + i + j) h}^{(T_m + \tau + i + j - 2) h} \left( \sigma_{e,s} - \sigma_{e,(T_m + \tau + i - 2)h} \right) dW_{e,s} \right]^{r/2}
\leq K_r \left( \log \left( \frac{T_n}{n_T} \right) \right)^{1/2} \left( n_T \sqrt{h} \right)^{-1} \left( \int_{(T_m + \tau + i - 1) h}^{(T_m + \tau + i + n_T - 1) h} \left( \mathbb{E} \left[ \phi_{\sigma,n_T,h,N} \right] \right)^{2/r} ds \right)^{r/2}
\leq K_r \left( \log \left( \frac{T_n}{n_T} \right) \right)^{1/2} h^{-1/3 + r/3 + \epsilon} \to 0.
$$

Altogether, these arguments can be used to verify the result of the lemma. □

**Lemma C.1.21.** As $h \downarrow 0$, $(\log \left( \frac{T_n}{n_T} \right) )^{1/2} \left( \text{MB}_{\max,h}^0 (T_n, \tau) - \text{MV}_{\max,h}^0 (T_n, \tau) \right) \xrightarrow{P} 0$.

**Proof.** The proof follows exactly the same steps as in the proof of Lemma C.1.17-C.1.18. Since some of the bounds need to be adjusted to account for the maximum being over $i = n_T, \ldots, T_n - n_T$, we can use the same argument as in the previous lemma. Then, all the quantities generalizing the expressions in the proofs of Lemma C.1.17-C.1.18 are controlled thereby yielding

$$(\log \left( \frac{T_n}{n_T} \right) )^{1/2} \left( \text{MB}_{\max,h}^0 (T_n, \tau) - \text{MB}_{\max,h}^0 (T_n, \tau) \right) \xrightarrow{P} 0,$$

and $(\log \left( \frac{T_n}{n_T} \right) )^{1/2} \left( \text{MV}_{\max,h} (T_n, \tau) - \text{MB}_{\max,h}^0 (T_n, \tau) \right) \xrightarrow{P} 0$. □

**Proof of Theorem 3.3.2-(ii).** From Lemma C.1.19-C.1.21,

$$
\sqrt{\log (T) n_T} \left( \text{MG}_{\max,h} (T_n, \tau) - \text{MV}_{\max,h} (T_n, \tau) \right) = o_P (1).
$$

As for the non-overlapping case, we deduce the limit distribution of $\text{MV}_{\max,h}$ from
that of \( \text{MB}_{\text{max},T_n} \) derived in Lemma C.1.10. Let
\[
\tilde{U}_{T_m+\tau+j-1} \triangleq \begin{cases} 
\frac{u_{T_m+\tau+j-1} - \mu_{u,T_m+\tau+i-nT-1}}{\sigma_u(T_m+\tau+i-nT-1) h} & \text{for } j = i + 1, \ldots, i + n_T \\
\frac{u_{T_m+\tau+j-1} - \mu_{u,T_m+\tau+i-nT-1}}{\sigma_u(T_m+\tau+i-nT-1) h} & \text{for } j = i - n_T + 1, \ldots, i.
\end{cases}
\]

Then, we have \( \mathbb{E}\left( \tilde{U}_{T_m+\tau+j-1} \right) = 0 \), \( \text{Var}\left( \tilde{U}_{T_m+\tau+j-1} \right) = 1 \) and the \( \tilde{U}_{T_m+\tau+j-1} \)'s are independent across \( j \). \( \text{MV}_{\text{max},h}(T_n, \tau) \) now corresponds to \( \text{MB}_{\text{max},T_n} \) from Lemma C.1.10. Thus, we can deduce the final result from Lemma C.1.10 since Assumption C.2 and Condition 3.3.1 holds. \( \Box \)

C.1.4.3.3 Negligibility of the drift term under general loss functions

The reasoning is similar to the quadratic loss case. We only show that the drift component \( \mu_{e,t} \) is of higher order. Without estimation uncertainty our tests statistics are simply functions of local averages of \( g(\Delta_h e^*_k; \beta^*) \), where \( g(\cdot, \cdot) \) is smooth. Note that conditional on \( \{\mu_{e,t}\}_{t \geq 0} \) and \( \{\sigma_{e,t}\}_{t \geq 0} \),
\[
\begin{align*}
&h^{-1/2} \Delta_h e^*_{T_m+\tau+bn_T+j-1} = h^{-1/2} \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \mu_{e,s} h^{-\vartheta} ds \\
&\quad \quad + h^{-1/2} \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \sigma_{e,s} dW_{e,s} \\
&\quad \quad = O(h^{1-\vartheta-1/2}) + h^{-1/2} \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \sigma_{e,s} dW_{e,s}.
\end{align*}
\]

Since \( \vartheta \in [0, 1/8) \) and \( \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \sigma_{e,s} dW_{e,s} \approx N\left(0, \int_{(T_m+\tau+bn_T+j-2)h}^{(T_m+\tau+bn_T+j-1)h} \sigma_{e,s}^2 ds\right) \), it follows that the first term above is of higher order and should not play any role for the asymptotic results of Lemma C.1.12-C.1.13.
C.1.4.4 Proof of Corollary 3.3.2

Proof. It follows the same arguments as for Corollary 3.3.1. \(\square\)

C.1.5 Proofs of Section 3.4

C.1.5.1 Proof of Theorem 3.4.1

Proof. See Theorem 3 in Wu and Zhao (2007). \(\square\)

C.1.5.2 Proof of Theorem 3.4.2

The initial step in the proof uses a uniform strong approximation result which essentially extends the strong invariance principle of Wu (2007) to our setting. The idea behind the proof is similar to that of Theorem 2.1 in Zhao and Li (2013). Before giving the result, we need to recall the more general framework of Wu (2007).

Let \(\{\xi_k\}_{k=1}^{T_n}\) be a sequence of zero-mean independent random variables with \(\text{Var} (\xi_k) = \sigma_k^2\) satisfying \(c_- \leq \min_{k \geq 1} \{\sigma_k\}\) and \(c_+ \geq \max_{k \geq 1} \{\sigma_k\}\) with \(0 < c_- < c_+ < \infty\). Let \(z_j \triangleq j^{-1} \sum_{k=1}^{j} \xi_k\), \(G_{\xi,j} \triangleq j z_j\) and \(V_{\xi,j} \triangleq \sum_{k=1}^{j} (\xi_k - z_j)^2\). Let \(\{\bar{B}_t\}_{t \geq 0}\) and \(\{\tilde{B}_t\}_{t \geq 0}\) denote two independent one-dimensional standard Wiener processes which need not be defined on the same probability space. Finally, let \(a_{T_n} \triangleq |\sigma_{T_n}| + \sum_{k=2}^{T_n} |\sigma_k - \sigma_{k-1}|\), \(c_{T_n} \triangleq \sigma_{T_n}^2 + \sum_{k=2}^{T_n} |\sigma_k^2 - \sigma_{k-1}^2|\), \(\Xi_j \triangleq \sum_{k=1}^{j} \sigma_k^2\) and \(\tilde{\Xi}_j \triangleq \sum_{k=1}^{j} \sigma_k^4\).

We begin with the following lemma involving a strong invariance principle for the process \(\{\xi_k\}\) and an uniform approximation for \(\{V_{\xi,k}\}\). Without loss of generality assume that \(\xi_k = \sigma_k \epsilon_k\), with \(\{\epsilon_k\}\) being a zero-mean stationary process with \(\mathbb{E} (\epsilon_k^2) = 1\). Further, denote by \(\varrho^2\) the long-run variance of \(\epsilon_k\) : \(\varrho^2 \triangleq \gamma_0 + 2 T_n^{-1} \sum_{i=1}^{T_n} \gamma_i\), where \(\gamma_i \triangleq \text{Cov} (\epsilon_{k+i}, \epsilon_k)\). Define similarly the long-run variance of \(\{\epsilon_k^2 - 1\}\) and denote it
by $\tilde{\varrho}^2$. Next, let

$$S_k = \sum_{j=1}^{k} \epsilon_j, \quad \tilde{S}_k = \sum_{j=1}^{k} (\epsilon_j^2 - 1), \quad k = 1, \ldots, T_n,$$

with the convention $S_0 = \tilde{S}_0 = 0$. Then we have the following strong invariance principles [cf. Wu (2007)]:

$$\max_{1 \leq k \leq T_n} |S_k - \varrho \mathbb{B}_k| = o_{a.s.} (\Delta T_n) \quad \text{and} \quad \max_{1 \leq k \leq T_n} |\tilde{S}_k - \tilde{\varrho} \tilde{\mathbb{B}}_k| = o_{a.s.} (\Delta T_n), \quad (C.1.65)$$

where $\Delta T_n$ is an approximation error that satisfies $\Delta T_n \to \infty$. Under our context, the order of $\Delta T_n$ is given by the following assumption

**Assumption C.3.** Assume $0 < \varrho, \tilde{\varrho} < \infty$. The relationships in (C.1.65) holds with $\Delta T_n = T_n^{1/4} \log (T_n)$.

**Lemma C.1.22.** Given Assumption C.3, for any $\eta \in (0, 1]$,

(i) $\max_{T_n \eta \leq j \leq T_n} |G_{\xi,j} - \varrho \sum_{k=1}^{j} \sigma_k (\mathbb{B}_k - \mathbb{B}_{k-1})| = O_{a.s.} (\Delta T_n)$;

(ii) $\max_{T_n \eta \leq j \leq T_n} |V_{\xi,j} - \Xi_j| = O_{a.s.} \left(\Delta T_n + \tilde{\Xi}_j + (\Delta^2 T_n + \Xi_j) / T_n \right)$.

**Proof.** To prove part (ii) one needs part (i). However, the same steps in the initial part in the proof of (ii) can be used to prove part (i) as we explain below. Thus, we only prove part (ii). After some simple algebraic manipulations one can verify the decomposition $V_{\xi,j} - \Xi_j = U_j - G_{\xi,j}^2 / j$ where $U_j = \sum_{k=1}^{j} \sigma_k^2 (\epsilon_j^2 - 1)$. By Abel’s formula and $\epsilon_j^2 - 1 = S_j^* - S_{j-1}^*$, we have

$$U_j = \sum_{k=1}^{j} \sigma_k^2 (\tilde{S}_k - \tilde{S}_{k-1}) = (\sigma_k^2 \tilde{S}_j - \sigma_0^2 \tilde{S}_0) - \sum_{k=1}^{j-1} (\sigma_{k+1}^2 - \sigma_k^2) \tilde{S}_k$$

$$= \sigma_k^2 \tilde{S}_j - \sum_{k=1}^{j-1} (\sigma_{k+1}^2 - \sigma_k^2) \tilde{S}_k, \quad \text{(C.1.66)}$$
and by the rightmost approximation in (C.1.65) it follows that

\[ U_j = \sigma_j^2 \tilde{\varrho} \tilde{B}_j - \sum_{k=1}^{j-1} \left( \sigma_{k+1}^2 - \sigma_k^2 \right) \tilde{\varrho} \tilde{B}_k + O_{\text{a.s.}}(\Delta T_n) \]

\[ = \tilde{\varrho} \sum_{k=1}^{j} \sigma_k^2 \left( \tilde{B}_k - \tilde{B}_{k-1} \right) + O_{\text{a.s.}}(\Delta T_n). \]  

(C.1.67)

Next, by Kolmogorov’s maximal inequality for independent random variables [cf. Theorem 22.4 in Billingsley (1995)], we have for \( C > 0 \),

\[ \mathbb{P} \left[ \max_{1 \leq j \leq T_n} \left| \tilde{\varrho} \sum_{k=1}^{j} \sigma_k^2 \left( \tilde{B}_k - \tilde{B}_{k-1} \right) \right| \geq C \tilde{\Xi} T_n \right] \leq \left( C \tilde{\Xi} T_n / \tilde{\varrho} \right)^{-2} \left[ \left( \sum_{k=1}^{T_n} \sigma_k^2 \left( \tilde{B}_k - \tilde{B}_{k-1} \right) \right)^2 \right] = \left( \tilde{\varrho} / C \right)^{-2}. \]

Thus, choosing \( C \) large enough shows that \( \max_{1 \leq k \leq T_n} \tilde{\Xi} T_n / \tilde{\varrho} \sum_{k=1}^{T_n} \sigma_k^2 \left( \tilde{B}_k - \tilde{B}_{k-1} \right) < \infty \). Use this result into (C.1.67) to verify that \( \max_{1 \leq j \leq T_n} |U_j| = O_{\mathbb{P}}(\tilde{\Xi} T_n + \Delta T_n) \).

Using the same steps as in (C.1.66)-(C.1.67), one verifies

\[ \max_{1 \leq j \leq T_n} |G_{\xi,j}| = O_{\mathbb{P}} \left( \sqrt{\tilde{\Xi} T_n + \Delta T_n} \right). \]

Hence,

\[ V_{\xi,j} - \Xi_j = U_j - G_{\xi,j}^2 / j = O_{\text{a.s.}} \left( \Delta T_n + \tilde{\Xi}_j + \left( \Delta T_n^2 + \tilde{\Xi}_n \right) / T_n \right), \]

uniformly in \( 1 \leq j \leq T_n \), which proves part (ii). \( \square \)

The first part of the proof uses Lemma C.1.22 applied to the sequence of normalized forecast losses \( \{L_{\psi,kh}(\beta^*)\}_{k=T_m}^{T_m + T_n} \). We provide the proof directly for a general loss function; the case of the quadratic loss function follows as a special case.
Since we have already dealt with the discretization error above and have shown that 
\( \mu_{e,s}h^{-\vartheta} \) is negligible for \( \vartheta \in [0, 1/8) \), in this section we assume for simplicity that 
\( L_{\psi,kh}(\beta^*) = g(\Delta_h\tilde{e}_k^*; \beta^*) \), where \( \Delta_h\tilde{e}_k^* = \sigma_{e,(k-1)h}h^{-1/2}\Delta_hW_{e,k} \). Let 
\[
\mu_{T_m+\tau+(b+1)n_T-1} \triangleq \mathbb{E}\left(L_{\psi,(T_m+\tau+(b+1)n_T+j-1)h}(\beta^*)\right),
\]
for \( j = 1, \ldots, n_T \), which is justified by the fact that these variables are in the same window.

**Proof of Theorem 3.4.2.** We shall use Lemma C.1.22-(ii). Let 
\[
\xi_j \triangleq g\left(\sigma_{e,(T_m+\tau+(b+1)n_T-1)h}h^{-1/2}\Delta_hW_{e,T_m+\tau+(b+1)n_T+j-1}\right) - \mu_{T_m+\tau+(b+1)n_T-1}
\]
for \( j = 1, \ldots, n_T \). Using basic arguments, we also have 
\[
V_{h,b}^* \triangleq n_T^{-1} \sum_{j=1}^{n_T} \left( g\left(\sigma_{e,(T_m+\tau+(b+1)n_T-1)h}h^{-1/2}\Delta_{T_m+\tau+(b+1)n_T+j-1}W_e\right) - \mu_{T_m+\tau+(b+1)n_T-1}\right)^2
\]
e.g., use the initial lemmas from the proof of Theorem 3.3.2 and note that \( L_{\psi,b}(\beta^*) - \mu(T_m+\tau+(b+1)n_T-1)h = \mathcal{O}_P\left(1/\sqrt{n_T}\right) \) by a basic central limit theorem for i.i.d. variables. Then, by Lemma C.1.22-(ii) we have 
\[
\max_{0 \leq b \leq \lfloor T_n/n_T \rfloor - 2} \left| n_TV_{h,b}(\beta^*) - \Sigma_{h,b}^* \right| = \mathcal{O}_P\left(\Delta_{T_n} + \Sigma_{h,b}^*\right),
\]
where $\Delta T_n = T_n^{1/4} \log (T_n)$. Let

$$d_{h,b} \triangleq n_T V_{h,b} \left( \hat{\beta} \right) / \Sigma_{h,b}^* - 1,$$

and note that $d_{h,b} = O_P \left( \left( \Delta T_n + \hat{\Sigma}_{h,b}^* \right) / \Sigma_{h,b}^* \right)$ by proceeding as in the proof of Lemma C.1.12 and thus using $\hat{\beta}_h - \beta^* = O_P \left( T_n^{-1/2} \right)$ uniformly by Assumption 3.8. By definition $\tilde{\Sigma}_{h,b}^*/\Sigma_{h,b}^* = O_P \left( n^{-1/2} \right)$ while by Condition 3 $\Delta T_n / \Sigma^* h,b \to 0$ so that we deduce $d_{h,b} = o_P(1)$ uniformly over $b$. Let $\{B_t\}_{t \geq 0}$ be a standard Wiener process and define

$$z_{h,b} \triangleq (\Sigma^*_b)^{-1/2} \sum_{j=1}^{n_T} \sigma_{L,T_m+(b+1)n_T+j-1} \left( B(b+1)n_T+j - B(b+1)n_T+j-1 \right).$$

By the law of iterated logarithms [cf. Billingsley (1995), Theorem 9.5 in Ch. 1], $\max_{0 \leq b \leq [T_n/n_T] - 2} |z_{h,b}| = O_P \left( \sqrt{\log (T_n)} \right)$. Under $H_0$, by the Lipschitz continuity of $\mu$ we have $\mu_{T_m+\tau+(b+1)n_T-1} - \mu_{T_m+\tau+bn_T-1} = O_P \left( n_T h \right)$. This together with applying multiple times the bounds used at the beginning of the proof concerning terms involving $\hat{\beta}$ and $\beta^*$ allows us to that $\zeta_{h,b} \left( \hat{\beta} \right)$ can be approximated by

$$\zeta_{h,b}^* \triangleq \sqrt{n_T} \left( A_{h,b}(\beta^*) - \mu_{T_m+\tau+(b+1)n_T-1} - (A_{h,b-1}(\beta^*) - \mu_{T_m+\tau+bn_T-1}) \right) / V_{h,b}$$

because for small $\epsilon > 0$, $\sqrt{n_T} \left( \mu_{T_m+\tau+(b+1)n_T-1} - \mu_{T_m+\tau+bn_T-1} \right) = h^\epsilon \to 0$. Let

$$\tilde{A}_{h,b}(\beta^*) \triangleq n_T \left( A_{h,b}(\beta^*) - \mu_{T_m+\tau+(b+1)n_T-1} \right).$$
Then,

\[
\zeta_{h,b}^* = \frac{\tilde{A}_{b,h}(\beta^*)}{\sqrt{\Sigma_{h,b}^*} \left( \sqrt{n_T V_{h,b}/\Sigma_{h,b}^*} \right)} - \frac{\tilde{A}_{b-1,h}(\beta^*)}{\sqrt{\Sigma_{h,b}^*} \left( \sqrt{n_T V_{h,b}/\Sigma_{h,b}^*} \right)},
\]

\[
= \frac{\tilde{A}_{b,h}(\beta^*)}{\sqrt{\Sigma_{h,b}^*} (1 + d_{h,b})} - \frac{\tilde{A}_{b-1,h}(\beta^*)}{\sqrt{\Sigma_{h,b}^*} (1 + d_{h,b}) \left( \Sigma_{h,b-1}^*/\Sigma_{h,b-1}^* \right)},
\]

and given \( d_{h,b} \convergesP 0 \) we know that \( \sqrt{1 + d_{h,b}} = 1 + O_P(d_{h,b}) \). In view of Lemma C.1.22-(i) we have

\[
|\tilde{A}_{h,b}| \leq O_a.s. (\Delta_{T_n})
\]

Therefore, the last inequality leads to

\[
\frac{\tilde{A}_{h,b}(\beta^*)}{\sqrt{\Sigma_{h,b}^*} (1 + d_{h,b})} = (1 + d_{h,b})
\]

\[
\times \left[ \frac{\nu_L \sum_{j=1}^{n_T} \sigma_L(T_{m+\tau+(b+1)n_T+j-1}h) \left( \mathbb{B}_{(b+1)n_T+j} - \mathbb{B}_{(b+1)n_T+j-1} \right)}{\sqrt{\Sigma_{h,b}^*}} \right] + O_a.s. \left( \frac{\Delta_{T_n}}{\sqrt{\Sigma_{h,b}^*}} \right).
\]

(C.1.68)

A similar argument can be used for the second term while in addition for the denominator we use the fact that \( \Sigma_{h,b-1}^* \) is Lipschitz continuous and therefore \( \Sigma_{h,b}^* - \Sigma_{h,b-1}^* = O_P(n_T h) \), which then gives \( \sqrt{1 + d_{h,b} \sqrt{1 + O_P(n_T h)}} = 1 + O_P(d_{h,b}) \). Let \( \{B_t\}_{t \geq 0} \) be
a standard Wiener process. We can then deduce that

\[
\zeta_{b,h}^* = \frac{\nu L \sum_{j=1}^{n_T} \sigma_L(T_m + \tau + (b+1)n_T + j-1)h \left( B(b+1)n_T + j - B(b+1)n_T + j-1 \right)}{\sqrt{\Sigma_{h,b}^*}} \\
+ \frac{\nu L \sum_{j=1}^{n_T} \sigma_L(T_m + \tau + bn_T + j-1)h \left( Bbn_T + j - Bbn_T + j-1 \right)}{\sqrt{\Sigma_{h,b-1}^*}} \\
+ (1 + d_{h,b}) O_{a.s.} \left( \frac{\Delta T_n}{\sqrt{\Sigma_{h,b-1}^*}} \right).
\]

The stochastic order term in the last equation is, for some small \( \epsilon > 0 \),

\[
O_{\mathbb{P}} \left( \log (T_n) / (T_n^{\epsilon/2}) \right) \to 0,
\]

where we have used \( \Delta T_n = T_n^{1/4} \log (T_n) \), Condition 3 and \( \Sigma_{h,b}^* = O_{\mathbb{P}}(n_T) \). Using the properties of the Wiener process, we have

\[
\left( \zeta_{b,h}^* \right)^2 = \frac{\nu_L^2 \sum_{j=1}^{n_T} \sigma_L^2(T_m + \tau + (b+1)n_T + j-1)h}{\sqrt{\Sigma_b^*}} \\
+ \frac{\nu_L^2 \sum_{j=1}^{n_T} \sigma_L^2(T_m + \tau + bn_T + j-1)h}{\sqrt{\Sigma_{b-1}^*}} + O_{\mathbb{P}} \left( \log (T_n) / T_n^{\epsilon/2} \right) \\
= 2\nu_L^2 + O_{\mathbb{P}} \left( (\log (T_n))^2 / T_n^{\epsilon} \right),
\]

and therefore, \( \tilde{\nu}_L^2 = \nu_L^2 + O_{\mathbb{P}} \left( (\log (T_n))^2 / T_n^{\epsilon} \right). \) \( \square \)

**C.1.6 Proofs of Section 3.5**

**C.1.6.1 Proof of Theorem 3.5.1**

C.1.6.2 Proof of Theorem 3.5.2

Let $\Delta_h \tilde{e}_k \triangleq \Delta_h Y_k - \beta'_k \Delta_h X_{k-\tau}$, where $\beta_k = \beta^* + \mu \beta_{k\tau} / (\log(T_n) n_T)^{1/4}$. Let

$$\tilde{MQ}_{\max, h}^* (T_n, \tau) \triangleq \nu_L^{-1} \max_{i = n_T, \ldots, T_n - n_T} \left| \left( \sum_{j = i+1}^{i+n_T} (SL_{\psi, T_m + \tau + j - 1} (\beta^*) - \zeta_{\mu, j, +}) - \sum_{j = i-n_T+1}^{i} (SL_{\psi, T_m + \tau + j - 1} (\beta^*) - \zeta_{\mu, j, -}) \right) \right|. \tag{C.1.69}$$

Our final goal is to show that $(\log(T_n) n_T)^{1/2} \left( V_{\max, h} (T_n, \tau) - \tilde{MQ}_{\max, h}^* (T_n, \tau) \right) \overset{P}{\to} 0$, where

$$V_{\max, h} (T_n, \tau) = \nu_L^{-1} \max_{i = n_T, \ldots, T_n - n_T} \left| \left( \sum_{j = i+1}^{i+n_T} (\Delta_h \tilde{e}_{T_m + \tau + j - 1}^2 - \sum_{j = i-n_T+1}^{i} (\Delta_h \tilde{e}_{T_m + \tau + j - 1}^2) \right) \right|. \tag{C.1.69}$$

The result of the theorem then follows from Corollary 3.3.2.

**Lemma C.1.23.** $(\log(T_n) n_T)^{1/2} \left( \tilde{MQ}_{\max, h}^* (T_n, \tau) - \tilde{MQ}_{\max, h}^* (T_n, \tau) \right) \overset{P}{\to} 0$.

**Proof.** We have

$$\tilde{MQ}_{\max, h}^* (T_n, \tau) - \tilde{MQ}_{\max, h}^* (T_n, \tau) \leq \nu_L^{-1} \max_{i = n_T, \ldots, T_n - n_T} \left| \left( \sum_{j = i+1}^{i+n_T} (SL_{\psi, T_m + \tau + j - 1} (\beta^*) - SL_{\psi, T_m + \tau + j - 1} (\tilde{\beta}_{T_m + j - 1}) - \zeta_{\mu, j, +}) \right) \right| + \nu_L^{-1} \max_{i = n_T, \ldots, T_n - n_T} \left| \left( \sum_{j = i-n_T+1}^{i} (SL_{\psi, T_m + \tau + j - 1} (\beta^*) - SL_{\psi, T_m + \tau + j - 1} (\tilde{\beta}_{T_m + j - 1}) - \zeta_{\mu, j, -}) \right) \right|. \tag{C.1.69}$$
Note that for any \( j = i + 1, \ldots, i + n_T \),

\[
SL_{\psi,T_m+\tau+j-1}(\beta^*) - SL_{\psi,T_m+\tau+j-1}(\hat{\beta}_{T_m+j-1}) = L_{\psi,T_m+\tau+j-1}(\beta^*) - L_{\psi,T_m+\tau+j-1}(\hat{\beta}_{T_m+j-1}) + o_P\left(1/(\log (T_n)n_T)^{1/2}\right),
\]

where the \( o_P\left(1/(\log (T_n)n_T)^{1/2}\right) \) term arises from Lemma C.1.24 below. Focusing on the first two terms we have

\[
L_{\psi,T_m+\tau+j-1}(\beta^*) - L_{\psi,T_m+\tau+j-1}(\hat{\beta}_{T_m+j-1}) - \zeta_{\mu,j,+} = (\beta^* - \hat{\beta}_{T_m+j-1})' \Delta_h \tilde{X}_{T_m+j-1} \Delta_h \tilde{X}'_{T_m+j-1} (\beta^* - \hat{\beta}_{T_m+j-1}) - \zeta_{\mu,j,+} + 2\sigma_e(T_m+\tau+i)h\left(h^{-1/2} \Delta_h W_{e,T_m+\tau+j-1}\right) (\beta^* - \hat{\beta}_{T_m+j-1})' \Delta_h \tilde{X}_{T_m+j-1}.
\]

(C.1.70)

We deal explicitly with the first term on the right-hand side above in (C.1.77) below and show that it is \( o_P\left(1/\log (T_n)n_T\right) h^{1/4} \). Moving to the second term, by Theorem 13.3.7 in Jacod and Protter (2012) we have

\[
n^{-1/2} \sum_{j=i+1}^{i+n_T} h^{-1/2} \Delta_h W_{e,T_m+\tau+j-1} \Delta_h \tilde{X}_{T_m+j-1} = O_P(1).\]

Using Assumption 3.10, \( \hat{\beta}_{T_m+j-1} - \beta^* = O_P\left(1/\log (T_n)n_T\right)^{-1/4} \) uniformly in \( j \) and
thus, for any $\varepsilon > 0$,

$$\mathbb{P}\left( \max_{i=n_T, \ldots, T_n - n_T} \nu^{-1}_L \left( \log \left( T_n \right) n_T \right)^{1/2} \cdot \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( SL_{T_m+r+j-1} (\beta^*) - SL_{T_m+r+j-1} (\hat{\beta}_{T_m+j-1}) \right) \right| > \varepsilon \right)$$

$$= \left( \frac{1}{\nu_L \varepsilon} \right)^r \left( \log \left( T_n \right) n_T \right)^{r/2} \sum_{i=n_T}^{T_n - n_T} \left[ \nu^{-1}_L \left( \log \left( T_n \right) n_T \right)^{r/4} \sum_{j=i+1}^{i+n_T} \left( SL_{T_m+r+j-1} (\beta^*) - SL_{T_m+r+j-1} (\hat{\beta}_{T_m+j-1}) \right) \right]^\gamma$$

$$\leq \left( \frac{C}{\nu_L \varepsilon} \right)^r \left( \log \left( T_n \right) n_T \right)^{r/4} O_p \left( n_T^{-r/2} \right) \to 0,$$

for $r > 0$ sufficiently large. The same bound applies to the term in (C.1.69) and this proves the claim of the lemma. □

**Lemma C.1.24.** As $h \downarrow 0$,

$$\nu^{-1}_L \max_{i=n_T, \ldots, T_n - n_T} \left( \log \left( T_n \right) n_T \right)^{1/2} \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( L_{\psi,T_{m+j-1}} (\beta^*) \right) - \sum_{j=i-n_T+1}^{i} \left( \nu^{-1}_L \left( \log \left( T_n \right) n_T \right)^{1/2} \right) \right| \to 0,$$

and

$$\nu^{-1}_L \max_{i=n_T, \ldots, T_n - n_T} \left( \log \left( T_n \right) n_T \right)^{1/2} \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( L_{\psi,T_{m+j-1}} (\beta^*) - L_{\psi,T_{m+j-1}} (\hat{\beta}_{T_m+j-1}) \right) \right| \to 0.$$
Proof. By definition,

\[ n^{-1}_T \sum_{j=i+1}^{i+n_T} T_{\psi,T_{m+j-1}}(\beta^*) - \sum_{j=i-n_T+1}^{i} T_{\psi,T_{m+j-1}}(\beta^*) \]

\[ \leq \left| n^{-1}_T \sum_{j=i+1}^{i+n_T} \sum_{l=1}^{T_{m+j-1}} \left( \Delta_h \tilde{e}_{l+\tau} \right)^2 \frac{1}{T_{m+j-1} - 1} - n^{-1}_T \sum_{j=i-n_T+1}^{i} \sum_{l=1}^{T_{m+j-1}} \left( \Delta_h \tilde{e}_{l+\tau} \right)^2 \frac{1}{T_{m+j-1} - 1} \right| \]

\[ + \left| n^{-1}_T \sum_{j=i+1}^{i+n_T} \sum_{l=1}^{T_{m+j-1}} \mu'_{\beta,(l+\tau)} \Delta_h \tilde{X}_l \Delta_h \tilde{X}'_l \mu_{\beta,(l+\tau)} \frac{1}{(T_{m+j-1} - 1) (\log T_n n_T)^{1/2}} \right| \]

\[ - n^{-1}_T \sum_{j=i-n_T+1}^{i} \sum_{l=1}^{T_{m+j-1}} \mu'_{\beta,(l+\tau)} \Delta_h \tilde{X}_l \Delta_h \tilde{X}'_l \mu_{\beta,(l+\tau)} \frac{1}{(T_{m+j-1} - 1) (\log T_n n_T)^{1/2}} \]

\[ + 2 \left| n^{-1}_T \sum_{j=i+1}^{i+n_T} \sum_{l=1}^{T_{m+j-1}} \Delta_h \tilde{e}_{l+\tau} \mu'_{\beta,(l+\tau)} \Delta_h \tilde{X}_l \Delta_h \tilde{X}'_l \mu_{\beta,(l+\tau)} \frac{1}{(T_{m+j-1} - 1) (\log T_n n_T)^{1/2}} \right| \]

\[ - n^{-1}_T \sum_{j=i-n_T+1}^{i} \sum_{l=1}^{T_{m+j-1}} \Delta_h \tilde{e}_{l+\tau} \mu'_{\beta,(l+\tau)} \Delta_h \tilde{X}_l \Delta_h \tilde{X}'_l \mu_{\beta,(l+\tau)} \frac{1}{(T_{m+j-1} - 1) (\log T_n n_T)^{1/2}} \]

\[ \triangleq A_{1,h} + A_{2,h} + A_{3,h}. \]

Observe that the leading term is \( A_{1,h} \) and thus it is sufficient to establish a bound for it. By proceeding as in the proof of Lemma C.1.3 we shall use

\[ (T_m + i - 1)^{-1} \sum_{l=1}^{T_{m+i-1-\tau}} (\Delta_h \tilde{e}_{l+\tau})^2 = O_T(1). \]

Then, \( A_{1,h} \) is less than

\[ \left| n^{-1}_T \sum_{j=i+1}^{i+n_T} \sum_{l=1}^{T_{m+j-1-\tau}} \frac{\left( \Delta_h \tilde{e}_{l+\tau}^2 \right)^2}{T_{m+j-1}} \right| \]

\[ - n^{-1}_T \sum_{j=i-n_T+1}^{i} \sum_{l=1}^{T_{m+j-1-\tau}} \frac{\left( \Delta_h \tilde{e}_{l+\tau}^2 \right)^2}{T_{m+j-1}} \left( \frac{1}{T_{m+j} - 1} - \frac{1}{T_{m+j-1}} \right) \]

\[ = \frac{2n_T}{T_m} O_T(1) + \frac{n_T}{T_m} O_T(1). \]
This leads to

$$\left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \mathcal{L}_{\psi,T_m+j-1} (\beta^*) - n_T^{-1} \sum_{j=i-n_T+1}^{i} \mathcal{L}_{\psi,T_m+j-1} (\beta^*) \right| \leq CO_{\bar{\mathcal{P}}} \left( \frac{n_T}{T_m} \right).$$

Thus, for any $\varepsilon > 0$,

$$\mathbb{P} \left( \nu_L^{-1} \max_{i=n_T, \ldots, T_n-n_T} (\log (T_n) n_T)^{1/2} \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \mathcal{L}_{\psi,T_m+j-1} (\beta^*) - n_T^{-1} \sum_{j=i-n_T+1}^{i} \mathcal{L}_{\psi,T_m+j-1} (\beta^*) \right| > \varepsilon \right) \leq \varepsilon^{-r} \sum_{i=n_T}^{T_n-n_T} \mathbb{E} \left[ (\log (T_n) n_T)^{r/2} \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \mathcal{L}_{\psi,T_m+j-1} (\beta^*) - n_T^{-1} \sum_{j=i-n_T+1}^{i} \mathcal{L}_{\psi,T_m+j-1} (\beta^*) \right|^r \right] \leq \varepsilon^{-r} C (\log (T_n) n_T)^{r/2} O_{\bar{\mathcal{P}}} \left( n_T^{-1/\tau} \right) \to 0,$$  

for $r > 0$ sufficiently large in view of Condition 2 and that $T_n = O(T_m)$. For the second claim of the lemma, note that

$$\mathcal{L}_{\psi,T_m+j-1} (\beta) - \mathcal{L}_{\psi,T_m+j-1} (\beta^*)$$

$$= \frac{T_m + j - 1 - \tau}{T_m + j - 1 - \tau} - \frac{\Delta_h \tilde{e}_{l+\tau}}{T_m + j - 1 - \tau} + \frac{(\Delta_h \tilde{e}_{l+\tau})^2}{T_m + j - 1 - \tau}$$

$$= \frac{1}{T_m + j - 1 - \tau} \sum_{l=1}^{T_m+j-1-\tau} (\tilde{\beta}_{T_m+j-1} - \beta^*)' (\Delta_h \tilde{X}_l \Delta_h \tilde{X}'_l) \tilde{e}_{l+\tau}.$$  

By Lemma C.1.1, $(T_m + j - 1)^{-1} \sum_{l=1}^{T_m+j-1-\tau} \Delta_h \tilde{X}_l \Delta_h \tilde{X}'_l$ is $O_{\bar{\mathcal{P}}} (1)$ while by Assumption 3.10, $\tilde{\beta}_k - \beta^* = O_{\bar{\mathcal{P}}} \left( 1 / (\log (T_n) n_T)^{1/4} \right)$ uniformly in $k$. It follows that the term
in equation (C.1.72) is \( O_P \left( (\log (T_n) n_T)^{1/2} \right) \) whereas the term in (C.1.73) is such that

\[
\frac{2}{T_m + j - 1} \sum_{l=1}^{T_m + j - 1 - \tau} \Delta_h \tilde{e}_{l+\tau} \left( \hat{\beta}_{T_m + j - 1} - \beta^* \right)' \Delta_h \tilde{X}_l \leq 2C \sup_j \left\| \hat{\beta}_{T_m + j - 1} - \beta^* \right\| \frac{1}{T_m + j - 1} \sum_{l=1}^{T_m + j - 1 - \tau} \Delta_h \tilde{e}_{l+\tau}' \Delta_h \tilde{X}_l \]

\[
= o_P \left( T_m^{-1/2} (\log (T_n) n_T)^{1/4} \right),
\]

where \( \iota \) is a \( q \times 1 \) unit vector and we have used the central limit theorem in Lemma C.1.1-(iii). Therefore, upon using the same argument that led to (C.1.71) and Condition 2 we have the last claim of the lemma. \( \square \)

**Lemma C.1.25.** \((\log (T_n) n_T)^{1/2} \left( V_{\text{max},h} (T_n, \tau) - \hat{\mathbf{MQ}}_{\text{max},h}^* - (T_n, \tau) \right) \xrightarrow{p} 0.\)**

**Proof.** Note that \( SL_{T_m + \tau + j - 1} (\beta^*) \) can be expanded as follows:

\[
SL_{T_m + \tau + j - 1} (\hat{\beta}) = L_{\psi,T_m + \tau + j - 1} (\hat{\beta}) - L_{\psi,T_m + \tau + j - 1} (\hat{\beta}) \]

\[
= (\Delta_h \tilde{e}_{T_m + \tau + j - 1})^2 + (\beta^* - \hat{\beta}_{T_m + j - 1})' \Delta_h \tilde{X}_{T_m + j - 1} \Delta_h \tilde{X}_{T_m + j - 1}' (\beta^* - \hat{\beta}_{T_m + j - 1}) \]

\[
- 2\Delta_h \tilde{e}_{T_m + \tau + j - 1} (\beta^* - \hat{\beta}_{T_m + j - 1})' \Delta_h \tilde{X}_{T_m + j - 1} - L_{T_m + \tau + j - 1} (\hat{\beta}).
\]
Then we can write (omitting the index from $\tilde{\beta}$),

\[
(\log (T_n n_T)^{1/2} \left( V_{\text{max},h} (T_n, \tau) - \bar{\mathcal{Q}}^*_{\text{max},h} - (T_n, \tau) \right) \right) \leq \\
\max_{i=n_T, \ldots, T_n-n_T} (\log (T_n n_T)^{1/2} \nu_L^{-1})
\times \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( (\beta^* - \tilde{\beta}) \Delta_h \bar{X}_{T_m+j-1} \Delta_h \bar{Y}_{T_m+j-1} (\beta^* - \tilde{\beta}) - \zeta_{\mu,j,+} \right) \right|
\times \left| n_T^{-1} \sum_{j=i-n_T+1}^{i} \left( (\beta^* - \tilde{\beta}) \Delta_h \bar{X}_{T_m+j-1} \Delta_h \bar{Y}_{T_m+j-1} (\beta^* - \tilde{\beta}) - \zeta_{\mu,j,-} \right) \right|
+ \max_{i=n_T, \ldots, T_n-n_T} (\log (T_n n_T)^{1/2} \nu_L^{-1})
\times \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} T_{\psi,T_m+j-1} (\beta^*) - n_T^{-1} \sum_{j=i-n_T+1}^{i} T_{\psi,T_m+j-1} (\beta^*) \right|
= A_{1,h} + A_{2,h} + A_{3,h}.
\]

Our goal is to show that $A_{l,h} \xrightarrow{P} 0$ for $l = 1, 2, 3$. By Lemma C.1.24 we know that $A_{3,h} \xrightarrow{P} 0$. Let us focus on $A_{1,h}$. Note that

\[
A_{1,h} \leq \max_{i=n_T, \ldots, T_n-n_T} (\log (T_n n_T)^{1/2} \nu_L^{-1})
\times \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \left( (\beta^* - \tilde{\beta}_{T_m+j-1}) \Delta_h \bar{X}_{T_m+j-1} \Delta_h \bar{Y}_{T_m+j-1} (\beta^* - \tilde{\beta}_{T_m+j-1}) - \zeta_{\mu,j,+} \right) \right|
+ \max_{i=n_T, \ldots, T_n-n_T} (\log (T_n n_T)^{1/2} \nu_L^{-1}) \times
\left| n_T^{-1} \sum_{j=i-n_T+1}^{i} \left( (\beta^* - \tilde{\beta}_{T_m+j-1}) \Delta_h \bar{X}_{T_m+j-1} \Delta_h \bar{Y}_{T_m+j-1} (\beta^* - \tilde{\beta}_{T_m+j-1}) - \zeta_{\mu,j,-} \right) \right|.
\]
We have \( \beta^* - \tilde{\beta}_{T_n + j - 1} = \mu_{\beta,(T_n + j - 1)h} / (\log (T_n) n_T)^{1/4} \) and

\[
h^{-1/4} \left( \sum_{j=1+1}^{i+n_T} \Delta_h \tilde{X}_{T_n + j - 1} \Delta_h \tilde{X}'_{T_n + j - 1} - \Sigma_X(T_{n+i})h \right) = O_P(1),
\]

by Theorem 13.3.7 in Jacod and Protter (2012). Upon using the property of the trace operator we have

\[
\frac{(\log (T_n) n_T)^{-1/2}}{n_T} \sum_{j=i+1}^{i+n_T} \mu'_{\beta,(T_n + \tau + j - 1)h} \left( \Delta_h \tilde{X}_{T_n + j - 1} \Delta_h \tilde{X}'_{T_n + j - 1} - \Sigma_X(T_{n+i})h \right) \mu_{\beta,(T_n + \tau + j - 1)h} = O_P \left( (\log (T_n) n_T)^{-1/2} h^{1/4} \right).
\]

Thus, the first term on the right-hand side of (C.1.76) is less than

\[
C_r \left( \frac{1}{\nu_L} \right)^r T_n^{-n_T} \sum_{i=n_T}^{n_T} \mathbb{E} \left[ \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} l' \left( \Delta_h \tilde{X}_{T_n + j - 1} \Delta_h \tilde{X}'_{T_n + j - 1} - \zeta_{\mu,j,\tau} \right) l \right|^r \right] \leq C_r \left( \frac{1}{\nu_L} \right)^r T_n h^{r/4} \rightarrow 0,
\]

for \( r > 0 \) sufficiently large given that \( h = O(T^{-1}) = O(T_n^{-1}) \) and \( \varepsilon > 0 \). The same argument can be applied to the second term of (C.1.76) which then yields \( A_{1,h} = o_P(1) \). It remains to consider \( A_{2,h} \). It is sufficient to show that

\[
\max_{i=n_T, \ldots, n_T^{-1}} (\log (T_n) n_T)^{1/2} \nu_L^{-1} \left| 2n_T^{-1} \sum_{j=i+1}^{i+n_T} \Delta_h \tilde{e}_{T_n + \tau + j - 1} \left( \beta^* - \tilde{\beta}_{T_n + j - 1} \right)' \Delta_h \tilde{X}_{T_n + j - 1} \right| \overset{p}{\rightarrow} 0.
\]
By Theorem 13.3.7 in Jacod and Protter (2012) we now have

\[
n_T^{-1/2} \sum_{j=i+1}^{i+n_T} \Delta_h \tilde{e}_{T_{m+r+j-1}} \Delta_h \tilde{X}_{T_{m+j-1}} < \infty.
\]

By Marlov’s inequality, for any \( \varepsilon > 0 \) we have

\[
\mathbb{P} \left( \max_{i=n_T, \ldots, T_n} (\log (T_n) n_T)^{1/2} \nu_L^{-1} \left| 2n_T^{-1} \sum_{j=i+1}^{i+n_T} \Delta_h \tilde{e}_{T_{m+r+j-1}} \left( \beta^* - \beta_{T_{m+j-1}} \right)' \Delta_h \tilde{X}_{T_{m+j-1}} \right| > \varepsilon \right)
\]

\[
\leq C_r \left( \frac{2}{\nu_L \varepsilon} \right)^r (\log (T_n) n_T)^{r/2-r/4} \times \sum_{i=n_T}^{T_n-n_T} \mathbb{E} \left[ n_T^{-1} \sum_{j=i+1}^{i+n_T} \Delta_h \tilde{e}_{T_{m+r+j-1}} \Delta_h \tilde{X}_{T_{m+j-1}} \right]^r
\]

\[
\leq C_r \left( \frac{2}{\nu_L \varepsilon} \right)^r (\log (T_n) n_T)^{r/2-r/4} \mathbb{O}_P \left( T_n^{-r/2} n_T^{-r/2} \right) \to 0
\]

\( r > 0 \) sufficiently large. Thus \( A_{2,h} \xrightarrow{\mathbb{P}} 0 \) which in turn concludes the proof. □

*Proof of Theorem 3.5.2.* From Lemma C.1.23-C.1.25

\[
(\log (T_n) n_T)^{1/2} \left( V_{\text{max,}h} (T_n, \tau) - \tilde{M}Q_{\text{max,}h} - (T_n, \tau) \right) \xrightarrow{\mathbb{P}} 0.
\]

The result then follows from Corollary 3.3.1. □

**C.1.6.3 Proof of Corollary 3.5.1**

*Proof.* Since the statistic \( \tilde{M}Q_{\text{max,}h} (T_n, \tau) \) admits a limit theorem by Theorem 3.5.2, it is sufficient to show that, conditional on \( \{\sigma_{X,t}\}_{t \geq 0} \), for all \( i = n_T, \ldots, T_n - n_T \),

\[
(\log (T_n) n_T)^{1/2} c \left| n_T^{-1} \sum_{j=i+1}^{i+n_T} \zeta_{\mu,j,+} - n_T^{-1} \sum_{j=i-n_T+1}^{i} \zeta_{\mu,j,-} \right| \to \infty,
\]
or that

\[
\begin{aligned}
&\sum_{j=i+1}^{i+n_T} \mu'_{\beta_i(T_m+\tau+j-1)h} \sigma_{X_i(T_m+\tau+i-1)h} \mu'_{\beta_i(T_m+\tau+j-1)h} \\
&= n_T^{-1} \sum_{j=i-n_T+1}^{i} \mu'_{\beta_i(T_m+\tau+j-1)h} \sigma_{X_i(T_m+\tau+i-n_T-1)h} \mu'_{\beta_i(T_m+\tau+j-1)h}
\end{aligned}
\]  

(C.1.79)

for all \(i = n_T, \ldots, T_n - n_T\) does not hold. Suppose by contradiction that (C.1.79) holds. Due to the block-wise structure of the statistic, we know that \(\sigma_{X_i(T_m+\tau+i-1)h} = \sigma_{X_i(T_m+\tau+i-1)h}\) for all \(j = i + 1, \ldots, i + n_T\) and \(\sigma_{X_i(T_m+\tau+j-1)h} = \sigma_{X_i(T_m+\tau+i-n_T-1)h}\) for all \(j = i - n_T, \ldots, i\). Then, (C.1.79) implies

\[
\begin{aligned}
\mu'_{\beta_i(T_m+\tau+i-1)h} \sigma_{X_i(T_m+\tau+i-1)h} \mu'_{\beta_i(T_m+\tau+i-1)h} \\
&= \mu'_{\beta_i(T_m+\tau+i-n_T-1)h} = \sigma_{X_i(T_m+\tau+i-n_T-1)h} \mu'_{\beta_i(T_m+\tau+i-n_T-1)h}
\end{aligned}
\]

for all \(i\). This holds if and only if the process \(\left\{z_i\right\}_{i=n_T}^{T_n-n_T}\) defined by

\[
z_i \triangleq \mu'_{\beta_i(T_m+\tau+i-1)h} \sigma_{X_i(T_m+\tau+i-1)h} \mu'_{\beta_i(T_m+\tau+i-1)h}
\]

is constant. But this is a contradiction because it is non-smooth by assumption (if only \(\mu'_{\beta_i(T_m+\tau+i-1)h}\) is non-smooth then \(z_i\) is still non-smooth because \(\sigma_{X_i(T_m+\tau+i-1)h} > 0\) \(\mathbb{P}\)-a.s. by assumption.) □
C.2 Additional Figures Related to Section 3.7

![Graphs showing power functions for model P1b with T = 100 and T = 150.](image)

**Figure C.1:** Power functions for model P1b with $T = 100$ and $T = 150$

Small sample power functions for model P1b: $Y_t = 2.73 - 0.44X_{t-1} + \delta X_{t-1} 1\{t > T_0^b\} + e_t$, where $X_{t-1} \sim \text{i.i.d.} \mathcal{N}(1, 1)$, $e_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$, and $T_0^b = T \lambda_0$. The sample size is $T = 100$ (left panels) and $T = 150$ (right panels). The fractional break date is $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = 0.4T$ while out-of-sample size is $T_n = 0.6T$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t^{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure C.2: Power functions for model P1b with $T = 200$ and $T = 300$
Small sample power functions for model P1b. The sample size is $T = 200$ (left panels) and $T = 300$ (right panels). The notes of Figure C.1 apply.
Figure C.3: Power functions for model P2 with $T = 100$ and $T = 150$

Small sample power functions for model P2: $Y_t = X_{t-1} + \delta X_{t-1} 1 \{t > T_0^b\} + e_t$ where $X_t$ is a Gaussian AR(1) with autoregressive coefficient 0.4 and unit variance, and $e_t \sim \text{i.i.d. } \mathcal{N}(0, 0.49)$. The sample size is $T = 100$ (left panel) and $T = 200$ (right panel). The fractional break date is $\lambda_0 = 0.7$ (top panel) and $\lambda_0 = 0.8$ (bottom panel). In-sample size is $T_m = 0.4T$ while out-of-sample size is $T_n = 0.6T$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure C.4: Power functions for model P4 with $T = 200$ and $T = 300$
Small sample power functions for model P4 (recurrent break in mean): $Y_t = \beta_t + \epsilon_t$, where $\beta_t$ switches between $\delta$ and 0 every $p$ periods and $\epsilon_t \sim \text{i.i.d.} \mathcal{N}(0, 0.64)$. We set $(T, p) = \{(200, 30), (300, 40)\}$. The fractional break date is $\lambda_0 = 0.5$ (top panels) and $\lambda_0 = 0.6$ (bottom panels). In-sample size is $T_m = T\lambda_0$ while out-of-sample size is $T_n = T(1 - \lambda_0)$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure C.5: Power functions for model P3 with $T = 400$ and $T = 500$
Small sample power functions for model P3. We set $(T, p) = \{(400, 30), (500, 40)\}$. The fractional break date is $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = T\lambda_0$ while out-of-sample size is $T_n = T(1 - \lambda_0)$. The notes of Figure C.4 apply.
Figure C.6: Power functions for model P6 with $T = 200$ and $T = 300$

Small sample power functions for model P6 (recursive break in variance): $Y_t = \mu + (1 + \beta_t) e_t$, where $\beta_t$ switches between $\delta$ and 0 every $p$ periods and $e_t \sim \text{i.i.d.}\ N(0, 0.49)$. We set $(T, p) = \{(200, 30), (300, 40)\}$. The fractional break date is $\lambda_0 = 0.5$ (top panels) and $\lambda_0 = 0.6$ (bottom panels). In-sample size is $T_m = T\lambda_0$ while out-of-sample size is $T_n = T\lambda_0$. The green and blue broken lines correspond to $B_{\text{max}, h}$ and $Q_{\text{max}, h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure C.7: Power functions for model P4 with $T = 400$ and $T = 500$
Small sample power functions for model P4 (single break in variance). The sample size is $T = 400$ (left panels) and $T = 500$ (right panels). The notes of Figure C.6 apply.
Small sample power functions for model P5 (recursive break in variance). We set \((T, p) = \{(200, 30), (300, 40)\}\). The fractional break date is \(\lambda_0 = 0.7\) (top panels) and \(\lambda_0 = 0.8\) (bottom panels). The notes of Figure C.6 apply.
Figure C.9: Power functions for model P5 with $T = 300$ and $T = 400$
Small sample power functions for model P5 (recurrent break in variance). The sample size is $T = 300$ (left panels) and $T = 400$ (right panels). The fractional break date is $\lambda_0 = 0.6$ (top panels) and $\lambda_0 = 0.7$ (bottom panels). The notes of Figure C.6 apply.
Figure C.10: Power functions for model P6 with $T = 200$ and $T = 300$

Small sample power functions for model P6 (lagged dependent variables): $Y_t = \delta 1 \{t > T^0_b\} + 0.3Y_{t-1} + e_t$, $e_t \sim \text{i.i.d.} \mathcal{N}(0, 0.49)$. The sample size is $T = 200$ (left panels) and $T = 300$ (right panels). The fractional break date is $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = 0.4T$ while out-of-sample size is $T_n = 0.6T$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
**Figure C.11:** Power functions for model P6 with $T = 200$ and $T = 300$

Small sample power functions for model P8 (autocorrelated errors): $Y_t = 1 + X_{t-1} + \delta X_{t-1} I\{t > T_0\} + e_t$, where $X_{t-1} \sim \text{i.i.d.} \mathcal{N}(0, 1.4)$ and $e_t = 0.4u_{t-1} + u_t$, $u_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$. The sample size is $T = 200$ (left panels) and $T = 300$ (right panels). The fractional break date is $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = 0.5T$ while out-of-sample size is $T_n = 0.5T$. The green and blue broken lines correspond to $B_{\text{max,}h}$ and $Q_{\text{max,}h}$, respectively. The red and orange broken lines correspond to the $t^{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure C.12: Power functions for model P1b with $T = 100$ and $T = 150$

Small sample power functions for model P1b with short-term instability: $Y_t = 2.73 - 0.44X_{t-1} + \delta X_{t-1}1 \{T_b^0 < t \leq T_b^0 + p\} + e_t$, where $X_{t-1} \sim \text{i.i.d.} \mathcal{N}(1, 1)$, $e_t \sim \text{i.i.d.} \mathcal{N}(0, 1)$, and $T_b^0 = T\lambda_0$. We set $(T, p) = \{(100, 20), (150, 25)\}$. The fractional break date is $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = 0.4T$ while out-of-sample size is $T_n = 0.6T$. The green and blue broken lines correspond to $B_{\text{max,}h}$ and $Q_{\text{max,}h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Figure C.13: Power functions for model P1b with $T = 100$ and $T = 200$

Small sample power functions for model P2 with short-term instability: $Y_t = X_{t-1} + \delta X_{t-1} 1 \{ T_b^0 < t \leq T_b^0 + p \} + e_t$, where $X_{t-1}$ is a Gaussina AR(1) with autoregressive coefficient 0.4 and unit variance, and $e_t \sim \text{i.i.d.} \mathcal{N}(0, 0.49)$, and $T_b^0 = T \lambda_0$. We set $(T, p) = \{(100, 20), (200, 30)\}$. The fractional break date is $\lambda_0 = 0.7$ (top panels) and $\lambda_0 = 0.8$ (bottom panels). In-sample size is $T_m = 0.4T$ while out-of-sample size is $T_n = 0.6T$. The green and blue broken lines correspond to $B_{\text{max},h}$ and $Q_{\text{max},h}$, respectively. The red and orange broken lines correspond to the $t_{\text{stat}}$ of Giacomini and Rossi (2009), respectively, the uncorrected and corrected version.
Appendix D

Supplement to Chapter 4: Generalized Laplace Inference in Multiple Change-Points Models

D.1 Mathematical Appendix

The mathematical appendix is structured as follows. Section D.1.2 presents some preliminary lemmas which will be used in the sequel. The proofs of the theoretical results in the paper can be found in Section D.1.3-D.1.5.

D.1.1 Additional Notation

The $(i, j)$ element of $A$ are denoted by $A^{(i,j)}$. For a matrix $A$, the orthogonal projection matrices $P_A$, $M_A$ are defined as $P_A = A (A' A)^{-1} A'$ and $M_A = I - P_A$, respectively. Also, for a projection matrix $P$, $\|PA\| \leq \|A\|$. We denote the $d$-dimensional identity matrix by $I_d$. When the context is clear we omit the subscript notation in the projection matrices. We denote the $(i, j)$-th element of a matrix $A$ as $(A)_{i,j}$ and the $i \times j$ upper-left (resp., lower-right) sub-block of $A$ as $[A]_{i \times j}$ (resp., $[A]_{j \times i}$). Note that the norm of $A$ is equal to the square root of the maximum eigenvalue of $A'A$, and thus, $\|A\| \leq [\text{tr} (A'A)]^{1/2}$. For a sequence of matrices $\{A_T\}$, we write $A_T = o_p(1)$ if
each of its elements is $o_{p}(1)$ and likewise for $O_{p}(1)$. For a random variable $\xi$ and a number $r \geq 1$, $\|\xi\|_{r} = (\mathbb{E} \|\xi\|^{r})^{1/r}$. $K$ is a generic constant that may vary from line to line; we may sometime write $K_{r}$ to emphasize the dependence of $K$ on a number $r$. For two scalars $a$ and $b$ the symbol $a \wedge b = \inf \{a, b\}$. We may use $\sum_{k}$ when the limit of the summation are clear from the context. Unless otherwise sated $A^{c}$ denotes the complementary set of $A$.

D.1.2 Preliminary Lemmas

In this subsection, we present results related to the extremum criterion function $Q_{T}(\delta(T_{b}), T_{b})$ under the following assumption.

Assumption D.1. We consider model (4.2.3) with Assumption 4.1-4.4 and 4.8-4.9. Also Assumption 4.7 holds with $T^{1/4} v_{T} \rightarrow C < \infty$ where $C > 0$.

Lemma D.1.1. The following inequalities hold $\mathbb{P}$-a.s.:

\[
\begin{align*}
(Z_{0}'M Z_{0}) - (Z_{0}'M Z_{2})(Z_{2}'M Z_{2})^{-1}(Z_{2}'M Z_{0}) & \geq D'(X_{\Delta}X_{\Delta})(X_{\Delta}'X_{\Delta})^{-1}(X_{0}'X_{0}) D, & T_{b} < T_{b}^{0} \\
(Z_{0}'M Z_{0}) - (Z_{0}'M Z_{2})(Z_{2}'M Z_{2})^{-1}(Z_{2}'M Z_{0}) & \geq D'(X_{\Delta}X_{\Delta})(X'X - X_{\Delta}'X_{\Delta})^{-1}(X'X - X_{0}'X_{0}) D, & T_{b} \geq T_{b}^{0}
\end{align*}
\]


Recall that $Q_{T}(\delta(T_{b}), T_{b}) = \delta(T_{b}) (Z_{2}'M Z_{2}) \delta(T_{b})$. We decompose $Q_{T}(\delta(T_{b}), T_{b}) - Q_{T}(\delta(T_{b}^{0}), T_{b}^{0})$ into a “deterministic” and a “stochastic” component. It follows by definition that,

\[
\delta(T_{b}) = (Z_{2}'M Z_{2})^{-1}(Z_{2}'M Y) = (Z_{2}'M Z_{2})^{-1}(Z_{2}'M Z_{0}) \delta T + (Z_{2}'M Z_{2})^{-1}Z_{2}M e,
\]
\[ \hat{\delta}(T^0_b) = (Z'_0MZ_0)^{-1}(Z'_0MY) = \delta_T + (Z'_0MZ_0)^{-1}(Z'_0Me). \]

Therefore

\[ Q_T(\hat{\delta}(T_b), T_b) - Q_T(\hat{\delta}(T^0_b), T^0_b) \]

\[ = \hat{\delta}(T_b)'(Z'_2MZ'_2)\hat{\delta}(T_b) - \hat{\delta}(T^0_b)'(Z'_0MZ_0)\hat{\delta}(T^0_b) \]

\[ \triangleq g_d(\delta_T, T_b) + g_e(\delta_T, T_b), \]

where

\[ g_d(\delta_T, T_b) = \delta_T' \left\{ (Z'_0MZ_2)(Z'_2MZ_2)^{-1}(Z'_2MZ_0) - Z'_0MZ_0 \right\} \delta_T, \]

(D.1.4)

and

\[ g_e(\delta_T, T_b) = 2\delta_T'(Z'_0MZ_2)(Z'_2MZ_2)^{-1}Z_2Me - 2\delta_T'(Z'_0Me) \]

\[ + e'MZ_2(Z'_2MZ_2)^{-1}Z_2Me - e'MZ_0(Z'_0MZ_0)^{-1}Z_0Me. \]

(D.1.5)

(D.1.4) constitutes the deterministic component while \( g_e(\delta_T, T_b) \) the stochastic one.

Denote

\[ X_\Delta \triangleq X_2 - X_0 = \begin{pmatrix} 0, \ldots, 0, x(T_b+1), \ldots, x_{(T_b+1)h}, 0, \ldots, x_{h} \end{pmatrix}', \quad \text{for } T_b < T^0_b \]

\[ X_\Delta \triangleq -(X_2 - X_0) = \begin{pmatrix} 0, \ldots, 0, x(T^0_b+1), \ldots, x_{T^0_bh}, 0, \ldots, x_{h} \end{pmatrix}', \quad \text{for } T_b > T^0_b \]

whereas \( X_\Delta \triangleq 0 \) when \( T_b = T^0_b \). Observe that \( X_2 = X_0 + X_\Delta \text{sign}(T^0_b - T_b) \). When the sign is immaterial, we simply write \( X_2 = X_0 + X_\Delta \). Next, let \( Z_\Delta = X_\Delta D \), and
define
\[ \overline{g}_d (\delta_T, T_b) \triangleq \frac{g_d (\delta_T, T_b)}{|T_b - T_b^0|}. \] (D.1.7)

We arbitrarily define \( \overline{g}_d (\delta^0, T_b) = \delta^0_T \delta_T \) when \( T_b = T_b^0 \) since both the numerator and denominator of \( \overline{g}_d (\delta_T, T_b) \) are zero. Observe that \( \overline{g}_d (\delta_T, T_b) \) is non-negative because the matrix inside the braces in (D.1.4) is negative semidefinite. (D.1.3) can be written as
\[ Q_T \left( \hat{\delta} (T_b), T_b \right) - Q_T \left( \hat{\delta} (T_b^0), T_b^0 \right) = -|T_b - T_b^0| \overline{g}_d (\delta_T, T_b) + g_e (\delta_T, T_b), \] for all \( T_b \).

We use the notation \( u = T \| \delta_T \|^2 (\lambda_b - \lambda_0) \) and \( T_b = T \lambda_b \). For \( \eta > 0 \), let \( B_{T,\eta} \triangleq \{ T_b : |T_b - T_b^0| \leq T \eta \} \), \( B_{T,K} \triangleq \{ T_b : |T_b - T_b^0| \leq K / \| \delta_T \|^2 \} \) and
\[ B_{T,K}^c \triangleq \{ T_b : T \eta \geq |T_b - T_b^0| > K / \| \delta_T \|^2 \}, \]
with \( K > 0 \). Note that \( B_{T,\eta} = B_{T,K} \cup B_{T,K}^c \). Let \( B_{T,K}^c \triangleq \{ T_b : |T_b - T_b^0| > T \eta \} \).

**Lemma D.1.2.** Under Assumption D.1,
\[ Q_T (\delta (T_b), T_b) - Q_T (\delta (T_b^0), T_b^0) = -\delta_T^T \Sigma_\Delta \Sigma_\Delta^\prime \delta_T + 2 \text{sgn} (T_b^0 - T_b) \delta_T^T \Sigma_\Delta^\prime \Sigma_\Delta e + o_P (1), \]
uniformly on \( B_{T,K} \) for \( K \) large enough.

**Proof.** It follows from Lemma A.5 in Bai (1997). \( \square \)

**Lemma D.1.3.** Under Assumption D.1, for \( T_b = T_b^0 + \left[ u / \| \delta_T \|^2 \right] \), we have \( \delta_T^T \Sigma_\Delta^\prime \Sigma_\Delta \delta_T = \delta_T^T \sum_{l=T_b+1}^{T_b^0} z_l z_l^T \delta_T - p \). Let \( \overline{V} \approx V_1 \) if \( u \leq 0 \) and \( \overline{V} = V_2 \) if \( u > 0 \).
Proof. It follows from basic arguments (cf. Assumptions 4.8-4.9). □

**Lemma D.1.4.** Under Assumption D.1, for any $\epsilon > 0$ there exists a $C < \infty$ and a positive sequence $\{\nu_T\}$, with $\nu_T \to \infty$ as $T \to \infty$, such that

$$\liminf_{T \to \infty} \mathbb{P} \left[ \sup_{K \leq |u| \leq \eta T \|\delta_T\|^2} Q_T (\delta (T_b), T_b) - Q_T (\delta (T_0^b), T_0^b) < -C\nu_T \right] \geq 1 - \epsilon,$$

for all sufficiently large $K$ and a sufficiently small $\eta > 0$.

Proof. Note that on $\{K \leq |u| \leq \eta T \|\delta_T\|^2\}$ we have $K / \|\delta_T\|^2 \leq |T_b - T_0^b| \leq \eta T$. In view of (D.1.7), the statement $Q_T (\delta (T_b), T_b) - Q_T (\delta (T_0^b), T_0^b) < -C\nu_T$ follows from showing that as $T \to \infty$,

$$\mathbb{P} \left( \sup_{T_b \in B_{K,T}} \kappa \|\delta_T\| \geq \inf_{T_b \in B_{K,T}} \left| T_b - T_0^b \right| \mathcal{g}_d (\delta_T, T_b) \right) < \epsilon,$$

where $\kappa \in (1/2, 1)$. Suppose $T_b < T_0^b$. We show that

$$\mathbb{P} \left( \sup_{T_b \in B_{K,T}} \frac{\|\delta_T\|}{K} \kappa \|\delta_T\| \geq \frac{1}{\|\delta_T\|^{2\kappa - 1}} \left( \frac{1}{K} \right)^{1-\kappa} \inf_{T_b \in B_{K,T}} \mathcal{g}_d (\delta_T, T_b) \right) < \epsilon. \quad (D.1.9)$$

Lemma D.1.5-(ii) stated below implies that $\inf_{T_b \in B_{T,K}} \mathcal{g}_d (\delta_T, T_b)$ is bounded away from zero as $T \to \infty$ for large $K$ and small $\eta$. Next, we show that

$$\sup_{T_b \in B_{K,T}} K^{-1} \|\delta_T\| \kappa \|\delta_T\| = o_P (1). \quad (D.1.10)$$
Consider the first term of (D.1.5),

\[ 2\delta_T' (Z_0'MZ_2) (Z_2'MZ_2)^{-1} Z_2 Me = 2\delta_T' (Z_0'MZ_2/T) (Z_2'MZ_2/T)^{-1} Z_2 Me = 2C \|\delta_T\| O_p (1) O_p (1) O_p \left( T^{1/2} \right) = CO_p \left( \|\delta_T\| T^{1/2} \right). \]

When multiplied by \( \|\delta_T\| / K \), this term is \( O_p \left( \|\delta_T\|^2 T^{1/2} / K \right) \) which goes to zero for large \( K \) in view of Assumption D.1. The second term in equation (D.1.5), when multiplied by \( \|\delta_T\| / K \), is

\[ 2K^{-1} \|\delta_T\| \delta_T' (Z_0'Me) = K^{-1} \|\delta_T\| O_p \left( \|\delta_T\| T^{1/2} \right) = K^{-1} O_p \left( \|\delta_T\|^2 T^{1/2} \right), \]

which converges to zero using the same argument as for the first term. Consider now the first term of (D.1.6), \( T^{-1/2} \epsilon'MZ_2 (Z_2'MZ_2/T)^{-1} T^{-1/2} Z_2 Me = O_p (1) \). A similar argument can be used for the second term which is also \( O_p (1) \). Each of the latter two terms when multiplied by \( \|\delta_T\| / K \) is \( O_p (\|\delta_T\| / K) = o_p (1) \). This proves (D.1.10) and thus (D.1.9). To conclude the proof, note that \( \kappa \in (1/2, 1) \) implies \( \|\delta_T\|^{-\left(2\kappa-1\right)} \rightarrow \infty \), so that we can choose \( \nu_T = \left( \|\delta_T\|^2 / K \right)^{-(1-\kappa)}. \)

**Lemma D.1.5.** Let \( \tilde{g}_d \triangleq \inf_{|T_b-T_b^0|>K\|\delta_T\|} \mathcal{g}_d (\delta_T, T_b) \). Under Assumption D.1,

(i) for any \( \epsilon > 0 \) there exists some \( C > 0 \) such that \( \liminf_{T \to \infty} \mathbb{P} \left( \tilde{g}_d > C \|\delta_T\|^2 \right) \leq 1 - \epsilon; \)

(ii) with \( B^{\epsilon}_{T,K} = \left\{ T_b : T \eta \geq |T_b - T_b^0| \geq K/\|\delta_T\|^2 \right\} \), for any \( \epsilon > 0 \) there exists a \( C > 0 \) such that \( \liminf_{T \to \infty} \mathbb{P} \left( \inf_{T_b \in B^{\epsilon}_{T,K}} \mathcal{g}_d (\delta_T, T_b) > C \right) \leq 1 - \epsilon. \)

**Proof.** Part (i) was proved in Lemma A.2 of Bai (1997). As for part (ii), by Lemma
D.1.1,

\[ g_d \left( \delta^0, T_b \right) \geq \delta_T D' \frac{X'_\Delta X_\Delta}{T_b^0 - T_b} (X'_2 X_2)^{-1} (X'_0 X_0) D \delta_T \geq \lambda_{J,T_b}, \]

where \( \lambda_{J,T_b} \) is the minimum eigenvalue of \( D' J (T_b) D \), with

\[ J (T_b) \triangleq \| \delta_T \|_2^2 \left( T_b^0 - T_b \right)^{-1} X'_\Delta X_\Delta (X'_2 X_2)^{-1} (X'_0 X_0). \]

It is sufficient to show that, for \( T_b \in B_{T,K}^c \), \( \lambda_{J,T_b} \) is bounded away from zero with large probability for large \( K \) and small \( \eta \). We have

\[ \left\| J^{-1} (T_b) \right\| \leq \left\| \| \delta_T \|_2^2 \left( T_b^0 - T_b \right)^{-1} X'_\Delta X_\Delta \right\|^{-1} \left( X'_2 X_2 \right)^{-1} \left( X'_0 X_0 \right)^{-1}, \]

and by Assumptions 4.3-4.4 \( \left\| (X'_2 X_2) (X'_0 X_0)^{-1} \right\| \leq \| X' X \| \left\| (X'_0 X_0)^{-1} \right\| \) is bounded. Next, note that \( (T_b^0 - T_b)^{-1} X'_\Delta X_\Delta = (T_b^0 - T_b)^{-1} \sum_{t=T_b+1}^{T_b^0} x_t x'_t \) is larger than

\[ (T \eta)^{-1} \sum_{t=T_b^0 - \left\lfloor K/\| \delta_T \|_2^2 \right\rfloor}^{T_b^0} x_t x'_t \]

on \( B_{T,K}^c \), and for all \( K \), \( \left( \| \delta_T \|_2^2 / K \right) \sum_{t=T_b^0 - \left\lfloor K/\| \delta_T \|_2^2 \right\rfloor}^{T_b^0} x_t x'_t \) is positive definite with large probability as \( T \to \infty \). Now, \( (K/T \eta) \left( \| \delta_T \|_2^2 / K \right) \sum_{t=T_b^0 - \left\lfloor K/\| \delta_T \|_2^2 \right\rfloor}^{T_b^0} x_t x'_t = O_P (1) \), by choosing sufficiently large \( K \) and small \( \eta \). Thus, \( \left\| \| \delta_T \|_2^2 (T_b^0 - T_b)^{-1} X'_\Delta X_\Delta \right\|^{-1} \) is bounded with large probability for such large \( K \) and small \( \eta \), which in turn implies that \( \| J^{-1} (T_b) \| \) is bounded. Since \( D \) has full column rank, \( \lambda_{J,T_b} \) is bounded away from zero for sufficiently large \( K \) and small \( \eta \). \( \square \)
Lemma D.1.6. Under Assumption D.1, for any $\epsilon > 0$ there exists a $C > 0$ such that

$$\liminf_{T \to \infty} \mathbb{P} \left[ \sup_{|u| \geq T \|\delta_T\|^2 \eta} Q_T (\delta (T_b), T_b) - Q_T \left( \delta \left( T^0_b \right), T^0_b \right) < -C \nu_T \right] \geq 1 - \epsilon,$$

for every $\eta > 0$, where $\nu_T \to \infty$.

Proof. Fix any $\eta > 0$. Note that on $\{ |u| \geq T \|\delta_T\|^2 \eta \}$ we have $|T_b - T^0_b| \geq T \eta$. We proceed in a similar manner to Lemma D.1.4. Let $B_{T, \eta} \triangleq \{ T_b : |T_b - T^0_b| \geq T \eta \}$ and recall (D.1.7). First, using a similar argument as in Lemma D.1.5-(i), we have $\inf_{T_b \in B_{T, \eta}} g_e (\delta_T, T_b) \geq C \|\delta_T\|^2$ with large probability for some $C > 0$. Noting that $T \eta \inf_{T_b \in B_{T, \eta}} g_d (\delta_T, T_b)$ diverges at rate $\tau_T = T \|\delta_T\|^2$, the claim follows if we can show that $g_e (\delta_T, T_b) = O_P (\nu_T^{\varpi})$, with $0 \leq \varpi < 1$ uniformly on $B_{T, \eta}$. This is shown in Lemma D.1.7 below, which suggests $\varpi \in (1/2, 1)$. Then, choose $\nu_T = \left( T \|\delta_T\|^2 \right)^{1-\varpi}$.

□

Lemma D.1.7. Under Assumption D.1, uniformly on $B_{T, \eta}$,

$$|g_e (\delta_T, T_b)| = O_P \left( \|\delta_T\| T^{1/2} \log T \right).$$

Proof. We show that $T^{-1} |g_e (\delta^0, T_b)| = O_P \left( \|\delta_T\| T^{-1/2} \log T \right)$ uniformly on $B_{T, \eta}$. Note that

$$\sup_{T_b \in B_{T, \eta}} |g_e (\delta_T, T_b)| \leq \sup_{q \leq T_b \leq T - q} |g_e (\delta_T, T_b)|,$$

and recall that $q = \dim (z_t)$ is needed for identification. Observe that

$$\sup_{q \leq T_b \leq T - q} \left\| (Z'_2 M Z_2)^{-1/2} Z'_2 M e \right\| = O_P (\log T), \quad (D.1.11)$$
by the law of iterated logarithms [cf. Billingsley (1995), Ch. 1, Theorem 9.5]. Next,

$$
\sup_{q \leq T_b \leq T-q} T^{-1/2} (Z_0' M Z_2) (Z_2' M Z_2)^{-1/2} = O_P(1),
$$

(D.1.12)

which can be proved using the inequality 

$$
(Z_0' M Z_2) (Z_2' M Z_2) (Z_0' M Z_0) = O_P(T) \text{ (which is valid for all } T_b). \text{ Thus, by (D.1.11) and (D.1.12), the first term on the right-hand side of (D.1.5) when multiplied by } T^{-1} \text{ is such that}
$$

$$
\sup_{q \leq T_b \leq T-q} 2\delta_T T^{-1} (Z_0' M Z_2) (Z_2' M Z_2)^{-1} Z_2' M e = O_P \left( \|\delta_T\| T^{-1/2} \log T \right).
$$

(D.1.13)

The second term on the right-hand side of (D.1.5) is $2\delta_T Z_0' M e = O_P \left( \|\delta_T\| T^{1/2} \right)$. Using (D.1.11), and dividing by $T$, the first term of (D.1.6) is $O_P \left( (\log T)^2 / T \right)$ while the last term is $O_P \left( T^{-1} \right)$. When divided by $T$, they are of order $O_P((\log T)^2 / T)$ and $O_P(T^{-1})$, respectively. Therefore, $|g_e(T_b, \delta^0)| = O_P \left( \|\delta_T\| T^{1/2} \log T \right)$, uniformly on $B^c_{T,0}$. □

**D.1.3 Proofs of Section 4.3**

We denote by $P$ the class of polynomial functions $p : \mathbb{R} \to \mathbb{R}$. Let

$$
U_T = \{ u \in \mathbb{R} : \lambda^0 + u/\psi_T \in \Gamma^0 \},
$$

$$
\Gamma_{T,0} = \{ u \in \mathbb{R} : |u| \leq \psi_T \}, \quad \Gamma_{T,0}^c = \mathbb{R} - \Gamma_{T,0}, \text{ and } \tilde{U}_T^c = U_T - \Gamma_{T,0}. \text{ For } u \in \mathbb{R}, \text{ let}
$$

$$
R_{T,v}(u) \triangleq Q_{T,v}(u) - A^0(u) \text{ and } G_{T,v}(u) \triangleq \sup_{\tilde{v} \in \mathbb{V}} \tilde{G}_{T,v}(u, \tilde{v}). \text{ The generic constant } 0 < C < \infty \text{ used below may change from line to line. Finally, let } \tilde{\gamma}_T \triangleq \gamma_T / T \|\delta_T\|^2.
$$

**D.1.3.1 Proof Theorem 4.3.1**

We start with the following lemmas.
Lemma D.1.8. For any \(a \in \mathbb{R}, |c| \leq 1\), and integer \(i \geq 0\),

\[
\left| \exp (ca) - \sum_{j=0}^{i} (ca)^j / j! \right| \leq |c|^{i+1} \exp (|a|).
\]

Proof. The proof is immediate and the same as the one in Jun et al. (2015). Using simple manipulations,

\[
\left| \exp (ca) - \sum_{j=0}^{i} (ca)^j / j! \right| \leq \left| \sum_{j=i+1}^{\infty} \frac{(ca)^j}{j!} \right| \leq |c|^{i+1} \left| \sum_{j=i+1}^{\infty} \frac{(a)^j}{j!} \right| \leq |c|^{i+1} \exp (|a|).
\]

□

Lemma D.1.9. \(\tilde{G}_{T,v}(u, \tilde{v}) \Rightarrow \mathcal{W}(u)\) in \(D_{b}(C \times V)\), where \(C \subset \mathbb{R}\) and \(V \subseteq \mathbb{R}^{p+2q}\) are both compact sets, and

\[
\mathcal{W}(u) \triangleq \begin{cases} 
2 \left( (\delta^0)' \Sigma_1 \delta^0 \right)^{1/2} W_1 (-u), & \text{if } u < 0 \\
2 \left( (\delta^0)' \Sigma_2 \delta^0 \right)^{1/2} W_2 (u), & \text{if } u \geq 0.
\end{cases}
\]

Proof. Consider \(u < 0\). According to the expansion of the criterion function given in Lemma D.1.2, for any \((u, \tilde{v}) \in C \times V\), \(\tilde{G}_{T,v}(u, \tilde{v})\) reduces to \(2 \text{sgn} (T^0_b - T^b_b (u)) \delta_T Z_{\Delta} e + o_P (1)\). Then, \(\delta_T Z_{\Delta} e = (\delta^0)' v_T \sum_{t=\left[u/v^2\right]}^{T^0_b} z_t e_t \Rightarrow (\delta^0)' \mathcal{G}_1 (-u)\), where \(\mathcal{G}_1\) is a multivariate Gaussian process. In particular, \((\delta^0)' \mathcal{G}_1 (-u)\) is equivalent in law to \(\left( (\delta^0)' \Sigma_1 \delta^0 \right)^{1/2} W_1 (-u)\), where \(W_1 (\cdot)\) is a standard Wiener process on \([0, \infty)\). Similarly, for \(u \geq 0\), \(\delta_T Z_{\Delta} e \Rightarrow \left( (\delta^0)' \Sigma_2 \delta^0 \right)^{1/2} W_2 (u)\), where \(W_2 (\cdot)\) is another standard Wiener process on \([0, \infty)\) which is independent of \(W_1\). Hence, \(\tilde{G}_{T,v}(u, \tilde{v}) \Rightarrow \mathcal{W}(u)\) in \(D_{b}(C \times V)\). □

Lemma D.1.10. Fix any \(a > 0\) and let \(\varpi \in (1/2, 1]\). (i) For any \(\nu > 0\) and any
\[ \varepsilon > 0, \]
\[ \lim_{i \to \infty} P \left[ \sup_{u \in \Gamma_{T,v}} \left\{ G_{T,v}(u) - a \| \delta^0 \|_2 \right\} > \nu \right] < \varepsilon. \]

(ii) For \( \tilde{u} \in \mathbb{R}_+ \) let \( \widetilde{\Gamma} \triangleq \{ u \in \mathbb{R} : |u| > \tilde{u} \} \). Then, for every \( \epsilon > 0 \),

\[ \lim_{u \to \infty} \lim_{i \to \infty} P \left[ \sup_{u \in \Gamma} \left\{ G_{T,v}(u) - a \| \delta^0 \|_2 \right\} > \nu \right] = 0. \]

Proof. We begin with part (i). Upon using Lemma D.1.9 and the continuous mapping theorem, with any nonnegative integer \( i \),

\[ \lim_{i \to \infty} P \left[ \sup_{u \in \Gamma_{T,v}} \left\{ G_{T,v}(u) - a \| \delta^0 \|_2 \right\} > \nu \right] \leq \lim_{T \to \infty} P \left[ \sup_{|u| \geq i} \left\{ G_{T,v}(u) - a \| \delta^0 \|_2 \right\} > \nu \right] \]

\[ \leq \lim_{T \to \infty} P \left[ \sup_{|u| \geq i} \left\{ G_{T,v}(u) > a \| \delta^0 \|_2 \right\} > \nu \right] \]

\[ \leq P \left[ \sup_{|u| \geq i} \left\{ |W(u)| - a \| \delta^0 \|_2 \right\} > \nu \right] \]

\[ \leq \sum_{r=i+1}^{\infty} P \left[ \sup_{r-1 \leq |u| < r} \left\{ |W(u)| - a \| \delta^0 \|_2 \right\} > \nu \right]. \]

Then,

\[ \sum_{r=i+1}^{\infty} P \left[ \sup_{r-1 \leq |u| < r} \frac{1}{\sqrt{r}} |W(u)| > \inf_{r-1 \leq |u| < r} a \frac{1}{\sqrt{r}} \| \delta^0 \|_2 \right] \]

\[ = \sum_{r=i+1}^{\infty} P \left[ \sup_{1+1/r \leq |u|/r \leq 1} |W(u/r)| > \inf_{1-1/r \leq |u|/r \leq 1} a \left( \frac{r}{r} \right)^{\omega - 1/2} \frac{|u|^{\omega}}{\sqrt{r}} \| \delta^0 \| \right] \]

\[ = \sum_{r=i+1}^{\infty} P \left[ \sup_{1+1/r \leq s \leq 1} |W(s)| > \inf_{c \leq s \leq 1} a r^{\omega - 1/2} C \| \delta^0 \| \right] \]

\[ = \sum_{r=i+1}^{\infty} P \left[ \sup_{s \leq 1} |W(s)| > r^{\omega - 1/2} C \| \delta^0 \| \right], \quad \text{(D.1.14)} \]
where $0 < c \leq 1$. By Markov’s inequality,

$$\sum_{r=1}^{\infty} \mathbb{P} \left[ \sup_{r < s \leq 1} |\mathcal{W}(s)|^4 > C^4 \|\delta^0\|^4 r^{4(\omega-1/2)} e^{4\omega} \right] (D.1.15)$$

$$\leq \frac{C}{\|\delta^0\|^4} \mathbb{E} \left( \sup_{s \leq 1} |\mathcal{W}(s)|^4 \right) \sum_{r=1}^{\infty} r^{-(4\omega-2)}.$$

By Proposition A.2.4 in van der Vaart and Wellner (1996) $\mathbb{E} \left( \sup_{s \leq 1} |\mathcal{W}(s)|^4 \right) \leq C \mathbb{E} \left( \sup_{s \leq 1} |\mathcal{W}(s)|^4 \right)^4$, for some $C < \infty$, which is finite by Corollary 2.2.8 in van der Vaart and Wellner (1996). Choose $K$ (thus $\pi$) large enough such that the right-hand side in (D.1.15) can be made arbitrarily smaller than $\varepsilon > 0$. The proof of the second part is similar and omitted. $\square$

**Lemma D.1.11.** Fix any $a > 0$. For any $\varepsilon > 0$ there exists a $C < \infty$ such that

$$\mathbb{P} \left[ \sup_{u \in \mathbb{R}} \left\{ \overline{G}_{T,v}(u) - a \|\delta^0\|^2 |u| \right\} > C \right] < \varepsilon, \quad \text{for all } T.$$

**Proof.** For any finite $T$, $\overline{G}_{T,v}(u) \in \mathcal{D}_b$ by definition. As for the limiting case, fix any $0 < \pi < \infty$,

$$\limsup_{T \to \infty} \mathbb{P} \left[ \sup_{u \in \mathbb{R}} \left\{ \overline{G}_{T,v}(u) - a \|\delta^0\|^2 |u| \right\} > C \right] \leq \limsup_{T \to \infty} \mathbb{P} \left[ \sup_{|u| \leq \pi} \overline{G}_{T,v}(u) > C \right]$$

$$+ \limsup_{T \to \infty} \mathbb{P} \left[ \sup_{|u| > \pi} \overline{G}_{T,v}(u) > a \|\delta^0\|^2 \pi \right].$$

The second term converges to zero letting $\pi \to \infty$ from Lemma D.1.10-(ii). For the first term, let $C \to \infty$, use the continuous mapping theorem and Lemma D.1.9 to deduce that it converges to zero by the properties of $\mathcal{W} \in \mathcal{D}_b$. $\square$
Lemma D.1.12. Consider $A_1(u, \tilde{v})$ and $A_2(u, \tilde{v})$ as defined in (D.1.16) below. For $m \geq 0$,

$$\liminf_{T \to \infty} \mathbb{P} \left[ \sup_{\tilde{v} \in V} \left| \int_{\Gamma_{T,\psi}^c} (A_1(u, \tilde{v}) - A_2(u, \tilde{v})) \right| < \epsilon \right] \geq 1 - \epsilon.$$

Proof. We consider each integrand $A_i(u, \tilde{v})$ ($i = 1, 2$) separately on $\Gamma_{T,\psi}^c$. Let us consider $A_1$ first. Lemma D.1.4 yields that whenever $\tilde{\gamma}_T \to \kappa_\gamma < \infty$, $A_1(u, \tilde{v}) \leq C_1 \exp \left( -C_2 \nu_T \right)$ where $0 < C_1, C_2 < \infty$ and $\nu_T$ is a divergent sequence. Note that the number $C_1$ follows from Assumption 4.6 (cf. $\pi(\cdot) < \infty$). The argument for $A_2(u, \tilde{v})$ relies on Lemma D.1.10-(i), which shows that $G_{T,v}(u, \tilde{v})$ is always less than $C |u|^{\varpi}$ uniformly on $\Gamma_{T,\psi}^c$, with $C > 0$ and $\varpi \in (1/2, 1)$. Thus, $A_2(u, \tilde{v}) = o_P(1)$ uniformly on $V$. □

Let $\Gamma_{T,K} \triangleq \{ u \in \mathbb{R} : |u| \leq K, K > 0 \}$, and $\Gamma_{T,\eta} \triangleq \{ u \in \mathbb{R} : K \leq |u| \leq \eta \psi_T, K, \eta > 0 \}$.

Lemma D.1.13. For any polynomial function $p \in P$ and any $C < \infty$, let

$$D_T \triangleq \sup_{\tilde{v} \in W} \int_{\Gamma_{T,K}} |p(u)| \exp \left( C \Gamma_{T,v}(u, \tilde{v}) \right) \left| \exp \left( R_{T,v}(u) \right) - 1 \right| \exp \left( -A^0(u) \right) du = o_P(1).$$

Proof. Let $0 < \epsilon < 1$. We shall use Lemma D.1.8 with $i = 0$, $a = R_{T,v}(u)/c$, and $c = \epsilon$ to deduce $D_T = O_P(\epsilon)$ and then let $\epsilon \to 0$. Note that

$$\epsilon^{-1} D_T \leq C \int_{\Gamma_{T,K}} |p(u)| \exp \left( C \Gamma_{T,v}(u, \tilde{v}) \right) \left| \exp \left( -1 \right) \exp \left( -A^0(u) \right) du. \right.$$

By definition $K \geq u = \|\delta_T\|^2 (T_b - T_b^0)$ on $\Gamma_{T,K}$. By Lemma D.1.2-D.1.3, on $\Gamma_{T,K}$ we have $R_{T,v}(u) = O_P \left( \|\delta_T\|^2 \right)$ for each $u$. Thus, for large enough $T$, the right-hand side
above is $O_P(1)$ and does not depend on $\epsilon$. Thus, $D_T = \epsilon O_P(1)$. The claim of the lemma follows by letting $\epsilon$ approach zero. □

Lemma D.1.14. For $p \in P$,

$$D_T \triangleq \sup_{\tilde{v} \in \mathbf{V}} \int_{\Gamma_{T,\eta}} |p(u)| \exp \left\{ \tilde{\gamma}_T \tilde{G}_{T,v}(u, \tilde{v}) \right\} \exp \left\{ -A^0(u) \right\} |\pi_{T,v}(u) - \pi^0| du = o_P(1).$$

Proof. By the differentiability of $\pi(\cdot)$ at $\lambda^0_b$ (cf. Assumption 4.6), for any $u \in \mathbb{R}$, $|\pi_{T,v}(u) - \pi^0| \leq |\pi (\lambda^0_b (v)) - \pi^0| + C \psi^{-1} |u|$, with $C > 0$. The first term on the right-hand side is $o(1)$ and does not depend on $u$. Recalling that $\tilde{G}_{T,v}(u, \tilde{v}) = \sup_{\tilde{v} \in \mathbf{V}} |\tilde{G}_{T,v}(u, \tilde{v})|$, 

$$D_T \leq K \left[ o(1) \int_{\Gamma_{T,\eta}} d_T(u) du + \psi^{-1} \int_{\Gamma_{T,\eta}} |u| d_T(u) du \right]$$

$$\leq K \left[ o(1) O_P(1) + \psi^{-1} O_P(1) \right],$$

where $d_T(u) \triangleq |p(u)| \exp \left\{ \tilde{\gamma}_T \tilde{G}_{T,v}(u, \tilde{v}) \right\} |\exp (-A^0(u))|$ and the $O_P(1)$ terms follows from Lemma D.1.11 and $\tilde{\gamma}_T \rightarrow \kappa_\gamma < \infty$. Since $\psi_T \rightarrow \infty$, we have $D_T = o_P(1)$. □

Lemma D.1.15. For any $p \in P$ and constants $C_1, C_2 > 0$,

$$\int_{\Gamma_{T,\psi}} |p(u)| \exp \left\{ C_1 \tilde{G}_{T}(u) - C_2 |u| \right\} du = o_P(1).$$

Proof. It follows from Lemma D.1.6. □

Lemma D.1.16. For $p \in P$ and constants $a_1, a_2, a_3 \geq 0$, with $a_2 + a_3 > 0$, let

$$D_T = \int_{\mathbf{U}_{T,\psi}} |p(u)| \exp \left\{ \tilde{\gamma}_T \left\{ a_1 \tilde{G}_{T,v}(u) + a_2 Q_{T,v}(u) - a_3 A^0(u) \right\} \right\} du = o_P(1).$$

Proof. It follows from Lemma D.1.10. □
Lemma D.1.17. For any integer \( m \geq 0 \),

\[
\sup_{\tilde{v} \in V} \left[ \pi_{T,v}(u) \exp(Q_{T,v}(u)) - \pi^0 \exp(-A^0(u)) \right] du = o_P(1).
\]

Proof. Let

\[
A_1(u, \tilde{v}) = u^m \pi_{T,v}(u) \exp\left(\tilde{\gamma}_T \tilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u)\right)
\]

\[
A_2(u, \tilde{v}) = u^m \pi^0 \exp\left(\tilde{\gamma}_T \tilde{G}_{T,v}(u, \tilde{v}) - \Lambda_0(u)\right).
\]

By Assumption 4.6, \( A_1(u, \tilde{v}) = 0 \) for \( u \in \Gamma_{T,\psi}^c - \tilde{U}_T^c \). Then, by omitting arguments, we can write,

\[
\sup \left| \int_R (A_1 - A_2) \right| \leq \sup \left| \int_{\Gamma_{T,\psi}} (A_1 - A_2) \right| + \sup \left| \int_{\Gamma_{T,\psi}} A_2 \right| + \sup \left| \int_{\tilde{U}_T^c} A_1 \right|.
\]

The first right-hand side term above converges in probability to zero by Lemmas D.1.13-D.1.14. The second and the last terms are each \( o_P(1) \) by, receptively, Lemma D.1.15 and Lemma D.1.16. \( \square \)

We are now in a position to conclude the proof of Theorem 4.3.1.

Proof. Let \( V \subset \mathbb{R}^{p+2q} \) be a compact set. From (4.3.9),

\[
\psi_T\left(\lambda^{GL,*}_b(\tilde{v}, v) - \lambda^0_{b,T}(v)\right) = \frac{\int_R u \exp\left(\tilde{\gamma}_T \left[ \tilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u) \right]\right) \pi_{T,v}(u) du}{\int_R \exp\left(\tilde{\gamma}_T \left[ \tilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u) \right]\right) \pi_{T,v}(u) du}.
\]

For a large enough \( T \), by Lemma D.1.17 the right-hand is uniformly in \( \tilde{v} \in V \) equal
to
\[
\frac{\int_{\mathbb{R}} u \exp \left( \tilde{\gamma}_T \tilde{G}_{T,v} (u, \tilde{v}) \right) \exp \left( -\kappa_{\gamma} \Lambda^0 (u) \right) du}{\int_{\mathbb{R}} \exp \left( \tilde{\gamma}_T \tilde{G}_{T,v} (u, \tilde{v}) \right) \exp \left( -\kappa_{\gamma} \Lambda^0 (u) \right) du} + o_P (1).
\]

The first term is integrable with large probability by Lemma D.1.10-D.1.11. Thus, by Lemma D.1.9 and the continuous mapping theorem,
\[
T \| \delta_T \|^2 \left( \hat{\lambda}_b^{GL,v} (\tilde{v}, v) - \lambda^0_{b,T} (v) \right) = \frac{\int_{\mathbb{R}} u \exp \left( \kappa_{\gamma} \mathcal{W} (u) \right) \exp \left( -\kappa_{\gamma} \Lambda^0 (u) \right) du}{\int_{\mathbb{R}} \exp \left( \kappa_{\gamma} \mathcal{W} (u) \right) \exp \left( -\kappa_{\gamma} \Lambda^0 (u) \right) du}.
\]

\[\Box\]

D.1.3.2 Proof of Proposition 4.3.1

We first need to introduce further notation. For a scalar \( u > 0 \) define \( \Gamma_u \triangleq \{ u \in \mathbb{R} : |u| \leq u \} \). Note that \( \tilde{\gamma}_T^{-1} = o (1) \). We shall be concerned with the asymptotic properties of the following statistic:
\[
\xi_T (\tilde{v}) = \frac{\int_{\Gamma_{\pi}} u \exp \left( \tilde{\gamma}_T \left( \tilde{G}_{T,v} (u, \tilde{v}) + Q_{T,v} (u) \right) \right) \pi_{T,v} (u) du}{\int_{\Gamma_{\pi}} \exp \left( \tilde{\gamma}_T \left( \tilde{G}_{T,v} (u, \tilde{v}) + Q_{T,v} (u) \right) \right) \pi_{T,v} (u) du}.
\]

Furthermore, for every \( \tilde{v} \in \mathcal{V} \), let \( \xi_0 (\tilde{v}) = \arg \max_{u \in \Gamma_{\pi}} \mathcal{Y} (u) \). It turns out that \( \xi_0 (\tilde{v}) \) is flat in \( \tilde{v} \) and thus we write \( \xi_0 = \xi_0 (\tilde{v}) \). Finally, recall that \( u = T \| \delta_T \|^2 \left( \hat{\lambda}_b - \lambda^0_{b,T} (v) \right) \).

Lemma D.1.18. Let \( \Gamma^c_{T,\pi} = U_T - \Gamma_{\pi} \). Then for any \( \epsilon > 0 \) and \( m = 0, 1 \),
\[
\lim_{\pi \to \infty} \lim_{T \to \infty} P \left( \frac{\sup_{\tilde{v} \in \mathcal{V}} \int_{\Gamma^c_{T,\pi}} |u|^m \exp \left( \tilde{\gamma}_T \left( \tilde{G}_{T,v} (u, \tilde{v}) + Q_{T,v} (u) \right) \right) \pi_{T,v} (u) du}{\sup_{\tilde{v} \in \mathcal{V}} \int_{\mathbb{R}} \exp \left( \tilde{\gamma}_T \left( \tilde{G}_{T,v} (u, \tilde{v}) + Q_{T,v} (u) \right) \right) \pi_{T,v} (u) du} > \epsilon \right) = 0.
\]

Proof. Let \( J_1 \) and \( J_2 \) denote the numerator and denominator, respectively, in the
display of the lemma. Then,

$$\mathbb{P}(J_1/J_2 > \epsilon) \leq \mathbb{P}(J_2 \leq \exp(-\bar{a}\bar{\gamma}_T)) + \mathbb{P}(J_1 > \epsilon \exp(-\bar{a}\bar{\gamma}_T)),$$

(D.1.18)

for any constant $\bar{a} > 0$. Let us consider the second term term in (D.1.18). For an arbitrary $a > 0$, let $H(\bar{a}, a) = \{u \in \Gamma^c_T, \bar{\pi} : \sup_{v \in V} |\bar{G}_{T,v}(u, \bar{v})| \leq a |u|\}$. Let $\bar{x} = 2 \sup_{b \in \Gamma^0} |\lambda_b|$. Note that $\bar{x} < 2$ and $\sup_{u \in H(\bar{a}, a)} |u| \leq \bar{x} T \|u\|^2$. By Assumption 4.4 and 4.8, and Lemma D.1.6, $Q_{T,v}(u) \leq -\min(\Lambda_0(u)/2, \eta \bar{x} T \|\delta_T\|^2 T)$ uniformly for all large $T$ where $\eta > 0$. Thus,

$$\sup_{u \in H(\bar{a}, a)} \sup_{v \in V} \exp \left( \bar{\gamma}_T \left[ a |u| - A^0(u)/4 + Q_{T,v}(u) \right] \right) \leq \sup_{u \in H(\bar{a}, a)} \sup_{v \in V} \exp \left( \bar{\gamma}_T \left[ a |u| - A^0(u) - \min(\Lambda^0(u)/4, \Lambda^0(u)/4 + \eta \|\delta_T\|^2 T) \right] \right)$$

$$\leq \sup_{u \in H(\bar{a}, c)} \exp \left( \bar{\gamma}_T \left[ a |u| - C_2 |u| \right] \right) + \exp \left( \bar{\gamma}_T \left[ a \bar{x} - \eta C \right] \right)$$

$$\leq \sup_{u \in H(\bar{a}, c)} \exp \left( \bar{\gamma}_T \left[ a - C_2 \right] \right) + \exp \left( \bar{\gamma}_T \left[ a \bar{x} - \eta C \right] \right) = o(\exp(-\bar{\gamma}_T a_1)),$$

when $a > 0$ is chosen sufficiently small and for some $a_1 > 0$. Furthermore, by Lemma D.1.10-(ii) below with $\varpi = 1$,

$$\lim_{\bar{a} \to \infty} \lim_{T \to \infty} \mathbb{P} \left( u \in \left\{ \Gamma^c_T, -H(\bar{a}, c) \right\} \right) \leq \lim_{\bar{a} \to \infty} \lim_{T \to \infty} \mathbb{P} \left( \sup_{|u| > \bar{a}} |\bar{G}_{T,v}(u, \bar{v})| > a \right) = 0.$$

(D.1.20)

By combining equations (D.1.19)-(D.1.20), $\mathbb{P}(J_1 > \epsilon \exp(-\bar{a}\bar{\gamma}_T)) \to 0$ as $T \to \infty$. Next, we consider the first right-hand side term in (D.1.18). Recall the definition of $\lambda_+$ from Assumption 4.9 and let $0 < b \leq \bar{a}/4\lambda_+$. Note that for $G_{T,v}(b) \triangleq$
\[
\sup_{|u| \leq b} \sup_{v \in V} |G_{T,v} (u, \tilde{v})|, \text{ then}
\]

\[
P \left( J_2 \leq \exp (-\overline{\alpha} \tilde{\gamma}_T) \right) \leq P \left( G_{T,v} (b) \leq \overline{\alpha}, J_2 \leq \exp (-\overline{\alpha} \tilde{\gamma}_T) \right) + P \left( G_{T,v} (b) > \overline{\alpha} \right).
\]

(D.1.21)

Under Assumption 4.6 and the second part of Assumption 4.9, using the definition of \( b \),

\[
P \left( G_{T,v} (b) \leq \overline{\alpha}, J_2 \leq \exp (-\overline{\alpha} \tilde{\gamma}_T) \right)
\leq P \left( C_{\pi} \int_{|u| \leq b} \exp (\tilde{\gamma}_T (-\overline{\alpha}/2 - \lambda_+) b) \, du \leq \exp (-\overline{\alpha} \tilde{\gamma}_T) \right)
\leq P \left( C_{\pi} \eta \exp (\overline{\alpha} \tilde{\gamma}_T / 2) \leq 1 \right) \to 0,
\]

as \( T \to \infty \). We shall use the uniform convergence in Lemma D.1.9 for the second right-hand side term in (D.1.21) to deduce that (recall that \( \overline{\alpha} \) was chosen sufficiently small and \( b \leq \overline{\alpha}/4\lambda_+ \)),

\[
\lim_{b \to 0} \lim_{T \to \infty} P \left( G_{T,v} (b) > \overline{\alpha} \right) \leq \lim_{b \to 0} \lim_{T \to \infty} P \left( \sup_{|u| \leq b} |W (u)| > \overline{\alpha} \right) = 0.
\]

\[ \square \]

Lemma D.1.19. As \( T \to \infty \), \( \xi_T (\tilde{v}) \Rightarrow \xi_0 \) in \( D_b (V) \).

Proof. Let \( B = \Gamma_{\pi} \times V \). For any fixed \( \pi \), Lemma D.1.9 and the result

\[
\sup_{(u, \tilde{v}) \in B} \left| Q_{T,v} (u) - \Lambda^0 (u) \right| = o_{\mathbb{P}} (1)
\]

(cf. Lemma D.1.3), imply that \( \overline{Q}_T \Rightarrow \mathcal{V} \) in \( D_b (B) \). By the Skorokhod representation theorem [cf. Theorem 6.4 in Billingsley (1999)] we can find a probability space
on which there exist processes $\tilde{Q}_T(u, \tilde{v})$ and $\tilde{\mathcal{V}}(u)$ which have the same law as $Q_T(u, \tilde{v})$ and $\mathcal{V}(u)$, respectively, and with the property that

$$\sup_{(u, \tilde{v}) \in \mathcal{B}} |\tilde{Q}_T(u, \tilde{v}) - \tilde{\mathcal{V}}(u)| \to 0 \quad \tilde{\mathbb{P}} - \text{a.s.} \quad (D.1.22)$$

Let

$$\tilde{\xi}_T(\tilde{v}) \triangleq \frac{\int_{\Gamma} u \exp (\tilde{\gamma}_T \tilde{Q}_T(u, \tilde{v})) \pi_T(u) du}{\int_{\Gamma} \exp (\tilde{\gamma}_T \tilde{Q}_T(u, \tilde{v})) \pi_T(u) du},$$

and $\tilde{\xi}_0 \triangleq \arg \max_{u \in \Gamma} \tilde{\mathcal{V}}(u)$. We shall rely on (D.1.22) to establish that

$$\sup_{\tilde{v} \in \mathcal{V}} |\tilde{\xi}_T(\tilde{v}) - \tilde{\xi}_0| \to 0 \quad \tilde{\mathbb{P}} - \text{a.s.} \quad (D.1.23)$$

Let us indicate any pair of sample paths of $\tilde{Q}_T(u, \tilde{v})$ and $\tilde{\mathcal{V}}$, for which (D.1.22) holds with a superscript $\omega$: $\tilde{Q}^\omega_{T,v}$ and $\tilde{\mathcal{V}}^\omega$, respectively. For arbitrary sets $S_1, S_2 \subset \mathcal{B}$, let

$$\tilde{\rho}(S_1, S_2) \triangleq \text{Leb}(S_1 - S_2) + \text{Leb}(S_2 - S_1)$$

where Leb(A) is the Lebesgue measure of the set A. Further, for an arbitrary scalar $c > 0$ and function $\Upsilon: \mathcal{B} \to \mathbb{R}$, define

$$S(\Upsilon, c) \triangleq \{(u, \tilde{v}) \in \mathcal{B}: |\Upsilon(u, \tilde{v}) - \tilde{\mathcal{V}}^\omega_M| \leq c\}$$

where $\tilde{\mathcal{V}}^\omega_M \triangleq \max_{u \in \Gamma} \tilde{\mathcal{V}}^\omega(u)$. The first step is to show that

$$\tilde{\rho}(S(\tilde{Q}^\omega_{T,v}, c), S(\tilde{\mathcal{V}}^\omega, c)) = o(1). \quad (D.1.24)$$

Let $S_{1, T}(c) = S(\tilde{Q}^\omega_{T,v}, c) - S(\tilde{\mathcal{V}}^\omega, c)$ and $S_{2, T}(c) = S(\tilde{\mathcal{V}}^\omega, c) - S(\tilde{Q}^\omega_{T,v}, c)$. We first establish that $\text{Leb}(S_{2, T}(c)) = o(1)$. For an arbitrary $\tau > 0$, define the set

$$\tilde{S}_T(\tau) \triangleq \{(u, \tilde{v}) \in \mathcal{B}: |\tilde{Q}^\omega_{T,v}(u, \tilde{v}) - \tilde{\mathcal{V}}^\omega(u)| \leq \tau\}$$

and its complement (relative to
B) $\bar{S}_t^c (\bar{c}) \triangleq \{(u, \bar{v}) \in B : |\tilde{Q}_{T,v}^\omega (u, \bar{v}) - \tilde{V}_M (u)| > \bar{c}\}$. We have

$$\text{Leb} (S_{2,T} (c)) = \text{Leb} \left( S_{2,T} (c) \cap \bar{S}_T (\bar{c}) \right) + \text{Leb} \left( S_{2,T} (c) \cap \bar{S}_T^c (\bar{c}) \right) \leq \text{Leb} \left( S_{2,T} (c) \cap \bar{S}_T (\bar{c}) \right) + \text{Leb} \left( \bar{S}_T^c (\bar{c}) \right).$$

Note that $\text{Leb} \left( \bar{S}_T^c (\bar{c}) \right) = o (1)$ since the path $\omega$ satisfies (D.1.22). Furthermore, $S_{2,T} (c) \cap \bar{S}_T (\bar{c}) \subset C_T (c, \bar{c})$ where

$$C_T (c, \bar{c}) \triangleq \{(u, \bar{v}) \in B : c \leq |\tilde{Q}_{T,v}^\omega (u, \bar{v}) - \tilde{V}_M | \leq c + \bar{c}\}.$$

In view of (D.1.22),

$$\lim_{\varepsilon \downarrow 0} \lim_{T \to \infty} \text{Leb} \left( C_T (c, \bar{c}) \right) = \lim_{\varepsilon \downarrow 0} \text{Leb} \left\{ (u, \bar{v}) \in B : c \leq |\tilde{V}_M (u)| \leq c + \bar{c} \right\} = \text{Leb} \left\{ (u, \bar{v}) \in B : |\tilde{V}_M (u)| = c \right\} = 0,$$

by the path properties of $\tilde{V}^\omega$. Since $\text{Leb} (S_{1,T} (c)) = o (1)$ can be proven in a similar fashion, (D.1.24) holds. For $m = 0, 1, C_1 < \infty$ and by Assumption 4.6 we know there exists some $C_2 < \infty$ such that

$$\sup_{\bar{v} \in V} \int_{\bar{S}^c (\tilde{Q}_{T,v}^\omega (u, \bar{v}), c)} |u|^m \exp \left( \tilde{\gamma}_T \left( \tilde{Q}_{T,v}^\omega (u, \bar{v}) - \tilde{V}_M \right) \right) \pi_{T,v} (u) du \leq C_1 \exp (-c\tilde{\gamma}_T) C_2 \int_{\Gamma} |u|^m du = o (1),$$

since $\{u \leq \bar{u}\}$ on $\Gamma$ and recalling that $\tilde{\gamma}_T \to \infty$. Then,

$$\sup_{\bar{v} \in V} \int_{\Gamma} u \exp \left( \tilde{\gamma}_T \tilde{Q}_{T,v}^\omega (u, \bar{v}) \right) \pi_{T,v} (u) du \leq \text{ess sup} S \left( \tilde{Q}_{T,v}^\omega, c \right) + o (1).$$
By (D.1.22) we deduce $\text{ess sup} \mathbf{S} \left( \bar{Q}_{T,v}, c \right) + o(1) = \text{ess sup} \mathbf{S} \left( \bar{\mathcal{Y}}^\omega, c \right) + o(1)$. The same argument yields

$$\inf_{\bar{\nu} \in \mathbf{V}} \int_{\Gamma_{\bar{\nu}}} u \exp \left( \bar{\gamma}_T \bar{Q}_{T,v}^\omega (u, \bar{v}) \right) \pi_{T,v} (u) du \geq \text{ess inf} \mathbf{S} \left( \bar{\mathcal{Y}}^\omega, c \right) + o(1).$$

Since almost every path $\omega$ of the Gaussian process $\bar{\mathcal{Y}}$ achieves its maximum at a unique point on compact sets [cf. Bai (1997) and Lemma 2.6 in Kim and Pollard (1990)], we have

$$\lim_{c \downarrow 0} \text{ess inf} \mathbf{S} \left( \bar{\mathcal{Y}}^\omega, c \right) = \lim_{c \downarrow 0} \text{ess sup} \mathbf{S} \left( \bar{\mathcal{Y}}^\omega, c \right) = \arg \max_{u \in \Gamma_{\bar{\nu}}} \bar{\mathcal{Y}}^\omega (u).$$

Hence, we have proved (D.1.23) which by the dominated convergence theorem then implies the weak convergence of $\bar{\xi}_T$ toward $\bar{\xi}_0$. Since the law of $\bar{\xi}_T$ ($\bar{\xi}_0$) under $\bar{\mathbb{P}}$ is the same as the law of $\xi_T$ ($\xi_0$) under $\mathbb{P}$, the claim of the Lemma follows. \(\Box\)

We are now in a position to conclude the proof of Proposition 4.3.1.

**Proof.** For a set $T \subset \mathbb{R}$ and $m = 0,1$ we define

$$J_m (T) \triangleq \int_T u^m \exp \left( \bar{\gamma}_T \left( \bar{G}_{T,v} (u, \bar{v}) + Q_{T,v} (u) \right) \right) \pi_{T,v} (u) du.$$

Hence, with this notation equation (4.3.9) can be rewritten as

$$T \| \delta_T \|^2 \left( \lambda_{b}^{\text{GL}*} (\bar{v}, \bar{v}) - \lambda_{b,T}^{0} (v) \right) = J_1 (\mathbb{R}) / J_0 (\mathbb{R}).$$
Applying simple manipulations, we obtain,

\[
J_1(\mathbb{R}) / J_0(\mathbb{R}) = \frac{J_1(\Gamma_\pi) + J_1(\Gamma_{\pi,T})}{J_0(\Gamma_\pi) + J_0(\Gamma_{\pi,T})} = \frac{J_1(\Gamma_\pi)}{J_0(\Gamma_\pi)} \left[ 1 - \frac{J_0(\Gamma_{\pi,T})}{J_0(\mathbb{R})} \right] + \frac{J_1(\Gamma_{\pi,T})}{J_0(\mathbb{R})}.
\]

(D.1.25)

By Lemma D.1.18, \( J_m(\Gamma_{\pi,T}) / J_0(\mathbb{R}) = o_p(1) \) \((m = 0, 1)\) uniformly in \( \tilde{v} \in V \). By Lemma D.1.19, with \( \xi_T(\tilde{v}) = J_1(\Gamma_\pi) / J_0(\Gamma_\pi) \), the first right-hand side term in (D.1.25) converges weakly to \( \arg \max_{u \in \mathbb{R}} \mathcal{V}(u) \) in \( D_b(V) \). □

D.1.3.3 Proof of Corollary 4.3.1

The proof involves a simple change in variable. We refer to Proposition 3 in Bai (1997).

D.1.3.4 Proof of Theorem 4.3.2

We begin by introducing some notation. Since \( l \in L \), for all real numbers \( B \) sufficiently large and \( \vartheta \) sufficiently small the following relationship holds

\[
\inf_{|u| > B} l(u) - \sup_{|u| \leq B^\vartheta} l(u) \geq 0.
\]

(D.1.26)

Let \( \zeta_{T,v}(u, \tilde{v}) = \exp(G_{T,v}(u, \tilde{v}) - \Lambda^0(u)) \), \( \Gamma_T \triangleq \{u \in \mathbb{R} : \lambda_u \in \Gamma^0\} \) and

\[
\Gamma_M = \{u \in \mathbb{R} : M \leq |u| < M + 1\} \cap \Gamma_T,
\]

and define

\[
J_{1,M} \triangleq \int_{\Gamma_M} \zeta_{T,v}(u, \tilde{v}) \pi_{T,v}(u) \, du, \quad J_2 \triangleq \int_{\Gamma_T} \zeta_{T,v}(u, \tilde{v}) \pi_{T,v}(u) \, du.
\]

(D.1.27)
In some steps in the proof we shall be working with elements of the following families of functions. A function \( f_T : \mathbb{R} \to \mathbb{R} \) is said to belong to the family \( F \) if it satisfies the following properties: (1) For fixed \( T \), \( f_T(x) \) increases monotonically to infinity with \( x \in [0, \infty) \); (2) For any \( b < \infty \), \( x^b \exp(-f_T(x)) \to 0 \) as both \( T \) and \( x \) diverge to infinity.

**Proof.** The random variable \( T \|\delta_T\|^2 (\tilde{\lambda}_{b}^{GL} - \lambda_0) = \tilde{\tau}_T \) is a minimizer of the function

\[
\Psi_{l,T}(s) = \int_{\Gamma_T} l(s-u) \frac{\exp \left( \tilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u) \right) \pi_{T,v}(u)}{\int_{\Gamma_T} \exp \left( \tilde{G}_{T,v}(w, \tilde{v}) + Q_{T,v}(w) \right) \pi_{T,v}(w) dw} du.
\]

Observe that Lemma D.1.13-D.1.17 apply to any polynomial \( p \in P \); therefore, they are still valid for \( l \in L \). We then have that the asymptotic behavior of \( \Psi_{l,T}(s) \) only matters when \( u \) (and thus \( s \)) varies on \( \Gamma_K = \{ u \in \mathbb{R} : u \leq K \} \). By Lemma D.1.24-D.1.25, for any \( \vartheta > 0 \), there exists a \( T \) such that for all \( T > T \),

\[
\mathbb{E} \left[ \int_{\Gamma_K} \frac{\exp \left( \tilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u) \right)}{\int_{\Gamma_T} \exp \left( \tilde{G}_{T,v}(w, \tilde{v}) + Q_{T,v}(w) \right) \pi_{T,v}(w) dw} du \right] \leq \frac{c_\vartheta}{K^\vartheta}. \tag{D.1.28}
\]

Therefore, for all \( T > T \),

\[
\Psi_{l,T}(s) = \int_{|w| \leq K} l(s-u) \frac{\exp \left( \tilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u) \right) du}{\int_{|w| \leq K} \exp \left( \tilde{G}_{T,v}(w, \tilde{v}) + Q_{T,v}(w) \right) dw} + o_P(1), \tag{D.1.29}
\]

where the \( o_P(1) \) term is uniform in \( T > T \) as \( K \) increases to infinity. By Assumption (4.6), \( |\pi_{T,v}(u) - \pi^0| \leq |\pi(\tilde{\lambda}_{b}^{GL}(v)) - \pi^0| + C \psi_T^{-1} |u| \), with \( C > 0 \). On \( \{|u| \leq K\} \), the first term on the right-hand side is \( o(1) \) and does not depend on \( u \). The second term is negligible when \( T \) is large. Thus, without loss of generality we set \( \pi_{T,v}(u) = 1 \) for all \( u \) in what follows.

Next, we show the convergence of the marginal distributions of the estimate
\( \Psi_{l,T}(s) \) to the marginals of the random function \( \Psi_l(s) \), where the region of integration in the definition of both the numerator and denominator of \( \Psi_{l,T}(s) \) and \( \Psi_l(s) \) is restricted over \( \{|u| \leq K\} \) only, in view of (D.1.29). For a finite integer \( n \), choose arbitrary real numbers \( a_j \) \( (j = 0, \ldots, n) \) and introduce the following estimate:

\[
\sum_{j=1}^{n} a_j \int_{|u| \leq K} l(s_j - u) \zeta_{T,v}(u, \tilde{v}) \, du + a_0 \int_{|u| \leq K} l(s_0 - u) \zeta_{T,v}(u, \tilde{v}) \, du. \tag{D.1.30}
\]

By Lemma D.1.21 and D.1.27, we can invoke Theorem I.A.22 in Ibragimov and Has’minskiï (1981) which gives the convergence in distribution of the estimate in (D.1.30) towards the distribution of the following random variable:

\[
\sum_{j=1}^{n} a_j \int_{|u| \leq K} l(s_j - u) \exp(\mathcal{V}(u)) \, du + a_0 \int_{|u| \leq K} l(s_0 - u) \exp(\mathcal{V}(u)) \, du.
\]

By the Cramer-Wold Theorem [cf. Theorem 29.4 in Billingsley (1995)] this suffices for the convergence in distribution of the vector

\[
\int_{|u| \leq K} l(s_i - u) \zeta_{T,v}(u, \tilde{v}) \, du, \ldots, \int_{|u| \leq K} l(s_n - u) \zeta_{T,v}(u, \tilde{v}) \, du, \int_{|u| \leq K} l(s_0 - u) \zeta_{T,v}(u, \tilde{v}) \, du,
\]

to the distribution of the vector

\[
\int_{|u| \leq K} l(s_i - u) \exp(\mathcal{V}(u)) \, du, \ldots, \int_{|u| \leq K} l(s_n - u) \exp(\mathcal{V}(u)) \, du, \int_{|u| \leq K} l(s_0 - u) \exp(\mathcal{V}(u)) \, du.
\]
As a consequence, for any $K_1, K_2 < \infty$, the marginal distributions of
\[
\int_{|u| \leq K_1} l(s - u) \exp\left(\widetilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u)\right) \, du
\]
\[
\int_{|w| \leq K_2} \exp\left(\widetilde{G}_{T,v}(w, \tilde{v}) + Q_{T,v}(w)\right) \, dw,
\]
converge to the marginals of
\[
\int_{|u| \leq K_1} l(s - u) \exp\left(\mathcal{V}(u)\right) \, du / \left(\int_{|w| \leq K_2} \exp\left(\mathcal{V}(w)\right) \, dw\right).
\]
The same convergence result extends to the distribution of
\[
\int_{M \leq |u| < M+1} \exp\left(\widetilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u)\right) \, du,
\]
towards the distribution of $\int_{M \leq |u| < M+1} \exp\left(\mathcal{V}(u)\right) \, du / \int_{|w| \leq K_2} \exp\left(\mathcal{V}(w)\right) \, dw$. By choosing $K_2 > M + 1$ we deduce
\[
\mathbb{E} \left[ \int_{M \leq |u| < M+1} \frac{\exp\left(\mathcal{V}(u)\right)}{\int_{\mathbb{R}} \exp\left(\mathcal{V}(w)\right) \, dw} \, du \right]
\]
\[
\leq \lim_{T \to \infty} \mathbb{E} \left[ \int_{M \leq |u| < M+1} \exp\left(\widetilde{G}_{T,v}(u, \tilde{v}) + Q_{T,v}(u)\right) \, du \right] \leq c_\theta M^{-\theta},
\]
in view of (D.1.28). This leads to
\[
\Psi_l(s) = \int_{|u| \leq K} l(s - u) \frac{\exp\left(\mathcal{V}(u)\right) \, du}{\int_{|w| \leq K} \exp\left(\mathcal{V}(w)\right) \, dw} + o_P(1),
\]
where the $o_P(1)$ term is uniform as $K$ increases to infinity. We then have the convergence of the finite-dimensional distributions of $\Psi_l(T) (s)$ toward $\Psi_l(s)$. Next, we need to prove the tightness of the sequence $\{\Psi_l(T) (s), T \geq 1\}$. More specifically, we shall show that the family of distributions on the space of continuous functions $\mathbb{C}_b(K)$
generated by the contractions of $\Psi_{l,T}(s)$ on $\{|s| \leq K\}$ are dense. For any $l \in L$ the inequality $l(u) \leq 2^r \left(1 + |u|^2\right)^r$ holds for some $r$. Let

$$\Upsilon_K(\varpi) \triangleq \int_{\mathbb{R}} \sup_{|s| \leq K, |y| \leq \varpi} |l(s + y - u) - l(s - u)| \left(1 + |u|^2\right)^{-r-1} du.$$ 

Fix $K < \infty$. We show $\lim_{\varpi \downarrow 0} \Upsilon_K(\varpi) = 0$. Note that for any $\kappa > 0$, we can choose a $M$ such that

$$\hat{|u|} > M \sup_{|s| \leq K, |y| \leq \varpi} |l(s + y - u) - l(s - u)| \left(1 + |u|^2\right)^{-r-1} du < \kappa.$$

We now use Lusin’s Theorem [cf. Section 3.3 in Royden and Fitzpatrick (2010)]. Since $l(\cdot)$ is measurable, there exists a continuous function $g(u)$ in the interval $\{u \in \mathbb{R} : |u| \leq K + 2M\}$ which agrees with $l(u)$ except on a set whose measure does not exceed $\kappa \left(2\overline{L}\right)^{-1}$, where $\overline{L}$ is the upper bound of

$$\{u \in \mathbb{R} : |u| \leq K + 2M\}$$

Denote the modulus of continuity of $g(\cdot)$ by $w_g(\varpi)$. Without loss of generality assume $|g(u)| \leq \overline{L}$ for all $u$ satisfying $|u| \leq K + 2M$. Then,

$$\int_{|u| > M} \sup_{|s| \leq K, |y| \leq \varpi} |l(s + y - u) - l(s - u)| \left(1 + |u|^2\right)^{-r-1} du$$

$$\leq \int_{\mathbb{R}} \sup_{|s| \leq K, |y| \leq \varpi} |l(s + y - u) - l(s - u)| \left(1 + |u|^2\right)^{-r-1} du$$

$$\leq w_g(\varpi) \int_{\mathbb{R}} \sup_{|s| \leq K, |y| \leq \varpi} \left(1 + |u|^2\right)^{-r-k} du + 2\overline{L}\text{Leb}\{u \in \mathbb{R} : |u| \leq K + 2M, l \neq g\},$$

and $\overline{L} \leq Cw_g(\varpi) + \kappa$ for some $C$. Hence, $\Upsilon_K(\varpi) \leq Cw_g(\varpi) + 2\kappa$ since $\kappa$ can be chosen arbitrary small and (for each fixed $\kappa$) $w_g(\varpi) \to 0$ as $\varpi \downarrow 0$ by definition. By
Assumption 4.11, there exists a number $C < \infty$ such that

$$
\begin{align*}
\mathbb E \left[ \sup_{|s| \leq K, |y| \leq \omega} |\Psi_{l,T}(s + y) - \Psi_{l,T}(s)| \right] \\
\leq \int_{\mathbb R} \sup_{|s| \leq K, |y| \leq \omega} |l(s + y - u) - l(s - u)| \\
\mathbb E \left( \frac{\exp \left( \bar{G}_{T,v}(u, \bar{v}) + Q_{T,v}(u) \right)}{\int_{U_T} \exp \left( \bar{G}_{T,v}(w, \bar{v}) + Q_{T,v}(w) \right) dw} \right) du \\
\leq C \Upsilon_K (\omega).
\end{align*}
$$

Markov’s inequality together with the above bound establish that the family of distributions generated by the contractions of $\Psi_{T,l}$ is dense in $C_b(K)$. Since the finite-dimensional convergence in distribution was demonstrated above, we can deduce the weak convergence $\Psi_{l,T} \Rightarrow \Psi_l$ in $D_b(V)$ uniformly in $\lambda_0^0 \in K$. Finally, we examine the oscillations of the minimum points of the sample criterion $\Psi_{l,T}$. Consider an open bounded interval $A$ that satisfies $\mathbb P \{ \xi_0^0 \in b(A) \} = 0$, where $b(A)$ denotes the boundary of the set $A$. Choose a real number $K$ sufficiently large such that $A \subset \{ s : |s| \leq K \}$ and define for $|s| \leq K$ the functionals $H_A(\Psi) = \inf_{s \in A} \Psi_l(s)$ and $H_{A^c}(\Psi) = \inf_{s \in A^c} \Psi_l(s)$. Let $M_T$ denote the set of minimum points of $\Psi_{l,T}$. We have

$$
\begin{align*}
\mathbb P [M_T \subset A] = \mathbb P [H_A(\Psi) < H_{A^c}(\Psi), M_T \subset \{ s : |s| \leq K \}] \\
\geq \mathbb P [H_A(\Psi) < H_{A^c}(\Psi)] - \mathbb P [M_T \notin \{ s : |s| \leq K \}].
\end{align*}
$$

Therefore,

$$
\liminf_{T \to \infty} \mathbb P [M_T \subset A] \geq \mathbb P [H_A(\Psi) < H_{A^c}(\Psi)] - \sup_T \mathbb P [M_T \notin \{ s : |s| \leq K \}],
$$

and

$$
\limsup_{T \to \infty} \mathbb P [M_T \subset A] \leq \mathbb P [H_A(\Psi) < H_{A^c}(\Psi)].
$$

Moreover, the minimum of
the population criterion $\Psi_l(\cdot)$ satisfies $\mathbb{P} [\xi_l^0 \in A] \leq \mathbb{P} [H_A (\Psi) < H_{A^c} (\Psi)]$ and

$$\mathbb{P} [\xi_l^0 \in A] + \mathbb{P} [\|\xi_l^0\| > K] \geq \mathbb{P} [H_A (\Psi) \leq H_{A^c} (\Psi)].$$

Lemma D.1.26 shall be used to deduce that the following relationship holds,

$$\limsup_{T \to \infty} \mathbb{E} \left[ l \left( T \|\delta_T\|^2 \left( \hat{\lambda}^{GL}_b - \lambda^0_b \right) \right) \right] < \infty,$$

for any loss function $l \in L$. Hence, the set $M_T$ of absolute minimum points of the function $\Psi_{l,T}(s)$ are uniformly stochastically bounded for all $T$ large enough: $\lim_{K \to \infty} \mathbb{P} [M_T \not\subset \{s : |s| \leq K\}] = 0$. The latter result together with the uniqueness assumption (cf. Assumption 4.11) yield

$$\lim_{K \to \infty} \left\{ \sup_{T} \mathbb{P} [M_T \not\subset \{s : |s| \leq K\}] + \mathbb{P} [\|\xi_l^0\| > K] \right\} = 0.$$

Hence, we have

$$\lim_{T \to \infty} \mathbb{P} [M_T \subset A] = \mathbb{P} [\xi_l^0 \in A]. \quad (D.1.32)$$

The last step involves showing that the length of the set $M_T$ approaches zero in probability as $T \to \infty$. Let $A_d$ denote an interval in $\mathbb{R}$ centered at the origin and of length $d < \infty$. Equation (D.1.32) guarantees that $\lim_{d \to \infty} \sup_{T} \mathbb{P} [M_T \not\subset A_d] = 0$. Choose any $\epsilon > 0$ and divide $A_d$ into admissible subintervals whose lengths do not
exceed $\epsilon/2$. Then,

$$
\mathbb{P} \left[ \sup_{s_i, s_j \in \mathcal{M}_T} |s_i - s_j| > \epsilon \right]
\leq \mathbb{P} [\mathcal{M}_T \not\subset \mathcal{A}_d] + (1 + 2d/\epsilon) \sup \mathbb{P} [H_\mathcal{A} (\Psi_{l,T}) = H_{\mathcal{A}^c} (\Psi_{l,T})],
$$

where the term $1 + 2d/\epsilon$ is an upper bound on the admissible number of subintervals and the supremum in the second term is over all possible open bounded subintervals $\mathcal{A} \subset \mathcal{A}_d$. The weak convergence result implies

$$
\mathbb{P} [H_\mathcal{A} (\Psi_{l,T}) = H_{\mathcal{A}^c} (\Psi_{l,T})] \rightarrow \mathbb{P} [H_\mathcal{A} (\Psi_l) = H_{\mathcal{A}^c} (\Psi_l)],
$$
as $T \to \infty$. Since $\mathbb{P} [H_\mathcal{A} (\Psi_l) = H_{\mathcal{A}^c} (\Psi_l)] = 0$ and $\mathbb{P} [\mathcal{M}_T \not\subset \mathcal{A}_d] \rightarrow 0$ for large $d$, then $\mathbb{P} \left[ \sup_{s_i, s_j \in \mathcal{M}_T} |s_i - s_j| > \epsilon \right] = o(1)$. Since $\epsilon > 0$ can be chosen arbitrary small we deduce that the limiting distribution of $T \|\delta_T\|^2 \left( \lambda_0^{GL} - \lambda_0^0 \right)$ converges to the distribution of $\xi_0^0$.

**Lemma D.1.20.** Let $u_1, u_2 \in \mathbb{R}$ be of the same sign with $0 < |u_1| < |u_2|$. For any integer $r > 0$ and some constants $c_r, C_r$ which depends on $r$ only, we have uniformly in $\tilde{v} \in \mathcal{V}$,

$$
\mathbb{E} \left[ \left( \zeta_{T,v}^{1/2r} (u_2, \tilde{v}) - \zeta_{T,v}^{1/2r} (u_1, \tilde{v}) \right)^{2r} \right] \leq c_r \left( |u_2 - u_1| \Sigma_i \right)^r \leq C_r |u_2 - u_1|^r,
$$

where $\Sigma_i$ is defined in Assumption 4.9 and $i = 1$ if $u_1 < 0$ and $i = 2$ if $u_1 > 0$.

**Proof.** The proof is given for the case $u_2 > u_1 > 0$. The other case is similar and thus omitted. We follow closely the proof of Lemma III.5.2 in Ibragimov and Has’minskiî (1981). Let $\mathcal{V} (u_i) = \exp (\mathcal{V} (u_i)), i = 1, 2$. We have $\mathbb{E} \left[ \left( \mathcal{V}^{1/2r} (u_2) - \mathcal{V}^{1/2r} (u_1) \right)^{2r} \right] = \sum_{j=0}^{2r} \binom{2r}{j} (-1)^j \mathbb{E}_{u_1} \left[ \mathcal{V}_{u_1}^{j/2r} (u_2) \right]$, where $\mathcal{V}_{u_1} (u_2) \triangleq \exp (\mathcal{V} (u_2) - \mathcal{V} (u_1))$. Using the
Gaussian property of \( \mathcal{V}(u) \), for each \( u \in \mathbb{R} \), we have

\[
\mathbb{E}_{u_1} [ \mathcal{V}^{j/2r}(u_2) ] = \exp \left( \frac{1}{2} \left( \frac{j}{2r} \right)^2 4 (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 - \frac{j}{2r} |A^0(u_2) - A^0(u_1)| \right).
\]

(D.1.33)

We then have

\[
\mathbb{E} \left[ \left( \mathcal{V}^{j/2r}(u_2) - \mathcal{V}^{j/2r}(u_1) \right)^2 \right] = \sum_{j=0}^{2r} \left( \frac{2r}{j} \right) (-1)^j d^{j/2r}
\]

with

\[
d \triangleq \exp \left( \frac{j}{2r} 2 (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 - |A^0(u_2) - A^0(u_1)| \right).
\]

Let \( B \triangleq 2 (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 - |A^0(u_2) - A^0(u_1)| \). There are different cases to be considered:

1. \( B < 0 \). Note that

\[
d = \exp \left( \frac{j}{2r} 2 (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 - \left( (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 \right) + B \right)
\]

\[
= \exp \left( - \frac{2r - j}{r} (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 \right) e^B,
\]

which then results in

\[
\mathbb{E} \left[ \left( \mathcal{V}^{j/2r}(u_2) - \mathcal{V}^{j/2r}(u_1) \right)^2 \right] \leq p_r(a),
\]

where \( p_r(a) \triangleq \sum_{j=0}^{2r} \left( \frac{2r}{j} \right) (-1)^j a^{(2r-j)} \) and \( a = e^{B/2r} \exp(-r^{-1} (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0) \).

2. \( 2 (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 = |A^0(u_2) - A^0(u_1)| \). This case is the same as the previous one but with \( a = \exp \left( -r^{-1} (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 \right) \).

3. \( B > 0 \). Upon simple manipulations,

\[
\mathbb{E} \left[ \left( \mathcal{V}^{j/2r}(u_2) - \mathcal{V}^{j/2r}(u_1) \right)^2 \right] \leq p_r(a),
\]
where

\[ p_r(a) = e^{-B/2r} \sum_{j=0}^{2r} \binom{2r}{j} (-1)^j a^{(2r-j)}, \]

with \( a = \exp \left( -r^{-1} (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 \right) \). We can thus proceed with the same proof for all the above cases. Let us consider the first case. We show that at the point \( a = 1 \), the polynomial \( p_r(a) \) admits a root of multiplicity \( r \). This can be established by verifying the equalities

\[
\begin{align*}
    p_r(1) &= p_r^{(1)}(1) = \cdots = p_r^{(r-1)}(1) = 0.
\end{align*}
\]

One then recognizes that \( p_r^{(i)}(a) \) is a linear combination of summations \( S_k (k = 0, 1, \ldots, 2r-2) \) given by

\[ S_k = e^B \sum_{j=0}^{2r} \binom{2r}{j} j^k. \]

Thus, one only needs to verify that \( S_k = 0 \) for \( k = 0, 1, \ldots, 2r-2 \). This follows because the expression for \( S_k \) is found by applying the operator \( e^B a (d/da) \) to the function \((1 - a^2)^{2r}\) and evaluating it at \( a = 1 \). Consequently, \( S_k = 0 \) for \( k = 0, 1, \ldots, 2r - 1 \). Using this result into (D.1.34) we find, with \( \bar{p}_r(a) \) being a polynomial of degree \( r^2 - r \),

\[
E \left[ (\mathcal{N}^{1/2r}(u_2) - \mathcal{N}^{1/2r}(u_1))^{2r} \right] = (1 - a)^r \bar{p}_r(a) \leq \left( r^{-1} (\delta^0)' (|u_2 - u_1| \Sigma_2) \delta^0 \right)^r \bar{p}_r(a),
\]

where the last inequality follows from \( 1 - e^{-c} \leq c \), for \( c > 0 \). Next, let \( \zeta_{T,v}^{1/2r}(u_2, u_1) = \zeta_{T,v}^{1/2r}(u_2) - \zeta_{T,v}^{1/2r}(u_1) \). By Lemma D.1.3 and D.1.9, the continuous mapping theorem and (D.1.35), \( \lim_{r \to \infty} \mathbb{E} \left[ \zeta_{T,v}^{1/2r}(u_2, u_1) \right] \leq (1 - a)^r \bar{p}_r(a) \), uniformly in \( v \in \mathbb{V} \). Noting that \( j \leq 2r \), we can set \( C_r = \max_{0 \leq a \leq 1} e^B \bar{p}_r(a) / r^r \) to prove the claim of the lemma.

\[ \square \]

**Lemma D.1.21.** For \( u_1, u_2 \in \mathbb{R} \) being of the same sign and satisfying \( 0 < |u_1| < \)
\(|u_2| < K < \infty\). Then, for all \(T\) sufficiently large, we have

\[
\mathbb{E} \left[ \left( \zeta_{T,v}^{1/4} (u_2, \bar{v}) - \zeta_{T,v}^{1/4} (u_1, \bar{v}) \right)^4 \right] \leq C_1 |u_2 - u_1|^2,
\]

where \(0 < C_1 < \infty\). Furthermore, for the constant \(C_1\) from Lemma D.1.20, we have

\[
\mathbb{P} \left[ \zeta_{T,v} (u, \bar{v}) > \exp (-3C_1 |u|/2) \right] \leq \exp (-C_1 |u|/4).
\]

Both relationships are valid uniformly in \(\bar{v} \in V\).

**Proof.** Suppose \(u > 0\). The relationship in (D.1.36) follows from Lemma D.1.20 with \(r = 2\). By Markov’s inequality and Lemma D.1.20,

\[
\mathbb{P} \left[ \zeta_{T,v} (u, \bar{v}) > \exp (-3C_1 |u|/2) \right] \leq \exp \left( 3C_1 |u|/4 - \left( \delta^0 \right) \left( |u| \Sigma_2 \right) \delta^0 \right) \leq \exp \left( -C_1 |u|/4 \right).
\]

\(\square\)

**Lemma D.1.22.** Under the conditions of Lemma D.1.21, for any \(\vartheta > 0\) there exists a finite real number \(c_\vartheta\) and a \(\overline{T}\) such that for all \(T > \overline{T}\),

\[
\sup_{\bar{v} \in V} \mathbb{P} \left[ \sup_{|u| > M} \zeta_{T,v} (u, \bar{v}) > M^{-\vartheta} \right] \leq c_\vartheta M^{-\vartheta}.
\]

**Proof.** It can be shown by using Lemma D.1.20-D.1.21. \(\square\)

**Lemma D.1.23.** For every sufficiently small \(\epsilon \leq \tau\), where \(\tau\) depends on the smoothness
of \( \pi (\cdot) \), there exists \( 0 < C < \infty \) such that

\[
\mathbb{P} \left[ \int_0^\epsilon \zeta_{T,v} (u, \tilde{v}) \pi (\lambda_b^0 + u/\psi_T) \, du < \epsilon \pi (\lambda_b^0) \right] < C \epsilon^{1/2}. \tag{D.1.38}
\]

**Proof.** Since \( \mathbb{E} (\zeta_{T,v} (0, \tilde{v})) = 1 \) and \( \mathbb{E} (\zeta_{T,v} (u, \tilde{v})) \leq 1 \) for sufficiently large \( T \), we have

\[
\mathbb{E} \left| \zeta_{T,v} (u, \tilde{v}) - \zeta_{T,v} (0, \tilde{v}) \right| \leq \left( \mathbb{E} \left| \zeta_{T,v}^{1/2} (u, \tilde{v}) + \zeta_{T,v}^{1/2} (0, \tilde{v}) \right|^2 \mathbb{E} \left| \zeta_{T,v}^{1/2} (u, \tilde{v}) - \zeta_{T,v}^{1/2} (0, \tilde{v}) \right|^2 \right)^{1/2} \leq C |u|^{1/2}
\]

by Lemma D.1.20 with \( r = 1 \). By Assumption 4.6,

\[
\left| \pi_{T,v} (u) - \pi^0 \right| \leq \pi \left( \lambda_{b,T} (v) \right) - \pi^0 \pi_T^{-1} |u|,
\]

with \( C > 0 \). The first term on the right-hand side is \( o (1) \) (and independent of \( u \)) while the second is asymptotically negligible for small \( u \). Thus, for a sufficiently small \( \epsilon > 0 \),

\[
\int_0^\epsilon \zeta_{T,v} (u, \tilde{v}) \pi_{T,v} (u) \, du > \frac{\pi^0}{2} \int_0^\epsilon \zeta_{T,v} (u, \tilde{v}) \, du.
\]

Next, using \( \zeta_{T,v} (0, \tilde{v}) = 1 \),

\[
\mathbb{P} \left[ \int_0^\epsilon \zeta_{T,v} (u, \tilde{v}) \pi_{T,v} (u) \, du < \epsilon/2 \right] \leq \mathbb{P} \left[ \int_0^\epsilon ( \zeta_{T,v} (u, \tilde{v}) - \zeta_{T,v} (0, \tilde{v}) ) \, du < -\epsilon/2 \right] \\
\leq \mathbb{P} \left[ \int_0^\epsilon |\zeta_{T,v} (u, \tilde{v}) - \zeta_{T,v} (0, \tilde{v})| \, du > \epsilon/2 \right],
\]

and by Markov’s inequality together with (D.1.39) the last expression is less than or
equal to

\[ (2/\epsilon) \int_0^\epsilon \mathbb{E} |\zeta_{T,v} (u, \bar{v}) - \zeta_{T,v} (0, \bar{v})| \, du < 2C\epsilon^{1/2}. \]

\[ \square \]

**Lemma D.1.24.** For \( f_T \in F \), and \( M \) sufficiently large, there exist constants \( c, C > 0 \) such that

\[ \mathbb{P} [J_{1,M} > \exp (-cf_T (M))] \leq C \left( 1 + M^C \right) \exp (-cf_T (M)), \quad (D.1.40) \]

uniformly in \( \bar{v} \in V \).

**Proof.** In view of the smootheness property of \( \pi (\cdot) \), without loss of generality we consider the case of the uniform prior (i.e., \( \pi_{T,v} (u) = 1 \) for all \( u \)). We begin by dividing the open interval \( \{u : M \leq |u| < M + 1\} \) into \( I \) disjoint segments denoting the \( i \)-th one by \( \Pi_i \). For each segment \( \Pi_i \) choose a point \( u_i \) and define

\[ J_{1,M}^I \triangleq \sup_{\bar{v} \in V} \sum_{i \in I} \zeta_{T,v} (u_i, \bar{v}) \text{Leb} (\Pi_i) = \sup_{\bar{v} \in V} \sum_{i \in I} \int_{\Pi_i} \zeta_{T,v} (u_i, \bar{v}) \, du. \]

Then,

\[ \mathbb{P} \left[ J_{1,M}^I > (1/4) \exp (-cf_T (M)) \right] \]

\[ \leq \mathbb{P} \left[ \max_{i \in I} \sup_{\bar{v} \in V} \zeta_{T,v}^{1/2} (u_i, \bar{v}) \text{Leb} (\Gamma_M)^{1/2} > (1/2) \exp (-f_T (M)/2) \right] \]

\[ \leq \sum_{i \in I} \mathbb{P} \left[ \zeta_{T,v}^{1/2} (u_i, \bar{v}) > (1/2) \text{Leb} (\Gamma_M)^{-1/2} \exp (-f_T (M)/2) \right] \]

\[ \leq 2I \left( \text{Leb} (\Gamma_M) \right)^{1/2} \exp (-f_T (M)/12), \quad (D.1.41) \]

where the last inequality follows from applying Lemma D.1.21 to each summand. Upon using the inequality \( \exp (-f_T (M)/2) < 1/2 \) (which is valid for sufficiently
large $M$), we have

$$
\mathbb{P} [J_{1,M} > \exp (-f_T (M) /2)] \leq \mathbb{P} \left[ |J_{1,M} - J_{1,M}^\Pi| > (1/2) \exp (-f_T (M) /2) \right] \\
+ \mathbb{P} \left[ J_{1,M}^\Pi > \exp (-f_T (M)) \right].
$$

Focusing on the first term,

$$
\mathbb{E} \left[ J_{1,M} - J_{1,M}^\Pi \right] \\
\leq \sum_{i \in I} \int_{\Pi_i} \mathbb{E} \left| \zeta_{T,v}^{1/2} (u, \bar{v}) - \zeta_{T,v}^{1/2} (u_i, \bar{v}) \right| du \\
\leq \sum_{i \in I} \int_{\Pi_i} \left( \mathbb{E} \left| \zeta_{T,v}^{1/2} (u, \bar{v}) + \zeta_{T,v}^{1/2} (u_i, \bar{v}) \right| \mathbb{E} \left| \zeta_{T,v}^{1/2} (u, \bar{v}) - \zeta_{T,v}^{1/2} (u_i, \bar{v}) \right| \right)^{1/2} du \\
\leq C (1 + M)^C \sum_{i \in I} \int_{\Pi_i} |u_i - u|^{1/2} du,
$$

where for the last inequality we have used Lemma D.1.21 since we can always choose the partition of the segments such that each $\Pi_i$ contains either positive or negative $u_i$. Since each summand on the right-hand side above is less than $C (MI^{-1})^{3/2}$ there exist numbers $C_1$ and $C_2$ such that

$$
\mathbb{E} \left[ J_{1,M} - J_{1,M}^\Pi \right] \leq C_1 \left( 1 + M^{-C_2} \right) I^{-1/2}.
$$

Using (D.1.41) and (D.1.42) we have

$$
\mathbb{P} [J_{1,M} > \exp (-f_T (M) /2)] \\
\leq C_1 \left( 1 + M^{-C_2} \right) I^{-1/2} + 2I \left( \text{Leb} (\Gamma_M) \right)^{1/2} \exp (-f_T (M) /12).
$$

The relationship in the last display leads to the claim of the lemma if we choose $I$ to satisfy $1 \leq I^{3/2} \exp (-f_T (M) /4) \leq 2$. \qed
Lemma D.1.25. For \( f_T \in F \), and \( M \) sufficiently large, there exist constants \( c, C > 0 \) such that

\[
\mathbb{E} \left[ J_{1,M}/J_2 \right] \leq C \left( 1 + M^C \right) \exp \left( -c f_T (M) \right),
\]  
(D.1.43)

uniformly in \( \tilde{v} \in V \).

Proof. Note that \( J_{1,M}/J_2 \leq 1 \). Thus, for any \( \epsilon > 0 \),

\[
\mathbb{E} \left[ J_{1,M}/J_2 \right] \leq \mathbb{P} \left[ J_{1,M} > \exp \left( -c f_T (M) /2 \right) \right] + (4/\epsilon) \exp \left( -c f_T (M) \right) \leq \mathbb{P} \left[ \zeta_{T,v} (u, \tilde{v}) du < \epsilon /4 \right] + (4/\epsilon) \exp \left( -c f_T (M) \right).
\]

By Lemma D.1.24, the first term is bounded by \( C \left( 1 + M^C \right) \exp \left( -c f_T (M) /4 \right) \) while for the last term we can use (D.1.38) to deduce

\[
\mathbb{E} \left[ J_{1,M}/J_2 \right] \leq C \left( 1 + M^C \right) \exp \left( -c f_T (M) \right) + (4/\epsilon) \exp \left( -c f_T (M) \right) + C\epsilon^{1/2}.
\]

Finally, choose \( \epsilon = \exp \left( (-2c/3) f_T (M) \right) \) to verify the claim of the lemma. \( \square \)

Lemma D.1.26. For \( l \in L \) and and any \( \vartheta > 0 \),

\[
\lim_{B \to \infty} \lim_{T \to \infty} B^d \mathbb{P} \left[ \psi_T \left( \lambda^0_{b} - \lambda^0_{b} \right) > B \right] = 0.
\]

Proof. Let \( p_T (u) \triangleq p_{1,T} (u) / p_T \) where \( p_{1,T} (u) = \exp \left( \tilde{G}_{T,v} (u, \tilde{v}) + Q_{T,v} (u) \right) \) and \( p_T \triangleq \int_{U_T} p_{1,T} (w) dw \). By definition, \( \lambda^0_{b} \) is the minimum of the function

\[
\int_{F_0} l \left( T \| \delta_T \|^2 (s - u) \right) p_{1,T} (u) \pi_{T,v} (u) du,
\]
with \( s \in \Gamma^0 \). Upon using a change in variables,

\[
\int_{\Gamma^0} l \left( T \| \delta_T \|^2 (s - u) \right) p_{1,T}(u) \pi_{T,v}(u) \, du = \left( T \| \delta_T \|^2 \right)^{-1} p_T \int_{U_T} l \left( T \| \delta_T \|^2 (s - \lambda^0_b) - u \right) p_T \left( \lambda^0_{b,T}(v) + (T \| \delta_T \|^2)^{-1} u \right)
\times \pi_{T,v} \left( \lambda^0_{b,T}(v) + (T \| \delta_T \|^2)^{-1} u \right) \, du.
\]

Thus, \( \lambda_{\delta,T} \triangleq T \| \delta_T \|^2 \left( \lambda^0_{b,\text{GL}} - \lambda^0_b \right) \) is the minimum of the function

\[
S_T(s) \triangleq \int_{U_T} l(s - u) \times \frac{p_T \left( \lambda^0_b + (T \| \delta_T \|^2)^{-1} u \right) \pi_{T,v} \left( \lambda^0_b + (T \| \delta_T \|^2)^{-1} u \right)}{\int_{U_T} p_T \left( \lambda^0_b + (T \| \delta_T \|^2)^{-1} u \right) \pi_{T,v} \left( \lambda^0_b + (T \| \delta_T \|^2)^{-1} u \right) \, du}
\]

where the optimization is over \( U_T \). The random function \( S_T(\cdot) \) converges with probability one in view of Lemmas D.1.24-D.1.25 together with the properties of the loss function \( l \) [cf. (D.1.29) and the discussion surrounding it]. Therefore, we shall show that the random function \( S_T(s) \) is strictly larger than \( S_T(0) \) on \( \{|s| > B\} \) with high probability as \( T \to \infty \). This reflects that

\[
\mathbb{P} \left[ T \| \delta_T \|^2 \left( \lambda^0_{b,\text{GL}} - \lambda^0_b \right) > B \right] \leq \mathbb{P} \left[ \inf_{|s| > B} S_T(s) \leq S_T(0) \right]. \tag{D.1.44}
\]

We present the proof for the case \( \pi_{T,v}(u) = 1 \) for all \( u \). The general case follows with no additional difficulties due to the assumptions satisfied by the prior \( \pi(\cdot) \). By the properties of the family \( L \) of loss functions, we can find \( \overline{u}_1, \overline{u}_2 \in \mathbb{R} \), with \( 0 < \overline{u}_1 < \overline{u}_2 \)
such that as $T$ increases,

$$l_{1,T} \triangleq \sup \{ l(u) : u \in \Gamma_{1,T} \} < l_{2,T} \triangleq \inf \{ l(u) : u \in \Gamma_{2,T} \},$$

where $\Gamma_{1,T} \triangleq U_T \cap (|u| \leq \overline{u}_1)$ and $\Gamma_{2,T} \triangleq U_T \cap (|u| > \overline{u}_2)$. With this notation,

$$S_T(0) \leq l_{1,T} \int_{\Gamma_{1,T}} p_T(u) \, du + \int_{U_T \cap (|u| > \overline{u}_1)} l(u) p_T(u) \, du.$$

Furthermore, if $l \in L$, then for sufficiently large $B$ the following relationships hold: (i) $l(u) - \inf_{|u| > B/2} l(v) \leq 0$; (ii) $|u| \leq (B/2)^\vartheta$, $\vartheta > 0$. We shall assume that $B$ is chosen so that $B > 2\overline{u}_2$ and $(B/2)^\vartheta > \overline{u}_2$ remain satisfied. Let $\Gamma_{T,B} \triangleq \{ u : (|u| > B/2) \cap U_T \}$.

Then, whenever $|s| > B$ and $|u| \leq B/2$, we have,

$$|u - s| > B/2 > \overline{u}_2 \quad \text{and} \quad \inf_{u \in \Gamma_{T,B}} l(u) \geq l_{2,T}. \quad (D.1.45)$$

With this notation,

$$\inf_{|s| > B} S_T(s) \geq \inf_{u \in \Gamma_{T,B}} l_T(u) \int_{(|w| \leq B/2) \cap U_T} p_T(w) \, dw \geq l_{2,T} \int_{(|u| \leq B/2) \cap U_T} p_T(w) \, dw,$$

from which it follows that

$$S_T(0) - \inf_{|s| > B} S_T(s) \leq -\varpi \int_{\Gamma_{1,T}} p_T(u) \, du$$

$$+ \int_{U_T \cap (|u| > B/2)^\vartheta \geq |u| \geq \overline{u}_1} \left( l(u) - \inf_{|s| > B/2} l_T(s) \right) p_T(u) \, du$$

$$+ \int_{U_T \cap (|u| > (B/2)^\vartheta)} l(u) p_T(u) \, du.$$
where \( \varpi \triangleq l_{2,T} - l_{1,T} \). The last inequality can be manipulated further using (D.1.45),
\[
S_T(0) - \inf_{|s| > B} S_T(s) \leq -\varpi \int_{\Gamma_{1,T}} p_T(u) \, du.
\]
(D.1.46)
\[
+ \int_{U_T \cap \{|u| > (B/2)\}} l_T(u) \, p_T(u) \, du.
\]
Let \( B_\vartheta \triangleq (B/2)^\vartheta \) and fix an arbitrary number \( \overline{\varpi} > 0 \). For the first term of (D.1.46), Lemma D.1.23 implies that for sufficiently large \( T \), we have
\[
\mathbb{P} \left[ \int_{\Gamma_{1,T}} p_T(u) \, du < 2 \left( \varpi B^\vartheta \right)^{-1} \right] \leq c \left( \varpi B^\vartheta \right)^{-1/2},
\]
(D.1.47)
where \( c < \infty \) is positive. Next, let us consider the second term of (D.1.46). We show that for large enough \( T \), an arbitrary number \( \overline{\varpi} > 0 \), and for some \( 0 < c < \infty \),
\[
\mathbb{P} \left[ \int_{U_T \cap \{|u| > B_\vartheta\}} l(u) \, p_T(u) \, du > B^{-\overline{\varpi}} \right] \leq c B^{-\overline{\varpi}}.
\]
(D.1.48)
Since \( l \in L \), we have \( l(u) \leq |u|^a \), \( a > 0 \) when \( u \) is large enough. Choosing \( B \) large leads to
\[
\mathbb{E} \left[ \int_{U_T \cap \{|u| > B_\vartheta\}} l(u) \, p_T(u) \, du \right] \leq \sum_{i=0}^{\infty} (B_\vartheta + i + 1)^a \mathbb{E} \left( J_{1,B_\vartheta+i}/J_2 \right),
\]
where \( J_{1,B_\vartheta+i}, J_2 \) are defined as in (D.1.27). By Lemma D.1.25,
\[
\mathbb{E} \left( J_{1,B_\vartheta+i}/J_2 \right) \leq c (1 + (B_\vartheta + i)^a) \exp \left( -bf_T(B_\vartheta + i) \right),
\]
where \( f_T \in F \) and thus for some \( b, c < \infty \),

\[
\mathbb{E} \left[ \int_{U_T \cap \{|u| > B_\varphi\}} l(u) p_T(u) \, du \right] \\
\leq c \int_{B_\varphi}^{\infty} (1 + v^a) \exp(-bf_T(v)) \, dv \leq c \exp(-bf_T(B_\varphi)).
\]

By property (ii) of the function \( f_T \) in the class \( F \), for any \( d \in \mathbb{R} \),

\[
\lim_{v \to \infty} \lim_{T \to \infty} v^d e^{-bf_T(v)} = 0.
\]

Thus, we know that for \( T \) large enough and some \( c < \infty \),

\[
\mathbb{E} \left[ \int_{U_T \cap \{|u| > B_\varphi\}} l(u) p_T(u) \, du \right] \leq cB^{-2\pi},
\]

from which we deduce (D.1.48) after applying Markov’s inequality. Therefore, for sufficiently large \( T \) and large \( B \), combining equation (D.1.44), and (D.1.47)-(D.1.48), we have

\[
\mathbb{P} \left[ T \|\delta_T\|^2 \left( \lambda_{GL}^\varphi - \lambda_p^0 \right) > B \right] \\
\leq \mathbb{P} \left[ -w \int_{\Gamma_1,T} p_T(u) \, du + \int_{U_T \cap \{|u| > B_\varphi\}} l_T(u) p_T(u) \, du \leq 0 \right] \\
\leq \mathbb{P} \left[ \int_{\Gamma_1,T} p_T(u) \, du < 2 \left( \frac{wB^\pi}{(\varphi)} \right)^{-1} \right] + \mathbb{P} \left[ \int_{U_T \cap \{|u| > B_\varphi\}} l(u) p_T(u) \, du > B^{-\pi} \right] \\
\leq c \left( B^{-\pi/2} + B^{-\pi} \right),
\]

which can be made arbitrarily small by choosing \( B \) large enough. \( \square \)

**Lemma D.1.27.** As \( T \to \infty \), the marginal distributions of \( \zeta_{T,v}(u, \tilde{v}) \) converge to the marginal distributions of \( \exp(\mathcal{V}(u, \tilde{v}, v)) \).
Proof. The results follows from Lemma D.1.3, Lemma D.1.9 and the continuous mapping theorem. □

### D.1.4 Proofs of Section 4.4

#### D.1.4.1 Proof of Proposition 4.4.1

The preliminary lemmas below consider the Gaussian process \( W \) on the positive half-line with \( s > 0 \). The case \( s \leq 0 \) is similar and omitted. The generic constant \( C > 0 \) used in the proofs of this section may change from line to line.

**Lemma D.1.28.** For \( \varpi > 3/4 \), we have \( \lim_{T \to \infty} \lim \sup_{|s| \to \infty} |\widehat{W}_T(s)|/|s|^\varpi = 0 \), \( \mathbb{P} \)-a.s.

**Proof.** For any \( \epsilon > 0 \), if we can show that

\[
\sum_{i=1}^{\infty} \mathbb{P} \left[ \sup_{i-1 \leq |s| < i} \left| \frac{\widehat{W}_T(s)}{|s|^\varpi} \right| > \epsilon \right] < \infty, \tag{D.1.49}
\]

then by the Borel-Cantelli lemma, \( \mathbb{P} \left[ \lim \sup_{|s| \to \infty} \left| \frac{\widehat{W}_T(s)}{|s|^\varpi} \right| > \epsilon \right] = 0 \). Proceeding as in the proof of Lemma D.1.10,

\[
\mathbb{P} \left[ \sup_{i-1 \leq |s| < i} \left| \frac{\widehat{W}_T(s)}{|s|^\varpi} \right| > \epsilon \right] \leq \mathbb{P} \left[ \sup_{|s| \leq 1} \left| \frac{\widehat{W}_T(s)}{\epsilon i^{\varpi-1/2}} \right| > 1 \right] \leq \frac{1}{\epsilon^4} \mathbb{E} \left[ \mathbb{E} \left( \sup_{|s| \leq 1} \left| \frac{\widehat{W}_T(s)}{\epsilon i^{\varpi-1/2}} \right|^4 \mathbb{E} \right) \right] \frac{1}{i^{4\varpi-2}}.
\]

The series \( \sum_{i=1}^{\infty} i^{-p} \) is a Riemann’s zeta function and satisfies \( \sum_{i=1}^{\infty} i^{-p} < \infty \) if \( p > 1 \).
Then,
\[
\sum_{i=1}^{\infty} \mathbb{P} \left[ \sup_{|s| < i \leq |s| < i} \left| \tilde{W}_T(s) \right| / |s|^{2\varphi} > \varepsilon \right] \leq \left( C / \varepsilon^4 \right) \mathbb{E} \left[ \left( \sup_{|s| \leq 1} \left( \tilde{W}_T(s) \right)^4 \right) \left( \hat{\Sigma}_T \right) \right] \\
\leq \left( C / \varepsilon^4 \right) \mathbb{E} \left[ \left( \sup_{|s| \leq 1} \tilde{W}_T(s) \right) \left( \hat{\Sigma}_T \right) \right]^4,
\tag{D.1.50}
\]
where \( C > 0 \) and the last inequality follows from Proposition A.2.4 in van der Vaart and Wellner (1996). The process \( \tilde{W}_T \), conditional on \( \hat{\Sigma}_T \), is sub-Gaussian with respect to the semimetric \( d^2_{VW}(t, s) = \hat{\Sigma}_T(t, t) + \hat{\Sigma}_T(s, s) \), which by invoking Assumption 4.12-(ii,iii) is bounded by
\[
\hat{\Sigma}_T(t - s, t - s) \leq |t - s| \sup_{|s| = 1} \hat{\Sigma}_T(s, s).
\]
Theorem 2.2.8 in van der Vaart and Wellner (1996) then implies
\[
\mathbb{E} \left( \sup_{|s| \leq 1} \tilde{W}_T(s) \right) \leq C \sup_{|s| = 1} \hat{\Sigma}_T^{1/2}(s, s).
\]
The above inequality can be used into the right-hand side of (D.1.50) to deduce that the latter is bounded by \( C \mathbb{E} \left( \sup_{|s| = 1} \hat{\Sigma}_T^2(s, s) \right) \). By Assumption 4.12-(iv)
\[
C \mathbb{E} \left( \sup_{|s| = 1} \hat{\Sigma}_T^2(s, s) \right) < \infty,
\]
and the proof is concluded. \( \square \)

**Lemma D.1.29.** \( \{ \tilde{W}_T \} \) converges weakly toward \( \mathcal{W} \) on compact subsets of \( \mathbb{D}_b \).

**Proof.** By the definition of \( \tilde{W}_T(\cdot) \), we have the finite-dimensional convergence in distribution of \( \tilde{W}_T \) toward \( \mathcal{W} \). Hence, it remains to show the (asymptotic) stochastic equicontinuity of the sequence of processes \( \{ \tilde{W}_T, T \geq 1 \} \). Let \( C \subset \mathbb{R}_+ \) be any compact
set. Fix any $\eta > 0$ and $\epsilon > 0$. We show that for any positive sequence \(\{d_T\}\), with \(d_T \downarrow 0\), and for every $t, s \in \mathbb{C}$,

\[
\limsup_{T \to \infty} \mathbb{P} \left( \sup_{|t-s| < d_T} |\mathcal{W}_T(t) - \mathcal{W}_T(s)| > \eta \right) < \epsilon.
\] (D.1.51)

By Markov’s inequality,

\[
\mathbb{P} \left( \sup_{|t-s| < d_T} |\mathcal{W}_T(t) - \mathcal{W}_T(s)| > \eta \right) \leq \mathbb{E} \left( \sup_{|t-s| < d_T} |\mathcal{W}_T(t) - \mathcal{W}_T(s)| \right) / \eta.
\]

Let $\hat{\mathcal{Y}}_T(t, s)$ denote the covariance matrix of $\left(\mathcal{W}_T(t), \mathcal{W}_T(s)\right)'$ and $\mathcal{N}$ be a two-dimensional standard normal vector. Letting $\iota \triangleq \begin{bmatrix} 1 & -1 \end{bmatrix}'$, we have

\[
\left[ \mathbb{E} \sup_{|t-s| < d_T} |\mathcal{W}_T(t) - \mathcal{W}_T(s)| \right]^2 = \left[ \mathbb{E} \sup_{|t-s| < d_T} |\iota' \hat{\mathcal{Y}}_T^{1/2}(t, s) \mathcal{N}| \right]^2
\leq \mathbb{E} \left[ \sup_{|t-s| < d_T} \iota' \hat{\mathcal{Y}}_T(t, s) \iota \right]
= \mathbb{E} \left[ \sup_{|t-s| < d_T} \hat{\Sigma}_T(t - s, t - s) \right]
\leq d_T \mathbb{E} \left[ \sup_{|s|=1} \hat{\Sigma}_T(s, s) \right],
\]

and so $\mathbb{E} \left[ \sup_{|t-s| < d_T} \hat{\Sigma}_T(t - s, t - s) \right] \leq 2d_T \mathbb{E} \left[ \sup_{|s|=1} \hat{\Sigma}_T(s, s) \right]$ where we have used Assumption 4.12-(iii) in the last step. As $d_T \downarrow 0$ the right-hand side goes to zero since $\mathbb{E} \left[ \sup_{|s|=1} \hat{\Sigma}_T(s, s) \right] = O(1)$ by Assumption 4.12-(iv).

\section*{Lemma D.1.30.}

Fix $0 < a < \infty$. For any $p \in \mathbb{P}$ and for any positive sequence $\{a_T\}$ satisfying $a_T \xrightarrow{\mathbb{P}} a$,

\[
\int_{\mathbb{R}} |p(s)| \exp(\mathcal{W}_T(s)) \exp(-a_T |s|) \, ds \xrightarrow{d} \int_{\mathbb{R}} |p(s)| \exp((\mathcal{W}(s))) \exp(-a |s|) \, ds.
\]
Proof. Let $B_+^+$ be a compact subset of $\mathbb{R}_+/\{0\}$. Let

$$G = \{(W, a_T) \in D_b(\mathbb{R}, \mathcal{B}, \mathbb{P}) \times B_+ : \quad \limsup_{|s| \to \infty} |W(s)|/|s|^{\varpi} = 0, \quad \varpi > 3/4, \quad a_T = a + a_T(1)\};$$

and denote by $f : G \to \mathbb{R}$ the functional given by

$$f(G) = \int |p(s)| \exp(W(s)) \exp(-a_T|s|) ds.$$

In view of the continuity of $f(\cdot)$ and $a_T \xrightarrow{\mathbb{P}} a$, the claim of the lemma follows by Lemma D.1.28-D.1.29 and the continuous mapping theorem. □

We are now in a position to conclude the proof of Proposition 4.4.1.

Proof. Suppose $\gamma_T = CT \|\delta_T\|^2$ for some $C > 0$. From equation (4.4.1),

$$\tilde{\xi}_T = \frac{\int u \exp\left(\hat{W}_T(u) - \hat{A}_T(u)\right) du}{\int \exp\left(\hat{W}_T(u) - \hat{A}_T(u)\right) du}.$$

By Lemma D.1.30, we deduce that $\tilde{\xi}_T$ converges in law to the distribution stated in (4.3.10). □

D.1.5 Proofs of Section 4.5

Rewrite the GL estimator $\hat{\lambda}^{GL,*}_b(\hat{\theta}) = \hat{\lambda}^{GL,*}_b(\tilde{v}, v)$ as the minimizer of

$$\Psi_{l,T}(s; \tilde{v}, v) \triangleq \int \psi(s - \lambda_b) \times$$

$$\frac{\exp\left(G_T(\theta^0 + \tilde{v}/r_T, \lambda_b) + Q^0_T(\theta^0 + v/r_T, \lambda_b)\right) \pi(\lambda_b)}{\int \exp\left(G_T(\theta^0 + \tilde{v}/r_T, \lambda_b) + Q^0_T(\theta^0 + v/r_T, \lambda_b)\right) \pi(\lambda_b) d\lambda_b} d\lambda_b.$$ (D.1.52)
Since the $\hat{\lambda}_i$'s are asymptotically distinct, the proof of Theorem 4.3.2 can be repeated for each $i = 1, \ldots, m$ separately.


