2014

Robust parameter estimation and pivotal inference under heterogeneous and nonstationary processes

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

ROBUST PARAMETER ESTIMATION AND PIVOTAL INFERENCE UNDER HETEROGENEOUS AND NONSTATIONARY PROCESSES

by

JIE HOU

B.S., Tsinghua University, 2003
M.A., Tsinghua University, 2006
M.Sc., The Hong Kong University of Science and Techonology, 2007

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2014

Approved by

First Reader

Pierre Perron, Ph.D
Professor of Economcs

Second Reader

Zhongjun Qu, Ph.D
Associate Professor of Economcs

Third Reader

Ivan Fernandez-Val, Ph.D
Associate Professor of Economics

Acknowledgements

First and foremost, I wish to express my great gratitude to my advisor, Professor Pierre Perron, for his insightful understanding, invaluable support, and encouragement throughout the whole course of my study. It is he who guided me throughout my journey from owning little knowledge to Econometrics to be at least comfortable with most papers. He is like a light in the darkness leading me to find among myriad possible directions the one to success. He holds high standards on his students as he does on himself, and in the mean time he cares for his students to a degree no one can surpass. His academic professionalism and everlasting strive for perfection inspire me greatly. It is with his help that I distinguish myself a little from a careless student and be a little closer to a researcher. Without his help this thesis would have been no difference from a scratch paper.

I am also specifically grateful to Professor Zhongjun Qu, my paper reader and another great scholar in our field. He is extremely gifted, with deep understanding and extensive knowledge on econometric problems as well as mathematics, and always very friendly and helpful to all students. From many discussions with him I benefited immensely.

I also received enormously helpful comments about my research project and topics from discussions with Professor Ivan Fernandez-Val and Professor Hiroaki Kaido. I also would like to say thanks to all faculty members with whom I sought help and talked, as well as all who have been my instructor, and all the faculty and staff in our department. It is the work of all of them that brings a wonderful academic environment and makes the Deparment of Economics in Boston University one of the top research centers in Economics and the best places to pursue a PhD study.

ROBUST PARAMETER ESTIMATION AND PIVOTAL INFERENCE UNDER HETEROGENEOUS AND NONSTATIONARY PROCESSES

(Order No.)

JIE HOU

Boston University Graduate School of Arts and Sciences, 2014

Major Professor: Pierre Perron, Professor of Economics

ABSTRACT

Robust parameter estimation and pivotal inference is crucial for credible statistical conclusions. This thesis addresses these issues in three contexts: long-memory parameter estimation robust to low frequency nonstationary contamination, long-memory properties of financial time series, and inference on structural changes in a joint segmented trend with heterogeneous noise.

Chapter 1 considers robust estimation of the long-memory parameter allowing for a wide collection of contamination processes, in particular low frequency nonstationary processes such as random level shifts. We propose a robust modified local-Whittle estimator and show it has the usual asymptotic distribution. We also provide modifications to further account for short-memory dynamics and additive noise. The proposed estimator provides substantial efficiency gains compared to existing methods in the presence of contaminations, without sacrificing efficiency when these are absent.

Chapter 2 applies the modified local-Whittle estimator to various volatilities series for stock indices and exchange rates to robustly estimate the long-memory parameter. Our findings suggest that all series are a combination of long and short-memory processes and random level shifts, with the magnitude of each component varying across series. Our results contrast with the view that long-memory is the dominant feature.

Chapter 3 is concerned with pivotal inference about structural changes in a joint segmented trend with heterogeneous noise. We provide tests for changes in the slope and the variance of the noise valid when both may be present, each allowed to occur at different dates. We suggest procedures for four testing problems.

Contents

List of Tables

List of Abbreviations

AR(FI)MA	(Fractionally Integrated) Auto-Regressive Moving-Average
CLT	Central Limit Theorem
DGP	Data Generating Process
FCLT	Functional Central Limit Theorem
FGLS	Feasible GLS
GLS	Generalized Ordinary Least Square
HAC	Heterskdasticity Auto-correlation Consistent
IC	Information Criterion
LFC	Low Frequency Contamination
LLN	Law of Large Numbers
(Trimmed) LP	(Trimmed) Log-Periodogram Estimator
LR	Likelihood Ratio
(Trimmed) LW	(Trimmed) Local-Whittle Estimator
LW(P)LFC	(Perturbed) Modified Local-Whittle Memory Parameter Estimator
OLS	Ordinary Least Square
(Q)MLE	(Quasi) Maximum Likelihood Estimator
(Q)MLF	(Quasi) Maximum Likelihood Function
RLS	Random Level Shift
RSS, or SSR	Sum of Squared Residuals
$\mathbb{R}^T$	T-dimensional linear metric space in real field
Ω^T	Probability space of events of T-dimensional random variables
Ω^∞	Probability space of events of continuous stochastic process
$\stackrel{d}{\to}$	Convergence in distribution
$\Longrightarrow$	Weak convergence in Skorokhod Topology
$\stackrel{p}{\to}$	Convergence in probability
$\to$	Limit as $T \to \infty$ (unless otherwise stated)
$a \curlyvee b$	Maximum of a and b
$x \sim y$	$x/y \stackrel{p}{\to} 1$

Chapter 1

Modified Local Whittle Estimator for Long Memory Processes in the Presence of Low Frequency (and Other) Contaminations (joint with Pierre Perron)

1.1 Introduction

Processes that are persistent in the sense that the serial correlation between distant observations decay hyperbolically are called long memory processes. They have found extensive use in capturing the behavior of many observed series since their introduction by Hurst (1951). A long memory process is also characterized in the frequency domain by a spectral density function proportional to λ^{-2d} as the frequency λ approaches zero at a rate dictated by the the memory parameter d. In terms of parametric modeling, Granger and Joyeux (1980) and Hosking (1981) introduced the fractionally integrated $ARFIMA(p, d, q)$ model, a long-memory generalization of the short-memory $ARMA(p, q)$ process.

The estimators of the memory parameter are divided into parametric and semi-parametric ones. The theory of parametric estimators was developed by Fox and Taqqu (1986) and Dahlhaus (1989), among others. Semiparametric estimators of the memory parameter have become popular since they do not require knowing the specific form of the short memory structure. They are based on the periodograms of the series, and can be categorized into two types: the log-periodogram (LP) estimator first proposed by Geweke and Porter-Hudak (1983) and the local-Whittle (LW) estimator which is credited to Kunsch (1987). The LP estimator is akin to OLS and the LW estimator to the MLE in the frequency domain. Robinson (1995a,b) analyzed the asymptotic properties of these two types of estimators. He showed that they are asymptotically normal, have the same convergence rate and that the asymptotic variance of the LW estimator is smaller than that of the LP estimator.

There are, however, so-called contaminations that have an effect on the bias and efficiency of these semi-parametric estimators, either in finite samples or even asymptotically. Much of the literature so far has focused on providing methods to mitigate the effect of additive noise and/or short-memory dynamics, which have only a finite sample effect. In the case of additive noise or so-called perturbed fractional processes, although both the LW and LP estimators preserve consistency and asymptotic normality, as shown by Deo and Hurvich (2001) and Arteche (2004), they can be severely biased. Hurvich and Ray (2003), Hurvich et al. (2005) and Arteche (2006), among others, have proposed estimators that can reduce the effect of noise by introducing an additive constant or polynomial term in the spectral density function. These methods are all based on local Whittle estimators, given their flexibility in accommodating more structures in the specified data-generating process. The estimators are also strongly biased when substantial short-memory dynamics are present. Among others, Andrews and Sun (2004) considered an adaptive local polynomial Whittle estimator. By substituting a polynomial structure for the constant term used to approximate the behavior of the short memory component near frequency zero in the local Whittle estimator, they showed that their estimator has considerable efficiency gains compared to classic LW and LP estimators under the presence of short memory dynamics. Recently, Frederiksen et al. (2012) combined the two methods and proposed estimators that can simultaneously reduce the bias and mean squared error caused by short memory dynamics and noise perturbation.

There are other low frequency contaminations (denoted as LFC) that can have a more serious effect causing outright inconsistent estimates. They may be important enough to induce researchers to mistakenly conclude that a short memory process with low frequency contaminations is actually a long memory process. Such an effect is often called "spurious long memory". These low frequency contaminations include, but are not confined to, random level shifts, deterministic level shifts and deterministic trends. A short-memory process contaminated by those components will exhibit hyperbolically decaying autocorrelations as well as a pole in its spectral density function at frequency zero, which are characteristics of a long memory process.

Among others, Diebold and Inoue (2001), Granger and Hyung (2004), Mikosch and Stărică (2004) and Perron and Qu (2010) provide theoretical explanations for and simulation evidence of this spurious long memory effect. It has also been argued that models incorporating a short memory process with such low frequency contaminations provide a better in-sample fit and, in particular, forecast better compared to models assuming a pure long memory process. Various studies reported evidence that these forms of data contaminations are in fact very likely present in the volatility of asset prices and considerably weakens the evidence of pure long-memory; see, e.g., Granger and Hyung (2004), Mikosch and Stărică (2004), Stărică and Granger (2005), Perron and Qu (2010), Lu and Perron (2010), Qu and Perron (2013), Varneskov and Perron (2013) and Xu and Perron (2013).

Recent work by Dolado et al. (2005), Ohanissian et al. (2008), Perron and Qu (2010) and Qu (2011) proposed tests in both the time and frequency domain with varying degrees of success. Many have argued that the long-memory properties of many economic time series are indeed spurious. These tests focus on distinguishing between a short memory process affected by low frequency contaminations from a true long memory process. So they do not offer methods to estimate the memory parameter in the presence of low frequency contaminations when the true signal may be of long or short memory.

Recently, attention focused on providing modified LP or LW estimators to account for low frequency contaminations. McCloskey and Perron (2013) proposed trimmed LP estimators that have desirable asymptotic and finite sample properties in the presence of low frequency contaminations. Using a similar trimming technique, McCloskey and Hill (2013) proposed trimmed frequency domain quasi maximum likelihood estimator estimators for short-memory time series models (e.g., ARMA, GARCH and stochastic volatility models) that may be contaminated by low frequency movements. McCloskey (2013) considered a trimmed frequency domain quasi maximum likelihood estimator that can be used to consistently estimate the parameters of a long-memory stochastic volatility model in the presence of low frequency contamination assuming the signal to be an $ARFIMA(p, d, q)$ process. Iacone (2010) considered trimmed LW estimators.

We propose modified LW estimators that work under all kinds of contaminations: low frequency, additive noise and short memory dynamics. Our emphasis is on accounting for low frequency contaminations and we show how to further modify the estimator to account for the other types. It adopts techniques used in Andrews and Sun (2004), Hurvich et al. (2005) and Frederiksen et al. (2012) to introduce additive terms in the frequency domain quasi maximum likelihood function to capture the effect of the low frequency contaminations, based on results of Perron and Qu (2010) and McCloskey and Perron (2013) showing the spectral density function of low frequency contaminations to be of order $O_p(T^{-1}\lambda_k^{-2})$ near frequency zero. To account for additive noise, we follow Hurvich et al. (2005). Interestingly, our modification for low frequency contaminations also reduces the finite sample bias induced by short-memory dynamics, so that no further modification is necessary for this case.

Our modified estimators have the following advantages: being semiparametric, they do not require knowing the structure of the short memory process; they do not require trimming so all data is used; unlike the trimmed LP estimator, they do not require the underlying process to be Gaussian; they have the same asymptotic variance as the standard LW estimator when no contamination is present; without low frequency contaminations, they are asymptotically equivalent to the standard LW estimator that does not account for low frequency contaminations so that no efficiency loss is incurred by incorporating our modifications; they can easily be extended to a full parametric case. When low frequency contaminations are present, it has, in most cases, the smallest bias and mean-squared error amongst all existing estimators designed to control for low frequency contaminations, whether or not other types of contaminations are present. To our knowledge, our contribution is the first to provide an estimator with good properties under all previously considered contaminations: low frequency, additive noise and short-memory dynamics.

The structure of the paper is as follows. Section 2 presents the model and some preliminary results. Section 3 motivates and introduces our modified LW estimator that accounts for possible low frequency contaminations. Section 4 presents results about the consistency and limit distribution. Section 5 discusses how to extend the estimator to account for additive noise and short-memory dynamics. Section 6 presents the results of simulations to assess the finite sample properties under a variety of possible scenarios. Section 7 provides brief concluding remarks. All technical derivations are collected in a mathematical appendix.

The following notation is used throughout: "$\xrightarrow{d}$" stands for convergence in distribution, "$\xrightarrow{p}$" for convergence in probability, "$\rightarrow$" for the limit as $T \rightarrow \infty$ (unless otherwise stated), "$a \curlyvee b$" denotes the maximum of a and b, "$x \sim y$" means that $x/y \xrightarrow{p} 1$.

1.2 The Model and Preliminary Results

We start with some basic definitions of a long memory process. Let $\{y_t\}_{t=1}^{T}$ be a stationary time series with spectral density function $f_y(\lambda)$ at frequency λ given by

$$f_y(\lambda) = G(\lambda)\lambda^{-2d} \text{ as } \lambda \rightarrow 0_+ \tag{1}$$

with $G(\lambda)$ a slowly varying function as $\lambda \rightarrow 0_+$ (i.e., for any real t, $G(t\lambda)/G(\lambda) \rightarrow 1$ as $\lambda \rightarrow 0_+$). When $d > 0$, y_t is a long-memory process with a spectral density function that increases for frequencies that get close to zero. The rate of divergence to infinity depends on the parameter d. Under some general conditions, this low-frequency definition is equivalent to the following long-lag autocorrelation definition (Beran, (1995)). Let $\gamma_y(\tau)$ be the autocorrelation function of y_t. If $\gamma_y(\tau) = c(\tau)\tau^{2d-1}$ as $\tau \rightarrow \infty$, with $c(\tau)$ a slowly varying function as $\tau \rightarrow \infty$, the process is said to have long memory. For $0 < d < 1/2$, this implies that the autocorrelations decreases to zero at a slow hyperbolic rate which depends on the parameter d, in contrast to the fast geometric rate of decay that applies to a short-memory process. Examples of long-memory processes include the popular class of fractionally integrated autoregressive moving average models, though in what follows we shall remain agnostic about the nature of the short-memory component imposing only high level assumptions. When $d = 0$, y_t is a short-memory process.

The Data Generating Process (DGP) considered is one where the series of interest, z_t, is a long or short-memory process plus some low frequency contamination, viz.,

$$z_t = c + y_t + u_t \tag{2}$$

where y_t is a process with memory parameter $d \in [0, 1/2)$ and c a constant. Note that the value $d = 0$ is allowed so the DGP includes a short-memory process contaminated by some low frequency component. The process u_t is the low frequency contamination which will be defined below. We suppose that a sample of size T is available. We define the periodograms of the processes $\{z_t, y_t, u_t\}$ to be, for some frequency ordinate λ_k, $I_{z,k} = I_k = I_z(\lambda_k)$, $I_{y,k} = I_y(\lambda_k)$ and $I_{u,k} = I_u(\lambda_k)$ where $I_w(\lambda) = (2\pi T)^{-1}|\Sigma_{t=1}^T w_t e^{it\lambda}|^2$ for $w = z, y, u$, and their spectral density functions by $f_{z,k} = f_k = f_z(\lambda_k)$, $f_{y,k} = f_y(\lambda_k)$ and $f_{u,k} = f_u(\lambda_k)$. Semiparameteric frequency domain estimators for non-contaminated fractional processes are all based on the local approximation (1) and are robust to the nature of the short memory dynamics since they only use information from periodogram ordinates near the origin.

The local Whittle (LW) estimation method of Kunsch (1987) and Robinson (1995a) has become popular because of its likelihood interpretation, nice asymptotic properties (smaller asymptotic variance compared to log-periodogram estimators), mild assumptions (e.g., no need for a normality assumption) and most importantly in our case, the possibility to easily modify it to accommodate the presence of contaminations. It is defined as the minimizer of the (negative) local Whittle likelihood function in the frequency domain

$$Q(G_0, d) = \frac{1}{m}\sum_{j=1}^{m}[\log(G_0\lambda_j^{-2d}) + I_z(\lambda_j)/(G_0\lambda_j^{-2d})]$$

where $G_0 = G(0)$, $m = m(T)$ is the bandwidth which goes to infinity as $T \to \infty$ but at a slower rate than T, $\lambda_j = 2\pi j/T$ are the Fourier frequencies. Concentrating with respect to G_0, the estimator of d is $\hat{d}_{LW} = \arg\min_d [\log \hat{G}_0(d) - 2dm^{-1}\sum_{j=1}^{m} \log \lambda_j]$, where $\hat{G}_0(d) = m^{-1}\sum_{j=1}^{m} \lambda_j^{2d} I_z(\lambda_j)$. The types of processes considered for the low frequency contamination (LFC) component u_t are laid out in the following definition.

Definition 1 *The low frequency contamination component u_t is generated by one of the following processes. 1) Random level shifts (RLS): $u_t = \Sigma_{t=1}^T \delta_{T,t}$ where $\delta_{T,t} = \pi_{T,t}\eta_t$ with $\eta_t \sim$ i.i.d.*

$N(0,\sigma_\eta^2)$ and $\pi_{T,t} \sim$ i.i.d. Bernoulli$(p/T,1)$ for some $p \geq 0$. The components $\pi_{T,t}$, η_t are mutually independent. 2) Deterministic level shifts: $u_t = \Sigma_{i=1}^{B} c_i \chi(T_{i-1} < t \leq T_i)$ where B is the (fixed) number of regimes ($B-1$ is the number of breaks), $0 < |c_i| < \infty$, $\chi(\cdot)$ is the indicator function, $0 = T_0 < T_1 < ... < T_{B-1} < T_B = T$ and $T_i/T \to \tau_i$ with $0 <.\tau_1 < ... < \tau_{B-1} < 1$. 3) Deterministic trends: $u_t = h(t/T)$ where $h(\cdot)$ is a deterministic non-constant function on $[0,1]$ that is either Lipschitz continuous or monotone with $h(1) = 0$ [1]. 4) Fractional trends: $u_t = O((t+1)^{\phi-1/2})$, $u_0 = 0$, $|u_{t+1} - u_t| = O(|u_t|/t)$ where $\phi \in (-1/2, 1/2)$.

Note that the probability of a level shift in the RLS model is sample size dependent. If this were not the case, u_t would have properties similar to that of a random walk. A defining characteristic of the RLS model is that the average number of level shifts p remains constant as the sample size grows. Note that p can be zero so that the assumption nests the no level shift or no contamination case as well. Perron and Qu (2010) considered the asymptotic properties of the periodogram of this type of process contaminating a short memory process and showed that, for any $k = 1, ..., [T/2]$, $(k^2/T)E(I_{u,k}) \to (p\sigma_\eta^2)/(4\pi^3)$ as $T \to \infty$. Mikosch and Stărică (2004) considered the asymptotic properties of the periodogram for a deterministic level shift component when $B = 2$ (one level shift), with the addition of a short-memory component, and showed that $E(I_{u,k}) = O(T/k^2)$. Kunsch (1987, Lemma 2) considered the asymptotic properties of the periodogram of a short-memory process contaminated by a bounded monotone trend. Qu (2011, Lemma 1) extended Kunsch's results to the Lipschitz continuous case and showed that $E(I_{u,k}) = O(T/k^2)$. Iacone (2010) discussed the order of the periodogram of in the case of a fractional trend and showed that $E(I_{u,k}) = O_p(T/k^2)$.

The common feature of these contaminating processes is that the mean of their periodogram near frequency zero is of order $O(T/k^2)$, or equivalently of order $O(T^{-1}\lambda_k^{-2})$ since $\lambda_k = 2\pi k/T$ (note that the O term could be o since it is possible that $E[(I_{u,k})/(T/k^2)] \to 0$, a case we shall discuss further later). Processes with such LFC as additive components are non-stationary so they do not have the traditionally defined spectral density function. Following common practice in such cases, we define their spectral density function to be the expectation of their periodogram. Since the spectral density function of a long memory process near frequency zero

[1] This includes all cases for which $h(\cdot)$ is monotonic and bounded because we can simply substract $h(1)$ from $h(\cdot)$ and add $h(1)$ to c in (2) to have the same DGP.

is of order $O(\lambda_k^{-2d})$, in general the spectral density function of such contaminating components dominates that of a long memory process at low frequencies and vice-versa at high frequencies. Note that in the representation (1), when the process is contaminated by such LFC, we have $G_u \equiv G_u(0) = \lim_{T\to\infty}(k^2/T)E(I_{u,k})$.

Remark 1 *Definition 1 could be replaced by the condition* $E(I_{u,k}) = G_u(k)(T/k^2)(1 + O(1))$ *with* $G_u(k) \leq B$ *where* B *is a fixed bounded positive constant. Hence,* $E(I_{u,k})/(T/k^2)$ *need not converge to a constant, it only needs to be bounded as* $T \to \infty$*. All LFC in Definition 1 satisfy this property and all results to be presented remain valid under this general condition.*

Unlike short memory dynamics or contaminating noise, which cause only finite sample biases to the memory parameter estimator, the bias caused by LFC usually remains asymptotically. To see when this applies, let $A_k = (k^2/T)E(I_{u,k})$, then one can show that $\lambda_k^{2d} I_k = \lambda_k^{2d} I_{y,k} + A_k O_p(T^{1-2d}/k^{2-2d})$. So the bias introduced by LFC is of order $O_p(m^{-1}T^{1-2d}\sum_{k=1}^m(A_k/k^{2-2d}))$. The following definition will be useful.

Definition 2 *A LFC is said to be non-degenerate if* $\lim_{T\to\infty}\{(k^2/T)E(I_{u,k})\} > 0$ *for every* k*. Otherwise it is said to be degenerate.*

An example of a non-degenerate LFC is a RLS model, in which case $\lim_{T\to\infty}(k^2/T)E(I_{u,k}) = (p\sigma_\eta^2)/(4\pi^3)$. An example of a degenerate LFC is a monotone deterministic trend. The bias caused by a non-degenerate LFC remains asymptotically while the bias caused by a degenerate LFC can either remain or vanish asymptotically, with the degree of the (potentially asymptotic or finite sample) bias depending on d and the bandwidth m.

1.3 The Modified Local Whittle Estimator

Let the Fourier transform of the process z_t be $h_z(\lambda_j) = (2\pi T)^{-1/2}(\sum_{t=1}^T z_j e^{-it\lambda_j})$ so that $f_z(\lambda_k) = E(I_z(\lambda_k)) = E(h_z(\lambda_k)h_z(\lambda_k)^*)$, where "$*$" denotes the complex conjugate value. One may then define the frequency domain pseudo Quasi Maximum Likelihood Function (QMLF) for $h_z(\lambda_k)$ as $\varphi_k = \log(f_z(\lambda_k)) + I_z(\lambda_k)/f_z(\lambda_k)$. When there is no contamination in the data, $f_z(\lambda_k)$ reduces to $f_y(\lambda_k)$ and the standard LW estimator is the minimizer of the pseudo-QMLF. With low frequency contamination given by u_t, a problem is how to construct a useful approximation to $f_z(\lambda_k)$ in such cases. Because the periodogram of u_t is of order $O_p(T^{-1}\lambda_k^{-2})$, a sensible strategy is to add a term $(G_u/T)\lambda_k^{-2}$ to the spectral density function of y_t to control for the low frequency contamination. Accordingly, we consider the pseudo spectral density function $f_k \triangleq f_z(\lambda_k) = G_0\lambda_k^{-2d} + G_u\lambda_k^{-2}/T$. Let $\theta = (G_u/G_0)$ be the signal to noise ratio, the pseudo spectral density function of the observed process is then:

$$f_k \triangleq f_z(\lambda_k) = G_0\lambda_k^{-2d} + G_u\lambda_k^{-2}/T = G_0(\lambda_k^{-2d} + (G_u/G_0)\lambda_k^{-2}/T) = G_0(\lambda_k^{-2d} + \theta\lambda_k^{-2}/T) = G_0 g_k$$

where $g_k = (\lambda_k^{-2d} + \theta\lambda_k^{-2}/T)$.

Remark 2 *f_k is the "pseudo spectral density function" in the sense that it is not the true spectral density function of the data, but an artificial construct aimed at providing a good approximation to the behavior of the generalized spectral density function (i.e., the expectation of the periodogram) and an extended LW type estimator with desirable properties.*

This pseudo spectral density function can then be used to approximate $E(I_{z,k})$ and the pseudo frequency domain QMLF is $\varphi(G, d, \theta) = m^{-1}\sum_{k=1}^{m}\varphi_k(G, d, \theta)$. Using the same technique as in Robinson (1995a), we can concentrate G out of the QMLF using $\hat{G} = m^{-1}\sum_{k=1}^{m}(I_k/g_k)$. Hence, the local Whittle (frequency domain QMLE) estimator applicable under LFC, denoted as the LWLFC estimator, is $(\hat{d}_m, \hat{\theta}_m) = \arg\min_{(d,\theta)} J_m(d, \theta)$, where

$$J_m(d, \theta) = \log(m^{-1}\sum_{k=1}^{m}(I_k/g_k) + m^{-1}\sum_{k=1}^{m}\log(g_k)$$

Remark 3 *The component θ is an "auxiliary variable" in the sense that it is not a parameter of primary interest but is introduced as a tool used to control the influence of the contaminations at low frequencies. Intuitively, θ is the appropriate signal to noise ratio to use as it measures the average of the relative magnitude of the contaminations across all frequencies. For the case of RLS contamination, we have an expression for θ in terms of the parameters of the model, given by $\theta = (G_u/G_0) \sim (2\pi\sigma_\eta^2/\sigma_\varepsilon^2)$; see Perron and Qu (2010).*

Remark 4 *The method can be extended to the case with a parametric specification for the long-memory process. For example, if y_t is assumed to follow the $ARFIMA(p, d, q)$ process $(1 - L)^d y_t = \tilde{y}_t$, where $A(L)\tilde{y}_t = B(L)\varepsilon_t$ and $\varepsilon_t \sim i.i.d.\ N(0, \sigma_\varepsilon^2)$, then we simply replace $G_0\lambda_k^{-2d}$ by $\sigma_\varepsilon^2(|B(e^{-i\lambda})|^2/|A(e^{-i\lambda})|^2)[2\pi|1 - e^{-i\lambda}|^{2d}]^{-1}$ in the objective function $J_m(d, \theta)$.*

1.4 Asymptotic Properties

We start by introducing the assumptions required to obtain the consistency result for the LWLFC estimator. Many are the same as in Robinson (1995a), but some are added or modified to accommodate the LFC components. Henceforth, we shall denote the true value of the long-memory parameter by d_0 and the true value of the signal-to-noise ratio by θ_0.

•Assumption A1. As $\lambda \to 0_+$, $f_y(\lambda) \sim G_0\lambda^{-2d_0}$ where $G_0 \in (0, \infty)$ and $d_0 \in [0, 1/2)$.

•Assumption A2. For λ in a neighborhood of 0, $f_y(\lambda)$ is differentiable and $d\log(f_y(\lambda))\,/d\lambda = O(\lambda^{-1})$.

•Assumption A3. y_t is stationary and admits an infinite MA representation: $y_t - E(y_t) = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{t-j}$ with $\sum_{j=0}^{\infty}\alpha_j^2 < \infty$ where $\{\varepsilon_t\}$ is a martingale difference sequence with $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\varepsilon^2$, $E(\varepsilon_t^3|\mathcal{F}_{t-1}) = \mu_3$, and $E(\varepsilon_t^4) = \mu_4$ where $\mathcal{F}_t$ is the σ-field generated by $\{\varepsilon_s; s \leq t\}$. Also, there exists a random variable ε such that $E(\varepsilon^2) < \infty$ and for all $\eta > 0$ and some $K > 0$, $P(|\varepsilon_t| > \eta) \leq KP(|\varepsilon| > \eta)$.

Remark 5 *We require ε_t to have finite fourth moment even to establish consistency to invoke a strong law of large numbers for $m^{-1}\sum_{k=1}^{m}(I_k/g_k(d, \theta))$ and show that the convergence of the memory parameter estimate does not depend on the signal to noise ratio.*

•Assumption A4. As $T \to \infty$, $T^{(1-(d_0^2-3d_0+9/4)^{-1})\curlyvee(1/2)}/m + m/T \to 0$.

Remark 6 *The requirement on the bandwidth to establish consistency departs from Robinson (1995a) who only requires that $(1/m) + (m/T) \to 0$. This is due to the need to suppress the impact of (I_k/g_k) at low frequencies, $k < T^{[(1-2d_0)/(2-d_0)]}$, in which case the periodogram of the LFC dominates that of the long memory process. With the addition of the term $(\theta/T)\lambda_k^{-2}$ in the QMLF, we can then bound $|I_k/g_k|$. However, to control the effect of $\{I_k/g_k\}$ at high frequencies where the periodogram of the long memory process dominates that of the LFC, we need a larger bandwidth to suppress the cumulative impact from the low frequencies. The closer is d_0 to 0, the higher is the required bandwidth because the contamination will then dominate at higher frequencies. The quantity $(1 - (d_0^2 - 3d_0 + 9/4)^{-1}) \curlyvee (1/2)$ achieves its maximum value $5/9$ when $d_0 = 0$. Hence, in practice with an unknown memory parameter d_0, we need to choose a bandwidth of order greater than $T^{5/9}$.*

•Assumption A5. u_t is one of the LFC as stated in Definition 1.

It will be useful to first establish a limit result pertaining to the estimate $\hat{\theta}_m$ of the signal to noise ratio. This will be used in the proof of the consistency of $\hat{d}_m$.

Lemma 1 *Under A1-A5, if a non-degenerate LFC is present, $\hat{\theta}_m$ is bounded above by zero.*

We now consider the consistency result and a preliminary bound on the convergence rate that will be used to establish the limit distribution of our estimator.

Theorem 1 *Under A1-A5: a) $\hat{d}_m \xrightarrow{p} d_0$ as $T \to \infty$; b) $|\hat{d}_m - d_0| = o_p((\log(m))^{-3})$.*

Note that this result does not require $\hat{\theta}_m$ to be a consistent estimate, all that is required is that if LFC components are present the probability limit of the estimate of $\hat{\theta}_m$ is bounded above by zero, which is guaranteed by Lemma 9. This implies that with probability arbitrarily close to one, $\hat{\theta}_m$ will be in a the set $(0, \infty)$ and we can consider analyzing the limit of $\hat{d}_m$ for any value or sequences of θ_m in the set $(0, \infty)$.

Before proceeding further, we need to discuss a property of the estimate of the signal-to-noise ratio $\hat{\theta}_m$ when there is no LFC present. This, in conjunction with Lemma 9, will allow us to derive the limit distribution of $\hat{d}_m$ for both cases with and without LFC. The required result is stated in the next lemma, which is of independent interest.

Lemma 2 *Suppose no LFC is present and that A1-A4 hold, then, as $T \to \infty$,*

$$\hat{\theta}_m = O_p(T^{-(1-2d_0)/(2-2d_0)}) \to 0$$

To prove the asymptotic normality of $\hat{d}_m$, further assumptions are needed, some of which are strengthened versions of Assumptions A1-A3.

•Assumption A6. For some $\tau \in (0, 2]$, $f_y(\lambda) \sim G_0 \lambda^{-2d_0}(1 + O(\lambda^{\tau}))$ as $\lambda \to 0_+$, where $G_0 \in (0, \infty)$ and $d_0 \in [0, 1/2)$.

•Assumption A7. In a neighborhood of the origin, $f_y(\lambda)$ is differentiable and $df_y(\lambda)/d\lambda = O(f_y(\lambda)/\lambda)$ as $\lambda \to 0_+$.

•Assumption A8. As $T \to \infty$, $m^{-1} + T^{-2\tau} m^{1+2\tau} (\log m)^2 \to 0$.

The following theorem states the asymptotic distribution of the estimate $\hat{d}_m$.

Theorem 2 *Under A1-A8:* $m^{1/2}(\hat{d}_m - d_0) \xrightarrow{d} N(0, 1/4)$ *as* $T \to \infty$.

Note that the asymptotic variance of our estimator is the same as that of the standard LW estimator of Robinson (1995a) applicable with no LFC. The intuitive reason is that, asymptotically, the additional term $G_u(\lambda_k^{-2}/T)$ controls the effect of LFC on the spectral density function well enough so that no efficiency loss ensues.

When the magnitude of the LFC is weak, the asymptotic distribution of Theorem 2 provides a good approximation to the finite sample distribution. However, when the magnitude of the LFC is substantial, $2m^{1/2}(\hat{d}_m - d_0)$ does converge to a normal distribution rapidly as T increases (even with T as small as 512) but the approach to a standard normal may be slow, i.e., the mean and variance of $2m^{1/2}(\hat{d}_m - d_0)$ may converge slowly to 0 and 1, respectively. Some approximate formulas to compute the finite sample bias and variance of $2m^{1/2}(\hat{d}_m - d_0)$ have been found in unreported simulations and they provide good approximations. Unfortunately, they all depend on θ_0, the signal to noise ratio which cannot be identified when it is greater than zero, rendering the corrections not applicable in practice. An important avenue of further research is to obtain a finite-sample scaling factor, say S, to replace m in order to obtain good finite sample coverage rates for the LWLFC estimate. A conjecture is that S should be a decreasing function of θ_0 to reflect the impact of LFC on the variance of the memory parameter estimate. But since $\hat{\theta}_m$ is not a consistent estimator of θ_0, it is unlikely that one can find a good applicable formula. This problem about the coverage rate is not unique to our method, and applies to all existing methods to estimate the memory parameter under some contamination. Alternative scaling factors have been proposed. For the log-periodogram estimator, Geweke and Porter-Hudak (1983) suggested using the scaling factor $S(l,m)^{1/2}$, where $S(l,m) = \sum_{j=1}^{m}(\log j - (m-l+1)^{-1}\sum_{\tau=l}^{m}\log\tau)^2$ for some lower trimming l, and its use was also discussed by Deo and Hurvich (2001). For local Whittle-type estimators, it was used by Hurvich et al. (2005) and Iacone (2010).

1.5 Extension to the Case of Additive Noise and Short Memory Dynamics

An advantage of LW-type estimators is that, since they use the QMLF in the frequency domain, they can easily be modified to accommodate more types of structures in the DGP, without the need to trim some of the low frequencies. We consider two extensions to account for additive noise and short-memory dynamics. These elements do not cause an asymptotic bias and, hence, the modifications are aimed solely at improving the finite sample performance. Consider first the case where both LFC and additive noise are to be accounted for. To be precise, instead of (2), the DGP is now $z_t = c + y_t + u_t + w_t$, where, following Assumption (H2) in Hurvich et, al (2005), the additive noise w_t is a zero mean white noise with variance σ_w^2, such that for each $s \neq t$, $E[w_s \varepsilon_t] = 0$ and for each t, $E[w_t \varepsilon_t] = \rho_w \sigma_w$, where ε_t is as defined in A3 and ρ_w is the correlation between w_t and ε_t, assumed to be constant. Also, w_t is independent of the LFC u_t. Following Hurvich et al. (2005), we add a constant term into the spectral density function, so that the modified pseudo spectral density function is:

$$\begin{aligned} f_k &\triangleq f_z(\lambda_k) = G_0 \lambda_k^{-2d} + G_w + G_u(\lambda_k^{-2}/T) = G_0(\lambda_k^{-2d} + (G_w/G_0) + (G_u/G_0)(\lambda_k^{-2}/T)) \\ &= G_0(\lambda_k^{-2d} + \theta_w + (\theta_u/T)\lambda_k^{-2}) = G_0 g_k \end{aligned} \tag{3}$$

where, with a slight abuse of notation relabeling $\theta_u = G_u/G_0$, $g_k = (\lambda_k^{-2d} + \theta_w + (\theta_u/T)\lambda_k^{-2})$ and the (approximate) frequency domain QMLF is $\varphi(G, d, \theta) = m^{-1}\sum_{k=1}^{m} \varphi_k(G, d, \theta)$ with $\theta = (\theta_w, \theta_u)'$. Concentrating G out of the QMLF, the estimate of G is $\hat{G} = m^{-1}\sum_{k=1}^{m}(I_k/g_k)$ and the local Whittle QMLE estimator under noise perturbations and low frequency contaminations, denoted as the LWPLFC estimator, is $(\hat{d}_m, \hat{\theta}_m) = \arg\min_{(d,\theta)} J_m(d, \theta)$, where

$$J_m(d, \theta) = \log\left(\frac{1}{m}\sum_{k=1}^{m} \frac{I_k}{g_k}\right) + \frac{1}{m}\sum_{k=1}^{m} \log(g_k)$$

For reasons discussed by Hurvich, et. al. (2005), the LWPLFC approach is expected to work when d_0 is not too close to zero. When $d_0 = 0$, the process is short-memory. We then have a combination of two additive short-memory processes which cannot be identified separately.

For the case of short memory dynamics plus LFC, we could follow the approach of Andrews and Sun (2004) who add a polynomial structure into G_0, i.e., replace G_0 in (3) by $G_0 \exp(-p_r(\lambda_j, \theta))$ where $p_r(\lambda_j, \theta) = \Sigma_{s=1}^r \theta_s \lambda_j^{2s}$ and $\theta = (\theta_1, ..., \theta_r)$. However, unreported simulations with $r = 1$, showed that doing so did not offer any gain in performance over our LWLFC estimator with a smaller value of the bandwidth (see the simulations in Section 6). This feature can be explained as follows. From simulations to be reported in the next section, under strong short memory dynamics and RLS, the LWLFC estimator constructed with a large bandwidth has substantial bias but very small variance, so that the overall MSE is almost entirely due to the bias. When a polynomial component is added, the upward bias is reduced but the variance is increased considerably so that the overall MSE is almost the same or larger than that of the LWLFC estimator. With no RLS, the increased variance is smaller so that the MSE is indeed reduced as reported by Andrews and Sun (2004). At the root of the issue is the fact that both RLS and short memory dynamics cause upward biases in the estimate of the memory parameter. Hence, there is a confounding effect so that the QMLF is flat with respect to the correction factors for short memory dynamics and LFC. In unreported simulations with both RLS and short memory dynamics, it was often found that either the coefficient to correct for short memory dynamics or the coefficient to account for LFC was very close to zero, despite having the true value of both coefficients greater than zero. As will be reported in the simulations, the best way to account for short memory dynamics and RLS is to use the LWLFC estimator with a small bandwidth.

When both additive noise and short-memory dynamics are to be accounted for, three approaches are possible. One is to use the LWLFC estimator with a small bandwidth, another is to use the LWPLFC with a large bandwidth, or we could follow the approach of Frederiksen et al. (2012) who add polynomials and a constant as additive terms in the QMLF. One drawback of the latter approach is that the increase in the number of parameters can induce an important increase in variance resulting in increased mean-squared error.

1.6 Finite Sample Properties

The Data Generating Process (DGP) used for the simulations is $z_t = y_t + u_t + w_t$, where y_t is an $ARFIMA(1, d, 0)$ process given by $(1 - \alpha L)(1 - L)^d y_t = e_t$ with $e_t \sim i.i.d.\ N(0, 1)$, u_t is a RLS process as described in Definition 1 with $\sigma_\eta^2 = 1$, and $w_t \sim i.i.d.\ N(0, \sigma_w^2)$ is the additive noise component. The values used are: $d = 0, 0.2, 0.45$; $\alpha = 0.0, 0.3, 0.6$ and $p = 0, 5, 10, 20$. The sample sizes are T = 256, 512, 1024, 2048 and 4096 in order to use of the fast Fourier transform algorithm with the whole data set. The estimate $\hat{d}_m$ is allowed to take values in the set $[-0.99, 0.99]$ when evaluating the maximizers of the objective function. The value of the bandwidth is set to $m = T^\beta$ for $\beta = 0.6, 0.7, 0.8$, the choice being dictated by the fact that β must be larger than 5/9. Throughout, 500 replications are used. These specifications were also used by McCloskey and Perron (2013) so that we can make direct comparisons of the relative performance of our estimators with theirs (the sample sizes they used are 1000 and 2000 but the minor differences in T should not be of concern given the rather large differences in performance). The trimmed LP estimator of McCloskey and Perron (2013) depends on a lower trimming and upper bandwidth, while ours depend on a bandwidth. We evaluate bias and Root Mean Squared Errors (RMSE). When making comparisons, we do so using the values of the bandwidth (and trimming for the LP estimator) that gives the best RMSE for each of the statistics. We focus on random level shifts as the contaminating component as this is arguably the most relevant in practice. The results are presented in Tables 1-3 for the cases with only RLS and RLS plus short-memory dynamics, for which we focus on the LWLFC estimator. Table 4 presents the results for the case of RLS plus additive noise, while Table 5 presents results when all three types of contaminations are present, in which cases we consider both the LWLFC and LWPLFC estimators. We do not make a direct comparison with the trimmed LW estimator of Iacone (2010). McCloskey and Perron (2013) performed a comparison between the trimmed LP and LW estimators. They concluded that the trimmed LP has generally smaller bias and the trimmed LW generally lower variance and concluded that the overall performance in the presence of RLS was comparable.

1.6.1 The case with only RLS

The results for the case with only RLS are presented in the first panels of Tables 1-3 corresponding to the case $\alpha = 0$. Note first that the best results in terms of RMSE are obtained with a large bandwidth using $\beta = 0.8$, though biases are slightly smaller with a smaller bandwidth. Second, the results show that our estimator performs better than McCloskey and Perron's (2013) trimmed LP estimator. When $d_0 = 0$, there is a 30-60% reduction in RMSE, when $d_0 = 0.2$ the reduction is in the range 30-40% while when $d_0 = 0.45$ it is in the range 5-20%. Hence, overall, the LWLFC estimator with a large bandwidth $\beta = 0.8$, shows smaller bias and RMSE than alternative estimators. When the process is uncontaminated ($p = 0$), the bias and RMSE of our estimator is small and close to that of the original LW estimator, so that very little efficiency loss is incurred when no contamination is present.

1.6.2 The case with RLS and short-run dynamics

We now consider the case with both RLS and short-run dynamics (presented in Tables 1-3 for non-zero values of α). In this case the best results for the LWLFC estimator are obtained with a small bandwidth, using $\beta = 0.6$, and more so as the magnitude of α increases. Compared to the trimmed LP estimator, the reduction in RMSE is very substantial especially for larger values of α. For example, with no RLS the reduction is around 65% when $d = 0$ and $\alpha = 0.6$, while it is around 40% when $d = 0.45$ and $\alpha = 0.6$. The LWLFC is able to reduce bias and variance when both RLS and short-run dynamics are present, even though it is designed to account only for LFC contamination. As discussed in Section 5, the approach of Andrews and Sun (2004) which adds a polynomial structure into G_0 does not offer additional improvement. As stated in the above discussion, the results show that the LWLFC estimator has indeed very small variance when both RLS and short-run dynamics are present.

1.6.3 The case with RLS and additive noise

The results for the case with RLS and additive noise are presented in Table 4 for the LWLFC (which accounts only for LFC) and LWPLFC estimators (which accounts for both). The variance of the noise is set to a large value $\sigma_w^2 = 4$. The results show that the LWPLFC estimator has very small biases irrespective of the choice of the bandwidth. The biases are indeed orders of magnitude smaller than those of the trimmed LP estimator which is severely affected by noise. The superiority of our estimator also holds when judged by the relative RMSE. According to the RMSE, the estimator performs best with a high bandwidth ($\beta = 0.8$). The LWLFC estimator shows higher bias (though still much smaller than that of the trimmed LP) but its variance is smaller. In three out of the four cases analyzed (the exception being $d = 0.2$ and $p = 20$) the reduction in variance is not big enough so that the LWPLFC estimator has overall a smaller RMSE when using a large bandwidth. As expected, the performance of the LWPLFC improves as d increases, for reasons explained in Section 5.

1.6.4 The case with all three types of contaminations

Table 5 presents results with all three types of contaminations. We consider strong short-memory dynamics ($\alpha = 0.6$) and a medium value for the average number of level shifts ($p = 10$). For the additive noise, we use $\sigma_w^2 = 1, 4$, and we set $d = 0.2, 0.45$. The results show that both the LWLFC and LWPLFC perform well. In general, the LWPLFC has better performance when a large bandwidth is used, while the LWLFC is better with a small bandwidth. For a large value of d_0 (0.45), the LWPLFC performs slightly better than the LWLFC under the optimal bandwidth applicable to each. When d_0 is small ($d_0 = 0.2$) the LWLFC has slightly better performance. This accords with Hurvich, et. al. (2005) who showed that the asymptotic variance of the LW estimator increases as d_0 decreases. Overall, the results show an advantage of using the LWPLFC with a large bandwidth. From unreported simulations, the performance of the LWLFC and LWPLFC deteriorates as α approaches 1 or with a moving-average parameter close to -1, with or without noise. This is a problem common to most, if not all, versions of LW or LP estimators, trimmed or not.

1.6.5 Overall summary and recommendations

The results showed that our estimators have good finite sample properties and offer improved methods of inference compared to what is available in the literature. As with all existing semi-parametric estimators of this type, the results can be sensitive to the choice of the bandwidth. In our case, a large bandwidth (e.g., $\beta = 0.8$) is preferable in most cases. One exception is when there is a strongly positively correlated short-memory component, in which case a smaller bandwidth ($\beta = 0.6$) is desirable. As of yet, there is no fully developed method to choose the bandwidth. But some approaches are possible for the practitioner to assess what is the best bandwidth to use. One is to estimate a preliminary parametric LFC model with an AR component for the noise. Upon obtaining a large estimate of the AR coefficient a smaller bandwidth is dictated and vice versa if the coefficient is small. While somewhat ad hoc, it should provide a useful guide.

1.7 Conclusions

We proposed a local-Whittle estimator of the memory parameter of a long memory time series process which has good properties under an almost complete collection of contamination processes that have been discussed in the literature. The estimator has many advantages: no assumption of Gaussianity is required unlike the trimmed log-periodogram estimator; there is no trimming involved so that all information from the low frequency components are retained; when there is no LFC, its performance is comparable to that of the standard LW estimator so that no asymptotic efficiency loss is incurred, with the loss of efficiency in finite sample being small as revealed by the simulations; with a proper choice of the bandwidth, the extended estimator has good finite sample properties with short-run dynamics and/or additive noise; it is semi-parametric so that there is no need for a full specification of the underlying short-memory structure, though it can also be extended to cover a fully specified parametric structure for the long-memory component such as an ARFIMA process.

It does, nevertheless, have some drawbacks. First, the performance of the estimator is sensitive to the choice of the bandwidth. An adaptive, data-dependant method to select the bandwidth is an important avenue for future research. Note, however, that all current semi-parametric estimators exhibit sensitivity to the bandwidth choice. Also, when the estimator is extended to account for noise, as in Hurvich et. al (2005), the RMSE is proportional to $(1/d_0)$ so that when the true parameter d_0 is close to zero the reduction in bias is offset by an increase in variance and a possible increase in the overall RMSE.

Appendix

We first introduce three lemmas which show that to some extent the pseudo spectral density function controls the periodogram of the process well, in the sense that the ratio $|I_k/f_k|$ is bounded and the average of $(I_k/f_k - 1)$ is $o_p(1)$.

Lemma A.1 *Let* $A_k = (2\pi T)^{-1/2}\sum_{t=1}^{T} z_t \cos(\lambda_k t)$, $B_k = (2\pi T)^{-1/2}\sum_{t=1}^{T} z_t \sin(\lambda_k t)$, *so that* $I_k = (A_k)^2 + (B_k)^2$, *and define the vector* $\gamma = ((f_k)^{-1/2}A_k, (f_k)^{-1/2}B_k, (f_j)^{-1/2}A_j, (f_j)^{-1/2}B_j)'$. *Let* $\kappa(X_1, X_2, X_3, X_4)$ *denote the joint cumulant of the random variables* X_1, X_2, X_3, X_4 *with* n_1, n_2, n_3, n_4 *nonnegative integers that sum to* n. *Then under Assumptions A1-A5, for* $\theta_0 > 0$ *and letting* $M_0 = \theta_0/(2\pi)^{2-2d_0}$, *for any sequences of positive integers* k *and* j *such that* $k > j$ *and* $k/T \to 0$, *the following result holds for* $n > 2$*:*

$$\begin{aligned}&\kappa(\gamma_1^{n_1}, \gamma_2^{n_2}, \gamma_3^{n_3}, \gamma_4^{n_4})\\ =\;& O\Big(\Big(\frac{T^{n/2-nd}}{k^{(n_1+n_3)(1-d_0)}j^{(n_2+n_4)(1-d_0)}}\Big)/(1+M_0\frac{T^{1-2d_0}}{k^{2-2d_0}})^{(n_1+n_3)}(1+M_0\frac{T^{1-2d_0}}{j^{2-2d_0}})^{(n_2+n_4)}\Big)^{1/2}\end{aligned}$$

which is $O(1)$ *if* $j \le T^{(1-2d_0)/(2-2d_0)}$ *and* $o(1)$ *if* $j > T^{(1-2d_0)/(2-2d_0)}$. *Similarly, for* $n > 2$, *the* n*-th cumulant of* $\tilde{\gamma} = (A_k/(f_k)^{1/2}, B_k/(f_k)^{1/2})'$ *are* $O((T^{n/2-nd_0}/k^{n(1-d_0)})/(1 + M_0(T^{1-2d_0}/k^{2-2d_0})^{n/2})$. *When* $\theta_0 = 0$, $M_0 = 0$ *and the result reduces to*

$$\kappa(\gamma_1^{n_1}, \gamma_2^{n_2}, \gamma_3^{n_3}, \gamma_4^{n_4}) = O(T^{n/(2-nd_0)}/[k^{(n_1+n_3)(1-d_0)}j^{(n_2+n_4)(1-d_0)}])$$

Proof. This lemma is a direct consequence of Lemma A.3 in McCloskey and Perron (2013), henceforth MP, and the definition of the pseudo spectral density function f_k. The difference in the results is simply due to the fact that we use $f_k = \lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}$, while MP use $f_k = \lambda_k^{-2d_0}$. Hence, a different expression is obtained when $\theta_0 > 0$. ■

Lemma A.2 *Under A1-A5, with* $I_k = \omega_k\omega_k^*$ *and* $M_0 = \theta_0/(2\pi)^{2-2d_0}$, *for* $1 \le j < k \le m$*:*

$$\begin{aligned}(i) E(I_k/f_k) &= 1 + [O(k^{-1}\log k) + O(k/T)^{1+2d_0}]\\ &\quad /[1 + M_0(T^{1-2d_0}/k^{2-2d_0})]\end{aligned}$$

$$\begin{aligned}(ii) E((\omega_k)^2/f_k) &= O(k^{-1}\log k) + O(T^{1-2d_0}/k^{2-2d_0})\\ &\quad /[1 + M_0(T^{1-2d_0}/k^{2-2d_0})]\end{aligned}$$

$$(iii)E(\frac{\omega_k\omega_j^*}{\sqrt{f_kf_j}}) = O(k^{-1}\log j)+O(T^{1-2d_0}/(k^{1-d_0}j^{1-d_0})) \\ /\sqrt{(1+M_0(T^{1-2d_0}/k^{2-2d_0}))(1+M_0(T^{1-2d_0}/j^{2-2d_0}))}$$

$$(iv)E(\frac{\omega_k\omega_j}{\sqrt{f_kf_j}}) = O(k^{-1}\log j)+O(T^{1-2d_0}/(k^{1-d_0}j^{1-d_0})) \\ /\sqrt{(1+M_0(T^{1-2d_0}/k^{2-2d_0}))(1+M_0(T^{1-2d_0}/j^{2-2d_0}))}$$

Proof. For part (i), we have $E(I_{u,k}/(T^{-1}\lambda_k^{-2}))=O_p(1)$. Hence, from Theorem 1 in MP,

$$\begin{aligned}
E(\frac{I_k}{f_k}) &= E(\frac{I_k}{f_{y,k}}\frac{f_{y,k}}{f_k}) \\
&= \frac{f_{y,k}}{f_k}E(\frac{I_k}{f_{y,k}}) = \frac{f_{y,k}}{f_k}E(\frac{I_{y,k}}{f_{y,k}}+\frac{I_{u,k}}{f_{y,k}}+\frac{2I_{yu,k}}{f_{y,k}}) \\
&= \frac{\lambda_k^{-2d_0}}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}(1+O(\frac{\log k}{k}+(\frac{k}{T})^2) \\
&\quad +M_0\frac{T^{1-2d_0}}{k^{2-2d_0}}+O(\frac{k^3}{T^2}\frac{T^{1-2d_0}}{k^{2-2d_0}})) \\
&= \frac{1}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}(1+M_0\frac{T^{1-2d_0}}{k^{2-2d_0}} \\
&\quad +O(\frac{\log k}{k}+(\frac{k}{T})^2)+O(\frac{k}{T})^{1+2d_0}) \\
&= 1+\frac{O(k^{-1}\log k)+O(k/T)^{1+2d_0}}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}
\end{aligned}$$

For part (ii),

$$\begin{aligned}
E(\frac{(\omega_k)^2}{f_k}) &= E(\frac{(\omega_k)^2}{f_{y,k}}\frac{f_{y,k}}{f_k}) = \frac{f_{y,k}}{f_k}E(\frac{(\omega_k)^2}{f_{y,k}}) \\
&= \frac{1}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}O(\frac{\log k}{k}+\frac{T^{1-2d_0}}{k^{2-2d_0}}) \\
&= O(\frac{\log k}{k})+\frac{O(T^{1-2d_0}/k^{2-2d_0})}{1+M_0(T^{1-2d_0}/k^{2-2d_0})}
\end{aligned}$$

For part (iii),

$$\begin{aligned}
E(\frac{\omega_k\omega_j^*}{\sqrt{f_k f_j}}) &= E(\frac{\omega_k\omega_j^*}{\sqrt{f_{y,k}f_{y,j}}}\frac{\sqrt{f_{y,k}f_{y,j}}}{\sqrt{f_k f_j}}) = \frac{\sqrt{f_{y,k}f_{y,j}}}{\sqrt{f_k f_j}}E(\frac{\omega_k\omega_j^*}{\sqrt{f_{y,k}f_{y,j}}}) \\
&= (1+M_0(T^{1-2d_0}/k^{2-2d_0})) \\
&\quad (1+M_0(T^{1-2d_0}/k^{2-2d_0}))^{-1/2}O(\frac{\log j}{k}+\frac{T^{1-2d_0}}{k^{1-d_0}j^{1-d_0}}) \\
&= \frac{O(T^{1-2d_0}/k^{1-d_0}j^{1-d_0})}{[(1+M_0(T^{1-2d_0}/k^{2-2d_0}))(1+M_0(T^{1-2d_0}/j^{2-2d_0}))]^{1/2}} \\
&\quad +O(\frac{\log j}{k})
\end{aligned}$$

and the proof is entirely analogous for part (iv). ■

Lemma A.3 *Under A1-A5: if a) $\theta = \theta_m$ is bounded away from zero or b) there is no LFC in data, then: 1) $|I_k/f_k|$ is bounded, and 2) $m^{-1}\sum_{k=1}^{m}(I_k/f_k - 1) = o_p(1)$.*

Proof. First,

$$\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{f_k}-1) = \frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}) + \frac{1}{m}\sum_{k=1}^{m}(\frac{I_{y,k}}{f_{y,k}}-1)$$

For the first term, we have:

$$\begin{aligned}
&\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}) \\
&= \frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}) + \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}})
\end{aligned}$$

whose first component is such that,

$$\begin{aligned}
&\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}}) \\
&= \frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_{y,k}}{f_{z,k}}-\frac{I_{y,k}}{f_{y,k}}+\frac{I_{u,k}}{f_{z,k}}+2\frac{I_{yu,k}}{f_{z,k}}) \\
&= \frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_{y,k}}{f_{y,k}}(-\frac{f_{u,k}}{f_{z,k}})+\frac{I_{u,k}}{f_{z,k}}+2\frac{I_{yu,k}}{f_{z,k}}) \\
&= \frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_{u,k}-f_{u,k}}{f_{z,k}}-(\frac{I_{y,k}}{f_{y,k}}-1)(\frac{f_{u,k}}{f_{z,k}})+2\frac{I_{yu,k}}{f_{z,k}})
\end{aligned}$$

Note that

$$E|\frac{I_{u,k}-f_{u,k}}{f_{z,k}}| = E|(\frac{I_{u,k}}{f_{u,k}}-1)/(\frac{f_{z,k}}{f_{u,k}})| = \frac{f_{u,k}}{f_{z,k}}E|\frac{I_{u,k}}{f_{u,k}}-1|$$

From MP (Lemma A.3) with $n_1 = n_2 = n_3 = n_4 = 1$ and $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = I_{u,k}/f_{y,k}$:

$$E|\frac{I_{u,k}}{f_{u,k}}-1| \leq E|\frac{I_{u,k}}{f_{u,k}}|+1 \leq [E(|\frac{I_{u,k}}{f_{u,k}}|^2)]^{1/2}+1 \leq C_1$$

So

$$\begin{aligned} & E|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_{u,k}-f_{u,k}}{f_{z,k}})| \\ \leq \quad & \frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}|\frac{f_{u,k}}{f_{z,k}}|E|\frac{I_{u,k}}{f_{u,k}}-1| \leq \frac{\sqrt{T}}{m}C_1 \to 0 \end{aligned}$$

if $\sqrt{T}/m \to 0$. We also have $E|m^{-1}\sum_{k=1}^{\sqrt{T}-1}(I_{y,k}/f_{y,k}-1)(f_{u,k}/f_{z,k})| \to 0$, since $|f_{u,k}/f_{z,k}| < 1$. From MP, Perron and Qu (2010) and Qu (2011): $I_{yu}(\lambda_k) = O_p(T^{-1/2}\lambda_k^{-(1+d_0)})$ and $f_k = f_{z,k} = f_{y,k} + f_{u,k} = G\lambda_k^{-2d_0} + G_u T^{-1}\lambda_k^{-2}$. Hence,

$$\begin{aligned} |\frac{I_{yu,k}}{f_{z,k}}| & \sim \frac{O_p(T^{-1/2}\lambda_k^{-(1+d_0)})}{O_p(\lambda_k^{-2d_0})+O_p(T^{-1}\lambda_k^{-2})} \\ & \sim \frac{1}{O_p(T^{1/2}\lambda_k^{1-d_0})+O_p(T^{-1/2}\lambda_k^{d_0-1})} < O_p(1) \end{aligned}$$

and

$$\begin{aligned} E|\frac{2}{m}\sum_{k=1}^{\sqrt{T}-1}\frac{I_{yu,k}}{f_{z,k}}| & \leq \frac{2}{m}\sum_{k=1}^{\sqrt{T}-1}E|\frac{I_{yu,k}}{f_{z,k}}| \\ & < \frac{2}{m}\sqrt{T}O_p(1) = O_p(\frac{\sqrt{T}}{m}) \to 0 \end{aligned}$$

if $\sqrt{T}/m \to 0$. Hence,

$$\begin{aligned}
& E|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}})| \\
= \; & E|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_{u,k}-f_{u,k}}{f_{z,k}}-(\frac{I_{y,k}}{f_{y,k}}-1)(\frac{f_{u,k}}{f_{z,k}})+2\frac{I_{yu,k}}{f_{z,k}})| \\
\leq \; & E[|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_{u,k}-f_{u,k}}{f_{z,k}})|]+E[|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_{y,k}}{f_{y,k}}-1)(\frac{f_{u,k}}{f_{z,k}})|] \\
& +E[|\frac{2}{m}\sum_{k=1}^{\sqrt{T}-1}\frac{I_{yu,k}}{f_{z,k}}|] \\
\to \; & 0
\end{aligned}$$

if $\sqrt{T}/m \to 0$. It is easy to show that $E|m^{-1}\sum_{k=\sqrt{T}}^{m}(I_k/f_k - I_{y,k}/f_{y,k})| \to 0$, and the fact that $E|m^{-1}\sum_{k=1}^{m}(I_{y,k}/f_{y,k}-1)| \to 0$ follows from Hurvich et. al. (2005). So

$$\begin{aligned}
& E|\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{f_k}-1)| \\
\leq \; & E[|\frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}})|]+E[|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}})|] \\
& +E[|\frac{1}{m}\sum_{k=1}^{m}(\frac{I_{y,k}}{f_{y,k}}-1)|] \\
\to \; & 0
\end{aligned}$$

if $\sqrt{T}/m \to 0$. Note that during the proof we also showed that $|I_k/f_k| \leq |I_k/f_k - I_{y,k}/f_{y,k}| + |I_{y,k}/f_{y,k}-1|+1$ is bounded. ■

Proof. of Lemma 1: Let $M_m = \hat{\theta}_m/(2\pi)^{2-2\hat{d}_m}$ and $M_0 = \theta_0/(2\pi)^{2-2d_0}$. We analyze the partial derivative of the objective function with respect to θ:

$$\begin{aligned}
& \frac{\partial}{\partial\theta} J_m(\hat{d}_m,\hat{\theta}_m) = \frac{1}{mT}[\sum_{k=1}^{m}\frac{1}{g_k(\hat{d}_m,\hat{\theta}_m)}\lambda_k^{-2} \\
& -(\frac{1}{m}\sum_{k=1}^{m}\frac{I_k}{g_k(\hat{d}_m,\hat{\theta}_m)})^{-1}\sum_{k=1}^{m}\frac{I_k}{(g_k(\hat{d}_m,\hat{\theta}_m))^2}\lambda_k^{-2}] \qquad \text{(A.1)} \\
= & \frac{1}{mT}[\sum_{k=1}^{m}(1-\frac{I_k}{G_0 g_k(\hat{d}_m,\hat{\theta}_m)}\frac{G_0}{m^{-1}\sum_{j=1}^{m}(I_j/g_j(\hat{d}_m,\hat{\theta}_m))})\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}] \\
= & \frac{1}{mT}[\sum_{k=1}^{m}(1-\frac{I_k}{f_k}\frac{G_0}{m^{-1}\sum_{j=1}^{m}(I_j/g_j(\hat{d}_m,\hat{\theta}_m))}\frac{g_k(d_0,\theta_0)}{g_k(\hat{d}_m,\hat{\theta}_m)}) \\
& \frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}] \qquad \text{(A.2)} \\
= & \frac{1}{mT}\{\sum_{k=1}^{m}(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}})[1-(m\frac{I_k}{f_k}(\frac{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}) \qquad \text{(A.3)} \\
& \backslash\sum_{j=1}^{m}\frac{I_j}{f_j}(\frac{\lambda_j^{-2d_0}+(\theta_0/T)\lambda_j^{-2}}{\lambda_j^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_j^{-2}}))]\}
\end{aligned}$$

Using summation by parts, (A.3) becomes:

$$\begin{aligned}
& \{\sum_{k=1}^{m}(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}})[1-(m\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d}/k^{2-2d})})) \\
& \backslash(\sum_{j=1}^{m}\frac{I_j}{f_j}\lambda_j^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})))]\} \\
= & (\frac{\lambda_m^{-2}}{\lambda_m^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_m^{-2}})\sum_{k=1}^{m}[1-(m\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})) \\
& \backslash(\sum_{j=1}^{m}\frac{I_j}{f_j}\lambda_j^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})))] \\
& +\sum_{j=1}^{m-1}[(\frac{\lambda_j^{-2}}{\lambda_j^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_j^{-2}})-(\frac{\lambda_{j+1}^{-2}}{\lambda_{j+1}^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_{j+1}^{-2}})] \\
& \{\sum_{k=1}^{j}(1-(m\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})) \\
& \backslash(\sum_{k=1}^{m}\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})))]\}
\end{aligned}$$

$$
\begin{aligned}
= \quad & \left(\frac{\lambda_m^{-2}}{\lambda_m^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_m^{-2}}\right)[m \\
& -m\sum_{k=1}^{m} \frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right) \\
& \backslash\left(\sum_{j=1}^{m} \frac{I_j}{f_j}\lambda_j^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)]
\end{aligned}
$$

$$
\begin{aligned}
& + \sum_{j=1}^{m-1}[\left(\frac{\lambda_j^{-2}}{\lambda_j^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_j^{-2}}\right) - \left(\frac{\lambda_{j+1}^{-2}}{\lambda_{j+1}^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_{j+1}^{-2}}\right)] \\
& [j - m\sum_{k=1}^{j}\left(\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right) \\
& \backslash\left(\sum_{k=1}^{m} \frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}\left(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\right)\right)]
\end{aligned}
$$

Now, suppose $\hat{\theta}_m \to 0$ with $(\partial/\partial\theta)J_m(\hat{d}_m, \hat{\theta}_m) = 0$. We define $h_k \equiv [1+M_0(T^{1-2d_0}/k^{2-2d_0})]/[1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})]$. We consider two cases. In the first, suppose $M_m \to 0$ at a slow rate such that for some small k, we still have $M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) \to \infty$. Let

$$
\tau_m = \inf_k \{M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) \to 0\}
$$

, then

$$
h_k \sim \begin{cases} \frac{M_0}{M_m}(\frac{T}{k})^{2(\hat{d}_m-d_0)} + \frac{1}{M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} \text{ when } k \leq \tau_m \\ 1+M_0(T^{1-2d_0}/k^{2-2d_0}) \text{ when } k > \tau_m \end{cases}
$$

Note that we must have either $(M_0/M_m)(T/k)^{2(\hat{d}_m-d_0)}$ or $M_0(T^{1-2d_0}/k^{2-2d_0})$ go to infinity for some small k. Also,

$$
\begin{aligned}
M_0(T^{1-2d_0}/k^{2-2d_0}) &= \frac{M_0}{M_m}\left(\frac{T}{k}\right)^{2(\hat{d}_m-d_0)}\left(M_m\left(\frac{T}{k}\right)^{-2(\hat{d}_m-d_0)}\frac{T^{1-2d_0}}{k^{2-2d_0}}\right) \\
&= \frac{M_0}{M_m}\left(\frac{T}{k}\right)^{2(\hat{d}_m-d_0)}\left(M_m\frac{T^{1-2\hat{d}_m}}{k^{2-2\hat{d}_m}}\right) \\
&= o_p\left(\frac{M_0}{M_m}\left(\frac{T}{k}\right)^{2(\hat{d}_m-d_0)}\right)
\end{aligned}
$$

when $k > \tau_m$. Hence,

$$\lambda_k^{2\hat{d}-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}) \sim \begin{cases} (M_0/M_m) \text{ when } k \leq \tau_m \\ o_p(M_0/M_m) \text{ when } k > \tau_m \end{cases}$$

Let

$$a_j = \frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})$$

then we know that $a_j = O_p(M_0/M_m)$ when $k \leq \tau_m$ and $a_j = o_p(M_0/M_m)$ when $k > \tau_m$. So, $\{a_j\}$ is a positive sequence whose first few terms have higher order than the rest. So we have

$$\begin{array}{c}(j/m) - \sum_{k=1}^{j}(\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})) \\ \backslash(\sum_{k=1}^{m}\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})) \leq C_j < 0\end{array} \tag{A.4}$$

where C_j is some constant. Under the second case, $M_m \to 0$ fast enough so that, for any k, $M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) \leq O_p(1)$. For this case, $h_k \sim 1 + M_0(T^{1-2d_0}/k^{2-2d_0})$ and

$$\begin{aligned} & \lambda_k^{2\hat{d}_m-2d_0}(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}) \\ \sim \; & \lambda_k^{2\hat{d}_m-2d_0}(1+M_0(T^{1-2d_0}/k^{2-2d_0}) \\ \sim \; & (T/k)^{2d_0-2\hat{d}_m} + M_0(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) \end{aligned}$$

If $d_0 \geq \hat{d}_m$, the last expression is decreasing in k for all $k = 1, ..., m$; if $d_0 < \hat{d}_m$, the first is increasing in k, but always smaller than 1, and the second is decreasing in k and goes to infinity when k is small. Hence, (A.4) still holds. Since for T large enough,

$$\lambda_j^{-2}(\lambda_j^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_j^{-2})^{-1} - \lambda_{j+1}^{-2}(\lambda_{j+1}^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_{j+1}^{-2})^{-1} \geq D_j > 0$$

where D_j is some constant, we have shown that $(\partial/\partial\theta)J_m(\hat{d}_m, \hat{\theta}_m) < 0$ if $\hat{\theta}_m \to 0$, which is a contradiction. So $\hat{\theta}_m$ has to be bounded from zero when $\theta_0 > 0$. ■

Proof. of Theorem 1: First, we consider the case when LFC indeed exists in the true DGP. The proof for the case with no LFC will follow with trivial modifications. Note that if LFC components are present, the probability limit of the estimate $\hat{\theta}_m$ is bounded above zero, by Lemma 1. This implies that with probability arbitrarily close to one, $\hat{\theta}_m$ will be in the set $(0, \infty)$ and, without loss of generality, we can consider analyzing the limit of $\hat{d}_m$ for any sequence or values of θ_m in the set $(0, \infty)$. Accordingly, we want to show that, with probability arbitrarily close to one for large T and m, if $\{\theta_m\}$ is a sequence bounded above from zero and if $\{\hat{d}_m\}$ minimizes $J_m(d, \theta_m)$ given $\{\theta_m\}$, then for $\hat{d}_m$ such that $|\hat{d}_m - d_0| \geq \delta$ for any $\delta > 0$, we have $J_m(\hat{d}_m, \theta_m) - J_m(d_0, \theta_m) > 0$, which delivers a contradiction showing that in the limit the minimizer of $J_m(d, \theta_m)$ must converge to d_0. Let $G(d, \theta_m) = m^{-1} \sum_{k=1}^m I_k / g_k$, where $g_k = (\lambda_k^{-2d} + (\theta_m/T)\lambda_k^{-2})$. We first have:

$$
\begin{aligned}
& J_m(\hat{d}_m, \theta_m) - J_m(d_0, \theta_m) \\
= \; & [\log G(\hat{d}_m, \theta_m) + \frac{1}{m} \sum_{k=1}^{m} \log(\lambda_k^{-2\hat{d}_m} (1 + \frac{\theta_m}{T} \lambda_k^{-2+2\hat{d}_m})) \\
& -[\log G(d_0, \theta_m) + \frac{1}{m} \sum_{k=1}^{m} \log(\lambda_k^{-2\hat{d}_m} (1 + \frac{\theta_m}{T} \lambda_k^{-2+2d_0}))]
\end{aligned}
$$

$$
\begin{aligned}
= \; & \log G(\hat{d}_m, \theta_m) - \log G(d_0, \theta_m) \\
& + \frac{1}{m} \sum_{k=1}^{m} \log(\lambda_k^{-2(\hat{d}_m - d_0)} (\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}})) \\
= \; & \log \frac{G(\hat{d}_m, \theta_m)}{G_0 (m^{-1} \sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})} - \log \frac{G(d_0, \theta_m)}{G_0} + \log(\frac{1}{m} \sum_{k=1}^{m} \lambda_k^{2(\hat{d}_m - d_0)}) \\
& - \frac{2(\hat{d}_m - d_0)}{m} \sum_{k=1}^{m} \lambda_k + \frac{1}{m} \sum_{k=1}^{m} \log(\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}})
\end{aligned}
$$

$$
\begin{aligned}
= \; & \log \frac{G(\hat{d}_m, \theta_m)}{G_0 (m^{-1} \sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})} - \log \frac{G(d_0, \theta_m)}{G_0} \\
& + \log(\lambda_k^{2(\hat{d}_m - d_0)} (2(\hat{d}_m - d_0) + 1)) - \log(2(\hat{d}_m - d_0) + 1)) \\
& -2(\hat{d}_m - d_0)[\frac{1}{m} \sum_{k=1}^{m} (\log k - \log m)] \\
& + \frac{1}{m} \sum_{k=1}^{m} \log(\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}})
\end{aligned}
$$

$$= \log \frac{G(\hat{d}_m, \theta_m)}{G_0(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})} - \log \frac{G(d_0, \theta_m)}{G_0}$$

$$+ \log(\frac{2(\hat{d}_m - d_0) + 1}{m} \sum_{k=1}^m (\frac{k}{m})^{2(\hat{d}_m - d_0)}) \quad \text{(A.5)}$$

$$-2(\hat{d}_m - d_0)[\frac{1}{m}\sum_{k=1}^m \log k - (\log m - 1)] + \frac{1}{m}\sum_{k=1}^m \log(\frac{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1 + (\theta_m/T)\lambda_k^{-2+2d_0}}) \quad \text{(A.6)}$$

$$- \log(1 + 2(\hat{d}_m - d_0)) + 2(\hat{d}_m - d_0)$$

Note that for the last term of (A.6), we have $-\log(1+2(\hat{d}_m - d_0)) + 2(\hat{d}_m - d_0) \geq (1/6)(\hat{d}_m - d_0)^2 \geq (1/6)\delta^2$. Hence, if we can show that the other five terms are $o_p(1)$, we can derive a contradiction. The third and fourth are $o_p(1)$ from Robinson (1995a).

Proof that the first term of (A.6) is $o_p(1)$. To show that, it is equivalent to prove that $G(\hat{d}_m, \theta_m)/[G_0(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})] - 1 = o_p(1)$. To that effect,

$$\frac{G(\hat{d}_m, \theta_m)}{G_0(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})} - 1$$

$$= (G_0(m^{-1}\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}))^{-1} \{\frac{1}{m}\sum_{k=1}^m \frac{I_k}{\lambda_k^{-2\hat{d}_m}(1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m})} - \frac{1}{m} G_0 \sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)}\}$$

$$= (\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})^{-1} \{\sum_{k=1}^m [\frac{I_k}{G_0 \lambda_k^{-2d_0}(1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m})} \lambda_k^{2(\hat{d}_m - d_0)} - \lambda_k^{2(\hat{d}_m - d_0)}]\}$$

$$= (\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)})^{-1} \{\sum_{k=1}^m \lambda_k^{2(\hat{d}_m - d_0)} [\frac{I_k}{f_k} \frac{1 + (\theta_0/T)\lambda_k^{-2+2d_0}}{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1]\}$$

$$= (\frac{1}{m}\sum_{k=1}^m (\frac{k}{m})^{2(\hat{d}_m - d_0)})^{-1} \{\frac{1}{m}\sum_{k=1}^m (\frac{k}{m})^{2(\hat{d}_m - d_0)} (\frac{I_k}{f_k} \frac{1 + (\theta_0/T)\lambda_k^{-2+2d_0}}{1 + (\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1)\}$$

From Robinson (1995a), $m^{-1}\sum_{k=1}^{m}(k/m)^{2(\hat{d}_m-d_0)} = o_p(1)$ if $\hat{d}_m - d_0 \neq -1/2$. When $k \geq \sqrt{T}$,

$$\frac{I_k}{f_k}\frac{1+(\theta_m/T)\lambda_k^{-2+2d_0}}{1+(\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1 = (\frac{I_k}{f_k} - 1) + \frac{I_k}{f_k}O((\frac{k}{T})^{2d_0}).$$

From Lemma 14, $m^{-1}\sum_{k=1}^{m}(I_k/f_k - 1) = o_p(1)$, hence $m^{-1}\sum_{k=1}^{m} I_k/f_k = 1 + o_p(1)$. Also,

$$\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k}{f_k} = \frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{I_k}{f_k} - \frac{I_{y,k}}{f_{y,k}}) + \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_{y,k}}{f_{y,k}}$$

The first term is $o_p(1)$ from Lemma A.3 and the second is $o_p(m^{-1}\sqrt{T}) = o_p(1)$ from Hurvich et. al. (2005). Combining these results, we have $m^{-1}\sum_{k=\sqrt{T}}^{m} I_k/f_k = 1 + o_p(1)$. Now,

$$\begin{aligned}
& \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k}\frac{1+(\theta_0/T)\lambda_k^{-2+2d_0}}{1+(\theta_m/T)\lambda_k^{-2+2\hat{d}_m}} - 1) \\
= \; & \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k} - 1) + \\
& \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}O(\frac{T}{k^2}(\frac{k}{T})^{2d_0})\frac{I_k}{f_k} \qquad \text{(A.7)} \\
= \; & \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k} - \frac{I_{y,k}}{f_{y,k}}) + \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_{y,k}}{f_{y,k}} - 1) \qquad \text{(A.8)} \\
& + \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2\hat{d}_m}O(\frac{T}{k^2}(\frac{m}{T})^{2d_0})(\frac{I_k}{f_k} - \frac{I_{y,k}}{f_{y,k}}) \\
& + \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2\hat{d}_m}O(\frac{T}{k^2}(\frac{m}{T})^{2d_0})\frac{I_{y,k}}{f_{y,k}} \qquad \text{(A.9)}
\end{aligned}$$

We will show that all four terms of (A.8) are $o_p(1)$ by showing that the expectations of their absolute values are $o_p(1)$. For the first term, we have from Lemma A.3,

$$E(|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k} - \frac{I_{y,k}}{f_{y,k}})|) \leq \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}E|(\frac{I_k}{f_k} - \frac{I_{y,k}}{f_{y,k}})|$$

From Lemma A.1, we know that $E|(I_k/f_k - I_{y,k}/f_{y,k})| \sim C(k/T)^{d_0} \leq C(m/T)^{d_0}$, where C is some constant not depending on T and m. We also have

$$\begin{aligned}
& \frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)} \\
= & \frac{\sqrt{T}}{m}\frac{1}{\sqrt{T}}\sum_{k=1}^{\sqrt{T}}(\frac{k}{\sqrt{T}})^{2(\hat{d}_m-d_0)}(\frac{\sqrt{T}}{m})^{2(\hat{d}_m-d_0)} \\
= & (\frac{\sqrt{T}}{m})^{1+2(\hat{d}_m-d_0)}\frac{1}{\sqrt{T}}\sum_{k=1}^{\sqrt{T}}(\frac{k}{\sqrt{T}})^{2(\hat{d}_m-d_0)} \\
= & O_p(\frac{\sqrt{T}}{m}) \to 0
\end{aligned}$$

where the last equality is from Robinson (1995a), Equation (3.7). Hence,

$$m^{-1}\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)} \sim \frac{1}{1+2(\hat{d}_m-d_0)}$$

, which shows that the first term is $o_p(1)$. For the second term,

$$\begin{aligned}
& E|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_{y,k}}{f_{y,k}}-1)| \\
= & E|\frac{1}{m}\sum_{k=1}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_{y,k}}{f_{y,k}}-1)-\frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_{y,k}}{f_{y,k}}-1)| \\
\leq & E|\frac{1}{m}\sum_{k=1}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_{y,k}}{f_{y,k}}-1)| \\
& +(\frac{\sqrt{T}}{m})^{1+2(\hat{d}_m-d_0)}E|\frac{1}{\sqrt{T}}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_{y,k}}{f_{y,k}}-1)| \\
= & o_p(1)
\end{aligned}$$

For the third term,

$$
\begin{aligned}
& E(|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2\hat{d}_m}O(\frac{T}{k^2}(\frac{m}{T})^{2d_0})(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}})|) \\
= \; & E|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2\hat{d}_m}O(\frac{m}{T})^{2d_0}(\frac{I_k}{f_k}-\frac{I_{y,k}}{f_{y,k}})| \\
\leq \; & E(\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}\frac{T}{k^2}O(\frac{m}{T})^{2d_0}C(\frac{k}{m})^{2d_0}) \\
\leq \; & O_p(\frac{T}{m^2}) = o_p(1)
\end{aligned}
$$

For the fourth term,

$$
\begin{aligned}
& E(|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2\hat{d}_m}O(\frac{T}{k^2}(\frac{m}{T})^{2d_0})\frac{I_{y,k}}{f_{y,k}}|) \\
= \; & E|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2\hat{d}_m}\frac{T}{k^2}O(\frac{m}{T})^{2d_0}\frac{I_{y,k}}{f_{y,k}}| \\
\leq \; & \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2\hat{d}_m}\frac{T}{k^2}O(\frac{m}{T})^{2d_0}E|\frac{I_{y,k}}{f_{y,k}}| \\
\leq \; & O_p(\frac{T}{m^2}) = o_p(1)
\end{aligned}
$$

according to Equation (3.15) in Robinson (1995a). Hence,

$$
\frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k}\frac{1+(\theta_0/T)\lambda_k^{-2+2d_0}}{1+(\theta_m/T)\lambda_k^{-2+2d}}-1) = o_p(1)
$$

Proof that the second term of (A.6) is $o_p(1)$. Note that

$$
(\theta_m/T)\lambda_k^{-2+2\hat{d}_m} = \theta_m(2\pi)^{-2+2\hat{d}_m}(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})
$$

and let $M_0 = \theta_0(2\pi)^{-2+2d_0}$, $M_m = \theta_m(2\pi)^{-2+2\hat{d}_m}$ and $\tilde{M} = \inf_{m\geq 1}\{\theta_m\}(2\pi)^{-2+2\hat{d}_m}$. Then

$$
\begin{aligned}
& \frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k}\frac{1+(\theta_0/T)\lambda_k^{-2+2d_0}}{1+(\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}-1) \\
= & \frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})+o_p(1)
\end{aligned}
$$

Suppose first that $\hat{d}_m \in [0, d_0-\delta)$, then $(1-2d_0)/(2-2d_0) < (1-2\hat{d}_m)/(2-2\hat{d}_m)$. When $M_0 > 0$ and $M > 0$, we have

$$
\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\left\{\begin{array}{c} \in [(1+\tilde{M})^{-1}, 1+M_0],\ \text{if } k \geq T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}} \\ = O_p(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m}) = o_p(1), \text{if } k \in (T^{\frac{1-2d_0}{2-2d_0}}, T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}}) \\ \leq 2(M_0/\tilde{M})(k/T)^{2(d_0-\hat{d}_m)},\ \text{if } k \leq T^{\frac{1-2d_0}{2-2d_0}} \end{array}\right|
$$

Hence,

$$
\begin{aligned}
& E|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})| \\
\leq & \frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}E|\frac{I_k}{f_k}|(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}) \\
\leq & \frac{C}{m}\sum_{k=1}^{T^{(1-2d_0)/(2-2d_0)}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}2\frac{M_0}{\tilde{M}}(\frac{k}{T})^{2(d_0-\hat{d}_m)} \\
& +\frac{1+M_0}{m}C\sum_{k=T^{(1-2d_0)/(2-2d_0)}}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)} \\
= & o_p(1)
\end{aligned}
$$

Second, suppose $\hat{d}_m \in (d_0+\delta, 1/2)$, then $(1-2\hat{d}_m)/(2-2\hat{d}_m) < (1-2d_0)/(2-2d_0)$, and

$$
\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\left\{\begin{array}{c} \in [(1+\tilde{M})^{-1}, 1+M_0],\ \text{if } k \geq T^{\frac{1-2d_0}{2-2d_0}} \\ = O_p(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m}) = o_p(1), \text{if } k \in (T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}}, T^{\frac{1-2d_0}{2-2d_0}}) \\ \leq 2(M_0/\tilde{M})(k/T)^{2(d_0-\hat{d}_m)},\ \text{if } k \leq T^{\frac{1-2\hat{d}_m}{2-2\hat{d}_m}} \end{array}\right|
$$

Hence,

$$
\begin{aligned}
& E|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{I_k}{f_k}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})})| \\
\leq\ & \frac{1}{m}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}E|\frac{I_k}{f_k}|(\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}) \\
\leq\ & 2\left(\frac{C}{m}\right)\sum_{k=1}^{T^{(1-2\hat{d}_m)/(2-2\hat{d}_m)}}\frac{M_0}{\tilde{M}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(\frac{k}{T})^{2(d_0-\hat{d}_m)} && \text{(A.10)} \\
& +2\left(\frac{C}{m}M_0\right)\sum_{k=T^{(1-2\hat{d}_m)/(2-2\hat{d}_m)}}^{T^{(1-2d_0)/(2-2d_0)}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}(T^{1-2d_0}/k^{2-2d_0}) && \text{(A.11)} \\
& +\left((1+M_0)\frac{C}{m}\right)\sum_{k=T^{(1-2d_0)/(2-2d_0)}}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)} && \text{(A.12)}
\end{aligned}
$$

Note that (A.10) is of order

$$T^{(1-2\hat{d}_m)/(2-2\hat{d}_m)+2(d-d_0)}/m^{1+2(\hat{d}_m-d_0)} = o_p(1)$$

and (A.11) is of order

$$T^{1-2d_0-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)}/m^{1+2(\hat{d}_m-d_0)} = o_p(1)$$

if

$$m/T^{[1-2d_0-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)]/[1+2(\hat{d}_m-d_0)]} \to \infty$$

Let

$$
\begin{aligned}
\beta_1(\hat{d}_m, d_0) &= [(1-2\hat{d}_m)/(2-2\hat{d}_m)+2(\hat{d}_m-d_0)]/[1+2(\hat{d}_m-d_0)], \\
\beta_2(\hat{d}_m, d_0) &= [1-2d_0-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)]/[1+2(\hat{d}_m-d_0)]
\end{aligned}
$$

Note that $(1-2\hat{d}_m)/(2-2\hat{d}_m)+2(\hat{d}_m-d_0) = (1-2d_0)-(1-2\hat{d}_m)^2/(2-2\hat{d}_m)$, so that $\beta_1(\hat{d}_m, d_0) = \beta_2(\hat{d}_m, d_0) \triangleq \beta(\hat{d}_m, d_0) = 1-(2(1-\hat{d}_m)(1-2d_0+2\hat{d}_m))^{-1}$. Tedious algebra shows that if $0 \leq d_0 < \hat{d}_m < 1/2$ (which holds since we are considering the case $\hat{d}_m \in (d_0+\delta, 1/2)$), then for a given d_0, the maximized value of $\beta(\hat{d}_m, d_0)$ is $1-(d_0^2-3d_0+9/4)^{-1}$.

So if $T^{1-(d_0^2-3d_0+9/4)^{-1}}/m \to 0$, which holds under Assumption A4, then (A.11) is $o_p(1)$. The arguments to show that (A.12) is $o_p(1)$ are similar but applied to the case $\hat{d}_m \in [0, d_0 - \delta)$.

Proof that the fifth term of (A.6) is $o_p(1)$. We have:

$$
\begin{aligned}
&\frac{1}{m}\sum_{k=1}^{m}\log(\frac{1+(\theta_m/T)\lambda_k^{-2+2\hat{d}_m}}{1+(\theta_m/T)\lambda_k^{-2+2d_0}}) \\
=\ &\frac{1}{m}\sum_{k=1}^{m}\log(\frac{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2d_0}}) \qquad (A.13)\\
=\ &\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(\frac{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2d_0}}) \qquad (A.14)\\
&+\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\log(\frac{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2d_0}}) \qquad (A.15)
\end{aligned}
$$

It is easy to show that the second term of (A.14) is $o_p(1)$. For the first term,

$$
\begin{aligned}
&|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(\frac{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}}{1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2d_0}})| \\
=\ &|\frac{1}{m}\sum_{k=1}^{\sqrt{T}}[\log(1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2\hat{d}_m}) - \log(1+(\theta_m T/(4\pi^2k^2))\lambda_k^{2d_0})]| \\
\leq\ &\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+\tilde{M}(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}) + \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+M_0(T^{1-2d_0}/k^{2-2d_0}) \\
\leq\ &\frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+\tilde{M}T^{1-2\hat{d}_m}) + \frac{1}{m}\sum_{k=1}^{\sqrt{T}}\log(1+M_0T^{1-2d_0}) \\
\sim\ &O_p(\frac{\sqrt{T}}{m}\log(T^{1-2\hat{d}_m})) + O_p(\frac{\sqrt{T}}{m}\log(T^{1-2d_0})) \\
\sim\ &O_p(\frac{\sqrt{T}\log T}{m}) + O_p(\frac{\sqrt{T}\log T}{m}) = o_p(1)
\end{aligned}
$$

This completes the proof for part (a) of Theorem 1. For part (b), note that $J_m(\hat{d}_m, \theta_m) - J_m(d_0, \theta_m) = O_p(m^{-1}T^{(1/2)(d_0^2-3d_0+9/4)\curlyvee(1/2)})$. So if $m \geq O_p(T^{\beta})$ with $\beta > (1/2)(d_0^2 - 3d_0 +$

$9/4) \curlyvee (1/2)$:

$$\begin{aligned} 0 &\geq O_p(T^{(1/2)(d_0^2-3d_0+9/4)\curlyvee(1/2)-\beta}) - (1/2)\log(1+2(\hat{d}_m - d_0)) + 2(\hat{d}_m - d_0) \\ &\geq O_p(T^{(1-(d_0^2-3d_0+9/4)^{-1})\curlyvee(1/2)-\beta}) + (1/6)(\hat{d}_m - d_0)^2 \end{aligned}$$

Hence $(1/6)(\hat{d}_m - d_0)^2 \leq O_p(T^{(1/2)(d_0^2-3d_0+9/4)\curlyvee(1/2)-\beta})$, so that $|\hat{d}_m - d_0| = o_p((\log m)^{-3})$ if $T^{(1-(d_0^2-3d_0+9/4)^{-1})\curlyvee(1/2)-\beta}) = o_p((\log m)^{-3})$ which is guaranteed if $\beta > (1-(d_0^2-3d_0+9/4)^{-1}) \curlyvee (1/2)$, by Assumption A4. This completes the proof of Theorem 1 for the case in which LFC are present. If no LFC is present $\theta_0 = 0$, in which case all proofs go through with no requirement on θ_m, since the ratio $(1+M_0(T^{1-2d_0}/k^{2-2d_0}))/(1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})) = 1/(1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m}))$ is bounded for any choice of $M_m \geq 0$. ■

Proof. of Lemma 2: Let $\bar{\theta} = \limsup \hat{\theta}_m$, $\bar{M} = (2\pi)^{2\hat{d}_m-2}\bar{\theta}$, $M_m = (2\pi)^{2\hat{d}_m-2}\hat{\theta}_m$, $T_{M_m} = \sup_k\{k|k^{2-2d_0}/T^{1-2d_0} \leq M_m\}$ and $T_{\bar{M}} = \sup_k\{k|k^{2-2d_0}/T^{1-2d_0} \leq \bar{M}\} = O_p(\bar{M}T^{(1-2d_0)/(2-2d_0)})$. Note that from Theorem 10, $\hat{d}_m \to d_0$. Then, (A.3) becomes

$$\begin{aligned} 0 &= \{\sum_{k=1}^{m}(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}})[1 - \\ &\quad (m\frac{I_k}{f_k}(\frac{\lambda_k^{-2d_0}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}}))\backslash(\sum_{j=1}^{m}\frac{I_j}{f_j}(\frac{\lambda_j^{-2d_0}}{\lambda_j^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_j^{-2}}))]\} \end{aligned}$$

$$
\begin{aligned}
&= \{\sum_{k=1}^{m}(\frac{\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}})[1-(m\frac{I_k}{f_k}\lambda_k^{2\hat{d}_m-2d_0}(\frac{1}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}))\\
&\quad \backslash(\sum_{j=1}^{m}\frac{I_j}{f_j}\lambda_j^{2\hat{d}_m-2d_0}(\frac{1}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}))]\}\\
&\sim (2\pi)^{2-2\hat{d}_m}\{\sum_{k=1}^{m}(\frac{T}{k})^{2-2\hat{d}_m}[1-\frac{I_k}{f_k}\frac{(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m})}{(k^{2-2\hat{d}_m}/T^{1-2\hat{d}_m})+M_m}]\}\\
&\geq \sum_{k=1}^{m}(\frac{T}{k})^{2-2\hat{d}_m}-\frac{1}{\bar{M}}\sum_{k=1}^{T_{M_m}}(\frac{T}{k})^{2-2\hat{d}_m}|\frac{I_k}{f_k}|\frac{k^{2-2\hat{d}_m}}{T^{1-2\hat{d}_m}}-\sum_{k=T_{M_m}+1}^{m}(\frac{T}{k})^{2-2\hat{d}_m}\frac{I_k}{f_k}\\
&\geq \sum_{k=1}^{T_M}(\frac{T}{k})^{2-2\hat{d}_m}-\frac{1}{\bar{M}}\sum_{k=1}^{T_{M_m}}T|\frac{I_k}{f_k}|-\sum_{k=T_{M_m}+1}^{m}(\frac{T}{k})^{2-2\hat{d}_m}|\frac{I_k}{f_k}-1|\\
&= T^{2-2d_0}(1-T_{\bar{M}}^{2d_0-1})-O_p(T^{1+(1-2d_0)/(2-2d_0)})-\sum_{k=T_{\bar{M}}+1}^{m}\frac{T^{2-2d_0}\log k}{k^{3-2d_0}}+o_p(1)
\end{aligned}
$$

If $\bar{\theta} > 0$, then $\bar{M} > 0$, and

$$
\begin{aligned}
&T^{2-2d_0}(1-T_{\bar{M}}^{2d_0-1})-O_p(T^{1+(1-2d_0)/(2-2d_0)})-\sum_{k=T_{\bar{M}}+1}^{m}\frac{T^{2-2d_0}\log k}{k^{3-2d_0}}\\
> &\ O_p(T^{1+(1-2d_0)})-O_p(T^{1+(1-2d_0)/(2-2d_0)})-O_p(T^{2-2d_0}\log m(T_M^{2d_0-2}-m^{2d_0-2}))\to\infty
\end{aligned}
$$

So the partial derivative with respect of θ will be always greater than zero and the objective function can not be minimized at $\hat{\theta}_m$, which is a contradiction. Hence, $\hat{\theta}_m \xrightarrow{p} 0$ when there is no LFC in the data. To complete the proof, note that:

$$
T^{2-2d_0}(1-T_{\bar{M}}^{2d_0-1})-O_p(T^{1+((1-2d_0)/(2-2d_0)})-\sum_{k=T_{\bar{M}}+1}^{m}\frac{T^{2-2d_0}\log k}{k^{3-2d_0}}<0
$$

so that $T_{\bar{M}}^{2d_0-1} \geq O_p(1)$. Hence,

$$
\begin{aligned}
T_{\bar{M}}^{1-2d_0} &= O_p(\bar{M}^{1-2d_0}T^{(1-2d_0)^2/(2-2d_0)})\\
&= O_p(\bar{\theta}^{1-2d_0}T^{(1-2d_0)^2/(2-2d_0)}) \leq O_p(1).
\end{aligned}
$$

which implies that $\bar{\theta} = \limsup \hat{\theta}_m \leq O_p(T^{-(1-2d_0)/(2-2d_0)})$ and proves the result. ■

Proof. of Theorem 2: The proof follows Robinson (1995a) with appropriate modifications to accommodate the extra term. Note that given Theorem 1 (b), we can restrict the analysis to values of $\hat{d}_m$ in the set $C_m(d) = \{\hat{d}_m : |\hat{d}_m - d_0| < \log(m)^{-3}\}$ and $\hat{\theta}_m$ in the set $(0, \infty)$ by Lemma 1 when LFC are present. We can write the objective function as

$$J_m(d,\theta) = \log(\frac{1}{m}\sum_{k=1}^{m}\frac{I_k}{\lambda_k^{-2d} + (\theta/T)\lambda_k^{-2}}) + \frac{1}{m}\sum_{k=1}^{m}\log(\lambda_k^{-2d} + (\theta/T)\lambda_k^{-2})$$

Since $\hat{G}(d,\theta) = m^{-1}\sum_{k=1}^{m}(I_k/(\lambda_k^{-2d} + (\theta/T)\lambda_k^{-2}))$, we have

$$\begin{aligned} J'_m(d,\theta) &= \frac{\partial}{\partial d}J_m(d,\theta) \\ &= \frac{1}{\hat{G}(d,\theta)}\frac{2}{m}\sum_{k=1}^{m}\frac{I_k}{g_k}\log(\lambda_k)\frac{g_{yk}}{g_k} - \frac{2}{m}\sum_{k=1}^{m}\log(\lambda_k)\frac{g_{yk}}{g_k} \end{aligned}$$

and the second order derivative is

$$\begin{aligned} J''_m(d,\theta) &= \frac{\partial^2}{\partial d^2}J_m(d,\theta) \\ &= -\frac{4}{m^2}\frac{1}{\hat{G}(d,\theta)^2}(\sum_{k=1}^{m}\frac{I_k}{g_k}\frac{\lambda_k^{-2d}}{g_k}\log(\lambda_k))^2 \\ &\quad +\frac{4}{\hat{G}(d,\theta)}\frac{1}{m}\sum_{k=1}^{m}(\frac{(\log(\lambda_k))^2}{(g_k)^3}I_k\lambda_k^{-2d}g_{yk}) \end{aligned}$$

$$-\frac{4}{\hat{G}(d,\theta)}\frac{1}{m}\sum_{k=1}^{m}(\frac{(\log(\lambda_k))^2}{(g_k)^3}I_k\lambda_k^{-2d}g_{uk}) + \frac{4}{m}\sum_{k=1}^{m}((\log(\lambda_k))^2\frac{\lambda_k^{-2d}g_{uk}}{(g_k)^2}) \tag{A.16}$$

We first show that when evaluated at $\hat{d}_m$ and $\hat{\theta}_m$ both terms of (A.16) are $o_p(1)$. For the first:

$$\begin{aligned} &-\frac{4}{\hat{G}(\hat{d}_m,\hat{\theta}_m)}\frac{1}{m}\sum_{k=1}^{m}(\frac{(\log(\lambda_k))^2}{(g_k)^3}I_k\lambda_k^{-2\hat{d}_m}g_{uk}) \\ \sim\ &\frac{1}{m}\sum_{k=1}^{m}(\log(\lambda_k)^2(\frac{I_k}{g_k^0})(\frac{g_k^0}{g_k})\frac{\lambda_k^{-2\hat{d}_m}g_{uk}}{(g_k)^2}) \\ \sim\ &\frac{1}{m}\sum_{k=1}^{m}((\log(\lambda_k)^2(\frac{I_k}{g_k^0})\lambda_k^{2(2\hat{d}_m - d_0)}) \end{aligned}$$

For $\hat{d}_m$ in $C_m(d)$, we have for T and m large enough, $2\hat{d}_m - d_0 \geq (1/2)d_0$, so that the first term is $o_p(1)$. It is trivial to show that the second is $o_p(1)$. Hence, the second derivative of the objective function evaluated at $(\hat{d}_m, \hat{\theta}_m)$ is such that:

$$\begin{aligned} J''_m(\hat{d}_m, \hat{\theta}_m) &= -\frac{4}{m^2} \frac{1}{\hat{G}(\hat{d}_m, \hat{\theta}_m)^2} (\sum_{k=1}^{m} \frac{I_k}{g_k} \frac{\lambda_k^{-2\hat{d}_m}}{g_k} \log(\lambda_k))^2 \\ &\quad + \frac{4}{\hat{G}(\hat{d}_m, \hat{\theta}_m)} \frac{1}{m} \sum_{k=1}^{m} (\frac{(\log(\lambda_k))^2}{(g_k)^3} I_k \lambda_k^{-2\hat{d}_m} g_{yk}) + o_p(1) \end{aligned}$$

Let $\hat{G}_l(\hat{d}_m, \hat{\theta}_m) = m^{-1} \sum_{k=1}^{m} (I_k/g_k)(\log(\lambda_k))^l (g_{yk}/g_k)^l$, then

$$J''_m(\hat{d}_m, \hat{\theta}_m) = \frac{4}{\hat{G}_0(\hat{d}_m, \hat{\theta}_m)} [\hat{G}_0(\hat{d}_m, \hat{\theta}_m) \hat{G}_2(\hat{d}_m, \hat{\theta}_m) - \hat{G}_1(\hat{d}_m, \hat{\theta}_m)^2] + o_p(1)$$

Defining $\tilde{G}_l(\hat{d}_m, \hat{\theta}_m) = m^{-1} \sum_{k=1}^{m} (I_k/g_k)(\log(\lambda_k))^l$, we will show that $\hat{G}_l(\hat{d}_m, \hat{\theta}_m) = \tilde{G}_l(\hat{d}_m, \hat{\theta}_m) + o_p(\tilde{G}_l(\hat{d}_m, \hat{\theta}_m))$, for $l = 0, 1, 2$. When $l = 1$,

$$\begin{aligned} \hat{G}_1(\hat{d}_m, \hat{\theta}_m) &= \frac{1}{m} \sum_{k=1}^{m} (\frac{I_k}{g_k}) \log(\lambda_k) (\frac{g_{yk}}{g_k}) \\ &= \frac{1}{m} \sum_{k=1}^{m} (\frac{I_k}{g_k}) \log(\lambda_k) - \frac{1}{m} \sum_{k=1}^{m} (\frac{I_k}{g_k}) \log(\lambda_k) (\frac{g_{uk}}{g_k}) \\ &= \tilde{G}_1(\hat{d}_m, \hat{\theta}_m) + o_p(\tilde{G}_1(\hat{d}_m, \hat{\theta}_m)) \end{aligned}$$

if $[m^{-1} \sum_{k=1}^{m} (I_k/g_k) \log(\lambda_k) (g_{uk}/g_k)] / [m^{-1} \sum_{k=1}^{m} (I_k/g_k) \log(\lambda_k)] \to 0$, which we now prove.

$$\begin{aligned} & \frac{m^{-1} \sum_{k=1}^{m} (I_k/g_k) \log(\lambda_k) (g_{uk}/g_k)}{m^{-1} \sum_{k=1}^{m} (I_k/g_k) \log(\lambda_k)} \\ = & \frac{m^{-1} \sum_{k=1}^{\sqrt{T}-1} (I_k/g_k) \log(\lambda_k) (g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m, \hat{\theta}_m)} \qquad \text{(A.17)} \\ & + \frac{m^{-1} \sum_{k=\sqrt{T}}^{m} (I_k/g_k) \log(\lambda_k) (g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m, \hat{\theta}_m)} \qquad \text{(A.18)} \end{aligned}$$

For the first term,

$$
\begin{aligned}
&|\frac{m^{-1}\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m,\hat{\theta}_m)}| \\
&= \frac{\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k^0)(g_k^0/g_k)(g_{uk}/g_k)\log(\lambda_k)}{\sum_{k=1}^{m}(I_k/g_k^0)(g_k^0/g_k)\log(\lambda_k)} \\
&= \frac{\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)(g_{uk}/g_k)}{\sum_{k=1}^{m}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)} \\
&\le \frac{\sum_{k=1}^{\sqrt{T}-1}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)}{\sum_{k=1}^{m}(I_k/g_k^0)(k/T)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(\lambda_k)}
\end{aligned}
$$

$$
\begin{aligned}
&= O_p\left(\frac{\log(T)\sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} - \sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(k)}{\log(T)\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})} - \sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\log(k)}\right) \\
&= O_p(\frac{\sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}}{\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}})
\end{aligned}
$$

From Robinson (1995a), Equation (3.7) and a result in the proof of consistency,

$$
\begin{aligned}
\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)} &\sim \sum_{k=1}^{m}(k/m)^{2(\hat{d}_m-d_0)} \\
&= m(1+2(\hat{d}_m-d_0))^{-1}+o_p(m)\sim m
\end{aligned}
$$

for $\hat{d}$ in $C_m(d)$. Hence,

$$
\begin{aligned}
&\frac{\sum_{k=1}^{\sqrt{T}-1}(k/m)^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}}{\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}} \\
&\sim \frac{1}{m}\sum_{k=1}^{\sqrt{T}-1}(\frac{k}{m})^{2(d-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}
\end{aligned}
$$

which is $o_p(1)$ under Assumption A4 from the proof of consistency. For the second term in

(A.17), using similar arguments, we have:

$$\begin{aligned} & \frac{m^{-1}\sum_{k=\sqrt{T}}^{m}(I_k/g_k)\log(\lambda_k)(g_{uk}/g_k)}{\tilde{G}_1(\hat{d}_m,\hat{\theta}_m)} \\ \sim \quad & \frac{1}{m}\sum_{k=\sqrt{T}}^{m}(\frac{k}{m})^{2(\hat{d}_m-d_0)}\frac{1+M_0(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})}\frac{(\hat{\theta}_m/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}} \end{aligned} \tag{A.19}$$

If $d_0 > 0$, then (A.19) is asymptotically equivalent to $m^{-1}\sum_{k=\sqrt{T}}^{m}(k/m)^{2(\hat{d}_m-d_0)}(k/T)^{2\hat{d}_m} = (m/T)^{2\hat{d}_m}m^{-1}\sum_{k=\sqrt{T}}^{m}(k/m)^{2(2\hat{d}_m-d_0)} \to 0$, for $\hat{d}_m$ in $C_m(d)$. If $d_0 = 0$, then (A.19) is asymptotically equivalent to $(T/m)\sum_{k=\sqrt{T}}^{m}(k/m)^4(\hat{\theta}_m/k^2) \sim (T/m^5)\sum_{k=\sqrt{T}}^{m}k^2 \sim (T/m^5)m^3 = (T/m^2) \to 0$. Hence, both terms of (A.17) are $o_p(1)$ and we have $\hat{G}_1(\hat{d}_m,\hat{\theta}_m) = \tilde{G}_1(\hat{d}_m,\hat{\theta}_m) + o(\tilde{G}_1(\hat{d}_m,\hat{\theta}_m))$. Similarly, $\hat{G}_2(\hat{d}_m,\hat{\theta}_m) = \tilde{G}_2(\hat{d}_m,\hat{\theta}_m) + o(\tilde{G}_2(\hat{d}_m,\hat{\theta}_m))$. Accordingly,

$$\begin{aligned} \frac{\partial^2}{\partial d^2}J_m(\hat{d}_m,\hat{\theta}_m) &\sim \frac{4}{\tilde{G}_0(\hat{d}_m,\hat{\theta}_m)^2}[\tilde{G}_0(\hat{d}_m,\hat{\theta}_m)\tilde{G}_1(\hat{d}_m,\hat{\theta}_m) - \tilde{G}_1(\hat{d}_m,\hat{\theta}_m)^2] \\ &= \frac{4}{(m^{-1}\sum_{k=1}^{m}(I_k/g_k))^2}[(\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{g_k}))(\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{g_k})\log^2(\lambda_k)) \\ &\quad -(\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{g_k})\log(\lambda_k))^2] \\ &= \frac{4}{(m^{-1}\sum_{k=1}^{m}(I_k/g_k))^2}[(\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{g_k}))(\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{g_k})\log^2(k)) \\ &\quad -(\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{g_k})\log(k))^2] \end{aligned}$$

Let $\hat{F}_l(\hat{d}_m,\hat{\theta}_m) = m^{-1}\sum_{k=1}^{m}(I_k/g_k)\log(k))^l$, $h_k = h_k(\hat{d}_m,\hat{\theta}_m) = 1 + (\hat{\theta}_m/T)\lambda_k^{-2+2\hat{d}_m} = 1 + M_m(T^{1-2\hat{d}_m}/k^{2-2\hat{d}_m})$, and $h_k^0 = h_k(d_0,\theta_0) = 1 + M_0(T^{1-2d_0}/k^{2-2d_0})$. Then

$$\begin{aligned} \frac{I_k}{g_k} &= \frac{I_k}{1+(\hat{\theta}_m/T)\lambda_k^{-2+2\hat{d}_m}}[k^{2\hat{d}_m}(\frac{2\pi}{T})^{2\hat{d}_m}] \\ &= \frac{I_k}{h_k}[k^{2\hat{d}_m}(\frac{2\pi}{T})^{2\hat{d}_m}]. \end{aligned}$$

For $\tau = 0, 1, 2,$

$$\begin{aligned}
& |\hat{F}_\tau(\hat{d}_m, \hat{\theta}_m) - \hat{F}_\tau(d_0, \theta_0)| \\
= & |\frac{1}{m}\sum_{k=1}^{m} I_k \log^\tau(k)[\frac{1}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} \qquad \text{(A.20)} \\
& -\frac{1}{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}]| \qquad \text{(A.21)} \\
= & |\frac{1}{m}\sum_{k=1}^{m} \frac{I_k \log^\tau(k)}{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}(\frac{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} - 1)| \\
\leq & |\frac{1}{m}\sum_{k=1}^{\sqrt{T}} \frac{I_k \log^\tau(k)}{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}(\frac{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} - 1)| \qquad \text{(A.22)} \\
& +|\frac{1}{m}\sum_{k=\sqrt{T}}^{m} \frac{I_k \log^\tau(k)}{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}(\frac{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} - 1)|
\end{aligned}$$

For the first term of (A.22), we have

$$\begin{aligned}
& |\frac{1}{m}\sum_{k=1}^{\sqrt{T}} \frac{I_k \log^\tau(k)}{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}(\frac{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} - 1)| \\
\leq & |\frac{1}{m}\sum_{k=1}^{\sqrt{T}} \frac{I_k \log^\tau(k)}{g_k^0}\frac{g_k^0}{g_k}| + |\frac{1}{m}\sum_{k=1}^{\sqrt{T}} \frac{I_k \log^\tau(k)}{g_k^0}| \\
\leq & |\frac{\log^\tau(m)}{m}\sum_{k=1}^{\sqrt{T}} \frac{I_k}{g_k^0}\lambda_k^{2(\hat{d}_m - d_0)}\frac{h_k(d_0,\theta_0)}{h_k(\hat{d}_m,\hat{\theta}_m)}| + |\frac{\log^\tau(m)}{m}\sum_{k=1}^{\sqrt{T}}\frac{I_k}{g_k^0}|
\end{aligned}$$

$$\sim |\frac{\log^\tau(m)}{m}(\frac{T}{m})^{2(d_0-\hat{d}_m)}\sum_{k=1}^{\sqrt{T}}(\frac{k}{m})^{2(\hat{d}_m-d_0)}\frac{h_k(d_0,\theta_0)}{h_k(\hat{d}_m,\hat{\theta}_m)}| + O_p(\frac{\log^\tau(m)}{m}\sqrt{T})$$

Note that from results in the proof for consistency and the fact that $\hat{d}_m$ is in $C_m(d)$, this last term is $o_p(1)$ if A4 holds. For the second term of (A.22), we have

$$\begin{aligned}
& |\frac{1}{m}\sum_{k=\sqrt{T}}^{m} \frac{I_k \log^\tau(k)}{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}(\frac{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} - 1)| \\
\sim & |\frac{1}{m}\sum_{k=\sqrt{T}}^{m} (\frac{\lambda_k^{-2d_0} + (\theta_0/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m} + (\hat{\theta}_m/T)\lambda_k^{-2}} - 1)\log^\tau(k)|
\end{aligned}$$

and the first derivative of the second term of (A.22) is

$$\begin{aligned}
& |\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\frac{g_k^0}{g_k}\log(\lambda_k)\frac{\lambda_k^{-2\hat{d}_m}}{g_k}(-2)\log^{\tau}(k)| \\
= \; & O_p(|\frac{1}{m}\sum_{k=\sqrt{T}}^{m}\frac{g_k^0}{g_k}(\log k-\log T)\log^{\tau}(k)|) \\
= \; & O_p(|\frac{\log T\log^{\tau}(m)}{m}\sum_{k=\sqrt{T}}^{m}\lambda_k^{2(\hat{d}_m-d_0)}\frac{I_k}{\lambda_k^{-2d_0}+(\theta_0/T)\lambda_k^{-2}}|) \\
\leq \; & |\frac{2\log m\log^{\tau}(m)}{m}\sum_{k=\sqrt{T}}^{m}(\frac{T}{k})^{2|\hat{d}_m-d_0|}(2\pi)^{2(\hat{d}_m-d_0)}| \\
= \; & O_p(\frac{\log^{\tau+1}(m)}{m}mT^{|\hat{d}_m-d_0|})\leq\log^{\tau+1}(m)m^{2|\hat{d}_m-d_0|}\to 0
\end{aligned}$$

since $\hat{d}_m$ in $C_m(d)$. Also, under A4 and for $\hat{d}_m$ in $C_m(d)$: $\log^{\tau+1}(m)m^{2|\hat{d}_m-d_0|}|\hat{d}_m-d_0| = o_p(\log^{\tau-2}(m)m^{2|\hat{d}_m-d_0|}) \leq o_p((\log m)^{\tau-2}m^{\log m-1}) \leq o_p((\log m)^{\tau-2}) = o_p(1)$. Hence, the second term of (A.22) converges to 0, so that $|\hat{F}_\tau(\hat{d}_m,\hat{\theta}_m)-\hat{F}_\tau(d_0,\theta_0)| \xrightarrow{p} 0$, and

$$\frac{\partial^2}{\partial d^2}J_m(\hat{d}_m,\hat{\theta}_m)\xrightarrow{p}\frac{4}{\hat{F}_0(d_0,\theta_0)^2}[\hat{F}_2(d_0,\theta_0)\hat{F}_0(d_0,\theta_0)-\hat{F}_1^2(d_0,\theta_0)]$$

We now show that $\hat{F}_\tau(d_0,\theta_0)=G_0m^{-1}\sum_{k=1}^{m}(\log k)^\tau+o_p(1)$. Using summations by parts,

$$|\hat{F}_\tau(d_0,\theta_0)-G_0\frac{1}{m}\sum_{k=1}^{m}(\log k)^\tau| = |\frac{1}{m}\sum_{k=1}^{m}(\frac{I_k}{g_k^0})\log(k)^\tau-G_0\frac{1}{m}\sum_{k=1}^{m}(\log k)^\tau|$$

$$\begin{aligned}
\leq \; & \frac{G_0}{m}\sum_{r=1}^{m-1}(|(\log r)^k-(\log(r+1))^k||\sum_{k=1}^{r}(\frac{I_k}{g_k^0}-1)|) \\
& +\frac{G_0}{m}(\log m)^\tau|\sum_{k=1}^{m}(\frac{I_k}{g_k^0}-1)| \\
\leq \; & \frac{G_0}{m}\sum_{r=1}^{m-1}(|(\log(r+1))^{\tau-1}||\frac{1}{r}\sum_{k=1}^{r}(\frac{I_k}{g_k^0}-1)|)+o_p(1) \\
= \; & \frac{G_0}{m}\sum_{r=1}^{T^{1/2+\varepsilon}}(|(\log(r+1))^{\tau-1}||\frac{1}{r}\sum_{k=1}^{r}(\frac{I_k}{g_k^0}-1)|)
\end{aligned}$$

$$
\begin{aligned}
& +\frac{G_0}{m}\sum_{r=T^{1/2+\varepsilon}+1}^{m-1}(|(\log(r+1))^{\tau-1}||\frac{1}{r}\sum_{k=1}^{r}(\frac{I_k}{g_k^0}-1)|)+o_p(1) \\
= & \; O_p(\frac{G_0}{m}\sum_{r=1}^{T^{1/2+\varepsilon}}(|(\log(r+1))^{\tau-1}|) \\
& +\frac{G_0}{m}\sum_{r=T^{1/2+\varepsilon}+1}^{m-1}(|(\log(r+1))^{\tau-1}|)o_p((\log(r+1))^{-2})+o_p(1) \\
= & \; O_p(G_0\frac{T^{1/2+\varepsilon}}{m}(\log(m+1))^{\tau-1})+o_p(G_0(\log(T^{1/2+\varepsilon}+1))^{\tau-2})+o_p(1) \\
= & \; o_p(1)
\end{aligned}
$$

So

$$
\frac{\partial^2}{\partial d^2}J_m(\hat{d}_m,\hat{\theta}_m)=4[\frac{1}{m}(\sum_{k=1}^{m}(\log k)^2)-(\frac{1}{m}\sum_{k=1}^{m}(\log k))^2]+o_p(1)\to 4
$$

Note that the derivations above are valid so long as the sequence $\{\hat{\theta}_m\}$ is bounded below from zero, which holds if LFC are present. Now because $\hat{G}(d_0,\hat{\theta}_m)=m^{-1}\sum_{k=1}^{m}I_k[\lambda_k^{-2d_0}+(\hat{\theta}_m/T)\lambda_k^{-2}]^{-1}\xrightarrow{p}G_0$, from the fact that the second term in (A.6) is $o_p(1)$ in the proof of Theorem 10, then using similar arguments as in Robinson (1995a), we have

$$
\begin{aligned}
m^{1/2}\frac{\partial}{\partial d}J_m(d_0,\hat{\theta}_m) &= m^{-1/2}\sum_{k=1}^{m}(\frac{I_k}{f_k}\frac{g_{y,k}}{g_k}+\frac{g_{u,k}}{g_k})\nu_k+o_p(1) \\
&= m^{-1/2}\sum_{k=1}^{m}((\frac{I_k}{f_k}-1)\frac{g_{y,k}}{g_k})\nu_k+o_p(1) \\
&= m^{-1/2}\sum_{k=1}^{m}((\frac{I_{yk}}{f_{yk}}-1)\frac{g_{y,k}}{g_k})\nu_k \\
&\quad +m^{-1/2}\sum_{k=1}^{m}((\frac{I_k}{f_k}-\frac{I_{yk}}{f_{yk}})\frac{g_{y,k}}{g_k})\nu_k+o_p(1)
\end{aligned}
$$

where $\nu_k=[\log k-(m^{-1}\sum_{j=1}^{m}\log j)]$. Using the same approach as in Robinson (1995a, pp. 1644-1653), the first part converges to a $N(0,4)$ (note that for the part involving the 4-th cumulant $cum(\omega_j/f_j,\omega_k/f_k,\bar{\omega}_j/f_j,\,\bar{\omega}_k/f_k)$ we need to use the results of Lemmas A.1-A.2 to get the corresponding results for the DGP with LFC). What remains to be shown is that the second part is $o_p(1)$. We have, where $\tilde{I}_{uk} \doteq I_k - I_{yk}$:

$$\begin{aligned}
& m^{-1/2}\sum_{k=1}^{m}((\frac{I_k}{f_k} - \frac{I_{yk}}{f_{yk}})\frac{g_{y,k}}{g_k})\nu_k \\
= & m^{-1/2}\sum_{k=1}^{m}((1 - \frac{I_{yk}}{f_{yk}})\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k})\nu_k \\
& +m^{-1/2}\sum_{k=1}^{m}((\frac{I_k - I_{yk} - f_{uk}}{f_{uk}})\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k})\nu_k \\
= & m^{-1/2}\sum_{k=1}^{m}((\frac{\tilde{I}_{uk} - f_{uk}}{f_{uk}})\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k})\nu_k + o_p(1) \\
= & m^{-1/2}\sum_{k=1}^{m}((\frac{\frac{k^2}{T}\tilde{I}_{uk} - M_m G_0}{M_m G_0})\frac{g_{u,k}}{g_k}\frac{g_{y,k}}{g_k})\nu_k + o_p(1) \\
= & \frac{1}{TG_0}m^{-1/2}\sum_{k=1}^{m}[(\frac{k^2}{T}\tilde{I}_{uk} - M_m G_0)(\frac{\lambda_k^{-2}}{g_k})^2(\lambda_k^{2-2d_0}\nu_k)] + o_p(1) \\
= & o_p(1)
\end{aligned}$$

using summations by parts and $M_m G_0 = [\sum_{k=1}^{m}(\lambda_k^{-2}/g_k)^2(k^2/T)\tilde{I}_{uk}/\sum_{k=1}^{m}(\lambda_k^{-2}/g_k)^2]$. Hence, a CLT can be applied so that $\sqrt{m}(\partial/\partial d)J_m(d_0, \hat{\theta}_m) \xrightarrow{d} N(0, 4)$. Thus from

$$\frac{\partial}{\partial d}J_m(d_0, \hat{\theta}_m) = \frac{\partial}{\partial d}J_m(\hat{d}_m, \hat{\theta}_m) + \frac{\partial^2}{\partial d^2}J_m(\hat{d}_m, \hat{\theta}_m)(\hat{d}_m - d_0)$$

and the fact that $(\partial/\partial d)J_m(\hat{d}_m, \hat{\theta}_m) = 0$, we have

$$\sqrt{m}(\hat{d}_m - d_0) \xrightarrow{d} \frac{\sqrt{m}\partial(J_m(d_0, \hat{\theta}_m))\partial d}{\partial^2(J_m(\hat{d}_m, \hat{\theta}_m))/\partial d^2} = N(0, 1/4)$$

This completes the proof of Theorem 2 for the case with LFC present. To complete the proof for the case with no LFC, we need to show that

$$m^{-1/2}\sum_{k=1}^{m}((\frac{I_k}{f_k} - \frac{I_{yk}}{f_{yk}})\frac{g_{yk}}{g_k})\upsilon_k = o_p(1). \tag{A.23}$$

Note that with no LFC, we have $I_k = I_{yk}$ and $\tilde{I}_{uk} = 0$. Hence,

$$m^{-1/2}\sum_{k=1}^{m}((\frac{I_k}{f_k} - \frac{I_{yk}}{f_{yk}})\frac{g_{yk}}{g_k})\upsilon_k = -m^{-1/2}\sum_{k=1}^{m}\frac{g_{uk}}{g_k}\frac{g_{yk}}{g_k}\upsilon_k + o_p(1)$$

So we want to show that $m^{-1/2}\sum_{k=1}^{m}(g_{uk}/g_k)(g_{yk}/g_k)\upsilon_k = o_p(1)$. To that effect, it suffices to show that $m^{-1/2}\sum_{k=1}^{m}(g_{uk}/g_k)\upsilon_k = o_p(1)$. To prove this, note that

$$\begin{aligned} m^{-1/2}\sum_{k=1}^{m}\frac{g_{uk}}{g_k}\upsilon_k &= m^{-1/2}\sum_{k=1}^{m}\frac{(\hat{\theta}_m/T)\lambda_k^{-2}}{\lambda_k^{-2\hat{d}_m}+(\hat{\theta}_m/T)\lambda_k^{-2}}\upsilon_k \\ &= m^{-1/2}\sum_{k=1}^{m}\frac{(\hat{\theta}_m/T)\lambda_k^{-2}}{\lambda_k^{-2d_0}+(\hat{\theta}_m/T)\lambda_k^{-2}}\upsilon_k + o_p(1) \\ &= m^{-1/2}\sum_{k=1}^{m}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k + o_p(1) \end{aligned}$$

since $\hat{d}_m = d_0 + O_p(m^{-1/2})$. Now, let $T_\theta = \sup_k\{k|M_m(k^{2-2d_0}/T^{1-2d_0}) = O_p(1)\}$. We have

$$\begin{aligned} T_\theta &= O_p(T^{(1-2d_0)/(2-2d_0)}\hat{\theta}_m^{1/(2-2d_0)}) \\ &\leq O_p(T^{(1-2d_0)/(2-2d_0)}T^{-[(1-2d_0)/(2-2d_0)]/(2-2d_0)}) \\ &= O_p(T^{((1-2d_0)/(2-2d_0))^2}) \end{aligned}$$

using Lemma 1. Then,

$$\begin{aligned} & m^{-1/2}\sum_{k=1}^{m}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k \\ =\; & m^{-1/2}\sum_{k=1}^{T_\theta}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k \\ & + m^{-1/2}\sum_{k=T_\theta+1}^{m}\frac{M_m(T^{1-2d_0}/k^{2-2d_0})}{1+M_m(T^{1-2d_0}/k^{2-2d_0})}\upsilon_k \\ \leq\; & m^{-1/2}T_\theta + m^{-1/2}\hat{\theta}_m T^{1-2d_0}T_\theta^{-(1-2d_0)} \\ =\; & O_p(m^{-1/2}T^{((1-2d_0)/(2-2d_0))^2}) = o_p(1) \end{aligned}$$

from Assumption A4. Hence, (A.23) holds when there is no LFC, which completes the proof of Theorem **2**. ■

Table 1.1: Bias and RMSE for a short memory process ARFIMA(a,d=0,0) with RLS

	p=0			p=5			p=10			p=20		
T\Beta	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
a) Bias						a=0						
256	-0.087	-0.047	-0.021	-0.021	-0.038	0.004	0.004	-0.017	-0.002	-0.011	-0.059	0.023
512	-0.052	-0.025	-0.008	-0.016	-0.007	0.005	-0.014	-0.028	-0.012	-0.086	-0.055	0.011
1024	-0.037	-0.016	-0.006	-0.029	-0.001	0.003	-0.018	-0.001	-0.004	0.008	-0.026	0.001
2048	-0.013	-0.009	-0.006	-0.012	0.001	0.001	0.016	-0.006	0.005	0.013	-0.011	-0.005
4096	-0.007	-0.004	-0.004	-0.009	-0.006	-0.001	-0.032	-0.009	0.003	0.003	-0.009	0.000
						a=0.3						
256	-0.022	0.070	0.168	0.009	0.122	0.246	0.215	0.413	0.524	-0.012	0.131	0.270
512	-0.029	0.041	0.134	0.015	0.116	0.194	0.166	0.354	0.491	0.059	0.142	0.242
1024	-0.012	0.025	0.118	-0.009	0.048	0.137	0.123	0.267	0.436	0.023	0.053	0.160
2048	-0.005	0.027	0.092	0.004	0.033	0.107	0.072	0.206	0.383	-0.012	0.040	0.013
4096	-0.007	0.014	0.073	0.018	0.015	0.085	0.044	0.145	0.329	-0.011	0.024	0.100
						a=0.6						
256	0.128	0.299	0.432	0.176	0.387	0.498	0.213	0.414	0.524	0.221	0.449	0.561
512	0.093	0.223	0.392	0.135	0.307	0.459	0.178	0.344	0.492	0.174	0.388	0.517
1024	0.052	0.170	0.347	0.109	0.250	0.406	0.069	0.202	0.380	0.108	0.298	0.463
2048	0.020	0.125	0.307	0.066	0.186	0.361	0.063	0.143	0.330	0.085	0.234	0.409
4096	0.013	0.084	0.267	0.036	0.131	0.311	0.014	0.101	0.281	0.035	0.161	0.349
b) RMSE						a=0						
256	0.191	0.111	0.078	0.423	0.291	0.140	0.546	0.327	0.162	0.656	0.421	0.231
512	0.132	0.075	0.046	0.353	0.169	0.082	0.402	0.197	0.121	0.575	0.362	0.104
1024	0.095	0.057	0.037	0.262	0.130	0.068	0.280	0.147	0.071	0.376	0.213	0.078
2048	0.068	0.041	0.026	0.215	0.089	0.044	0.274	0.108	0.046	0.316	0.131	0.052
4096	0.062	0.029	0.019	0.147	0.069	0.031	0.218	0.075	0.031	0.264	0.101	0.041
						a=0.3						
256	0.172	0.118	0.180	0.371	0.201	0.265	0.467	0.269	0.277	0.524	0.337	0.311
512	0.132	0.086	0.142	0.282	0.179	0.207	0.354	0.185	0.235	0.432	0.241	0.261
1024	0.082	0.058	0.122	0.157	0.087	0.142	0.203	0.096	0.154	0.276	0.117	0.172
2048	0.068	0.049	0.096	0.128	0.060	0.111	0.140	0.066	0.120	0.221	0.083	0.137
4096	0.049	0.035	0.076	0.102	0.041	0.088	0.115	0.047	0.094	0.150	0.060	0.102
						a=0.6						
256	0.194	0.312	0.437	0.357	0.407	0.504	0.385	0.442	0.530	0.482	0.493	0.569
512	0.145	0.233	0.395	0.237	0.322	0.463	0.300	0.370	0.497	0.340	0.414	0.522
1024	0.099	0.177	0.349	0.190	0.260	0.409	0.229	0.278	0.438	0.279	0.313	0.466
2048	0.066	0.132	0.308	0.136	0.194	0.362	0.167	0.216	0.385	0.203	0.245	0.411
4096	0.053	0.098	0.268	0.092	0.138	0.312	0.116	0.153	0.330	0.146	0.170	0.351

Table 1.2: Bias and RMSE for a long memory process ARFIMA(a,d=0.2,0) with RLS

	p=0			p=5			p=10			p=20		
T\Beta	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
a) Bias						a=0						
256	-0.106	-0.054	-0.033	-0.103	-0.05	-0.030	-0.093	-0.057	-0.053	-0.159	-0.094	-0.047
512	-0.050	-0.025	-0.020	-0.054	-0.011	-0.026	-0.107	-0.027	-0.025	-0.064	-0.052	-0.001
1024	-0.033	-0.019	-0.015	-0.025	-0.012	-0.036	-0.012	-0.014	-0.042	-0.054	-0.029	-0.039
2048	-0.016	-0.011	-0.012	-0.027	-0.008	-0.027	-0.016	-0.010	-0.031	-0.009	-0.013	-0.030
4096	-0.015	-0.008	-0.006	0.004	-0.009	0.020	-0.011	-0.006	-0.022	-0.006	-0.009	-0.022
						a=0.3						
256	-0.047	0.063	0.149	-0.080	0.101	0.203	0.021	0.114	0.214	-0.030	0.118	0.237
512	-0.017	0.034	0.126	-0.023	0.088	0.167	0.001	0.098	0.185	-0.019	0.102	0.220
1024	-0.014	0.019	0.102	-0.006	0.054	0.141	0.003	0.003	0.162	-0.035	0.073	0.172
2048	-0.022	0.023	0.088	-0.018	0.035	0.117	-0.020	0.040	0.128	-0.011	0.048	0.140
4096	-0.014	0.013	0.074	0.010	0.025	0.092	0.004	0.033	0.101	0.005	0.042	0.112
						a=0.6						
256	0.128	0.302	0.424	0.160	0.336	0.442	0.192	0.362	0.474	0.149	0.383	0.483
512	0.080	0.223	0.384	0.098	0.268	0.415	0.138	0.297	0.427	0.129	0.339	0.461
1024	0.045	0.170	0.348	0.078	0.209	0.365	0.091	0.227	0.380	0.118	0.251	0.412
2048	0.038	0.120	0.306	0.050	0.153	0.322	0.067	0.170	0.335	0.056	0.194	0.361
4096	0.015	0.086	0.267	0.031	0.114	0.277	0.039	0.125	0.292	0.026	0.137	0.309
b) RMSE						a=0						
256	0.229	0.126	0.082	0.409	0.213	0.139	0.508	0.298	0.191	0.635	0.398	0.191
512	0.132	0.083	0.058	0.302	0.136	0.091	0.414	0.198	0.103	0.514	0.273	0.116
1024	0.097	0.057	0.040	0.205	0.076	0.059	0.244	0.068	0.057	0.370	0.088	0.059
2048	0.065	0.041	0.028	0.163	0.051	0.044	0.194	0.043	0.039	0.265	0.057	0.039
4096	0.053	0.033	0.021	0.109	0.037	0.032	0.148	0.029	0.028	0.205	0.039	0.029
						a=0.3						
256	0.174	0.114	0.161	0.330	0.201	0.221	0.379	0.226	0.243	0.485	0.316	0.274
512	0.128	0.085	0.137	0.256	0.146	0.179	0.292	0.173	0.198	0.409	0.203	0.235
1024	0.085	0.056	0.107	0.200	0.099	0.147	0.256	0.115	0.169	0.329	0.140	0.180
2048	0.078	0.045	0.091	0.114	0.070	0.121	0.170	0.079	0.132	0.202	0.097	0.145
4096	0.056	0.033	0.077	0.096	0.048	0.095	0.118	0.066	0.104	0.143	0.070	0.115
						a=0.6						
256	0.216	0.314	0.429	0.264	0.350	0.447	0.323	0.380	0.479	0.382	0.434	0.503
512	0.137	0.233	0.387	0.218	0.279	0.418	0.231	0.311	0.430	0.265	0.352	0.465
1024	0.098	0.180	0.350	0.135	0.217	0.367	0.171	0.238	0.382	0.139	0.262	0.414
2048	0.075	0.126	0.307	0.106	0.160	0.323	0.122	0.177	0.337	0.139	0.203	0.363
4096	0.051	0.090	0.268	0.077	0.120	0.278	0.093	0.132	0.293	0.099	0.143	0.310

Table 1.3: Bias and RMSE for a long memory process ARFIMA (a,d=0.45,0) with RLS

	p=0			p=5			p=10			p=20		
T\Beta	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
a) Bias	a=0											
256	-0.148	-0.056	-0.065	-0.142	-0.082	-0.080	-0.144	-0.096	-0.062	-0.195	-0.090	-0.099
512	-0.073	-0.0382	-0.036	-0.075	-0.047	-0.045	-0.081	-0.041	-0.050	-0.097	-0.050	-0.063
1024	-0.049	-0.025	-0.026	-0.041	-0.039	-0.033	-0.036	-0.027	-0.081	-0.036	-0.027	-0.081
2048	-0.027	-0.018	-0.017	-0.060	-0.040	-0.030	-0.023	-0.014	-0.062	-0.023	-0.014	-0.062
4096	-0.016	-0.009	-0.010	-0.049	-0.035	-0.022	0.004	-0.013	-0.047	0.004	-0.013	-0.047
	a=0.3											
256	-0.071	0.069	0.142	-0.091	0.034	0.162	-0.060	0.079	0.170	-0.053	0.101	0.200
512	-0.056	0.043	0.125	-0.061	0.048	0.143	-0.054	0.078	0.144	-0.046	0.074	0.074
1024	-0.034	0.027	0.106	-0.014	0.044	0.118	-0.034	0.050	0.126	-0.005	0.054	0.140
2048	-0.001	0.015	0.091	-0.008	0.024	0.094	-0.014	0.033	0.105	-0.014	0.038	0.112
4096	-0.007	0.015	0.070	-0.006	0.019	0.078	0.002	0.020	0.081	-0.005	0.027	0.088
	a=0.6											
256	0.107	0.289	0.407	0.069	0.302	0.409	0.129	0.291	0.417	0.135	0.304	0.414
512	0.067	0.223	0.376	0.067	0.234	0.380	0.079	0.246	0.388	0.095	0.266	0.389
1024	0.042	0.169	0.341	0.049	0.169	0.342	0.048	0.185	0.344	0.061	0.201	0.355
2048	0.032	0.129	0.300	0.028	0.130	0.284	0.028	0.141	0.305	0.040	0.154	0.310
4096	0.019	0.089	0.263	0.014	0.071	0.227	0.022	0.100	0.234	0.009	0.107	0.269
b) RMSE	a=0											
256	0.324	0.160	0.129	0.405	0.236	0.172	0.456	0.340	0.151	0.561	0.391	0.258
512	0.238	0.105	0.069	0.288	0.169	0.090	0.327	0.167	0.107	0.404	0.253	0.143
1024	0.187	0.077	0.052	0.191	0.101	0.084	0.210	0.099	0.082	0.288	0.133	0.084
2048	0.147	0.064	0.039	0.130	0.078	0.064	0.142	0.055	0.059	0.158	0.089	0.047
4096	0.080	0.043	0.030	0.092	0.059	0.050	0.097	0.040	0.042	0.116	0.061	0.028
	a=0.3											
256	0.329	0.124	0.159	0.356	0.206	0.199	0.347	0.193	0.201	0.451	0.256	0.229
512	0.213	0.093	0.135	0.257	0.161	0.154	0.270	0.137	0.155	0.399	0.181	0.183
1024	0.133	0.066	0.111	0.144	0.090	0.124	0.215	0.094	0.133	0.227	0.114	0.146
2048	0.085	0.044	0.094	0.094	0.056	0.097	0.117	0.066	0.109	0.146	0.081	0.116
4096	0.060	0.033	0.072	0.063	0.042	0.081	0.082	0.048	0.083	0.097	0.055	0.091
	a=0.6											
256	0.244	0.307	0.413	0.324	0.320	0.414	0.250	0.325	0.422	0.289	0.385	0.421
512	0.195	0.234	0.380	0.196	0.243	0.383	0.215	0.257	0.391	0.239	0.278	0.392
1024	0.105	0.177	0.343	0.121	0.186	0.346	0.130	0.195	0.350	0.147	0.208	0.357
2048	0.073	0.135	0.301	0.088	0.135	0.284	0.091	0.148	0.310	0.101	0.160	0.311
4096	0.060	0.094	0.264	0.045	0.074	0.228	0.062	0.114	0.236	0.059	0.110	0.270

Table 1.4: Bias and RMSE for a long memory process ARFIMA(0.6,d,0) with RLS (p=10) and additive noise

	LWLFC						LWPLFC					
noise	var(noise)=1			var(noise)=4			var(noise)=1			var(noise)=4		
T\Beta	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
a) bias: d=0.2												
256	0.131	0.252	0.246	-0.015	0.060	0.038	0.271	0.293	0.310	0.045	0.117	0.183
512	0.106	0.212	0.251	0.013	0.080	0.054	0.231	0.239	0.260	0.060	0.136	0.140
1024	0.061	0.178	0.242	0.004	0.061	0.076	0.179	0.185	0.250	0.061	0.106	0.121
2048	0.045	0.128	0.226	0.008	0.050	0.070	0.135	0.139	0.227	0.103	0.076	0.092
4096	0.021	0.092	0.202	-0.011	0.029	0.069	0.113	0.107	0.204	0.127	0.044	0.071
8192	0.012	0.066	0.176	-0.009	0.013	0.062	0.112	0.076	0.182	0.101	0.031	0.046
b) bias: d=0.45												
256	0.057	0.167	0.145	-0.027	-0.064	-0.139	0.131	0.218	0.230	-0.014	0.072	0.115
512	0.032	0.171	0.184	-0.013	0.025	-0.051	0.112	0.185	0.215	0.05	0.082	0.113
1024	0.030	0.139	0.193	-0.023	0.033	-0.004	0.094	0.142	0.195	0.046	0.064	0.096
2048	0.015	0.112	0.186	-0.016	0.030	0.021	0.062	0.113	0.183	0.030	0.050	0.068
4096	0.010	0.083	0.174	-0.012	0.028	0.034	0.059	0.085	0.173	0.044	0.040	0.052
8192	0.005	0.059	0.156	-0.009	0.015	0.039	0.050	0.063	0.161	0.030	0.035	0.030
c) RMSE: d=0.2												
256	0.290	0.287	0.280	0.305	0.171	0.112	0.380	0.325	0.331	0.396	0.321	0.293
512	0.202	0.231	0.258	0.236	0.119	0.087	0.345	0.263	0.268	0.322	0.197	0.190
1024	0.144	0.189	0.246	0.142	0.089	0.088	0.291	0.199	0.254	0.282	0.149	0.147
2048	0.111	0.137	0.228	0.102	0.075	0.077	0.229	0.153	0.229	0.225	0.110	0.103
4096	0.088	0.100	0.204	0.077	0.050	0.072	0.206	0.121	0.205	0.231	0.078	0.076
8192	0.067	0.072	0.177	0.063	0.033	0.067	0.205	0.086	0.183	0.212	0.061	0.056
d) RMSE: d=0.45												
256	0.319	0.251	0.219	0.273	0.258	0.216	0.303	0.245	0.255	0.362	0.272	0.212
512	0.232	0.188	0.198	0.204	0.122	0.097	0.186	0.198	0.222	0.200	0.145	0.158
1024	0.142	0.150	0.196	0.160	0.075	0.058	0.145	0.157	0.199	0.162	0.100	0.121
2048	0.081	0.120	0.190	0.097	0.059	0.040	0.111	0.121	0.185	0.120	0.072	0.089
4096	0.061	0.089	0.175	0.061	0.043	0.041	0.099	0.091	0.175	0.094	0.060	0.060
8192	0.050	0.064	0.157	0.047	0.029	0.042	0.076	0.066	0.162	0.067	0.049	0.040

Table 1.5: Bias and RMSE for a long memory process with additive noise and RLS

	p=0						p=20					
	d=0.2			d=0.45			d=0.2			d=0.45		
T\Beta	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8	0.6	0.7	0.8
a) Bias of LWLFC												
256	-0.218	-0.180	-0.172	-0.349	-0.328	-0.342	-0.248	-0.177	-0.164	-0.403	-0.366	-0.368
512	-0.182	-0.165	-0.164	-0.284	-0.296	-0.324	-0.228	-0.164	-0.164	-0.332	-0.314	-0.344
1024	-0.161	-0.146	-0.152	-0.224	-0.254	-0.299	-0.138	-0.141	-0.151	-0.257	-0.282	-0.322
2048	-0.133	-0.138	-0.145	-0.183	-0.223	-0.279	-0.096	-0.123	-0.123	-0.163	-0.182	-0.232
4096	-0.124	-0.128	-0.137	-0.142	-0.197	-0.261	-0.102	-0.110	-0.121	-0.114	-0.152	-0.215
b) RMSE of LWLFC												
256	0.278	0.213	0.186	0.436	0.355	0.355	0.462	0.285	0.212	0.601	0.445	0.398
512	0.225	0.178	0.172	0.335	0.320	0.332	0.420	0.229	0.182	0.490	0.352	0.355
1024	0.191	0.156	0.156	0.261	0.266	0.303	0.276	0.181	0.162	0.358	0.307	0.328
2048	0.150	0.144	0.147	0.208	0.230	0.281	0.190	0.132	0.125	0.256	0.205	0.238
4096	0.135	0.131	0.139	0.158	0.202	0.263	0.157	0.109	0.119	0.157	0.163	0.218
c) Bias of LWPLFC												
256	-0.420	-0.296	-0.158	-0.310	-0.241	-0.228	0.236	0.248	0.195	0.028	-0.007	0.060
512	-0.450	-0.246	-0.171	-0.219	-0.144	-0.115	0.258	0.137	0.116	0.096	0.068	0.066
1024	-0.368	-0.184	-0.107	-0.094	-0.074	-0.033	0.279	0.155	0.104	0.134	0.082	-0.006
2048	-0.232	-0.083	-0.038	-0.074	-0.050	-0.044	0.190	0.163	0.131	0.092	0.051	-0.006
4096	-0.188	-0.048	-0.030	-0.036	-0.030	-0.026	0.136	0.136	0.051	0.081	0.009	0.009
d) RMSE of LWPLFC												
256	0.645	0.551	0.370	0.524	0.433	0.361	0.559	0.582	0.522	0.456	0.454	0.392
512	0.675	0.485	0.387	0.436	0.284	0.298	0.587	0.538	0.505	0.358	0.381	0.327
1024	0.613	0.431	0.258	0.288	0.184	0.163	0.563	0.564	0.496	0.338	0.272	0.252
2048	0.490	0.309	0.131	0.195	0.144	0.123	0.544	0.470	0.457	0.234	0.208	0.151
4096	0.447	0.231	0.105	0.160	0.113	0.090	0.531	0.419	0.341	0.145	0.153	0.117

Chapter 2

Robust Memory Parameter Estimates: A Re-examination of Daily and High-Frequency Asset Returns Volatility

2.1 Introduction

Time series that exhibit hyperbolic decay of serial correlation are called "long-memory" processes. (For relevant background material concerning long-memory processes, see Baillie (1996).) The estimation and modeling of long-memory processes date from the seminal contribution of Hurst (1951) in the context of hydrology, and have attracted a substantial amount of attention from researchers since the discovery of a long-memory property in most financial time series. Long-memory processes are usually characterized in the time domain by an autocorrelation function with a diverging summation of their absolute value, and an autocorrelation decaying hyperbolically at long lags. A long-memory process is charaterized in the frequency domain by having a spectral density function proportional to λ^{-2d} as the frequency λ approaches to zero, with d being known as the "memory parameter". Such a process is stationary when $d < 0.5$, with the special case of $d = 0$ nested as a short-memory process. A process with $d > 0$ has autocorrelations that hyperbolically decay and is called a genuine long-memory process. When $d \in (0, 0.5)$, the process still resides in the stationary region, while it is no longer stationary when $d \in (0.5, 1)$. The case of $d = 1$ coincides with that of a unit root process. If $d \in (-0.5, 0)$ the process is said to be anti-persistent and its inverse autocorrelations decay hyperbolically, which is of relatively less importance in practice. As in most of the previous literature, throughout the present paper we confine our attention to the stationary long-memory process with $d \in [0, 0.5)$, although simulation evidence shows that our Local-Whittle estimators are still consistent when $d \in [0.5, 1)$.

Independently, Granger and Joyeux (1980) and Hosking (1981) introduced the fractionally integrated ARFIMA(p,d,q) process, which then served as the most frequent examples of long-memory processes in simulations because of their easy generation. The ARFIMA(p,d,q) is a long-memory generalization of the I(0) ARMA(p; q) process. Under this ARFIMA(p,d,q) framework, full-parametric estimates of d requiring the correct specification of the entire spectral density function have been proposed by Fox and Taqqu (1986) and Dahlhaus (1989), among others. Semiparametric estimates of the memory parameter have become popular, as they do not require specification of the short-memory process underlying the long-memory process. The most widely applied semi-parametric estimators are the log-periodogram (LP) estimator of Geweke and Porter-Hudak (1983) and the local Whittle estimator proposed by Kunsch (1987).

The seminal work by Perron (1989, 1990) shows that unit roots ($d = 1$) and structural changes are easily confused in the sense that the sum of the autoregressive coefficients from a stationary process are biased toward 1 if the series is contaminated by shifts in the mean (see Perron (2006) for a comprehensive survey), and also raises the possibility that long-memory may be confused with a short-memory process contaminated by level shifts, now labelled as "spurious long-memory". Similar findings are demonstrated by Bhattacharya et al. (1983), with regard to deterministic trends. Following this lead, Lobato and Savin (1998), Diebold and Inoue (2001), Granger and Hyung (2004), and Perron and Qu (2007, 2010), among others, show theoretically, empirically, and through simulations that a short-memory process contaminated by level shifts will acquire hyperbolically decaying autocorrelations similar to those which characterizes genuine long-memory, and hence confirm the concept of spurious long-memory. Monte Carlo simulation studies by Haldrup and Nielsen (2007) reveals that both short- and long-memory processes under a broad range of d values can exhibit a largely upwardly biased memory parameter estimate when the mean of a time series undergoes random level shifts, with both parametric and semi-parametric estimators affected. The same conclusion arises from another branch of the literature that considers testing for spurious long-memory against the alternative of a short-memory process contaminated by level shifts; see Ohanissian, Russell and Tsay (2008), Qu (2011), and Perron and Qu (2010) for detailed accounts. Such contaminations, among which random level shifts (abbreviated to RLS throughout this paper) are the most common, are called "low frequency contaminations" (abbreviated as LFC throughout this paper).

Whether the theoretical possibility of LFC contaminations is relevant in economic and financial time series has been explored. For example, the almost universal presence of LFC in stock market volatility data has been reported by Granger and Hyung (2004), Mikosch and Starica (2004) and Perron and Qu (2010), among others. Garcia and Perron (1996) identified the large magnitude of level shifts in U.S. real interest rate series. Applying a test against spurious long-memory, Qu (2011) rejects the null hypothesis that a U.S. inflation rate series is a stationary short- or long-memory process. Recently, Varneskov and Perron (2013) propose a new forecasting approach modelling both short- and long-memory components as well as LFC and find improvements on the forecasting results compared to previous models.

The extensive and ever-growing evidence of the LFC naturally poses new tasks for econometricians, which includes, to name a few, distinguishing true from spurious long-memory, identifying the contaminations present in a given data series, and robustly estimating the true memory parameter under such contaminations. In terms of the robust estimation of a memory parameter under LFC, several semi-parametric estimators, (e.g. Smith (2005) and McCloskey and Perron (2013)) have been proposed. Using a similar trimming technique as in McCloskey and Perron (2013), McCloskey and Hill (2013) proposed a trimmed estimator for ARMA, GARCH, and stochastic volatility models that may be contaminated by low frequency movements, while assuming the true signal process to be short-memory. McCloskey (2013) considered a trimmed frequency domain quasi-maximum likelihood (trimmed FDQML) estimator that can be used to consistently estimate the parameters of a long-memory stochastic volatility model in the presence of low frequency contamination, assuming the signal to be an ARFIMA(p,d,q) process. Iacone (2010) considered trimmed LW estimators. However, although these estimators are robust under LFC, they are silent on the possibility of other contaminations present in the data, which as well-known in the literature (e.g, Deo and Hurvich (2001), Arteche (2004, 2006), Hurvich and Ray (2003), and Hurvich et.al. (2005), among others), can also have an significant effect on the bias and efficiency of memory parameter estimates. Among those other contaminations the most important is additive noise, which is shown to be prevalent in most long-memory processes. The aforementioned LFC-robust estimators, which do not consider additive noise, tend to exclude the existence of genuine long-memory in most daily financial series that show the high persistence of autocorrelation (by attributing such persistence to the effect of LFC), a conclusion that contrasts with those derived from other estimators that are not robust under LFC but may be robust to other contaminations.

Hou and Perron (2013) propose a modified semi-parametric Local-Whittle estimator (abbreviated to LWLFC throughout this paper) and its variants that are robust under LFC and have the following advantages: they do not require the knowledge of the structure of the short-memory process; they do not require trimming, so all data are used; they do not require a Gaussian assumption on the underlying process; they have the same asymptotic variance as the standard LW estimator when no contamination is present so that without low frequency contaminations, no efficiency loss is incurred; and they can easily be extended to a full-parametric case and to perturbed estimators to model the additive noise. When low frequency contaminations are present, the LWLFC has, in most cases, the smallest bias and mean-squared error amongst all existing estimators designed to control for LFC, whether or not other types of contaminations are present.

This paper extends the modified local-Whittle estimator of Hou and Perron (2013) to a broader class of full-parametric and perturbed estimators as they suggested. Those estimators include: an LWLFC estimator, perturbed LWPLFC (P denotes perturbed) estimator, and full-parametric ARFIMA(0,d,0)+LFC, ARFIMA(1,d,1)+LFC, PARFIMA(0,d,0)+LFC, and PARFIMA(1,d,1)+LFC. We perform simulations to measure the performance of these estimators with DGP settings pertaining to those reported in empirical settings. Then we apply the LWLFC and its variants to a collection of daily and high-frequency financial time series, including volatilities of stock indices and exchange rates. The estimators proposed in the aforementioned literature, such as the classical semi-parametric LW estimator, the LWP estimator modelling for additive noise, and full-parametric ARFIMA estimators are also used for comparison. A sufficient condition for the existence of a long-memory process in the data and a mixed procedure to implement it are provided which, to the best of the author's knowledge, is among the first empirical approaches to robustly estimate memory parameters allowing for the simultaneous coexistence of short-memory process, long-memory process, as well as low frequency contaminations such as level shifts and additive noises in the data. Our findings suggest that most low frequency daily financial time series consist of both long- and short-memory processes as well as low frequencies contaminations such as random level shifts. The relative magnitude of each of these components varies substantially according to the specific type of the low frequency data. On the other hand, our results also indicate the existence of a short-memory process in some measures of volatilities constructed from high-frequency data (e.g. 30-year T-Bonds), in contrast with the view that long-memory is the dominant feature for such measures.

The structure of the paper is as follows. Section 2 introduces the model of the long-memory process under low frequency contamination, the LWLFC estimator and its variants with preliminary results, as well as a brief description of the data, simulation, and estimation strategy. Section 3 carries out simulation studies on those estimators extended from LWLFC to modelling for additive noise or to full-parametric estimators; and introduces a sufficient condition to confirm the existence of a long-memory process in the data with a procedure to implement it. Section 4 considers the empirical application of our approach to financial time series and provides a detailed discussion of the findings. Finally, brief concluding remarks are afforded in Section 5.

The following abbreviations are adopted throughout this paper: LFC denotes Low Frequency Contamination; RLS denotes Random Level Shift; LWLFC denotes Modified Local-Whittle Memory Parameter Estimator robust under LFC; LWPLFC denotes Perturbed Modified Local-Whittle Memory Parameter Estimator robust under LFC; ARFIMA(p,d,q) denotes the Full-parametric Modified Local-Whittle Memory Parameter Estimator; ARFIMA(p,d,q) +LFC denotes the Full-parametric Modified Local-Whittle Memory Parameter Estimator robust under LFC; PARFIMA(p,d,q)+LFC (or ARFIMA(p,d,q)+LFC+noise) denotes the Perturbed Full-parametric Modified Local-Whittle Memory Parameter Estimator robust under LFC and additive noise; and LWP denotes Perturbed Local-Whittle Memory Parameter Estimator as in Hurvich et. al. (2007).

2.2 The Model of Long-Memory under LFC: Motivation and Specification

2.2.1 The Data Generating Process with Low Frequency Contamination

Consider the data generating process (DGP):

$$z_t = a + y_t + u_t + w_t$$

where a is a constant, y_t is a long-memory process, u_t is a low frequency contamination (LFC), mostly in the form of random level shift (RLS), and w_t is a noise. These components are defined in the following:

Definition 3 *1) The long-memory component y_t is given by the fractionally integrated process with memory parameter d (abbreviated to FI(d)) of the form*

$$(1 - L)^d y_t = \tilde{y}_t$$

where $\tilde{y}_t$ is a short-memory process (often called the underlying short-memory process for y_t), $d \in (-1/2, 1/2)$.
2) $\tilde{y}_t$ may have short-memory dynamics itself. Particularly, y_t is called an ARFIMA (p,d,q) process if $\tilde{y}_t$ an ARMA(p,q) process, i.e. if

$$A(L)\tilde{y}_t = B(L)\varepsilon_t,\ \varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2)$$

where $A(L) = \sum_{k=0}^{p} a_k L^k$, $B(L) = \sum_{k=0}^{p} b_k L^k$, $a_0 = b_0 = 1$, and $A(L), B(L)$ have no common roots. Also, the roots of $A(L)$ and $B(L)$ are outside the unit circle.

Definition 4 *The low frequency contamination (abbreviated to LFC) component u_t is generated by one of the following processes:*
1) Random level shifts (RLS): $u_t = \Sigma_{t=1}^T \delta_{T,t}$ where $\delta_{T,t} = \pi_{T,t}\eta_t$ with $\eta_t \sim$ i.i.d. $N(0, \sigma_\eta^2)$ and $\pi_{T,t} \sim$ i.i.d. Bernoulli.$(p/T, 1)$ for some $p \geq 0$. The components $\pi_{T,t}$, η_t are mutually independent.
2) Deterministic level shifts: $u_t = \Sigma_{i=1}^B c_i \chi(T_{i-1} < t \leq T_i)$, where B is the number of regimes (so that $B-1$ is the number of breaks), $0 < |c_i| < \infty$, $\chi(\cdot)$ is the indicator function, $0 = T_0 < T_1 < ... < T_{B-1} < T_B = T$, and $T_i/T \to \tau_i \in (0,1)$ for $i = 1, ..., B-1$.
3) Deterministic trends: $u_t = h(t/T)$ where $h(\cdot)$ is a deterministic non-constant function on $[0,1]$, that is either Lipschitz continuous or monotone with $h(1) = 0$.
4) Fractional trends: $u_t = O((t+1)^{\phi-1/2})$, $u_0 = 0$, $|u_{t+1} - u_t| = O(|u_t|/t)$, where $\phi \in (-1/2, 1/2)$.

The components y_t, u_t, and w_t are assumed to be mutually independent.

Remark 7 *The semi-parametric estimators can be applied to any long-memory process defined in Definition 1-1); but Definition 1-2) has to be imposed in order to implement full-parametric estimators. However, in the latter case, the normality of ε_t is not needed for consistency.*

Remark 8 *It is important to note that the probability of a level shift in the RLS model is sample size dependent. If this were not the case, u_t would have properties similar to that of a random walk. A defining characteristic of the RLS model is that the average number of level shifts p remains constant as the sample size grows. Note that p can be zero so that the assumption nests the no level shift or no contamination case as well.*

Remark 9 *An important, or defining characteristic of a LFC component u_t lies in its spectral density function:*

$$\frac{E(I_{u,k})}{T/k^2} = O_P(1)$$

where $I_{uk} = I_u(\lambda_k) = (2\pi T)^{-1}|\Sigma_{t=1}^T u_t e^{it\lambda}|^2$ is the periodogram of u_t. For RLS components, Perron and Qu (2010) showed that

$$\frac{E(I_{u,k})}{T/k^2} \to \frac{p\sigma_\eta^2}{4\pi^3}$$

as $T \to \infty$. For other types of LFC components, Mikosch and Stărică (2004) considered the asymptotic properties of the periodogram for a deterministic level shift component when $B = 2$ (one level shift), with the addition of a short-memory component. Of interest, they showed that $E(I_{u,k}) = O_p(T/k^2)$. Kunsch (1986, Lemma 2) considered the asymptotic properties of *the periodogram of a short-memory process contaminated by a bounded monotone trend. Qu (2008, Lemma 1) extended Kunsch's results to the Lipschitz continuous case and showed that* $E(I_{u,k}) = O_p(T/k^2)$*. Iacone (2010) discussed the order of the periodogram of in the case of a fractional trend and showed that* $E(I_{u,k}) = O_p(T/k^2)$*.*

Remark 10 *Unlike short-memory dynamics or contaminating noise, which introduce only finite sample biases to the memory parameter estimator, the bias caused by LFC usually remains asymptotically, as in Hou and Perron (2013). In this sense, LFC is a more potent form of contamination than additive noise and short-memory dynamics.*

Remark 11 *In the aforementioned literature,* w_t *is often referred to as an additive noise component when the primary concern is the long-memory property of the observed time series.* $\{w_t\}$ *comes into the data series mostly as idiosycrastic innovations of a stationary short-memory process and, when it coexists with a long-memory process, is known to cause downward bias to memory parameter estimates of the long-memory process if ignored by the model (note that this is different from the short memory dynamics of the long memory process* $\tilde{y}_t$*). Such bias is higher when using a larger bandwidth. Additive noise can be captured by perturbed LW estimators as in Hurvich. et. al. (2007), but not by LP estimators. We know from Hou and Perron (2013) that although a downward bias on memory parameter estimates is universal among models not capturing noise, the LWLFC estimator has a smaller bias caused by such contamination of additive noise than both trimmed and non-trimmed LP estimators.*

It is assumed throughout this Chapter that the regularity assumptions A1~A5 in Hou and Perron (2013) hold, in order to have consistency and the usual asymptotic distribution for the modified LW estimators.

2.2.2 The Modified Local-Whittle Estimators

We suppress the existence of noise processes in the DGP, i.e. assuming $w_t = 0$, until the introduction of perturbed LW estimators and start by giving the definition of these modified LW estimators below:

Definition 5 *The LWLFC estimator proposed in Hou and Perron (2013) is the minimizer* $(\hat{d}, \hat{\theta})$ *of the following objective function:*

$$J_m(d, \theta) = \log(\frac{1}{m}\sum_{k=1}^{m}\frac{I_k}{g_k}) + \frac{1}{m}\sum_{k=1}^{m}\log(g_k)$$

where $g_k = \lambda_k^{-2d} + (\theta/T)\lambda_k^{-2}$, $I_k = I_z(\lambda_k) = (2\pi T)^{-1}|\Sigma_{t=1}^{T} z_t e^{it\lambda}|^2$ *is the periodogram of the observed data* z_t, *and* $\theta = G_u/G_0$ *is the noise-signal ratio for LFC.*

Remark 12 *LFC is controlled because the spectral density function is of order* $(G_u/T)\lambda_k^{-2}$, *where* G_u *is the magnitude of the LFC. LWLFC is semi-parametric in the sense that it assumes the spectral density function of a long-memory process can be approximated by* $G_0\lambda_k^{-2d}$, *when* λ_k *is small enough and the approximated spectral density function for* z_t *is*

$$f_k = G_0\lambda_k^{-2d} + \frac{G_u}{T}\lambda_k^{-2}$$

which is called the pseudo spectral density function of z_t.

Definition 6 *The full-parametric ARFIMA(p,d,q)+LFC LW estimator is the minimizer* $\hat{\phi} = (\hat{d}, \hat{\theta}, \{\hat{a}_k\}_{k=1}^{p}, \{\hat{b}_k\}_{k=1}^{q}, \sigma_\varepsilon^2)$ *of the following objective function:*

$$J_m(\phi) = \log(\frac{1}{m}\sum_{k=1}^{m}\frac{I_k}{g_k}) + \frac{1}{m}\sum_{k=1}^{m}\log(g_k)$$

where

$$g_k = |1 - e^{-i\lambda}|^{-2d} + \frac{\theta}{T}\lambda_k^{-2}$$

and

$$G_0 = \frac{|B(0)|^2}{|A(0)|^2}\frac{\sigma_\varepsilon^2}{2\pi}, \theta = \frac{G_u}{G_0} = G_u\frac{|A(0)|^2}{|B(0)|^2}\frac{2\pi}{\sigma_\varepsilon^2}$$

It is the frequency domain MLE if we assume that the long-memory process y_t *follows an ARFIMA(p,d,q) process:*

$$(1 - L)^d y_t = \tilde{y}_t$$

where

$$A(L)\tilde{y}_t = B(L)\varepsilon_t$$

ε_t˜*i.i.d* $N(0, \sigma_\varepsilon^2)$, *so the spectral density function of* y_t *is*

$$\frac{|B(e^{-i\lambda})|^2}{|A(e^{-i\lambda})|^2}\frac{\sigma_\varepsilon^2}{2\pi|1 - e^{-i\lambda}|^{2d}}$$

and the pseudo-spectral density function of z_t *is given by*

$$f_k = \frac{|B(e^{-i\lambda})|^2}{|A(e^{-i\lambda})|^2}\frac{\sigma_\varepsilon^2}{2\pi|1 - e^{-i\lambda}|^{2d}} + \frac{G_u}{T}\lambda_k^{-2}$$

We now address the additive noise process w_t, a component which stimulated perturbed estimators. An advantage of LW-type estimators is that, since they use the QMLF in the frequency domain, they can easily be modified to accommodate more types of structures in the DGP, without the requirement to trim some of the low frequencies. The extension to account for additive noise, known as the perturbed modified Local-Whittle estimator or LWPLFC, is introduced below. The reader is referred to Hou and Perron (2013) for more details.

Definition 7 *The LWPLFC estimator is a perturbed LWLFC estimator, where both LFC and additive noise are accounted for: we add a constant term into the spectral density function, so that the modified pseudo spectral density function is:*

$$\begin{aligned} f_k &\triangleq f_z(\lambda_k) = G_0 \lambda_k^{-2d} + G_w + G_u \frac{\lambda_k^{-2}}{T} = G_0(\lambda_k^{-2d} + \frac{G_w}{G_0} + \frac{G_u}{G_0}\frac{\lambda_k^{-2}}{T}) \\ &= G_0(\lambda_k^{-2d} + \theta_w + \frac{\theta_u}{T}\lambda_k^{-2}) = G_0 g_k \end{aligned}$$

where, with a slight abuse of notation relabeling $\theta_u = G_u/G_0$,

$$g_k = (\lambda_k^{-2d} + \theta_w + \frac{\theta_u}{T}\lambda_k^{-2})$$

and the (approximate) frequency domain QMLF is, with $\theta = (\theta_w, \theta_u)'$,

$$\varphi(G, d, \theta) = \frac{1}{m}\sum_{k=1}^{m} \varphi_k(G, d, \theta)$$

Concentrating G out of the QMLF, the estimate of G is:

$$\hat{G} = \frac{1}{m}\sum_{k=1}^{m} \frac{I_k}{g_k}$$

and the local Whittle (frequency domain QMLE) estimator under noise perturbations and low frequency contaminations, denoted as the LWPLFC estimator, is:

$$(\hat{d}_m, \hat{\theta}_m) = \arg\min_{d,\theta} J_m(d, \theta)$$

Another direction of extension is assuming a full parametric structure on the long-memory process y_t, which results in the full-parametric modified Local-Whittle estimator:

Definition 8 *The full-parametric perturbed ARFIMA(p,d,q)+LFC LW estimator is the minimizer* $\hat{\phi} = (\hat{d}, \hat{\theta} = (\hat{\theta}_w, \hat{\theta}_u)', \{\hat{a}_k\}_{k=1}^{p}, \{\hat{b}_k\}_{k=1}^{q}, \sigma_\varepsilon^2)$ *of the following objective function:*

$$J_m(\phi) = \log(\frac{1}{m}\sum_{k=1}^{m}\frac{I_k}{g_k}) + \frac{1}{m}\sum_{k=1}^{m}\log(g_k)$$

where

$$g_k = |1 - e^{-i\lambda}|^{-2d} + \frac{\theta_u}{T}\lambda_k^{-2} + \theta_w$$

with

$$G_0 = \frac{|B(0)|^2}{|A(0)|^2}\frac{\sigma_\varepsilon^2}{2\pi}, \theta_u = \frac{G_u}{G_0} = G_u\frac{|A(0)|^2}{|B(0)|^2}\frac{2\pi}{\sigma_\varepsilon^2}, \theta_w = \frac{G_w}{G_0}$$

It is the frequency domain MLE if we assume that the long-memory process y_t *follows an ARFIMA(p,d,q) process:*

$$(1 - L)^d y_t = \tilde{y}_t$$

where

$$A(L)\tilde{y}_t = B(L)\varepsilon_t$$

ε_t˜*i.i.d* $N(0, \sigma_\varepsilon^2)$*, so the spectral density function of* y_t *is*

$$\frac{|B(e^{-i\lambda})|^2}{|A(e^{-i\lambda})|^2}\frac{\sigma_\varepsilon^2}{2\pi|1 - e^{-i\lambda}|^{2d}}$$

and the pseudo-spectral density function of z_t *is given by*

$$f_k = \frac{|B(e^{-i\lambda})|^2}{|A(e^{-i\lambda})|^2}\frac{\sigma_\varepsilon^2}{2\pi|1 - e^{-i\lambda}|^{2d}} + \frac{G_u}{T}\lambda_k^{-2} + G_w$$

Remark 13 *The full-parametric perturbed LW estimators are constructed as full-parametric LW estimators with a perturbation (noise) parameter. By modeling additive noise, the variance of those estimators are all proportional to the reciprocal of the true memory parameter in the long-memory process (see Hurvich et.,al. (2005) for details). Hence, care should be exercised when interpreting the numerical results from these types of estimators, and cross-checking with results from other estimators should be performed to determine the most likely true DGP of the data to avoid applying the perturbed estimators to DGPs with a true memory parameter close to zero.*

Remark 14 *An effective technique for distinguishing a process with both true long-memory and short-memory components from a true pure short-memory process is to first apply non-perturbed LFC-robust LW estimators, such as LWLFC, to get a lower bound of the true memory parameter since it is not upward-biased by LFC (and short memory dynamics with proper choice* *of bandwidth), but may be downwardly-biased by additive noise. If the LWLFC estimate of the memory parameter is above zero, then the true memory parameter of the long-memory process is greater than zero, so we can apply the perturbed LW estimators legitimately. However, an LWLFC estimate close to zero does not guarantee a pure short-memory process, since it may be the result of a large additive noise component. An important future research objective is to develop strategies applicable for this case.*

Remark 15 *For the bandwidth choice, as pointed out in Hou and Perron (2013), when there is additive noise, the performance of LWLFC is better when using a small bandwidth, while the performance of LWPLFC is better when using a large bandwidth. For the full-parametric estimators, ideally, the choice of bandwidth should not be a problem, and all frequencies should be taken into consideration; but when there is noise in the data, the QML function becomes flat at higher frequencies and hence the variance becomes large. To mitigate this, we report the truncated full-parametric estimation using truncation* $m = T^{0.8}$*, a bandwidth with good performance according to simulations.*

Remark 16 *Semi-parametric and full-parametric estimators have their own advantages and drawbacks. For semi-parametric estimators, the DGP is not required to follow a specific distribution; for LW based estimators even Gaussianity of the underlying process is not required, while LP based estimators do need Gaussianity. However, semi-parametric estimators uses an approximation of the spectral density function of the observed data when the frequency is close to zero, and hence cannot utilize the full range of frequencies. By using an approximation of the true spectral density function they will be less effective than full-parametric estimators if the latter are applied to a correctly specified DGP. Full-parametric estimators generally gives better results than semi-parametric estimators when applied to a correctly specified DGP, but may give biased and inconsistent estimates if the model is not correctly specified.*

2.2.3 The Data

The data collected in this paper consists of ten financial time series that have been identified by aforementioned researchers as either arising from or exhibiting properties of long-memory processes. [1] The first four time series we study are those examined by Lu and Perron (2010) and McCloskey and Perron (2011): log-squared daily returns series of the 1) S&P 500, 2) Dow Jones Industrial Average (DJIA), 3) NASDAQ, and 4) AMEX stock market indices. The S&P 500 series analyzed here starts on July 3, 1962 and ends on March 25, 2004, with T = 10504

[1] We are grateful to Adam McCloskey and Rasmus T. Varneskov for providing the first four and last six financial time data series, respectively.

observations; the DJIA series starts on March 4, 1957 and ends on October 30, 2002, with T = 11534 observations; the NASDAQ series starts on December 15, 1972 and ends on December 31, 2006, with T = 8592 observations; the AMEX series starts on July 3, 1962 and ends on December 31, 2006, with T = 11201 observations. As discussed in Lu and Perron (2010), log-squared returns are a common measure of volatility, and appear to resemble that of a long-memory process (Mikosch and Stărică (2004)), but at the mean time can also explained as an i.i.d. series affected by occasional shift in the unconditional variance (Stărică and Granger (2005)). When returns are zero or close to it, the log-absolute value transformation implies extreme negative values. To avoid this problem, we bound absolute returns away from zero by adding a small constant so-called offset parameter, a technique introduced to the stochastic volatility literature by Fuller (1996). Lu and Perron (2010) show that the results of this approach are robust to alternative specifications, for example using another value for offset parameter, deleting the zero observations, or replacing them by a small value.

The other six time series are those examined by Varneskov and Perron (2013): 5) realized volatility series constructed from tick-by-tick trades sampled on the SPY from January 1997 through July 2008, with T = 2914 observations (trading days); two series of realized volatility estimates for the period of 1982 to 2007 using five-minute returns on 6) S&P 500 with T = 6262 observations and 7) T-bond futures during trading hours with T = 5069 observations, respectively, after deleting missing entries; daily returns on the 7) Dollar-Aus and 8) Dollar-Yen exchange rates from January 4, 1971 to April 10, 2009, both with T = 9600 observations; and 10) S&P 500 daily returns dating from 1929 to 2004, with T = 20327 observations. As pointed out in Varneskov and Perron (2013), HF estimates of volatility are unbiased and highly efficient proxies of return volatility, thus permitting greater statistical precision. The reader is referred to Varneskov and Perron (2013) for details about construction of data series. For similar reasons mentioned in the construction of the first four time series, we remain concerned with the same logarithmic transformations of these last six measures as well.

Note that all these empirical time series are much longer than those studied in simulations. As a result, we expect bias and RMSE measures to be lower than those indicated by simulations, provided that the models adopted are consistent with true DGPs of the empirical data.

2.2.4 Simulation and Estimation Strategy

We briefly describe the simulation and estimation strategy adopted throughout this paper. In simulation we focus on a performance analysis of estimators under misspecification. For non-perturbed estimators, we study their performance under additive noise; for semi-parametric estimators, we study their performance under short-memory dynamics. We are particularly interested in the performance of estimators under DGPs with large additive noise, with close to zero or large memory parameters, since these features are pertinent to empirical data. We also study the general performance of full-parametric estimators, which is not covered by Hou and Perron (2012).

In empirical analysis, we first apply non-perturbed estimators and investigate their performance under different bandwidths. After finding evidence that the true memory parameter is above zero and is sufficiently large, we use perturbed estimators and compare the difference between results. This approach is defined as a mixed procedure, for reasons that will become apparent in Section 2.3.

2.3 Simulation Studies of Various Estimation Methods under Different Contaminations Magnitudes

In this Section we use simulations to illustrate the performance and effectiveness of various LW based estimators under a variety of different DGPs for which our robust estimators were designed. The parameter settings of simulations are adjusted to imitate those of particular interest in the current literature. The primitive shocks ε_t used to generate the long-memory process y_t are i.i.d Normal, and always has its variance normalized to unity: $var(\varepsilon_t) = 1$. The long-memory process y_t is generated using a Gaussian fractional ARFIMA(p,d,q) process with innovation variance $var(\varepsilon_t)$ set to unity as above, and reduces to a simple Gaussian fractional long-memory process when $p = q = 0$; y_t is said to be contaminated by short-memory dynamics if p or q are not equal to zero. The LFC u_t is exclusively considered as RLS for its proven prevalence in most financial time series in the contemporary literature. We report results for $d = 0$ as y_t being a short-memory process, 0.45 as a stationary long-memory process, and 0.7 as a nonstationary long-memory process. The last setting of d is motivated by the empirical findings, in which some perturbed estimators have a d estimate larger than 0.5 in certain data series.

The average number of level shifts per sample is set to $p = 10, 20$ and the distribution of the shifts is $\eta_t\tilde{}i.i.d.N(0, \sigma_\varepsilon^2)$, so $\theta_u = 2p(\sigma_\eta^2/\sigma_\varepsilon^2)$ is the noise-signal ratio for LFC. This setting is adopted by most authors studying RLS. Throughout the simulations, w_t is a white noise: $w_t\tilde{}i.i.d.N(0, \sigma_w^2)$, with variance σ_w^2 so the noise-signal ratio of additive noise is $\theta_w = \sigma_w^2/\sigma_\varepsilon^2$, and the noise-signal ratio θ_w is set to be either moderate ($\theta_w = 1, 4$) or large ($\theta_w = 30$), simulating series either composed of long-memory process and small additive noise, or series composed of a large noise and a moderate long-memory process.

The remaining parameter settings of the simulations are: bandwidth $\beta = 0.8, 0.7, 0.6, 0.5$; total observation T=3000, 5000; and number of replications N=500. Although θ_u cannot be consistently estimated when LFC exists (see Hou and Perron, (2013)), its value will be recorded on occasions where it may be of interest. It will be shown later that although $\hat{\theta}_u$ will converge to zero when there is no LFC, having a large $\hat{\theta}_u$ cannot be presented as evidence supporting the existence of LFC.

2.3.1 Performance of Semi-Parametric LW Estimators

Performance of LWLFC under additive noise, T=5000

We begin our simulations by studying the performance of LWLFC estimators with varying bandwidth. The DGP used here is the one with long-memory and additive noise:

$$z_t = y_t + u_t + w_t$$

where z_t is the observed data, $y_t\tilde{}$ARFIMA(0,d,0), u_t is a RLS process with an average number of shifts $p = 10$ and variance of shift σ_η^2 so $\theta_u = 2p(\sigma_\eta^2/\sigma_\varepsilon^2)$, and w_t is white noise with variance

σ_w^2 so $\theta_w = \sigma_w^2/\sigma_\varepsilon^2$. The results are recorded in Tables 2.1-2.3.

Tables 2.1-2.3 illustrate a substantial downward bias that even a moderate additive noise can cause in an LWLFC estimator. Even with a large number of observations (T = 5000), the downward bias is still strong enough to induce an erroneous identification of the process of interest as a short-memory process contaminated by LFC. Bias increases with the noise-signal ratio $\theta_w = \sigma_w^2/\sigma_\varepsilon^2$, although it still remains sizable when θ_w is small. Bias also increases with the bandwidth, an unwelcome feature that inhibits maximum usage of data. Nevertheless, LWLFC is still the estimator with lowest downward bias among all LFC-robust estimators as mentioned in Hou and Perron (2013).

Performance of LWLFC under short-memory dynamics, T=5000

Next we study LWLFC and LWPLFC under short-memory dynamics. The DGP used here is without additive noise:

$$z_t = y_t + u_t$$

where z_t is the observed data, $y_t\tilde{}$.ARFIMA(1,d,1), with a being the AR(1) coefficient and b being the MA(1) coefficient, u_t is an RLS process with an average number of shifts $p = 10$ and variance of shift σ_η^2 so $\theta_u = 2p(\sigma_\eta^2/\sigma_\varepsilon^2)$. The results are listed in Tables 2.4-2.5.

Upon examination of Tables 4-5, it is clear that a positive AR(1) coefficient in short-memory dynamics can cause a high upward bias when bandwidth is large, but its impact decays rapidly when bandwidth decreases. A positive MA(1) coefficient has almost no observable effect on LWLFC; but a negative MA(1) coefficient can cause a downward bias similarly to how additive noise does, and its impact also decays with a decreasing bandwidth, although the rate of decrease is slower than that with a positive AR(1) coefficient. An ARFIMA(1,d,1) process with both a positive AR(1) and a MA(1) coefficients will show primarily AR(1) feature and causes an upward bias. An ARFIMA(1,d,1) process with a positive AR(1) and a negative MA(1) coefficients will have the biases caused by its AR(1) and MA(1) coefficients cancel out each other to make the final magnitude of bias smaller than any of the two biases alone. As a special case, when AR(1)=-MA(1), these biases completely neutralize each other and an ARFIMA(1,d,1) process simplifies to an ARFIMA(0,d,0) process, which is a trival case that is not reported in the simulations. A DGP with a lower true memory parameter is more strongly biased by a MA, and as a result occasionally a negative value of the memory parameter estimates is recorded. Since applying LWLFC with a small bandwidth efficiently eliminates upward bias on the memory parameter caused by short memory dynamics as well as LFC, LWLFC with a small bandwidth gives a credible lower bound of the memory parameter.

2.3.2 A Sufficient Condition to Identify Long-memory: The Mixed Procedure

As pointed out by the literature cited above, the two primary contaminations that can cause substantial bias in memory parameter estimation in empirical studies are additive noise, which causes negative bias; and LFC, which causes positive bias. When additive noise is of concern, non-perturbed estimators are shown to have a large negative bias whether they are robust to LFC or not. On the other hand, although perturbed estimators are robust to additive noise in a genuine long-memory environment, their variance inflates to infinity as the memory parameter of the long-memory process approaches zero. These complications raise the natural question of whether there is a criterion that can robustly identify long-memory processes. In what follows we provide the first estimation strategy to address these two types of contaminations simultaneously and to establish a sufficient (but not necessary) condition to positively identify the existence of a long-memory process, even under large additive noise and LFC.

Recall that the DGP used here is

$$z_t = y_t + u_t + w_t$$

where z_t is the observed data, y_t˜ARFIMA(0,d,0), u_t is a RLS process at an average of shifts $p = 10$ and variance of shift σ_η^2 so $\theta_u = 2p(\sigma_\eta^2/\sigma_\varepsilon^2)$, and w_t is white noise with variance σ_w^2 so $\theta_w = \sigma_w^2/\sigma_\varepsilon^2$.

To illustrate the method, we first need a clear picture of the biases of LWLFC and LWPLFC under different settings, which is listed in Tables 2.6-2.7.

Table 2.6 provides estimates of d and θ_u for LWLFC and LWPLFC under large noise. As noted in the beginning of the simulations, here we report the non-consistent estimator $\hat{\theta}_u$, which is a random variable, to offer some insight. An interesting fact is that when the true DGP is the combination of a long-memory process (usually nonstationary as well) and a additive noise, the LWLFC estimator, without additive noise modeled, while giving a downwardly biased estimate of the memory parameter, will mostly give an upwardly biased estimate of θ_u. Because $\hat{\theta}_u$ will converge to zero when there is no LFC, a large $\hat{\theta}_u$ may induce a practitioner to falsely conclude evidence of LFC in the data. The intuitive explanation for this is that the objective function, failing to be consistent with true DGP, will try increasing $\hat{\theta}_u$ to "compensate" the bias on the memory parameter estimate. In this sense LWLFC tends to confuse a long-memory process component and an LFC component, and when only a long-memory process but no LFC exists, generates a memory parameter estimate value in between the long and short-memory parameters, while suggesting the false existence of LFC. This is a feature that we suspect is shared by all estimators that are robust under LFC but not under additive noise. On the other hand, the perturbed LWPLFC estimator removes the negative bias of d caused by noise, but still has large estimates of θ_u, indicating that LWPLFC may also lead to a false-positive existence of LFC despite its robustness under both noise and LFC. As a consequence, we recommend that in practice the large value of $\hat{\theta}_u$ cannot be used as evidence confirming the existence of LFC.

Table 2.7 reveals the degree of RMSE explosiveness when LWPLFC is applied to a DGP without long-memory. Interestingly, the bias of d is always positive and $\hat{\theta}_u$ is arbitrarily large. Such huge bias and RMSE of LWPLFC under a short-memory process casts a shadow on any results with a large $\hat{d}$, since it may be either caused by a genuine long-memory contaminated by additive noise, or by a genuine short-memory process without additive noise.

However, using a procedure that combines LWLFC and LWPLFC can establish a sufficient condition for the existence of a long-memory process and give a robust estimate of the memory parameter of the long-memory process once that condition is satisfied. In this procedure, the practitioner first applies LWLFC to the data series of interest, and as discussed above, receives a lower bound of d. If this value is sufficiently greater than zero, one can confidently conclude that the true d lies above zero, and hence validate an application of the LWPLFC estimator that will eliminate a negative bias on $\hat{d}$ of LWLFC and result in a robust estimate of d. We refer to this procedure as a "mixed procedure", and exploit it extensively in the empirical studies.

It is also of interest to explore the order of estimates given by LW, LWP, LWLFC, LWPLFC, and trimmed LP estimators applied on the same data. The classical LW estimator is subject to two types of biases: the upward bias from LFC and the downward bias from additive noise. LWP is subject to upward bias from LFC but not subject to downward bias from additive noise. On the other hand, estimators only controlling for LFC (trimmed LP, LWLFC) are not subject to an upward bias from LFC but to downward bias from additive noise. So in a DGP where both LFC and additive noise exist, the value of d estimates should be in the following order:

$$\{\hat{d}_{LWLFC}, \hat{d}_{trimmedLP}\} < \{\hat{d}_{LW}, \hat{d}_{LWPLFC}\} < \hat{d}_{LWP}$$

where estimates using estimators in the same group may have any order. From this comes the well known fact that classical LW or LP estimators that do not consider any contamination have larger d estimates than LWLFC and trimmed LP estimates. The perturbed estimators not robust under LFC, which are not biased downwardly by noise but biased upwardly by LFC, should always have the largest estimate of d.

Remark 17 *The mixed procedure can positively confirm the existence of long-memory. However, it cannot be employed to exclude the existence of long-memory in data. As of now, there is no known credible strategy that can generate a sufficient condition for absence of long-memory. Research on this problem may prove promising in the future.*

2.3.3 Performance of Full-parametric LWLFC Estimators

We now turn our sights to the parametric LWLFC estimator ARFIMA(p,d,q)+LFC. Our primary interest is the case in which $p = q = 0$ and $p = q = 1$, as simulations (not listed) having too many ARMA coefficients will greatly increase the time of computation, as well as the difficulty of finding the true global maximum of QMLF and the RMSE of all estimates. For

ARMA coefficients, we choose the most appropriate settings that may occur in our empirical data: when there is no short-memory dynamic ($a = 0$, $b = 0$); and when there is positive autocorrelation and a negative moving average ($a = 0.3$, $b = -0.7$). The memory parameter may take value of 0, 0.45, or 0.7, representing a short-memory process, and a stationary or nonstationary long-memory process, respectively. Departing from the previous analysis, we fix a bandwidth at $\beta = 0.8$ in all full-parametric simulations because 1) β is not required to be smaller than one since the true form of the QML function is adopted; 2) if the full range of frequencies is used ($\beta = 1$) the QML function becomes flat at higher frequencies and the estimate loses accuracy (as confirmed by unreported simulations).

The DGP used here is

$$z_t = y_t + u_t + w_t$$

where z_t is the observed data, y_t˜ARFIMA(0,d,0), u_t is a RLS process with an average number of shifts $p = 10$ and variance of shift σ_η^2 so $\theta_u = 2p(\sigma_\eta^2/\sigma_\varepsilon^2)$, and w_t is white noise with variance σ_w^2 so $\theta_w = \sigma_w^2/\sigma_\varepsilon^2$. Simulation results are presented in Tables 2.8-2.9 without noise and in Table 2.10 with noise.

The parameter values used in Tables 8-9 are: 1) $\sigma_\eta^2/\sigma_\varepsilon^2 = 1$, $p = 10$, $\theta_u = 20$, $\theta_w = 0$. 2) $\sigma_\eta^2/\sigma_\varepsilon^2 = 1$, $p = 10$, $\theta_u = 20$, $\theta_w = 0$. These two tables show that the ARFIMA+LFC estimators generally have good accuracy about the memory parameter, but large RMSE on the ARMA coefficients. Note that even without additive noise, $\hat{\theta}_u$ is still highly volatile.

Table 2.10, with parameters set to $\sigma_\eta^2/\sigma_\varepsilon^2 = 1$, $p = 10$, $\theta_u = 20$, $\theta_w = 10$, explores the performance of ARFIMA+LFC estimators under additive noise and attempts to check if an ARFIMA(1,d,1)+LFC model can absorb additive noise efficiently by adjusting its MA(1) coefficient. We can see that although an ARFIMA(1,d,1)+LFC model has lower downward bias on memory parameter estimates than an ARFIMA(0,d,0)+LFC model, it is still downwardly biased. So its MA(1) parameter absorbs a part, but not all, of the additive noise. The estimate of $\hat{\theta}_u$ is still inflated by additive noise, just as in the case of semi-parametric estimators with inaccuracy growing with d, further indicating that $\hat{\theta}_u$ cannot be used to confirm the existence of LFC using full-parametric estimators.

2.3.4 Performance of Full-parametric Perturbed LWLFC estimators

For the reason stated earlier, throughout this subsection we retain the setting of a fixed bandwidth $\beta = 0.8$ and sample size T=5000.

We now study the most complicated estimator: full-parametric perturbed LW estimators PARFIMA(0,d,0)+LFC. In such models, an important question arises: whether the benefit from the additional fit between a complex model and true DGP can compensate for the increased RMSE caused by the greater number of parameters introduced. The DGP used here is the same as that for the non-perturbed ARFIMA+LFC estimators:

$$z_t = y_t + u_t + w_t$$

where z_t is the observed data, y_t˜ARFIMA(0,d,0), u_t is a RLS process at an average of shifts $p = 10$ and variance of shift σ_η^2 so $\theta_u = 2p(\sigma_\eta^2/\sigma_\varepsilon^2)$, and w_t is white noise with variance σ_w^2 so $\theta_w = \sigma_w^2/\sigma_\varepsilon^2$. Results are shown in Tables 2.11-2.12 for PARFIMA(0,d,0)+LFC, and in Tables 2.13-2.19 for PARFIMA(1,d,1)+LFC.

The performance of full-parametric perturbed LW estimators PARFIMA(0,d,0)+LFC (Tables 2.11-2.12) reveals that when the true memory parameter is far from zero, even for data with large additive noise, perturbed full-parametric LW estimators can still capture the long-memory process with decent accuracy, but without good accuracy on parameters other than the memory parameter. However, as predicted by the theory of perturbed estimators, when the true memory parameter is zero, the bias and RMSE of perturbed estimates are large and the PARFIMA(0,d,0)+LFC estimator loses accuracy.

The performance of PARFIMA(1,d,1)+LFC (Tables 2.13-2.19) indicates that the bias and RMSE of our PARFIMA(1,d,1)+LFC estimator is comparable with the trimmed ARFIMA(1,d,0) estimator in McCloskey (2013), with our estimator having less bias, less RMSE for d, and a larger RMSE for the ARMA coefficient. Also there is trade-off effect between MA(1) and additive noise, since they both have large RMSE. Again, as predicted by the theory of perturbed estimators, when the true memory parameter is zero, the bias and RMSE of perturbed estimators inflate immensely and the PARFIMA(1,d,1)+LFC estimator loses accuracy.

2.4 Empirical Analysis

In this Section, we apply our estimators to various time series that have been identified in the literature as either arising from or exhibiting properties of long-memory processes. We examine the extent to which RLS or other LFC components may be playing a role in influencing the long-memory character of the data, and biasing the memory parameter estimate toward the conclusion that a genuine long-memory process is coexisting with them. We also investigate the variation of estimates according to bandwidth choice and report the random variable $\hat{\theta}_u$ for each time series under study, with the hope of deriving more insights about the most likely types of DGPs underlying the data. All statistics reported as zero are accurate to absolute values within 10^{-3}. It must be pointed out that the true DGP of the long or short-memory process may not follow an ARFIMA(p,d,q), so all the full-parametric estimators may be misspecified. For each data series, two tables are provided to present results from semi-parametric and full-parametric estimators, respectively. We also report LWLFC with more bandwidths in the last column of the table of semi-parametric estimators, to give more insight on the sensitivity of the d estimate to bandwidth choice, in order to investigate whether the mixed procedure is applicable.

The first data series we analyze is the S&P 500 daily volatility series (1962-2009), with the results listed in Tables 2.20-2.21. The classical LW estimator gives a memory parameter estimate from 0.261 to 0.495, increasing as bandwidth decreases, implying a long-memory process. The significant changes on parameter estimates with bandwidth indicates that the true DGP is not a pure long-memory process, and can be explained both as (1) the true DGP is a long-memory process, possibly accompanied by a noise process and contaminated by LFC; and (2) the true DGP consists of the noise and the RLS, and possibly some short-memory process. The LWP estimator has higher estimates of d, confirming the existence of a noise but, as we appeal to the case discussed earlier, excludes neither possibility (1) nor (2), since in condition (2) the LWP estimator loses its accuracy when applied to a pure short-memory process.

The LFC-robust estimators, such as LWLFC and trimmed LP, on the other hand, confirmed the existence of LFC by giving much lower memory parameter estimates than those using classical LW and LP estimators. It remains to be seen whether a true long-memory process exists or not. The LWLFC estimates increase as bandwidth decreases, which is consistent with the pattern of LFC-robust d estimates of a long-memory process contaminated by noise. And a d estimate of 0.183, 0.212, 0.237 when $\beta = 0.75, 0.73, 0.7$, respectively, is large enough to positively identify a pure long-memory process and invoke the mixed procedure. While we admit that at the highest bandwidth our LWLFC estimates (0.095 when $\beta = 0.8$), albeit still positive and higher than that of trimmed LP from McCloskey and Perron (2013) (-0.017 when $\beta = 0.8$), are not far enough from zero, we argue that this is most likely caused by the existence of noise, since the d estimate increases considerably with bandwidth.

Turning to full-parametric estimators, the ARFIMA(0,d,0)+LFC estimator produces estimates reminiscent of those from LWLFC when $\beta = 0.8$. The ARFIMA(1,d,1)+LFC estimator gives a high memory parameter estimate and strong mean reverting coefficient, suggesting the existence of a large additive noise and a long-memory process.

The perturbed models, both LWPLFC and full-parametric ones, all give similar results. As was noted in Sections 2 and 3, because noise in the data causes the frequency-domain QMLF to become flat at higher frequencies, using a bandwidth of $\beta = 0.8$ is better than using a full range of frequencies. These high d estimates ($\hat{d}_{LWPLFC} = 0.655$ when $\beta = 0.8$), larger than all that derived using LWLFC and less sensitive to bandwidth, are consistent with a long-memory process contaminated by a noise process.

We conclude that our findings for these data, besides confirming the existence of noise and a LFC, which is consistent with the previous literature, also suggests the evidence of a long-memory process. Although the LWLFC estimate of d at $\beta = 0.8$ is not far enough from zero, the fact that the LW and LWLFC d estimates increases rapidly as bandwidth decreases and that ARFIMA(1,d,1)+LFC yields a high memory parameter estimate and strong negative MA(1) coefficient indicates that the true d is indeed large enough to employ the mixed strategy.

At every bandwidth the trimmed LP gives a lower d estimate than the LWLFC does, reiterating the theoretical explanation in (Hou and Perron (2013) that under contamination of additive noise, the trimming technique aggravates negative bias by excluding low frequencies (at which the spectral density of a long-memory process with higher memory differs most from a short-memory process,) from data.

Most of the daily data studied- i.e. daily returns series of DJIA in Tables 2.22-2.23, NASDAQ in Tables 2.24-2.25, AMEX in Tables 2.26-2.27, another S&P 500 daily volatility series in Tables 2.30-2.31, and daily data on the Dollar-Yen exchange rate in Tables 2.34-2.35- show properties reminiscent to various degrees of those from the first data of the daily S&P 500. These properties are: (1) those d estimates that are not robust to LFC suggest a long-memory and even a nonstationary process; (2) those d estimates that are robust to LFC but not additive noise (such as trimmed LP) find no evidence of long-memory memory but strong existence of LFC; (3) those d estimates of non-perturbed Local-Whittle LFC-robust estimators are positive, below those d estimates not robust to LFC and above those previous d estimates robust to LFC, and although they may not be large enough at the largest bandwidth considered, increase rapidly as bandwidth decreases; (4) perturbed LFC-robust estimators give large d estimates; (5) the ARFIMA(1,d,1)+LFC estimator gives negative MA(1) coefficient estimates. These properties lead us to the conclusion that aside from LFC and noise previously identified in the data, a long-memory component resides in them as well.

Unlike the other daily series, daily data on the Dollar-Aus exchange rate follows a different pattern. Its LWLFC d estimate remains relatively steady across bandwidths above 0.7 and the MA(1) coefficient is positive and close to zero, implying the possibility of a true memory parameter close to zero and casting doubts on the credibility of perturbed estimators. This is a case for which the mixed procedure cannot be reliably applied. We conclude that the data reflect a short-memory process and LFC, possibly mixed with a weak long-memory process.

Presented in Tables 2.28-2.39 are the results for high-frequency data series, including HF data on SPY in Tables 2.28-2.29, HF data on the S&P 500 in Tables 2.36-2.37 and HF data on 30-Year T-Bonds are similar among themselves, but are quite different than those for daily data, with strong evidence of long-memory. All estimators have relatively stable d estimates across all bandwidths. Classic LW and LWP estimators do have higher estimates of d due to bias caused by LFC. For HF data on SPY and the S&P 500, perturbed estimators, validated by large d estimates from non-perturbed estimators, give d estimates not significantly larger than their non-perturbed counterparts and small estimates of θ_w, indicating that the presence of noise is not substantial using high-frequency data. Their true DGPs are best construed as a long-memory process with LFC. The case of HF data on 30-year T-Bonds is slightly different from the other two high-frequency data series, in that it shows the presence of additive noise supported by the facts the that d estimates of LWLFC increases as bandwidth decreases, that significant increases of d estimates are given by perturbed estimators, and that a negative MA(1) coefficient estimate is given by ARFIMA(1,d,1)+LFC. This shows that although additive noise is considered to be insubstantial in most high-frequency based series, it cannot be completely excluded from consideration.

2.5 Conclusions

In this paper, we provide the first empirical approach for robustly estimating the memory parameters of data series that allows for coexistence of both short-memory process and long-memory process, low frequency contaminations such as level shifts as well as additive noises. We provide a sufficient condition for the existence of long-memory and propose a mixed procedure that combines a modified Local-Whittle estimator and its perturbed and full-parametric variants to verify that sufficient condition in practice. We apply our estimation methodology to daily returns based volatilities on S&P 500, DJIA, NASDAQ, AMEX, Dollar-AUS and Dollar-Yen exchange rates, together with high-frequency based volatility for the bond and stock market data. These series have been the focus of research using a variety of methods, due to the fact that they exhibit evidence of long-memory that however also appears to be spurious if contaminations are taken into account. Our estimation results reveal that: 1) level shifts are indeed present in all series; 2) a genuine long-memory component is present in high-frequency based measures of volatility, where little evidence of noise is present, except for the realized volatility series of 30-year Treasury Bond futures; and 3) that noise is present in all daily data. These findings agree with the contemporary literature that studies LFC and spurious long-memory, such as McCloskey and Perron (2013). Through our mixed procedure, we contribute to the literature by finding evidence of long-memory processes in most low frequency daily measures, suggesting a combination of a long-memory processes, a noise, as well as a LFC in such data, with the relative magnitude of each of these components varying according to the specific series. We also perform simulations to show the finite sample properties of several modified LFC-robust LW estimators, including several perturbed and full-parametric estimators.

There appear to be many promising avenues for future research. The mixed procedure, although it can verify the sufficient condition for evidence of long-memory, loses its power when the lower bound of the d estimates given by non-perturbed estimators approach zero. Methods to identify a long-memory process in a DGP with a large noise component are desired. Estimation strategies that can model the correlation between different components in the same data should provide more reliable empirical results than current methods assuming the mutual independence of different components.

Table 2.1: LWLFC under large noise, d=0.45, T=5000

LWLFC, $\theta_u = 20$, d=0.45						
	$\theta_w = 5$		$\theta_w = 10$		$\theta_w = 20$	
β	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.8	-0.291	0.292	-0.345	0.346	-0.382	0.383
0.75	-0.263	0.265	-0.319	0.321	-0.368	0.370
0.7	-0.226	0.232	-0.295	0.298	-0.343	0.346
0.65	-0.188	0.200	-0.258	0.263	-0.323	0.328
0.6	-0.154	0.177	-0.245	0.258	-0.296	0.314
0.55	-0.150	0.194	-0.205	0.241	-0.277	0.301

Table 2.2: LWLFC under large noise, d=0.7, T=5000

LWLFC, $\theta_u = 20$, d=0.7						
	$\theta_w = 5$		$\theta_w = 10$		$\theta_w = 20$	
β	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.8	-0.467	0.469	-0.552	0.554	-0.599	0.600
0.75	-0.389	0.395	-0.479	0.483	-0.549	0.551
0.7	-0.300	0.314	-0.415	0.424	-0.496	0.501
0.65	-0.211	0.241	-0.321	0.338	-0.454	0.477
0.6	-0.167	0.236	-0.214	0.259	-0.408	0.451
0.55	-0.142	0.269	-0.183	0.255	-0.296	0.383

Table 2.3: LWLFC on short-memory process with short-memory dynamics, T=5000

$z_t = y_t + u_t$, LWLFC, $\theta_u = 20$, d=0							
	$a = 0.6, b = 0$		$a = 0, b = -0.6$		$a = 0, b = 0.6$	$a = 0.6, b = 0.6$	
β	Bias	RMSE	Bias	RMSE	RMSE	Bias	RMSE
0.8	0.318	0.320	-0.496	0.500	0.063	0.341	0.342
0.7	0.131	0.137	-0.246	0.282	0.052	0.129	0.134
0.6	0.044	0.112	-0.186	0.459	0.178	0.032	0.091

Table 2.4: LWLFC on long-memory process with short-memory dynamics, T=5000

$z_t = y_t + u_t$, LWLFC, $\theta_u = 20$, d=0.45								
	$a = 0.6, b = 0$		$a = 0, b = -0.6$		$a = 0, b = 0.6$		$a = 0.6, b = 0.6$	
β	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.8	0.256	0.257	-0.484	0.490	0.058	0.063	0.341	0.342
0.7	0.095	0.100	-0.202	0.222	0.013	0.052	0.129	0.134
0.6	0.021	0.059	-0.080	0.193	-0.016	0.178	0.032	0.091

Table 2.5: LWLFC on nonstationary long-memory process with short-memory dynamics, T=5000

$z_t = y_t + u_t$, LWLFC, $\theta_u = 20$, d=0.7								
	$a = 0.6, b = 0$		$a = 0, b = -0.6$		$a = 0, b = 0.6$		$a = 0.6, b = 0.6$	
β	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.8	0.248	0.249	-0.577	0.581	0.039	0.045	0.281	0.281
0.7	0.104	0.109	-0.205	0.243	-0.006	0.055	0.109	0.114
0.6	-0.010	0.115	-0.058	0.146	-0.043	0.161	0.002	0.096

Table 2.6: LWLFC, LWPLFC on a stationary long-memory process with large noise

$z_t = y_t + u_t + w_t$, $d_0 = 0.7$, $\theta_w = 30$, $\theta_u = 0$								
	LWLFC				LWPLFC			
	mean of $\hat{\theta}_u$		Bias of d		mean of $\hat{\theta}_u$		Bias of d	
β/T	3000	5000	3000	5000	3000	5000	3000	5000
0.8	11.2	17.0	-0.624	-0.617	22.9	18.0	-0.007	-0.001
0.75	10.6	16.5	-0.604	-0.580	18.3	18.6	0.005	-0.007
0.7	9.61	14.89	-0.550	-0.53	20.6	21.0	-0.005	0.006
0.65	9.86	16.2	-0.524	-0.479	22.9	20.9	-0.008	-0.03
0.6	11.1	16.98	-0.459	-0.417	18.7	24.2	-0.017	-0.031

Table 2.7: LWLFC, LWPLFC with pure short-memory process+RLS

$z_t = y_t + u_t$, $d_0 = 0, \sigma_e^2/\sigma_u^2 = 1, p = 10, \theta_u = 20$								
	LWLFC				LWPLFC			
	mean of $\hat{\theta}_u$		Bias of d		mean of $\hat{\theta}_u$		Bias of d	
β/T	3000	5000	3000	5000	3000	5000	3000	5000
0.8	76.8	44.8	-0.076	-0.043	1.9E9	4.7E5	0.722	0.745
0.75	34.3	40.9	-0.097	-0.055	1.4E9	1.6E6	0.604	0.664
0.7	27.4	40.9	-0.116	-0.067	1.7E6	3.6E8	0.666	0.657
0.65	23.7	38.4	-0.081	-0.068	5.0E6	1.1E10	0.711	0.729
0.6	33.9	18.1	-0.094	-0.091	2.7E6	6.2E5	0.644	0.721

Table 2.8: Full Parametric ARFIMA(0,d,0)+LFC

$\beta = 0.8$, $\frac{\sigma_\varepsilon^2}{\sigma_\eta^2} = 1$, $p = 10$, $\theta_u = 20$, $\theta_w = 0$, T=5000								
$z_t = y_t + u_t$, ARFIMA(0,d,0)+LFC								
	$a = 0$, $b = 0$				$a = 0.3$, $b = -0.7$			
	d=0		d=0.45		d=0		d=0.45	
	d	θ_u	d	θ_u	d	θ_u	d	θ_u
Bias	0.022	3.29	0.008	56.5	-0.028	1.425	0.067	42.3
RMSE	0.144	23.0	0.094	175	0.095	21.3	0.095	416

Table 2.9: Full Parametric ARFIMA(1,d,1)+LFC

$\beta = 0.8$, $\frac{\sigma_\varepsilon^2}{\sigma_\eta^2} = 1$, $p = 10$, $\theta_u = 20$, $\theta_w = 0$, T=5000									
$z_t = y_t + u_t$, ARFIMA(1,d,1)+LFC									
	$a = 0$, $b = 0$					$a = 0.3$, $b = -0.7$			
	d=0					d=0			
	d	θ_u	a	b		d	θ_u	a	b
Bias	-0.003	4.04	-0.092	0.150	Bias	-0.015	-3.64	0.057	0.217
RMSE	0.135	26.5	0.577	0.628	RMSE	0.069	15.4	0.161	0.575
	d=0.45					d=0.45			
	d	θ_u	a	b		d	θ_u	a	b
Bias	0.011	76.6	-0.144	0.016	Bias	0.057	27.9	-0.080	-0.279
RMSE	0.069	194	0.571	0.440	RMSE	0.109	233	0.247	0.390

Table 2.10: ARFIMA+LFC model with additive noise

$\beta = 0.8$, $\frac{\sigma_\varepsilon^2}{\sigma_\eta^2} = 1$, $p = 10$, $\theta_u = 20$, $\theta_w = 10$							
$z_t = y_t + u_t + w_t$, $d = 0.45$							
ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC				
	d	θ_u		d	θ_u	a	b
Bias	-0.381	-18.1	Bias	-0.303	-18.6	0.037	-0.238
RMSE	0.382	18.2	RMSE	0.315	18.7	0.508	0.482
	d=0.45			d=0.45			
	d	θ_u		d	θ_u	a	b
Bias	-0.584	6.19	Bias	-0.274	-8.87	0.036	-0.428
RMSE	0.585	12.1	RMSE	0.380	17.8	0.495	0.658

Table 2.11: PARFIMA(0,d,0)+LFC on long-memory process with LFC and noise

PARFIMA(0,d,0)+LFC, $\theta_w = 30$							
$z_t = y_t + u_t + w_t$, $p = 10$							
	$\theta_u = 0$				$\theta_u = 20$		
d=0.45							
	d	θ_u	θ_w		d	θ_u	θ_w
Bias	-0.094	1.05	-0.051	Bias	-0.086	-7.36	17.4
RMSE	0.221	2.30	44.5	RMSE	0.252	31.7	96.9
d=0.7							
	d	θ_u	θ_w		d	θ_u	θ_w
Bias	-0.066	33.5	0.490	Bias	-0.092	21.7	-3.81
RMSE	0.195	90.2	27.3	RMSE	0.218	84.6	23.4

Table 2.12: PARFIMA(0,d,0)+LFC on moderate or no additive noise

PARFIMA(0,d,0)+LFC, $\theta_u = 0$							
$z_t = y_t + u_t + w_t$, $p = 10$							
$\theta_w = 0$				$\theta_w = 4$			
d=0.45							
	d	θ_u	θ_w		d	θ_u	θ_w
Bias	0.014	1.74	0.122	Bias	-0.038	2.09	-0.606
RMSE	0.033	3.28	0.228	RMSE	0.081	4.01	2.20
d=0.7							
	d	θ_u	θ_w		d	θ_u	θ_w
Bias	0.026	23.9	0.109	Bias	-0.010	50.7	-0.052
RMSE	0.044	74.7	0.175	RMSE	0.091	125	1.71

Table 2.13: PARFIMA(0,d,0)+LFC on pure short-memory processes with LFC

ARFIMA(0,d,0)+LFC+noise							
$z_t = y_t + u_t + w_t$, $p = 10$, $d = 0$							
$\theta_u = 0$							
$\theta_w = 0$				$\theta_w = 4$			
	d	θ_u	θ_w		d	θ_u	θ_w
Bias	-0.064	0.018	12.0	Bias	-0.076	0.018	4.53
RMSE	0.251	0.067	68.5	RMSE	0.308	0.064	32.3
$\theta_u = 20$							
$\theta_w = 0$				$\theta_w = 4$			
	d	θ_u	θ_w		d	θ_u	θ_w
Bias	0.371	196	82.1	Bias	-0.020	9.77	28.4
RMSE	0.687	808	147	RMSE	0.469	88.5	87.7

Table 2.14: PARFIMA(1,d,1)+LFC on long memory process

PARFIMA(1,d,1)+LFC, $\beta = 0.8$, T=5000					
$z_t = y_t + w_t$, $\theta_w = 0$					
d=0.45					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	0.018	0.692	-0.033	0.043	0.334
RMSE	0.037	5.90	0.294	0.324	1.643
d=0.7					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	0.025	0.342	0.004	0.003	0.163
RMSE	0.042	2.16	0.240	0.284	0.634

Table 2.15: PARFIMA(1,d,1)+LFC on long memory process, and LFC

PARFIMA(1,d,1)+LFC, $\beta = 0.8$, T=5000					
$z_t = y_t + u_t + w_t$, $\theta_u = 20, \theta_w = 0$					
d=0.45					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	0.004	18.5	-0.031	0.180	0.518
RMSE	0.057	70.1	0.343	0.479	1.667
d=0.7					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	-0.016	156	0.065	0.081	0.192
RMSE	0.115	509	0.262	0.408	0.562

Table 2.16: PARFIMA(1,d,1)+LFC on long memory process and additive noise

PARFIMA(1,d,1)+LFC					
$z_t = y_t + w_t$, $\theta_w = 4$					
d=0.45					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	-0.023	1.18	-0.005	0.046	5.22
RMSE	0.087	5.41	0.304	0.499	19.5
d=0.7					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	0.022	0.128	-0.056	0.143	2.94
RMSE	0.058	0.265	0.279	0.536	7.82

Table 2.17: PARFIMA(1,d,1)+LFC on long memory process, LFC and additive noise

PARFIMA(1,d,1)+LFC, $\beta = 0.8$, T=5000					
$z_t = y_t + u_t + w_t$, $\theta_u = 20, \theta_w = 4$					
d=0.45					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	-0.050	21.3	-0.018	0.188	6.95
RMSE	0.129	78.3	0.311	0.594	17.7
d=0.7					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	-0.037	158.8	-0.071	0.136	3.15
RMSE	0.120	392	0.307	0.505	11.2

Table 2.18: PARFIMA(1,d,1)+LFC on short-memory process and LFC

PARFIMA(1,d,1)+LFC, $\beta = 0.8$, T=5000					
$d = 0, \theta_w = 0$					
$\theta_u = 0$					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	-0.137	00.008	0.011	0.001	24.6
RMSE	0.416	0.030	0.187	0.199	102
$\theta_u = 20$					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	0.961	-19.7	0.130	0.268	280
RMSE	0.966	19.8	0.469	0.457	932

Table 2.19: PARFIMA(1,d,1)+LFC on short-memory process with additive noise and LFC

PARFIMA(1,d,1)+LFC, $\beta = 0.8$, T=5000					
$\theta_w = 4, d = 0,$					
$\theta_u = 0$					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	-0.142	0.011	0.010	0.037	20.7
RMSE	0.434	0.038	0.210	0.220	95.2
$\theta_u = 20$					
	d	θ_u	AR(1)	MA(1)	θ_w
Bias	0.937	-19.7	-0.325	-0.254	441
RMSE	0.943	19.7	0.528	0.409	548

Table 2.20: SP 500, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.261	0.655	29.0	0.095	0.655	29.01	0.8	0.095
0.7	0.370	0.628	20.8	0.237	0.628	20.77	0.75	0.183
0.6	0.508	0.514	0	0.514	0.520	0.384	0.73	0.212
0.5	0.495	0.583	17.3	0.515	0.540	4.67	0.7	0.237

Table 2.21: SP 500, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC		
β		d	d	a		b
1		0.118	0.643	0.178		-0.805
0.8		0.097	0.537	0.516		-0.821
PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.690	54.9	0.484	0.232	–0.684	0.383
0.8	0.655	29.3	0.630	0.017	-0.591	3.48

Table 2.22: DJIA, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.241	0.654	32.4	0.065	0.654	32.4	0.8	0.065
0.7	0.363	0.575	12.5	0.277	0.575	12.5	0.75	0.148
0.6	0.476	0.486	0	0.480	0.500	0.783	0.73	0.223
0.5	0.441	0.485	3.06	0.475	0.485	3.06	0.7	0.277

Table 2.23: DJIA, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC			
β		d	d	a		b
1		0.105	0.407	0.289		-0.678
0.8		0.097	0.481	0.443		-0.780
PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.683	56.7	-	-	-	-
0.8	0.653	32.2	0.644	0.418	-0.568	15.9

Table 2.24: NASDAQ, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.269	0.631	23.9	0.110	0.606	20.9	0.8	0.110
0.7	0.379	0.588	14.0	0.259	0.444	4.27	0.75	0.208
0.6	0.451	0.668	48.7	0.294	0.572	21.9	0.73	0.212
0.5	0.538	0.667	51.8	0.407	0.407	0	0.7	0.259

Table 2.25: NASDAQ, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC			
β		d	d	a		b
1		0.124	0.357	0.288		-0.604
0.8		0.113	0.332	0.317		-0.603
PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.586	24.1	-	-	-	-
0.8	0.605	20.6	0.336	-0.310	-0.602	0.024

Table 2.26: AMEX, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.246	0.541	11.9	0.155	0.542	11.9	0.8	0.155
0.7	0.346	0.487	5.52	0.274	0.333	0.84	0.75	0.213
0.6	0.403	0.524	11.1	0.306	0.307	0	0.73	0.232
0.5	0.449	0.707	224	0.232	0.708	224	0.7	0.274

Table 2.27: AMEX, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC			
β		d	d	a		b
1		0.153	0.354	0.384		-0.635
0.8		0.158	0.320	-0.322		-0.576
PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.457	7.03	-	-	-	-
0.8	0.540	11.8	0.359	0	-0.468	0.296

Table 2.28: SPY, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.540	0.590	0.386	0.503	0.516	0.06	0.8	0.503
0.7	0.570	0.674	2.25	0.525	0.674	2.24	0.75	0.501
0.6	0.626	0.637	0	0.625	0.634	0	0.73	0.467
0.5	0.546	0.714	38.6	0.378	0.377	0	0.7	0.525

Table 2.29: SPY, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC			
β		d	d	a		b
1		0.466	0.486	0.355		-0.379
0.8		0.514	0.532	-0.063		-0.019
PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.472	0.02	-	-	-	-
0.8	0.515	0	0.533	-0.063	-0.017	0

Table 2.30: SP 500 (2), Semiparametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.206	0.608	50.4	0.095	0.604	49.2	0.8	0.095
0.7	0.308	0.566	29.1	0.187	0.503	15.5	0.75	0.148
0.6	0.403	0.578	36.3	0.282	0.283	0	0.73	0.169
0.5	0.468	0.723	398	0.250	0.723	398	0.7	0.187

Table 2.31: SP 500 (2), Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC			
β		d	d	a		b
1		0.090	0.288	0.396		-0.650
0.8		0.080	0.305	0.583		-0.763
PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.567	44.1	-	-	-	-
0.8	0.604	48.9	0.568	0.472	-0.681	12.4

Table 2.32: Dollar-AUS, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.349	0.781	51.6	0.094	0.488	12.6	0.8	0.094
0.7	0.456	0.832	93.5	0.121	0.655	41.5	0.75	0.120
0.6	0.596	0.864	147	0.215	0.864	147	0.73	0.119
0.5	0.768	0.832	36.5	0.641	0.810	16.0	0.7	0.121

Table 2.33: Dollar-AUS, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC		
β		d	d	a	b
1		0.124	0.357	0.288	-0.604
0.8		0.113	0.332	0.317	-0.603
PARFIMA(0,d,0)+LFC			**PARFIMA(1,d,1)+LFC**			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.586	24.1	-	-	-	-
0.8	0.605	20.6	0.336	-0.310	-0.602	0.024

Table 2.34: Dollar-Yen, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.317	0.594	8.80	0.170	0.594	8.8	0.8	0.170
0.7	0.401	0.616	12.2	0.325	0.616	12.1	0.75	0.278
0.6	0.513	0.531	0.604	0.508	0.508	0	0.73	0.290
0.5	0.543	0.568	0	0.567	0.567	0	0.7	0.325

Table 2.35: Dollar-Yen, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC		
β		d	d	a	b
1		0.194	0.401	0.263	-0.540
0.8		0.175	0.456	0.413	-0.673

PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.542	6.61	-	-	-	-
0.8	0.593	8.68	0.593	-0.006	-0.059	8.68

Table 2.36: SP 500 HF, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.438	0.438	0	0.438	0.439	0.03	0.8	0.438
0.7	0.384	0.420	0.598	0.369	0.378	0.10	0.75	0.404
0.6	0.411	0.481	2.85	0.412	0.485	3.01	0.73	0.386
0.5	0.435	0.463	0	0.463	0.462	0	0.7	0.369

Table 2.37: SP 500 HF, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC			ARFIMA(1,d,1)+LFC		
β		d	d	a	b
1		0.477	0.472	0.008	-0.001
0.8		0.443	0.410	0.006	-0.002

PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.477	0	-	-	-	-
0.8	0.440	0	0.441	-0.004	0.017	0.004

Table 2.38: 30-Year T-Bonds, Semi-Parametric Estimators

Classic LW		LWP		LWLFC	LWPLFC		LWLFC, more β	
β	d	d	θ_w	d	d	θ_w	β	d
0.8	0.317	0.585	5.84	0.200	0.585	5.84	0.8	0.200
0.7	0.416	0.531	2.62	0.409	0.532	2.63	0.75	0.339
0.6	0.448	0.457	0	0.457	0.457	0.013	0.73	0.373
0.5	0.408	0.437	0	0.436	0.613	38.1	0.7	0.409

Table 2.39: 30-Year T-Bonds, Full-Parametric Estimators

ARFIMA(0,d,0)+LFC		ARFIMA(1,d,1)+LFC		
β	d	d	a	b
1	0.183	0.427	0.215	-0.530
0.8	0.207	0.502	0.337	-0.644

PARFIMA(0,d,0)+LFC			PARFIMA(1,d,1)+LFC			
β	d	θ_w	d	AR(1)	MA(1)	θ_w
1	0.564	6.74	-	-	-	-
0.8	0.584	5.71	0.583	-0.015	-0.006	5.44

Chapter 3

Pivotal Inference on Structural Changes in Joint Trend Break Model with Heterogeneous Innovation (joint with Pierre Perron)

3.1 Introduction

Issues related to structural breaks have received a lot of attention in the statistics and econometrics literature (see Perron, 2006, for a survey). Substantial advances have been made to cover general models in the context of estimating and testing structural breaks in both single and multiple equations systems. In the single equation case, Bai (1997) studies the least squares estimation of a single change point in regressions involving stationary and/or trending regressors. He derives the consistency, rate of convergence and the limiting distributions of change point estimates under general conditions on the regressors and the errors. Bai and Perron (1998) extend the testing and estimation analysis to the case of multiple structure changes, while Bai and Perron (2003) present an efficient algorithm to obtain the break date estimates as the global minimizers of the overall sum of squared residuals. Andrews (1993), Bai and Perron (1998), and Altissimo and Corradi (2003) discuss issues related to testing for a single or multiple changes.

Much of the work in the literature concentrated on the case where the regressors and errors are stationary. Nevertheless, issues related to structural changes are also important in the context of trending regressors and non-stationary time series. Perron and Zhu (2005) consider a linear trend function subject to a one-time change in the parameters. They analyze the consistency, rate of convergence and limiting distributions of the parameters with errors that can be stationary or have an autoregressive unit root. They consider three different models: a "joint broken trend", a "local disjoint broken trend" and a "global disjoint broken trend". They show that each case involves different asymptotic results, in particular pertaining to the rate of convergence and the asymptotic distribution of the estimates of the break dates. The model we consider in this paper is the "joint broken trend" model, whereby the slope of the trend changes and the series is joined at the time of the break.

Advances have also been made for structural change problems in the context of testing for changes in the regression coefficients and the variance of the noise component. Building on the work of Perron and Qu (2007), Perron and Zhou (2008) provide a comprehensive treatment of the problem of testing jointly for structural changes in both the regression coefficients and the variance of the errors in a single equation involving stationary regressors, allowing the break dates for the two components to be different or coincide. They provide the required tools for addressing the following testing problems, among others: a) testing for given numbers of changes in regression coefficients and variance of the errors; b) testing for some unknown number of changes less than some pre-specified maximum; c) testing for changes in variance (regression coefficients) allowing for a given number of changes in regression coefficients (variance); and d) estimating the number of changes present.

A problem with the analysis of Perron and Zhou (2008) is that trends are not permitted, in particular those joined at the time of the break. Such breaking trends are very relevant in practice as evidenced by many series in macroeconomics, finance and even climate change (e.g., Estrada, Perron and Martinez-López, 2013). The latter case is indeed the motivation behind this paper as global and hemispheric temperatures as well as radiative forcings (e.g., greenhouse gases) are well approximated by a linear trend with a one-time change in slope near 1960 with the noise component being stationary. As shown in Perron and Zhu (2005), the limit results with joined segmented trends are very different from locally or globally disjoined trends as well as the stationary case. Hence, the need for a separate treatment.

The aim of this paper is to provide the relevant results concerning testing for changes in the slope of the trend and the variance of the noise. We start with a single possible break in each and address the following issues: 1) testing for a change in trend with or without a change in variance; b) testing for a change in variance with or without a change in trend. Asymptotically pivotal statistics are provided for each cases. We then generalize some results to the case of multiple changes. Our work is related to that of Li and Perron (2013) who analyzed the problem of testing for common breaks and forming confidence intervals when the breaks dates are locally ordered in a system of equations with joint-segmented trends.

The structure of this paper is as follows. Section 2 presents the model and the assumptions. Section 3 presents results about the limit distribution of the estimates of the trend function, while Section 4 does so for the estimate for the change in variance. Section 5 is devoted to the pivotal statistic to test for a trend change allowing for a variance break. Section 6 is concerned with pivotal statistics for the coefficient of the trend when there is a trend break allowing for a variance break. Section 7 offers some extensions for cases with multiple breaks. Section 8 provides brief concluding remarks and technical derivations are contained in an appendix.

The following notation is used throughout this paper: "$\overset{d}{\to}$" stands for convergence in distribution, "$\Longrightarrow$" for weak convergence in Skorokhod Topology, "$\overset{p}{\to}$" for convergence in probability, and "$\to$" for the limit as $T \to \infty$ (unless otherwise stated).

3.2 The Model and Assumptions

The model considered is a joint-segmented trend of the form:

$$y_t = \mu + \beta t + \delta B_t + u_t$$

where $B_t = t - T_B$ when $t > T_B$ and 0 otherwise. We let $\lambda_B = T_B/T$ and assume that $\epsilon < \lambda_B < 1 - \epsilon$ for some $\epsilon \in (0, 1/2)$. If there is no change in the slope of the trend, $\delta = 0$ and T_B is undefined. In vector form, we have:

$$Y = X(T_B)\kappa + U$$

where $Y = (y_1, ..., y_t)'$, $U = (u_1, ..., u_T)'$, $\kappa = (\mu, \beta, \delta)'$, $X(T_B) = (X_1, X_2, X(T_B)_3)$ with $X_1 = (1, ..., 1)'$, $X_2 = (1, 2, ..., T)'$ and $X(T_B)_3 = (0, ..., 0, 1, ...T - T_B)'$. If there is a change in the structure of u_t, it occurs at some date T_V with $\lambda_V = T_V/T$ where we assume that $\epsilon < \lambda_V < 1-\epsilon$ for some $\epsilon \in (0, 1/2)$. The true Data Generating Process will be denoted with a 0 superscript or subscript, so that

$$Y = X(T_B^0)\kappa^0 + U$$

where $\kappa^0 = (\mu^0, \beta^0, \delta^0)'$, $T_B^0 = [\lambda_B^0/T]$ and $T_V^0 = [\lambda_V^0/T]$ with $[\cdot]$ denoting the integer argument. In order to state the conditions pertaining to the noise component u_t, we first state the following condition.

Condition 1: Define the L_r-norm of a random matrix X as $\|X\|_r = \left(\sum_i \sum_j E\,|X_{ij}|^r\right)^{1/r}$ for $r \geq 1$ and $\mathcal{F}_t = \sigma\text{-}field\ \{..., x_{t-1}, x_t, u_{t-2}, u_{t-1}\}$. If $x_t u_t$ is weakly stationary within each segment, then (a) $\{x_t u_t, \mathcal{F}_t\}$ forms a strongly mixing (α-mixing) sequence with size $-4r/(r-2)$ for some $r > 2$, (b) $E(x_t u_t) = 0$ and $\|x_t u_t\|_{2r+C_0} < C_1 < \infty$ for some $C_0, C_1 > 0$, (c) Let $S_{k,1}(\ell) = \sum_{1+\ell}^{T_V^0+\ell+k} x_t U$ and $S_{k,2}(\ell) = \sum_{T_V^0+1+\ell}^{T+\ell+k} x_t U$, for each $e \in \mathbb{R}^T$ of length 1, $var(\langle e, S_{k,j}(0)\rangle) \geq v(k)$ for some function $v(k) \to \infty$ as $k \to \infty$ (with $\langle\cdot\rangle$, the usual inner product). If $x_t u_t$ is not weakly stationary within each segment, we assume that (a)-(c) holds, and in addition, that there exists a positive definite matrix W such that we have, uniformly in ℓ, $\left|k^{-1}E\left((S_{k,j}(\ell))_i, (S_{k,j}(\ell))_s\right) - (W)_{i,s}\right| \leq C_2 k^{-C_3}$, for some $C_2, C_3 > 0$.

We consider two assumptions about the noise component u_t.

- Assumption 1(a): Let $var(u_t) = \sigma_{10}^2$ if $t \leq T_V^0$ and $var(u_t) = \sigma_{20}^2$ if $t > T_V^0$. We assume that Condition 1 holds with z_t replaced by u_t or $u_t^2/\sigma_{j0}^2 - 1$ in each regime defined by $T_{V,j-1}^0 \leq t < T_{V,j}^0$ $(j = 1, 2)$, with the convention that $T_{V,0}^0 = 1$ and $T_{V,2}^0 = T$.

- Assumption 1(b): $\{u_t\}$ is an autoregressive process of order p given by $u_t = \sum_{i=0}^p c_i u_{t-i} + e_t$, where the roots of $C(L) = 1 - \sum_{i=0}^p c_i L^i$ are outside the unit circle and $e_t = \sigma_{j0}\varepsilon_t$, for $T_{V,j-1}^0 < t \leq T_{V,j}^0$ $(j = 1, 2)$, with ε_t a martingale-difference sequence satisfying Condition 1 in each regime.

The conditions are mild in the sense that they allow for substantial conditional heteroskedasticity and autocorrelation. Also, if no autocorrelation is present, i.e., $\{u_t\}$ is martingale difference sequences with respect to the filtration $\mathcal{F}_t$, then even the weak stationarity assumption can be dropped and u_t allowed to be unconditionally heteroskedastic with bounded fourth moments. Note that Condition 1 could be replaced by other sufficient conditions that can yield a strong invariance principle or Functional Central Limit Theorem (FCLT)[1].

Consider first Assumption 1(a). Here a change in the structure of the noise u_t corresponds to a change in the variance of u_t. Under the stated conditions, the following FCLTs hold for all $r \in [0, 1]$. First,

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[Tr]} u_t &= \begin{pmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} u_t \text{ when } r \leq \lambda_V^0 \\ T^{-1/2}\sum_{t=1}^{[T\lambda_V^0]} u_t + T^{-1/2}\sum_{t=[T\lambda_V^0]+1}^{[Tr]} u_t \text{ when } r > \lambda_V^0 \end{pmatrix} \\ &\Rightarrow \begin{pmatrix} \omega_{10}W(r) \text{ when } r \leq \lambda_V^0 \\ \omega_{10}W(\lambda_V^0) + \omega_{20}[W(r) - W(\lambda_V^0)] \text{ when } r > \lambda_V^0 \end{pmatrix} \equiv \phi(r) \end{aligned}$$

where $\omega_{10} = \lim_{T\to\infty} E((\sum_{t=1}^{[T\lambda_V^0]} u_t)^2)$ and $\omega_{20} = \lim_{T\to\infty} E((\sum_{t=[T\lambda_V^0]+1}^{T} u_t)^2)$, are the so-called long-run variances in each regime. Here, and throughout, "$\Rightarrow$" refers to weak convergence in distribution under the Skorohod topology. Let $\sigma_{t0} = \sigma_{10}$ for $t \leq T_V^0$ and $\sigma_{t0} = \sigma_{20}$ for $t > T_V^0$, then we also have,

$$T^{-1/2}\sum_{t=1}^{[Tr]}((\frac{u_t}{\sigma_{t0}})^2 - 1) \Rightarrow \psi B(r) \tag{1}$$

where $\psi = \lim_{T\to\infty} E((\sum_{t=1}^{T}((u_t/\sigma_{t0})^2 - 1))^2)$. Note that $W(r)$ and $B(r)$ are independent standard Wiener processes.

Under Assumption 1(a), what is being considered is a change in the variance of u_t. This implies that the change can be due to either the nature of the serial correlation or the underlying variance of the primitive shocks. In some cases, it may be desirable to test for a change in the variance of the primitive shocks assuming constant coefficients for the dynamics. This can be achieved adopting Assumption 1(b), which specifies an autoregressive structure for the dynamics. The conditions are obviously less general than those in Assumption 1(a) but this is inevitable given the need to model the dynamics. We assume for simplicity that the

[1] Examples of such conditions are discussed by Davidosn (1994), Dehling and Philipp (1982), Altissimo and Corradi (2003) and Lavielle and Moulines (2000), among others.

order of the autoregression is fixed for simplicity. Our results could be extended to more general processes having an infinite AR structure, which can be approximated by a sequence of increasing autoregressive orders, at the expense of considerable technical complexities. Under the conditions stated in Assmption 1(b), the following FCLTs hold for all $r \in [0,1]$. First,

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[Tr]} e_t &= \begin{pmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} e_t \text{ when } r \leq \lambda_V^0 \\ T^{-1/2}\sum_{t=1}^{[T\lambda_V^0]} e_t + T^{-1/2}\sum_{t=[T\lambda_V^0]+1}^{[Tr]} e_t \text{ when } r > \lambda_V^0 \end{pmatrix} \\ &\Rightarrow \begin{pmatrix} \sigma_{10}W(r) \text{ when } r \leq \lambda_V^0 \\ \sigma_{10}W(\lambda_V^0) + \sigma_{20}[W(r) - W(\lambda_V^0)] \text{ when } r > \lambda_V^0 \end{pmatrix} \equiv \phi(r) \end{aligned}$$

Also,

$$T^{-1/2}\sum_{t=1}^{[Tr]}((\frac{e_t}{\sigma_{t0}})^2 - 1) \Rightarrow \psi B(r) \tag{2}$$

where $\psi = \lim_{T\to\infty} E((\sum_{t=1}^{T}((e_t/\sigma_{t0})^2 - 1))^2)$. Again, $W(r)$ and $B(r)$ are independent standard Wiener processes. Note that if there is no break in the variance $\omega_{10} = \omega_{20} = \omega_0$ and $\phi(r) = \omega_0 W(r)$ under Assumption 1(a), while $\sigma_{10} = \sigma_{20} = \sigma_0$ and $\phi(r) = \sigma_0 W(r)$ under Assumption 1(b).

The goal is to derive pivotal test statistics for the following testing problems:

- TP-1) H_0 : no break in either trend or volatility versus H_1 : one break in trend and no break in volatility.

- TP-2) H_0 : no break in either trend or volatility versus H_1 : no breaks in trend and one break in volatility.

- TP-3) H_0 : one break in trend and no break in volatility versus H_1 : one break in trend and one break in volatility.

- TP-4) H_0 : no break in trend and one break in volatility versus H_1 : one break in trend and one break in volatility.

We return later to some extensions involving multiple changes.

3.3 Limit distributions of the estimates of the trend function

We start with results pertaining to the estimates of the trend function. We cover the cases with or without a break in the slope and for each cases with or without a change in the variance. We consider estimates of the parameters of the trend function obtained using a least-squares method. Hence, the estimate of the break date is given by:

$$\hat{T}_B = \arg\min_{T_B}\{Y(I - X(T_B)(X(T_B)'X(T_B))^{-1}X(T_B)')Y\}$$

where the minimization is taken over the set $\Lambda_\varepsilon = \{\lambda | \epsilon < \lambda < 1-\epsilon\}$ with $T_B = [T\lambda]$, while the estimates of the coefficients are:

$$\hat{\kappa} = (\hat{\mu}, \hat{\beta}, \hat{\delta})' = (X(\hat{T}_B)'X(\hat{T}_B))^{-1}X(\hat{T}_B)'Y$$

The estimate of the associated break fraction is $\hat{\lambda}_B = \hat{T}_B/T$. Also, the estimated residuals are $\hat{U} = (\hat{u}_1, ..., \hat{u}_T)$ defined by:

$$\hat{U} = Y - X(\hat{T}_B)\hat{\kappa} \tag{3}$$

The results are presented in the following Theorem.

Theorem 1 *Let*

$$\Sigma_a(\lambda)^{-1} = \begin{pmatrix} \frac{\lambda+3}{\lambda} & -\frac{3(\lambda+1)}{(\lambda)^2} & \frac{3}{(\lambda)^2(1-\lambda)} \\ -\frac{3(\lambda+1)}{(\lambda)^2} & \frac{3(3\lambda+1)}{(\lambda)^3} & -\frac{3(2\lambda+1)}{(\lambda)^3(1-\lambda)} \\ \frac{3}{(\lambda)^2(1-\lambda)} & -\frac{3(2\lambda+1)}{(\lambda)^3(1-\lambda)} & \frac{3}{(\lambda)^3(1-\lambda)^3} \end{pmatrix}$$

$$\zeta(\lambda) = (\int_0^1 dW(r), \int_0^1 rdW(r), \int_\lambda^1 (r-\lambda)dW(r))'$$

and

$$\xi(\lambda) = (\int_0^1 d\phi(r), \int_0^1 rd\phi(r), \int_\lambda^1 (r-\lambda)d\phi(r))'. \tag{4}$$

(1) When there is break in the trend at λ_B^0,

$$T^{3/2}(\hat{\lambda}_B - \lambda_B^0) \Rightarrow -\frac{4\chi}{\delta^0\lambda_B^0(1-\lambda_B^0)}$$

where

$$\chi = \int_0^{\lambda_B^0} \frac{\lambda_B^0 - (\lambda_B^0)^2 - 3r + 3r\lambda_B^0}{2\lambda_B^0} d\phi(r) + \int_{\lambda_B^0}^1 \lambda_B^0 \frac{2 + \lambda_B^0 - 3r}{2(1-\lambda_B^0)} d\phi(r)$$

with $\phi(r)$ defined in Section 2. Note that when there is no break in variance $\chi \sim N(0, \lambda_B^0(1-\lambda_B^0)/4)$ and $T^{3/2}(\hat{\lambda}_B - \lambda_B^0) \Rightarrow N(0, 4(\omega^0)^2/\lambda_B^0(1-\lambda_B^0)(\delta^0)^2)$ as in Perron and Zhu (2005). The limit distribution of the estimates of the coefficients is given by:

$$\begin{pmatrix} T^{-1/2}(\hat{\mu} - \mu^0) \\ T^{-3/2}(\hat{\beta} - \beta^0) \\ T^{-3/2}(\hat{\delta} - \delta^0) \end{pmatrix} \Rightarrow N(0, \Sigma_B) = N(0, \Sigma_a(\lambda_B^0)^{-1}\Sigma_d\Sigma_a(\lambda_B^0)^{-1})$$

where Σ_d is the variance-covariance matrix of

$$\int_0^{\lambda_B^0} \begin{pmatrix} \frac{-2\lambda_B^0 + 3(\lambda_B^0)^2 + 6r - 6r\lambda_B^0}{(\lambda_B^0)^2} \\ \frac{(\lambda_B^0)^3 - 2r(\lambda_B^0)^2 - \lambda_B^0 + 3r}{(\lambda_B^0)^2} \\ \lambda_B^0(\lambda_B^0 - 1) + 3r\frac{(1-\lambda_B^0)^2}{(\lambda_B^0)^2} \end{pmatrix} d\phi(r) + \int_{\lambda_B^0}^1 \begin{pmatrix} \frac{1+\lambda_B^0 - 2r}{-1+\lambda_B^0} \\ \frac{3\lambda_B^0 + (\lambda_B^0)^2 - 4r + 2 - 2r\lambda_B^0}{-1+\lambda_B^0} \\ -2 - 2\lambda_B^0 + 4r \end{pmatrix} d\phi(r).$$

When there is no break in variance,

$$\Sigma_d = \omega_0^2 \begin{pmatrix} \frac{4-\lambda_B^0}{\lambda_B^0} & \frac{4-4(\lambda_B^0)^2+\lambda_B^0}{2\lambda_B^0} & \frac{4+(\lambda_B^0)^3+2(\lambda_B^0)^2-7\lambda_B^0}{2\lambda_B^0} \\ \frac{4-4(\lambda_B^0)^2+\lambda_B^0}{2\lambda_B^0} & \frac{3-3(\lambda_B^0)^3-3(\lambda_B^0)^2+4\lambda_B^0}{3\lambda_B^0} & \frac{(\lambda_B^0)^4+6(\lambda_B^0)^3-9(\lambda_B^0)^2-4\lambda_B^0+6}{6\lambda_B^0} \\ \frac{4+(\lambda_B^0)^3+2(\lambda_B^0)^2-7\lambda_B^0}{2\lambda_B^0} & \frac{(\lambda_B^0)^4+6(\lambda_B^0)^3-9(\lambda_B^0)^2-4\lambda_B^0+6}{6\lambda_B^0} & \frac{-(\lambda_B^0)^4+6(\lambda_B^0)^2-8\lambda_B^0+3}{3\lambda_B^0} \end{pmatrix}$$

and the limit distribution reduces to that in Perron and Zhu (2005).
2) When there is no break in the trend, we have $\hat{\lambda}_B \Rightarrow \lambda_B^*$ where

$$\begin{aligned} \lambda_B^* &= \arg\max_{\lambda \in \Lambda_\varepsilon} \xi(\lambda)'\Sigma_a(\lambda)^{-1}\xi(\lambda) = \arg\max_{\lambda \in \Lambda_\varepsilon}\{\frac{(\int_0^1 g(\lambda, r)d\phi(r))^2}{\int_0^1 g^2(\lambda, r)dr}\} \quad (5) \\ &= \arg\max_{\lambda \in \Lambda_\varepsilon}\{\frac{[\omega_{10}\int_0^{\lambda_V^0} g(\lambda, r)dW(r) + \omega_{20}\int_{\lambda_V^0}^1 g(\lambda, r)dW(r)]^2}{\int_0^1 g^2(\lambda, r)dr}\} \quad (6) \end{aligned}$$

with the function $g(\lambda, r)$ given by

$$g(\lambda, r) = \mathbf{1(r >}\lambda)(r - \lambda) + \lambda^3 - 2\lambda^2 + \lambda - 2\lambda^3 r + 3\lambda^2 r - r \quad (7)$$

is the continuous analog of the residuals from a regression of $X(T_B)$ on X_1, X_2. The limit distribution of the coefficients $(\hat{\mu}, \hat{\beta}, \hat{\delta})$ is given by

$$(T^{1/2}(\hat{\mu} - \mu_0), T^{3/2}(\hat{\beta} - \beta_0), T^{3/2}\hat{\delta})' \Rightarrow \Sigma_a(\lambda_B^*)^{-1}\xi(\lambda_B^*) \quad (8)$$

If, in addition, there is no break in variance, $\xi(\lambda) = \omega_0\zeta(\lambda)$ and we have:

$$\frac{1}{\omega_0}(T^{1/2}(\hat{\mu} - \mu_0), T^{3/2}(\hat{\beta} - \beta_0), T^{3/2}\hat{\delta})' \Rightarrow \Sigma_a(\lambda_B^*)^{-1}\zeta(\lambda_B^*) \tag{9}$$

With $\hat{\omega}^2$ as consistent estimate of ω_0^2, note that when there is no break in variance the statistic $\hat{\delta}_{PIV} \triangleq T^{3/2}\hat{\delta}/\hat{\omega}$ can act as a test for a change in trend since it has a pivotal limit distribution given by (9). Tedious algebra leads to its limit distribution as

$$\hat{\delta}_{PIV} \Rightarrow [\frac{3}{(\lambda_B^*)^3(1-\lambda_B^*)^3}]^{1/2}[\int_0^1 g(\lambda_B^*, r)dW(r)] \tag{10}$$

This addresses the testing problem TP-1. While this approach was used in Sayginsoy and Vogelsang (2011), the critical values have never been published. For completeness, we present them in Table 1. See Remarks 1 below for the construction of the estimate $\hat{\omega}^2$.
3) When there is a break in variance, the limit distribution of $T^{3/2}\hat{\delta}$ is

$$\frac{\int_0^1 g(\lambda_B^*, r)d\phi(r)}{\int_0^1 g^2(\lambda_B^*, r)dr} = \frac{\omega_{10}\int_0^{\lambda_V^0} g(\lambda_B^*, r)dW(r) + \omega_{20}\int_{\lambda_V^0}^1 g(\lambda_B^*, r)dW(r)}{\int_0^1 g^2(\lambda_B^*, r)dr} \tag{11}$$

Allowing heterogeneous variance, $\Delta_{4,T} = T^{3/2}\hat{\delta}$ can be used as a test for TP-4): $H_0 : \delta^0 = 0$ versus $H_0 : \delta^0 \neq 0$ with limit distribution of (11) with λ_B^* having limit distribution given by (5).

Remark 18 *When there is a break in variance, the limit distribution (15) depends on nuisance parameters $(\lambda_V^0, \omega_{10}^2, \omega_{20}^2)$, and is non-standard. Hence, to carry proper inference, some modifications are necessary, among them using simulations to obtain the relevant quantiles or resorting to bootstrap. To use simulation to obtain the relevant quantiles, first estimate the heterogeneous trend break model to get estimates of nuisance parameters about innovation $(\hat{\lambda}_V, \hat{\omega}_1, \hat{\omega}_2)$, then invoke 11 to simulate its critical values that depends on these nuisance parameters, with the true value of nuisance parameters replaced by their estimates, respectively. We return to this problem in Section 5 for a pivotal statistic.*

Remark 19 *With homogeneous variance, (11) simplifies to the pivotal one given by (10).*

Remark 20 *The authors have simulation evidence in favor of a pivotal distribution for*

$$\begin{aligned}\hat{\delta}^2 &= (\int_0^1 g^2(\lambda_B^*, r)dr)^{-1/2}[\omega_{10}\int_0^{\lambda_V^0} g(\lambda_B^*, r)dW(r) \\ &\quad +\omega_{20}\int_{\lambda_V^0}^1 g(\lambda_B^*, r)dW(r)]^2\end{aligned}$$

. However, a rigorous proof is not readily accessible and is left to future work.

3.4 Limit distributions of the estimates for the change in variance

The test for a change in variance will involve different quantities depending on whether we deal with Assumption 1(a) or 1(b). In the case of Assumption 1(a), it will be based directly on the least-squares residuals $\hat{U} = (\hat{u}_1, ..., \hat{u}_T)$ defined by (3). The estimate of the break date and variance in each regime are then

$$(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{T}_V) = \arg\min_{\sigma_1,\sigma_2,T_V} \sum_{t=1}^{T_V} ((\hat{u}_t)^2 - \sigma_1^2)^2 + \sum_{t=T_V+1}^{T} ((\hat{u}_t)^2 - \sigma_2^2)^2$$

Under assumption 1(b) a second-stage OLS regression is needed, namely the following estimated AR(p)

$$\hat{u}_t = \sum_{j=0}^{p} \tilde{c}_j \hat{u}_{t-j} + \tilde{e}_t.$$

which yields the estimates $\tilde{c}_j$ $(j = 1, ..., p)$ and $\tilde{e}_t$ $(t = p+1, ..., T)$. The test will then be based on the residuals $\tilde{e}_t$ and the estimate of the break date and variance in each regime are

$$(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{T}_V) = \arg\min_{\sigma_1,\sigma_2,T_V} \sum_{t=p+1}^{T_V} ((\tilde{e}_t)^2 - \sigma_1^2)^2 + \sum_{t=T_V+1}^{T} ((\tilde{e}_t)^2 - \sigma_2^2)^2.$$

In both cases, the minimization problem is taken over the set $\Lambda_\varepsilon = \{\lambda_V | \epsilon < \lambda_V < 1 - \epsilon\}$.We start with the following lemma, which shows that the FCLTs (1) and (2) remain valid when the true residuals are replaced by estimates. The proof is straightforward given the rate of convergence of the estimates and, hence, omitted.

Lemma 1 *Under Assumption 1(a), with or without a break in the trend, we have for $s \in [0, 1]$:*

$$T^{-1/2} \sum_{t=1}^{[Ts]} (\hat{u}_t^2/\sigma_{t0}^2 - 1) \Rightarrow \psi B(s).$$

Similarly, under Assumption 1(b), with or without a break in the trend, we have for $s \in [0, 1]$:

$$T^{-1/2} \sum_{t=1}^{[Ts]} (\tilde{e}_t^2/\sigma_{t0}^2 - 1) \xrightarrow{d} \psi B(s).$$

Using this lemma, it is straightforward to prove the following results about the limit distribution of the estimates.

Theorem 2 *Under Assumption A1.(a) or 1(b), under the alternative of a change in variance:*

$$T^{1/2}(\tilde{\sigma}_1^2 - \sigma_{10}^2) \Rightarrow \sigma_{10}^2 \psi^{1/2} \frac{B(\lambda_V^0)}{\lambda_V^0}$$

$$T^{1/2}(\tilde{\sigma}_2^2 - \sigma_{20}^2) \Rightarrow \sigma_{20}^2 \psi^{1/2} \frac{B(1) - B(\lambda_V^0)}{1 - \lambda_V^0}$$

Under the null hypothesis of no change in varinace

$$T^{1/2}(\tilde{\sigma}^2 - \sigma_0^2) \Rightarrow \sigma_0^2 \psi^{1/2} B(1)$$

Remark 21 *A consistent estimate of ψ can be constructed using Andrews' (1991) kernel method, namely*

$$\hat{\psi} = T^{-1}\{\sum_{j=-(T-1)}^{(T-1)} K(j,m) \sum_{t=j+1}^{T} \hat{\eta}_t \hat{\eta}_{t-j}\}$$

where $K(j,m)$ is some kernel function with m the bandwidth. A common choice for the kernel function is the Quadractic Spectral and m is selected using Andrews (1991) method with an AR(1) approximation. Under Assumption 1(a) $\hat{\eta}_t = (\hat{u}_t/\hat{\sigma}_t)^2 - 1$ where $\hat{\sigma}_t^2 = \hat{\sigma}_1^2 = \hat{T}_V^{-1} \sum_{j=1}^{\hat{T}_V} \hat{u}_j^2$ if $1 \leq t \leq \hat{T}_V$, $\hat{\sigma}_t^2 = \hat{\sigma}_2^2 = (T - \hat{T}_V)^{-1} \sum_{j=\hat{T}_V+1}^{T} \hat{u}_j^2$ if $\hat{T}_V < t \leq T$. Alternatively, one could use an estimate that imposes the null hypothesis, namely $\hat{\sigma}_t^2 = T^{-1} \sum_{j=1}^{T} \hat{u}_j^2$ for all t. A method that offers better finite sample properties was suggested by Kejriwal (2009) (see also Kejriwal and Perron, 2010). It involves using the residuals under the null to construct $\hat{u}_t$ but using the residuals under the alternative to select the bandwidth parameter m. The same applies under Assumption 1(b) except that one uses $\tilde{e}_t^2$ instead of $\hat{u}_t^2$.

The result of Theorem 2 allows us to construct pivotal statistics for the testing problem TP-2 and TP-3, which pertain to testing for a change in variance with or without a change in trend. Let $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$ under Assumption 1(a) and $\hat{\sigma}^2 = T^{-1}\sum_{t=p+1}^{T}\tilde{e}_t^2$ under Assumption 1(b). Also let $\hat{\psi}^2$ be as defined in Remark 1, then we consider the statistic

$$\begin{aligned} F_T &= T^{1/2}(\tilde{\sigma}_1^2 - \tilde{\sigma}_2^2)/(\hat{\psi}\hat{\sigma}^2) \\ &\Rightarrow [\frac{B(\lambda_V^*)}{\lambda_V^*} - \frac{B(1)-B(\lambda_V^*)}{1-\lambda_V^*}] \end{aligned} \tag{12}$$

where

$$\lambda_V^* = \arg\min_{\lambda\in\Lambda_\varepsilon}[B(\lambda)^2 + (B(1)-B(\lambda))^2]$$

The limit distribution stated in (12) is non-standard but pivotal. The relevant quantiles can be obtained via simulations. The critical values are presented in Table 3.1.

3.5 Pivotal statistic to test for a trend change allowing for a variance break

We recall that from Theorem 1, when there is a variance break at λ_V^0 but no break in trend, we have

$$D_T^{1/2}(\hat{\kappa} - \kappa^0) \Rightarrow \Sigma_a(\lambda_B^*)^{-1}\xi(\lambda_B^*)$$

where where λ_B^*, $\xi(\cdot)$ is given by 5, 4 respectively, and $g(\lambda, r)$ defined by 7 is the residual function on $r \in [0,1]$ from a continuous time least-squares regression of $\mathbf{1}(\mathbf{r} > \lambda)$ on $[1, r]$. The limit distribution of regression coefficents $\hat{\kappa}$, and hence $\hat{\delta}$, is non-pivotal since $\phi(s)$ is no longer a homogeneous Brownian motion. Nevertheless, because innovation u_t is still stationary in each regime $[0, [\lambda_V^0 T]]$ and $[[\lambda_V^0 T]+1, T]$, trend break tests in each regime will stay pivotal, if a minimum distance is imposed between λ_V^0 and λ_B^0. This motivates us to consider the following two regression and testing problems:

$$y_{1,t} = \mu + \beta t + \delta B_t + u_{1,t}$$

and

$$y_{2,t} = \mu + \beta t + \delta B_t + u_{2,t}$$

where $B_t = t - T_B$ when $t > T_B$ and 0 otherwise. Or, in vector form:

$$Y_1 = X(T_B)\kappa + U_1$$

and

$$Y_2 = X(T_B)\kappa + U_2$$

where Y_1, Y_2, U_1, U_2, κ, $X(T_B) = (X_1, X_2, X(T_B)_3)$ defined similarly as the standard model, but with U_1, U_2 being stationary process with long-run variance $\omega_{10}^2, \omega_{20}^2$, respectively. The true Data Generating Process will be denoted with a 0 superscript or subscript, so that

$$Y_1 = X(T_B^0)\kappa^0 + U_1$$

and

$$Y_2 = X(T_B^0)\kappa^0 + U_2$$

where $\kappa^0 = (\mu^0, \beta^0, \delta^0)'$, $T_B^0 = [\lambda_B^0/T]$. So a test for trend break with heterogeneous innovation is disintegrated into two tests for trend break each with stationary innovation and independent of each other.

This lead us to the following result about the construction of a pivotal statistic for a change in trend under heterogeneous innovation. It bears resemblance to the supLR tests considered in Perron and Zhou (2008) in the sense that its limit distribution is the summation of several maximized square of Brownian motions.

Theorem 3 *Allowing heterogeneous variance, assuming $|\lambda_V^0 - \lambda_B^0| > \varepsilon$, a pivotal test for TP-4): $H_0 : \delta^0 = 0$ versus $H_0 : \delta^0 \neq 0$ that only depends on the trimming parameter ε can be defined as*

$$F_{4,T} = (\hat{\lambda}_V)^3 T^3 \frac{\hat{\delta}_1^2}{\hat{\omega}_{10}^2} + (1 - (\hat{\lambda}_V)^3) T^3 \frac{\hat{\delta}_2^2}{\hat{\omega}_{20}^2} \tag{13}$$

Its limit distribution is given by

$$\max_{\varepsilon < \lambda_1 < 1-\varepsilon} \{\frac{3}{(\lambda_1)^3(1-\lambda_1)^3}[\int_0^1 g(\lambda_1, r)dW(r)]^2\} + \max_{\varepsilon < \lambda_2 < 1-\varepsilon} \{\frac{3}{(\lambda_2)^3(1-\lambda_2)^3}[\int_0^1 g(\lambda_2, r)dW(r)]^2\} \tag{14}$$

where $(\hat{\lambda}_V, \hat{\omega}_{10}^2, \hat{\omega}_{20}^2)$ *are the estimates of the volatility break fraction and the long-run variance of the innovation, and* $\hat{\delta}_1, \hat{\delta}_2$ *are the trend break coefficients in the regime* $[0, [\hat{\lambda}_V T]]$ *and* $[[\hat{\lambda}_V T]+1, T]$*, respectively, estimated by*

$$\hat{T}_{i,B} = \arg\min_{T_B}\{Y_i(I - X(T_B)(X(T_B)'X(T_B))^{-1}X(T_B)')Y_i\}$$

where the minimization is taken over the set $\Lambda_\varepsilon = \{\lambda = T_B/T_i | \epsilon < \lambda < 1-\epsilon\}$ *with* $T_i = T\lambda_V^0$*,* $T_2 = T(1-\lambda_V^0)$*. The estimates of the coefficients in each regime are:*

$$\hat{\kappa}_i = (\hat{\mu}_i, \hat{\beta}_i, \hat{\delta}_i)' = (X(\hat{T}_{i,B})'X(\hat{T}_{i,B}))^{-1}X(\hat{T}_{i,B})'Y_i$$

Remark 22 *The limit distribution (14) is non-standard, but the relevant quantiles can be obtained via simulations. The relavant quantile values are presented in Table 3.1.*

3.6 Pivotal statistics with trend break allowing heterogeneous variance

With trend break, $\hat{\lambda}_B$ converges to λ_B^0 at rate $T^{3/2}$ and $\hat{\delta}$ has normal distribution. Hence we have the following result:

Theorem 4 *Allowing for a break in variance, when there is trend break, i.e.,* $\delta^0 \neq 0$*, 1) a pivotal test for* $H_0 : \delta^0 = \delta^1$ *versus* $H_0 : \delta^0 \neq \delta^1$ *that does not depend on the trimming parameter* ε *is:*

$$\hat{\delta}_{PIV} = T^{3/2} d_T^\delta(\hat{\omega}_1^2, \hat{\omega}_2^2, \hat{\lambda}_B, \hat{\lambda}_V)\hat{\delta} \Rightarrow N(0,1)$$

where the scalar $d_T^\delta(\hat{\omega}_1^2, \hat{\omega}_2^2, \hat{\lambda}_B, \hat{\lambda}_V)$ *is defined by*

$$\begin{aligned} d_T^\delta(\hat{\omega}_1^2, \hat{\omega}_2^2, \hat{\lambda}_B, \hat{\lambda}_V) &= \frac{(\hat{\lambda}_B)^3(1-\hat{\lambda}_B)^3}{3[\hat{\omega}_1^2 Var[\xi_{V1}(\hat{\lambda}_V, \hat{\lambda}_B)] + \hat{\omega}_2^2 Var[\xi_{V2}(\hat{\lambda}_V, \hat{\lambda}_B)]])]^{1/2}} \\ &\Rightarrow \frac{(\lambda_B^0)^3(1-\lambda_B^0)^3}{3[\omega_{10}^2 Var(\xi_{\delta 1}(\lambda_V^0, \lambda_B^0)) + \omega_{20}^2 Var(\xi_{\delta 2}(\lambda_V^0, \lambda_B^0))]^{1/2}} \end{aligned}$$

where $\xi_{\delta 1}(\lambda_V^0, \lambda_B^0)$ *and* $\xi_{\delta 2}(\lambda_V^0, \lambda_B^0)$ *are normal random variables which do not depend on the long run variances* ω_{10} *and* ω_{20}*, having variances*

$$\begin{aligned} & Var(\xi_{\delta 1}(\lambda_V^0, \lambda_B^0)) \\ = \; & [\lambda_V^0 \lambda_B^0 (1-\lambda_B^0)^4 - (\lambda_V^0)^2 \lambda_B^0 (2\lambda_B^0 + 1)(1-\lambda_B^0)^4 \\ & + \frac{(\lambda_V^0)^3}{3}(2\lambda_B^0+1)^2(1-\lambda_B^0)^4] - 2[\lambda_V^0 - \min\{\lambda_B^0, \lambda_V^0\}]^2 \lambda_B^0 (1-\lambda_B^0)^2 \\ & - \frac{1}{3}[\lambda_V^0 - \min\{\lambda_B^0, \lambda_V^0\}]^3 (4\lambda_B^0 - 6\lambda_B^0 + 1)\} \end{aligned}$$

and

$$\begin{aligned}
& Var(\xi_{\delta 2}(\lambda_V^0, \lambda_B^0)) \\
= \; & (1-\lambda_V^0)\lambda_B^0(1-\lambda_B^0)^4 - (1-(\lambda_V^0)^2)\lambda_B^0(2\lambda_B^0+1)(1-\lambda_B^0)^4 \\
& + \frac{1-(\lambda_V^0)^3}{3}(2\lambda_B^0+1)^2(1-\lambda_B^0)^4 \\
& -2[(1-\lambda_B^0)^2 - (\max\{\lambda_B^0, \lambda_V^0\} - \lambda_B^0\})^2]\hat{\lambda}_B(1-\lambda_B^0)^2 \\
& -\frac{1}{3}[(1-\lambda_B^0)^3 - (\max\{\lambda_B^0, \lambda_V^0\} - \lambda_B^0)^3](4\lambda_B^0 - 6\lambda_B^0 + 1)
\end{aligned}$$

The limit distribution of $\hat{\delta}_{PIV}$ is standard normal:

$$\hat{\delta}_{PIV} \Rightarrow \int_0^1 g(\lambda_B^0, r)dW(r)[\frac{3\omega_0^2}{(\lambda_B^0)^3(1-\lambda_B^0)^3}]^{1/2} = N(0,1) \tag{15}$$

(2) Generally, for each pair of parameters $\hat{\varkappa}$ whose elements is a subset of $\{\hat{\mu}, \hat{\beta}, \hat{\delta}\}$, we have a pivotal statistic $\hat{\varkappa}_{PIV}$ and a scaling matrix or scalar $d_T^{\varkappa}$ so that $\hat{\varkappa}_{PIV} = d_T^{\varkappa}\hat{\varkappa}$ is pivotal. In particular, for the whole vector of coefficient estimates $\hat{\kappa}$, we have that

$$\hat{\kappa}_{PIV} = d_T^{\kappa}(\hat{\omega}_{10}, \hat{\omega}_{20}, \hat{\lambda}_V, \hat{\lambda}_B)D_T^{1/2}(\hat{\kappa} - \kappa^0)$$

can be used to test a joint hypothesis of the form $H_0 : \kappa^0 = \kappa^1$. The specific form of the normalization statistic $d_T^{\varkappa}$ depends on the coefficients involved in the hypothesis testing. The limit distribution of $\hat{\varkappa}_{PIV}$ is standard normal.

Remark 23 *When there is no variance break, $\hat{\delta}_{PIV}$ simplifies to*

$$\hat{\delta}_{PIV} = T^{3/2}\hat{\delta}d_T^{\delta}(\hat{\omega}_0^2, \hat{\lambda}_B)$$

where

$$d_T^{\delta}(\hat{\omega}_0^2, \hat{\lambda}_B) = (\frac{(\hat{\lambda}_B)^3(1-\hat{\lambda}_B)^3}{3\hat{\omega}^2})^{1/2}.$$

Remark 24 *The construction of $d_T^{\delta}(\hat{\omega}_{10}^2, \hat{\omega}_{20}^2, \hat{\lambda}_B, \hat{\lambda}_V)$ depends, in general, on the number of variance breaks and only the case of a single variance break is shown here.*

Remark 25 *Each pivotal statistic can only be used to test an hypothesis corresponding to estimates included. For example, $\hat{\kappa}_{PIV}$ can only be used to test the joint hypothesis $H_0 : \kappa^0 = \kappa^1$ but not hypotheses pertaining to individual coefficients, e.g., $H_0 : \mu^0 = \mu^1$, $\beta^0 = \beta^1$ or $\delta^0 = \delta^1$. This is because the elements of $\hat{\kappa}_{PIV}$ are linear combinations of all three parameter estimates and, hence, they cannot be separated to test hypotheses about individual coefficients.*

3.7 Extensions to Multiple Breaks

We now consider extensions of the testing problems TP-2 and TP-3 to the case of multiple breaks in variance. While it is feasible to similarly extend the testing problems TP-1 and TP-4, the task is much more difficult and left for subsequent work.

Under both the null and alternative hypotheses there may be no break in trend (TP-2) or there may be a break in slope at date T_B^0 (TP-3). Under the null hypothesis there is no break in variance while under the alternative there are n breaks occurring at dates $\{T_{V,1}^0, ..., T_{V,n}^0\}$. Again, the break fractions are $\lambda_{V,i}^0 = T_{V,i}^0/T$ $(i = 1, ..., n)$. We use the convention that $T_{V,0}^0 = 1$, $T_{V,n+1}^0 = T$, $\lambda_{V,0}^0 = 0$ and $\lambda_{V,n+1}^0 = 1$. Under Assumption 1(a), the estimates are obtained as:

$$(\tilde{\sigma}_1, ..., \tilde{\sigma}_{n+1}, \tilde{T}_{V,1}, ..., \tilde{T}_{V,n}) = \underset{\sigma_1,...,\sigma_{n+1},T_{V,1},...,T_{V,n}}{\arg\min} \sum_{j=1}^{n+1} \sum_{t=T_{V,j-1}}^{T_{V,j}} (\hat{u}_t^2 - \sigma_j^2)^2$$

where $\hat{u}_t$ is defined by (3). Similarly, under assumption 1(b), we estimate the following AR(p)

$$\hat{u}_t = \sum_{j=0}^{p} \tilde{c}_j \hat{u}_{t-j} + \tilde{e}_t.$$

which yields the estimates $\tilde{c}_j$ $(j = 1, ..., p)$ and $\tilde{e}_t$ $(t = p+1, ..., T)$. The test will then be based on the residuals $\tilde{e}_t$ and the estimates of the break dates and variances in each regime given by

$$(\tilde{\sigma}_1, ..., \tilde{\sigma}_{n+1}, \tilde{T}_{V,1}, ..., \tilde{T}_{V,n}) = \underset{\sigma_1,...,\sigma_{n+1},T_{V,1},...,T_{V,n}}{\arg\min} \sum_{j=1}^{n+1} \sum_{t=T_{V,j-1}}^{T_{V,j}} (\tilde{e}_t^2 - \sigma_j^2)^2$$

where $T_{V,0}^0 = 0$ and $T_{V,n+1}^0 = T$. In both cases, the minimization problem is taken over the set

$$\Lambda_\varepsilon = \{\lambda_{V,1}, ..., \lambda_{V,n} | (\lambda_{V,j} - \lambda_{V,j-1}) > \epsilon\}.$$

The test is defined as

$$F_T^n = (\frac{T}{\hat{\psi}\hat{\sigma}^2}) \sum_{j=1}^{n_a} [(\tilde{\sigma}_j^2 - \tilde{\sigma}_{j+1}^2)^2]$$

The limit distribution of the test is presented in the following theorem, whose proof is standard and, hence, omitted.

Theorem 5 *Under Assumption 1(a) or 1(b), with or without a break in trend, under the null hypothesis of no break in variance, we have:*

$$F_T^n \Rightarrow \sum_{j=1}^{n}[\frac{B(\lambda_{V,j}^*) - B(\lambda_{V,j-1}^*)}{\lambda_{V,j}^* - \lambda_{V,j-1}^*}]^2$$

where

$$\{\lambda_{V,1}^*, ..., \lambda_{V,n}^*\} = \underset{\{\lambda_{V,1}, ..., \lambda_{V,n}\} \in \Lambda_\varepsilon}{\arg\min} \{\sum_{j=1}^{n+1}(B(\lambda_{V,j}) - B(\lambda_{V,j-1}))^2\}$$

3.8 Conclusions

We provide relevant results about testing for changes in the slope of the trend and the variance of the noise for a joint segmented trend. We start with a single possible break in each and address the following issues: 1) testing for a change in trend with or without a change in variance; 2) testing for a change in variance with or without a change in trend. We give results about the limit distribution of the estimates of the trend function allowing for a variance break, and the estimate for the change in variance allowing for a trend break. We propose asymptotically pivotal statistics for each cases. We also generalize some results to the case of multiple changes.

Appendix

Definition 9 *In a T-dimensional linear metric space $\mathbb{R}^T$, we define the following: a) the standard inner product: $\forall U, V \in \mathbb{R}^T$, $\langle U, V\rangle = U'V = \sum_{t=1}^{T} U_t V_t$; b) $||\cdot||$ as the Euclidian norm induced by the inner product $||U||^2 = \langle U, U\rangle = U'U$; c) the standard normalized orthogonal basis $\{E_t\}_{t=1}^T$ as $E_t = (0, ...0, 1, 0, ..0)'$ with its t-th element being 1 and the others being 0; d) the following vectors in $\mathbb{R}^T$: $\alpha_0 = (1, ..., 1)'$, $\alpha_1 = (1, ..., T)'$ and $\alpha_t = (0, .., 0, 1, ..., T - t + 1)'$. Also $\tilde{\alpha}_0, \tilde{\alpha}_t$ are the normalized vector of α_0, α_t: $\tilde{\alpha}_0 = \alpha_0/||\alpha_0||, \tilde{\alpha}_t = \alpha_t/||\alpha_t||$. Note that $\tilde{\alpha}_0 = T^{-1/2}\alpha_0$ and $\tilde{\alpha}_1 = \sqrt{3}T^{-3/2}\alpha_1$.*

Proof. of Theorem 1: The proof of Theorem **1** part 1) and 2) is a straightforward extension of some results in Perron and Zhu (2005) and, hence, omitted. To show 3), from regression by parts, we know that

$$\hat{\delta} = \frac{U' M_1 \alpha_{\hat{T}_B}}{\alpha'_{\hat{T}_B} M_1 \alpha_{\hat{T}_B}} = \frac{U' M_1 \alpha_{\hat{T}_B}}{SSR(\alpha_{\hat{T}_B}, \alpha_0, \alpha_1)}$$

where α_0 is the constant regressors, α_1 is the trend regressor and $M_1 = M(\alpha_0, \alpha_1) = (I - P(\alpha_0, \alpha_1))$ is the matrix that projects on the orthogonal complement to the range space of (α_0, α_1). Also, $SSR(\alpha_{\hat{T}_B}, \alpha_0, \alpha_1)$ is the sum of squared residuals from a regression of $\alpha_{\hat{T}_B}$ on (α_0, α_1). Denote the residuals from a regression of $\alpha_t = X(t)$ on (α_0, α_1) by $\tilde{X}_t = M(\alpha_0, \alpha_1)\alpha_t$. We have:

$$\begin{aligned}\langle \frac{\tilde{X}_t}{||\tilde{X}_t||}, U\rangle &= \int_0^1 g(\lambda, r) d\phi(r) \\ &= \omega_{10}[\int_0^{\lambda_V^0} g(\lambda_B^*, r) dW(r) + \omega_{20} \int_{\lambda_V^0}^1 g(\lambda_B^*, r) dW(r),\end{aligned}$$

and

$$\begin{aligned}\lambda_B^* &= \arg \max_{\lambda \in [\varepsilon, 1-\varepsilon]} \{\langle \frac{\tilde{X}_\lambda}{||\tilde{X}_\lambda||}, U\rangle^2\} \\ &= \arg \max_{\lambda \in [\varepsilon, 1-\varepsilon]} \{[\omega_{10}[\int_0^{\lambda_V^0} g(\lambda_B^*, r) dW(r) + \omega_{20} \int_{\lambda_V^0}^1 g(\lambda_B^*, r) dW(r)]^2\}\end{aligned}$$

. We elaborate on $\tilde{X}_t$ and its continuous analog $\tilde{X}_\lambda(r)$, which is the residual from regressing $\mathbf{1(r>}\lambda)(r-\lambda)$ continuously on $[1,r]$. The minimization problem is:

$$\tilde{X}_\lambda(r) = \min_{c_0,c_1} \int_0^1 [\mathbf{1(r>}\lambda)(r-\lambda) - c_0 - c_1 r]^2 dr$$

Tedious algebra and calculus lead us to its solution:

$$\begin{pmatrix} c_0 = -2\lambda(1-\lambda)^2 \\ c_1 = 2\lambda^3 - 3\lambda^2 + 1 \end{pmatrix}$$

and the residual function of $\tilde{X}_\lambda(r)$ is given by (7):

$$\begin{aligned} \tilde{X}_\lambda(r) &= g(\lambda, r) \\ &= \mathbf{1(r>}\lambda)(r-\lambda) + \lambda^3 - 2\lambda^2 + \lambda - 2\lambda^3 r + 3\lambda^2 r - r \end{aligned}$$

. Hence, with λ_B^* the limit of estimate of the trend break date fraction,

$$\begin{aligned} T^{3/2}\hat{\delta} &= \frac{U' M_1 \alpha_{\hat{T}_B}}{SSR(\alpha_{\hat{T}_B}, \alpha_0, \alpha_1)} \xrightarrow{d} \frac{\int_0^1 g(\lambda_B^*, r) d\phi(r)}{\int_0^1 g^2(\lambda_B^*, r) dr} \\ &= \frac{\omega_{10} \int_0^{\lambda_V^0} g(\lambda_B^*, r) dW(r) + \omega_{20} \int_{\lambda_V^0}^1 g(\lambda_B^*, r) dW(r)}{\int_0^1 g^2(\lambda_B^*, r) dr} \end{aligned}$$

■

Proof. Theorem 3: Denote the minimized sum of squared residuals from a regression of U on all regressors by $SSR(U)$ and the minimized sum of squared residuals from a regression of U on (α_0, α_1) by $SSR(U, \alpha_0, \alpha_1)$. We know that $SSR(U)$ can be generated by regressing the residuals from a regression of U on (α_0, α_1) on the residuals from a regression of $\alpha_{\hat{T}_B}$ on (α_0, α_1). Hence,

$$\begin{aligned} SSR(U) &= SSR(U, \alpha_0, \alpha_1) - U' P(\tilde{X}_{\hat{T}_B}) U \\ &= SSR(U, \alpha_0, \alpha_1) \\ &\quad - U'(\alpha_0, \alpha_1)\alpha_{\hat{T}_B} (\alpha'_{\hat{T}_B} M(\alpha_0, \alpha_1) \alpha_{\hat{T}_B})^{-1} \alpha'_{\hat{T}_B} (\alpha_0, \alpha_1) U \end{aligned}$$

with a variance break profile $\{\lambda_V^0, \omega_{10}, \omega_{20}\}$, the convergance rate for each parameter estimates remains invariant. Hence we can consistently estimate the parameters $\{\hat{\delta}, \hat{\beta}, \hat{\mu}, \hat{\lambda}_B, \hat{\lambda}_V, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\omega}_1, \hat{\omega}_2\}$. Because $\{u_t\}$ is stationary on each regime $\{1, ...[\lambda_V^0 T]\}$ and $\{[\lambda_V^0 T]+1, ..., T-[\lambda_V^0 T]\}$, we then do two re-searches of trend break dates using model of one trend break and no variance break in the regime $[0, \hat{\lambda}_V T]$ and $[\hat{\lambda}_V T, T]$ separately, assuming stationary innovation with variance $\hat{\sigma}_i$ and long-run variance $\hat{\omega}_i$, $i = 1, 2$ in each regime. Denote the estimates of the trend break date fraction and the trend break coefficient in each regime by $\hat{\lambda}_{B1}, \hat{\lambda}_{B2}$ and $\hat{\delta}_1, \hat{\delta}_2$, respectively, with $\varepsilon < \hat{\lambda}_{B1} < \hat{\lambda}_V - \varepsilon < \hat{\lambda}_V + \varepsilon < \hat{\lambda}_{B2} < 1 - \varepsilon$. Note that a minimum distance between λ_V^0 and λ_B^0 is required for this approach to work. Under the null hypothesis of no break in trends, we have

$$\hat{\lambda}_{B1} \implies \lambda_{B1}^* = \underset{\varepsilon<\lambda<\lambda_V^0-\varepsilon}{\arg\max} \frac{3}{(\lambda)^3(1-\lambda)^3}[\int_0^{\lambda_V^0} g(\lambda_{B1}^*, r)dW(r)]^2$$

$$\hat{\lambda}_{B2} \implies \lambda_{B2}^* = \underset{\lambda_V^0+\varepsilon<\lambda<1-\varepsilon}{\arg\max} \frac{3}{(\lambda)^3(1-\lambda)^3}[\int_{\lambda_V^0}^{1} g(\lambda_{B2}^*, r)dW(r)]^2$$

and

$$(\lambda_V^0)^3 T^3 \frac{\hat{\delta}_1^2}{\hat{\omega}_{10}^2} \implies \frac{3}{(\lambda_{B1}^*/\lambda_V^0)^3(1-\lambda_{B1}^*/\lambda_V^0)^3}[\int_0^1 g(\lambda_{B1}^*/\lambda_V^0, r)dW(r)]^2$$
$$\overset{d}{=} \max_{\varepsilon<\lambda_1<1-\varepsilon} \frac{3}{(\lambda_1)^3(1-\lambda_1)^3}[\int_0^1 g(\lambda_1, r)dW(r)]^2$$

$$(1-(\lambda_V^0)^3) T^3 \frac{\hat{\delta}_2^2}{\hat{\omega}_{20}^2} \implies \frac{3}{(\lambda_{B2}^*/\lambda_V^0)^3(1-\lambda_{B2}^*/\lambda_V^0)^3}[\int_0^1 g(\lambda_{B2}^*/\lambda_V^0, r)dW(r)]^2$$
$$\overset{d}{=} \max_{\varepsilon<\lambda_2<1-\varepsilon} \frac{3}{(\lambda_2)^3(1-\lambda_2)^3}[\int_0^1 g(\lambda_2, r)dW(r)]^2$$

Because $\hat{\delta}_1^2, \hat{\delta}_2^2$ are constructed on different regimes, they are independent so

$$\begin{aligned} F_{4,T} &= (\lambda_V^0)^3 T^3 \frac{\hat{\delta}_1^2}{\hat{\omega}_{10}^2} + (1-(\lambda_V^0)^3) T^3 \frac{\hat{\delta}_2^2}{\hat{\omega}_{20}^2} \\ &\Longrightarrow \max_{\varepsilon<\lambda_1<1-\varepsilon} \{ \frac{3}{(\lambda_1)^3(1-\lambda_1)^3} [\int_0^1 g(\lambda_1, r) dW(r)]^2 \} \\ &\quad + \max_{\varepsilon<\lambda_2<1-\varepsilon} \{ \frac{3}{(\lambda_2)^3(1-\lambda_2)^3} [\int_0^1 g(\lambda_2, r) dW(r)]^2 \} \end{aligned}$$

is a pivotal statistic for TP-4. ■

Proof. of Theorem 4: In what follows let $\{dW(r)\}_{r\in[0,1]}$ be a realized value of all differential increments $dW(r)$ and let Ω^∞ be the probability space with events that determine the value of $\{dW(r)\}_{r\in[0,1]}$. Recall from Theorem **1** that $|\hat{\lambda}_B - \lambda_B^0| = O_P(T^{-3/2})$ and

$$D_T^{1/2}(\hat{\kappa} - \kappa^0) \Rightarrow \Sigma_a(\lambda_B^0)^{-1}\xi(\lambda_B^0)$$

Recall that

$$\phi(r) = \begin{pmatrix} \omega_{10} W(r) \text{ when } r \leq \lambda_V^0 \\ \omega_{10} W(\lambda_V^0) + \omega_{20}[W(r) - W(\lambda_V^0)] \text{ when } r > \lambda_V^0 \end{pmatrix}$$

with

$$\xi(\lambda) = (\int_0^1 d\phi(r), \int_0^1 r d\phi(r), \int_\lambda^1 (r-\lambda) d\phi(r))'$$

To see how $d_T^\delta(\omega_{10}^2, \omega_{20}^2, \lambda_B^*, \lambda_V^0)$ is constructed, note that the estimate of the trend break coefficient $\hat{\delta}$, as the 3rd element of $D_T^{1/2}(\hat{\kappa} - \kappa^0)$, has the following limit distribution:

$$\begin{aligned} T^{3/2}\hat{\delta} &= [D_T^{1/2}(\hat{\kappa} - \kappa^0)]_3 \Rightarrow [\Sigma_a(\lambda_B^0)^{-1}\xi(\lambda_B^0)]_3 \\ &= 3\{\omega_{10} \int_0^{\lambda_V^0} [\frac{1}{(\lambda_B^0)^2(1-\lambda_B^0)} - \frac{(2\lambda_B^0+1)r}{(\lambda_B^0)^3(1-\lambda_B^0)} + \frac{1(r>\lambda_B^0)(r-\lambda_B^0)}{(\lambda_B^0)^3(1-\lambda_B^0)^3}] dW(r) \\ &\quad + \omega_{20} \int_{\lambda_V^0}^1 [\frac{1}{(\lambda_B^0)^2(1-\lambda_B^0)} - \frac{(2\lambda_B^0+1)r}{(\lambda_B^0)^3(1-\lambda_B^0)}] dW(r) \\ &\quad + \int_{\max\{\lambda_B^0, \lambda_V^0\}}^1 \frac{(r-\lambda_B^0)}{(\lambda_B^0)^3(1-\lambda_B^0)^3} dW(r)\} \\ &= \frac{3}{(\lambda_B^0)^3(1-\lambda_B^0)^3} \{\omega_{10}\xi_{V1}(\lambda_V^0, \lambda_B^0) + \omega_{20}\xi_{V2}(\lambda_V^0, \lambda_B^0)\} \end{aligned}$$

where

$$\begin{aligned}
\xi_{V1}(\lambda_V^0, \lambda_B^0) &= \int_0^{\lambda_V^0} [\lambda_B^0(1-\lambda_B^0)^2 - (2\lambda_B^0+1)(1-\lambda_B^0)^2 r + 1(r > \lambda_B^0)(r-\lambda_B^0)] dW(r) \\
&= \int_0^{\min\{\lambda_B^0, \lambda_V^0\}} [\lambda_B^0(1-\lambda_B^0)^2 - (2\lambda_B^0+1)(1-\lambda_B^0)^2 r] dW(r) \\
&\quad + \int_{\min\{\lambda_B^0, \lambda_V^0\}}^{\lambda_V^0} [\lambda_B^0(1-\lambda_B^0)^2 - (2\lambda_B^0+1)(1-\lambda_B^0)^2 r + (r-\lambda_B^0)] dW(r)
\end{aligned}$$

and

$$\begin{aligned}
\xi_{V2}(\lambda_V^0, \lambda_B^0) &= \int_{\lambda_V^0}^1 (\lambda_B^0(1-\lambda_B^0)^2 - (2\lambda_B^0+1)(1-\lambda_B^0)^2 r] dW(r) \\
&\quad + \int_{\max\{\lambda_B^0, \lambda_V^0\}}^1 (r-\lambda_B^0)] dW(r)\} \\
&= \int_{\lambda_V^0}^1 (\lambda_B^0(1-\lambda_B^0)^2 - (2\lambda_B^0+1)(1-\lambda_B^0)^2 r + \mathbf{1}(r > \lambda_B^0)(r-\lambda_B^0)] dW(r) \\
&= \int_{\lambda_V^0}^{\max\{\lambda_B^0, \lambda_V^0\}} [\lambda_B^0(1-\lambda_B^0)^2 - (2\lambda_B^0+1)(1-\lambda_B^0)^2 r] dW(r) \\
&\quad + \int_{\max\{\lambda_B^0, \lambda_V^0\}}^1 [\lambda_B^0(1-\lambda_B^0)^2 - (2\lambda_B^0+1)(1-\lambda_B^0)^2 r + (r-\lambda_B^0)] dW(r)
\end{aligned}$$

with the variance of $\xi_{V1}(\lambda_V^0, \lambda_B^0), \xi_{V2}(\lambda_V^0, \lambda_B^0)$ given by

$$\begin{aligned}
& Var[\xi_{V1}(\lambda_V^0, \lambda_B^0)] \\
= \; & \{[\lambda_V^0\lambda_B^0(1-\lambda_B^0)^4 - (\lambda_V^0)^2\lambda_B^0(2\lambda_B^*+1)(1-\lambda_B^0)^4 + \frac{(\lambda_V^0)^3}{3}(2\lambda_B^0+1)^2(1-\lambda_B^0)^4] \\
& -2[\lambda_V^0 - \min\{\lambda_B^0, \lambda_V^0\}]^2\lambda_B^0(1-\lambda_B^0)^2 - \frac{1}{3}[\lambda_V^0 - \min\{\lambda_B^0, \lambda_V^0\}]^3(4(\lambda_B^0)^3 - 6(\lambda_B^0)^2+1)]\}
\end{aligned}$$

and

$$
\begin{aligned}
& Var[\xi_{V2}(\lambda_V^0, \lambda_B^0)] \\
= \; & \{(1-\lambda_V^0)\lambda_B^0(1-\lambda_B^0)^4 - (1-(\lambda_V^0)^2)\lambda_B^0(2\lambda_B^0+1)(1-\lambda_B^0)^4 \\
& + \frac{1-(\lambda_V^0)^3}{3}(2\lambda_B^0+1)^2(1-\lambda_B^0)^4 \\
& -2[(1-\lambda_B^0)^2 - (\max\{\lambda_B^0, \lambda_V^0\} - \lambda_B^0\})^2]\lambda_B^0(1-\lambda_B^0)^2 \\
& -\frac{1}{3}[(1-\lambda_B^0)^3 - (\max\{\lambda_B^0, \lambda_V^0\} - \lambda_B^0)^3](4(\lambda_B^0)^3 - 6(\lambda_B^0)^2 + 1)\}
\end{aligned}
$$

This motivates defining the scaling statistic

$$
d_T^\delta(\omega_{10}^2, \omega_{20}^2, \lambda_B^0, \lambda_V^0) = \frac{(\lambda_B^0)^3(1-\lambda_B^0)^3}{3[\omega_{10}^2 Var[\xi_{V1}(\lambda_V^0, \lambda_B^0)] + \omega_{20}^2 Var[\xi_{V2}(\lambda_V^0, \lambda_B^0)])]^{1/2}},
$$

so that we have

$$
\hat{\delta}_{PIV} = T^{3/2} d_T^\delta(\omega_{10}^2, \omega_{20}^2, \lambda_B^0, \lambda_V^0)\hat{\delta} \xrightarrow{d} N(0,1)
$$

In practive, true values of parameters are replaced by their estimates. So

$$
\hat{d}_T^\delta(\hat{\omega}_1^2, \hat{\omega}_2^2, \hat{\lambda}_B, \hat{\lambda}_V) = \frac{(\hat{\lambda}_B)^3(1-\hat{\lambda}_B)^3}{3[\hat{\omega}_1^2 Var[\xi_{V1}(\hat{\lambda}_V, \hat{\lambda}_B)] + \hat{\omega}_2^2 Var[\xi_{V2}(\hat{\lambda}_V, \hat{\lambda}_B)])]^{1/2}}
$$

■

Table 3.1: Critical Values for Pivotal Tests

	CV for variance break test $F_{2,T}, F_{3,T}$			
ε\quantile	99%	97.5%	95%	90%
0.05	[-14.0,15.9]	[-12.4,13.7]	[-10.3, 10.72]	[-9.24, 9.37]
0.1	[-9.89,10.6]	[-9.18,8.87]	[-8.26, 7.96]	[-6.86, 6.82]
	CV for trend break test $F_{1,T}$			
ε\quantile	99%	97.5%	95%	90%
0.05	[-3.64,4.01]	[-3.35,3.76]	[-2.70,2.69]	[-2.28,2.31]
0.1	[-3.18,3.39]	[-2.79,3.13]	[-2.43,2.42]	[-2.17,2.13]
	CV for trend break test $F_{4,T}$			
ε\quantile	99%	97.5%	95%	90%
0.05	30.4	25.9	14.58	10.5
0.1	20.4	18.1	11.5	9.24

References

Altissimo, F., Corradi, V. (2003), "Strong Rules for Detecting the Number of Breaks in a Time Series," *Journal of Econometrics* 117, 207-244.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, 821-856.

Andrews, D.W.K., and Sun, Y. (2004), "Adaptive Local Polynomial Whittle Estimation of Long-range Dependence," *Econometrica* 72, 569-614.

Arteche, J. (2004), "Gaussian Semi-parametric Estimation in Long-memory in Stochastic Volatility and Signal Plus Noise Models," *Journal of Econometrics* 119, 131-154.

Arteche, J. (2006), "Semiparametric Estimation in Perturbed Long-memory Series," *Computational Statistics and Data Analysis* 51, 2118-2141.

Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics* 79, 551-63.

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Strucutral Changes," *Econometrica* 66, 47-78.

Bai, J. and Perron, P. , (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18, 1-22.

Baillie, R. T. (1996), "Long Memory Processes and Fractional Integration in Econometrics", *Journal of Econometrics* 73, 5-59.

Beran, J. (1995), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long-memory Autoregressive Integrated Moving Average Models", *Journal of the Royal Statistical Society* 57, 659-672.

Beran, J. (1998), *Statistics for Long-Memory Processes*, CRC Press.

Chen, C. and Tiao, G. C. (1990), "Random Level Shift Time Series Models, ARIMA Approximations, and Level-shift Detection", *Journal of Business and Economic Statistics 8*, 83-97.

Dahlhaus, R. (1989), "Efficient Parameter Estimation for Self Similar Processes," *The Annals of Statistics*, 17, 1749-1766.

Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press: Oxford.

Dehling, H. and Philipp, W. (1982), "Almost Sure Invariance Principles for Weakly Dependent Vector-Valued Random Variables," *The Annals of Probability* 10, 689-701.

Deo, R.S., and Hurvich, C.M. (2001), "On the Log Periodogram Regression Estimator of the Memory Parameter in Long-memory Stochastic Volatility Models," *Econometric Theory* 17, 686-710.

Diebold, F., and Inoue, A. (2001), "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131-159.

Dolado, J., Gonzalo, J., and Mayoral, L. (2005), "What Is What?: A Simple Test of Long-memory Versus Structural Breaks in the Time Domain," Unpublished Manuscript, Department of Economics, Universidad Carlos III de Madrid.

Deo, R. S. and Hurvich, C. M. (2001), "On the Log-periodogram Regression Estimator of the Memory Parameter in Long-memory Stochastic Volatility Models", *Econometric Theory* 17, 686-710.

Diebold, F. X. and Inoue, A. (2001), "Long Memory and Regime Switching", *Journal of Econometrics* 105, 131-159.

Estrada, F. P., Perron, P. and Martinez-López, B. (2013), "Statistically-derived Contributions of Diverse Human Influences to 20th Century Temperature Changes," *Nature Geoscience* 6, 1050-1055.

Fox, R., and Taqqu, M. (1986), "Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *The Annals of Statistics*, 14, 517-532.

Frederiksen, P., Nielsen, F. S., and Nielsen, M. O. (2012), "Local polynomial Whittle estimation of perturbed fractional processes," *Journal of Econometrics*, 167, 426-447.

Fuller, W. A. (1996), *Introduction to Time Series (2nd ed.)*, John Wiley, New York.

Geweke, J. and Porter-Hudak, S. (1983), "The Estimation and Application of Long-memory Time Series Models", *Journal of Time Series Analysis* 4, 221-238.

Granger, C. W. J. and Hyung, N. (2004), "Occasional Structural Breaks and Long-memory with An Application to the S and P 500 Absolute Stock Returns", *Journal of Empirical Finance* 11, 399-421.

Granger, C. W. J. and Joyeux, R. (1980), "Long Memory Relationships and the Aggregation of Dynamic Models", *Journal of Time Series Analysis* 1, 15-29.

Haldrup, N., and Nielsen, M. Ő. (2007), "Estimation of Fractional Integration in the Presence of Data Noise", *Computational Statistics and Data Analysis* 51, 3100-3114.

Hosking, J. (1981), "Fractional Differencing," *Biometrika*, 68, 165-176.

Hou, J., and and Perron, P. (2013), "Modified Local Whittle Estimator for Long Memory Processes in the Presence of Low Frequency (and Other) Contaminations", Unpublished Manuscript, Department of Economics, Boston University.

Hurst, H. (1951), "Long-term Storage Capacity of Reservoirs," *Transactions of the American Society of Civil Engineers*, 116, 770-799.

Hurvich, C.M., Moulines, E., and Soulier, P. (2005), "Estimating Long-memory in Volatility," *Econometrica*, 73, 1283-1328.

Hurvich, C.M., and Ray, B.K. (2003), "The Local Whittle Estimator of Long-memory Stochastic Volatility," *Journal of Financial Econometrics*, 1, 445-470.

Hurvich, C. M., and Moulines, E. and Soulier, P. (2005), "Estimating Long-memory in Volatility", *Econometrica* 78, 1283-1328.

Iacone, F. (2010), "Local Whittle Estimation of the Memory Parameter in Presence of Deterministic Components," *Journal of Time Series Analysis*, 31, 37-49.

Kejriwal, M. (2009), "Tests for a Mean Shift with Good Size and Monotonic Power," *Economics Letters* 102, 78-82.

Kejriwal, M., and Perron, P. (2010), "Testing for Multiple Structural Changes in Cointegrated Regression Models," *Journal of Business and Economic Statistics* 28, 503-522.

Kunsch, H. (1987), "Statistical Aspects of Self-similar Processes," *Proceedings of the First World Congress of the Bernoulli Society* Vol. 1. VNU Science Press, Utrecht, 67-74.

Leipus, I. N., and Viano, M. C. (2003), "Long Memory and Stochastic Trend", *Statistics and Probability Letters* 61, 177-190.

Lobato, I. and Savin, N. (1998), "Real and Spurious Memory Long-memory Properties of Stock Market Data", *Journal of Business and Economic Statistics* 16, 261-268.

Li, Y., and Perron. P. (2013), "Inference Related to Locally Ordered and Common Breaks in a Multivariate System with Joined Segmented Trends," Unpublished Manuscript, Department of Economics, Boston University.

Lu, Y. K., and Perron, P. (2010), "Modeling and Forecasting Stock Return Volatility Using A Random Level Shift Model", *Journal of Empirical Finance* 17, 138-156.

McCloskey, A. (2013), "Consistent Estimation of the Long-memory Stochastic Volatility Model Parameters That Is Robust to Level Shifts And Deterministic Trends," Unpublished Manuscript, Department of Economics, Brown University.

McCloskey, A., and Hill, J. B. (2013), "Parameter Estimation Robust to Low-frequency Contamination," Unpublished Manuscript, Department of Economics, Brown University.

McCloskey, A., and Perron, P. (2013), "Memory Parameter Estimation in the Presence of Level Shifts and Deterministic Trends," forthcoming in *Econometric Theory.*

Mikosch, T., and Stărică, C. (2004), "Nonstationarities in Financial Time Series, the Long-range Dependence, And the IGARCH Effects," *Review of Economics and Statistics*, 86, 378-390.

McCulloch, R. and Tsay, R. (1993), "Bayesian Inference And Prediction for Mean and Variance Shifts in Autoregressive Time Series", *Journal of the American Statistical Association* 88, 968-978.

Ohanissian, A., Russell, J. R. and Tsay, R. S. (2008), "True or Spurious Long-memory? A New Test", *Journal of Business and Economic Statistics* 26, 161-175.

Perron, P. (1989), "The Great Crash, the Oil Price Shock And the Unit Root Hypothesis", *Econometrica* 57, 1361-1401.

Perron, P. (1990), "Testing for A Unit Root in A Time Series Regression with Changing Mean", *Journal of Business and Economic Statistics* 8, 153-162.

Perron, P. (2006), "Dealing with Structural Breaks," in *Palgrave Handbook of Econometrics, Vol.1: Econometric Theory.*

Perron, P., and Qu, Z. (2007), "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica* 75, 459-502.

Perron, P., and Qu, Z. (2010), "Long-memory and Level Shifts in the Volatility of Stock Market Return Indices," *Journal of Business and Economic Statistics*, 28, 275-290.

Perron, P., and X. Zhu (2005), "Structural Breaks with Deterministic and Stochastic Trends," *Journal of Econometrics* 129, 65-119.

Qu, Z. (2011), "A Test against Spurious Long-memory", *Journal of Business and Economic Statistics* 29, 423-438.

Qu, Z., and Perron, P. (2013), "A Stochastic Volatility Model with Random Level Shifts And Its Application to S And P 500 And NASDAQ Return Indices," *Econometrics Journal* 16, 309-339.

Ray, B. K., and Tsay, R. S. (2002), "Bayesian Estimation for Change-point Detection in Long-range Dependent Processes", *Journal of Time Series Analysis* 23, 687-705.

Robinson, P. M. (1995a), "Gaussian Semi-parametric Estimation of Long Range Dependence", *The Annals of Statistics* 23, 1630-1661.

Robinson, P. M. (1995b), "Log-periodogram Regression of Time Series with Long Range Dependence", *The Annals of Statistics* 23, 1048-1072.

Sayginsoy, Ö., and T. J. Vogelsang (2011), "Testing for A Shift in Trend at An Unknown Date: A Fixed-b Analysis of Heteroskedasticity Autocorrelation Robust OLS-Based Tests," *Econometric Theory* 27, 992-1025.

Smith, A. (2005), "Level Shifts And the Illusion of Long-memory in Economic Time Series", *Journal of Business and Economic Statistics* 23, 355-389.

Stărică, C., and Granger, C.W.J. (2005), "Nonstationarities in Stock Returns," *The Review of Economics and Statistics* 87, 503-522.

Sun, Y., and Phillips, P. C. B. (2003), "Nonlinear Log-periodogram Regression for Perturbed Fractional Processes", *Journal of Econometrics* 115, 355-389.

Tsay, W.-J., and Hardle, W. K. (2009), "A Generalized ARFIMA Process with Markov-switching Fractional Differencing Parameter", *Journal of Statistical Computation and Simulation* 79, 731-745.

Varneskov, R.T., and Perron P. (2013), "Combining Long-memory And Level Shifts in Modeling and Forecasting the Volatility of Asset Returns," Unpublished Manuscript, Department of Economics, Boston University.

Xu, J., and Perron, P. (2012), "Modeling And Forecasting Stock Return Volatility: Level Shift Model with Time Varying Jump Probability And Mean Reversion," Unpublished Manuscript, Department of Economics, Boston University.

Curriculum Vitae

[redacted]