

2018

On some new advances in self-normalization approaches for inference on time series

<https://hdl.handle.net/2144/31689>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**ON SOME NEW ADVANCES IN SELF-NORMALIZATION
APPROACHES FOR INFERENCE ON TIME SERIES**

by

LILIYA LAVITAS

Specialist, Moscow State University, Russia, 2011

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

2018

© 2018 by
LILIYA LAVITAS
All rights reserved

Approved by

First Reader

Ting Zhang, PhD
Assistant Professor of Statistics

Second Reader

Mark Kon, PhD
Professor of Statistics

Third Reader

Ashis Gangopadhyay, PhD
Associate Professor of Statistics

Acknowledgments

I first would like to express my sincere gratitude to my advisor Professor Ting Zhang for his continuous support, guidance and encouragement over the last three years. Through the course of “A Seminar Course on Time Series Analysis” Professor Zhang taught and the countless meetings he have helped me to envision and to accomplish the work presented in this dissertation. I would also like to thank the members of my committee: Professor Mark Kon, Professor Ashis Gangopadhyay, Professor Uri Eden and Professor Daniel Sussman for their encouragement and insightful comments. I’d further like to thank Professor Murad Taqqu, Professor Eric D. Kolaczyk and Professor Mamikon Ginovyan for all they have taught me these last few years.

I would also like to thank my parents, friends and family for their support throughout the years.

Finally, I would like to thank my husband Pavel as without his love, support and encouragement none of this would be possible.

ON SOME NEW ADVANCES IN SELF-NORMALIZATION APPROACHES FOR INFERENCE ON TIME SERIES

LILIYA LAVITAS

Boston University, Graduate School of Arts and Sciences, 2018

Major Professor: Ting Zhang
Assistant Professor of Statistics

ABSTRACT

Statistical inference in time series analysis has been an important subject in various fields including climate science, economics, finance and industrial engineering among others. Numerous problems of research interest include statistical inference about unknown quantities, assessing structural stability and forecasting. These problems have been widely studied in the literature, but mainly for independent data, while in many applications involving time series data dependence is not unusual and in fact quite common. In this thesis, we incorporate serial dependence into the analysis by involving self-normalization in time series analysis.

We start with the problem of testing whether there are change-points in a given time series. The method we propose does not require the number of change-points to be predefined, and thus is unsupervised. It does not require any tuning parameters and can be applied to a wide class of quantities of interest. The asymptotic distribution of the test statistic is studied and an approximation scheme is proposed to reduce testing procedure complexity. We then consider the problem of construction of confidence intervals, for which the conventional self-normalizer exhibits certain degrees of asymmetry when applied to quantities other than the mean. The method we

propose provides a time-symmetric generalization to the conventional self-normalizer and leads to improved finite sample performance for quantities other than the mean.

Contents

1	Introduction	1
1.1	Self-Normalization for Independent Observations	1
1.2	Self-Normalization for Dependent Data	3
1.2.1	Self-Normalization in Change-Point Detection	7
1.2.2	Confidence Interval Construction for a Parameter in Time Series	14
1.3	Conclusion	16
2	Change point detection in the case of the mean function	18
2.1	Introduction and Problem Formulation	18
2.2	Proposed Test Statistics	21
2.3	Simulation Results	30
2.4	Conclusion	33
3	Alternative formulation for the test statistic	38
3.1	Introduction	38
3.2	Alternative formulation for the test statistic	39
3.3	Simulation Results	40
3.3.1	Computational complexity	40
3.3.2	Finite sample size and power properties	41
3.4	Conclusion	45
4	Change point detection in the case of other quantities of interest	47
4.1	Introduction	47

4.2	The Conventional Approach	49
4.3	A New Approach: The Contrast-Based Method	56
4.4	Implementation	59
4.4.1	An Approximation Scheme	59
4.5	Simulation Results	61
4.5.1	Comparison: The computationally efficient test and its full-scale counterpart	62
4.5.2	Comparison: The proposed contrast-based method and the conventional method	67
4.5.3	Comparison with the tests of Shao and Zhang (2010)	68
4.5.4	Testing for Change Points in Variance and Autocorrelation	70
4.6	A Real Data Analysis	74
4.7	Conclusion	76
5	A Time-Symmetric Self-Normalization Approach for Inference of Time Series	78
5.1	Introduction and Problem Formulation	78
5.2	T-Symmetric Self-Normalizer	81
5.3	Numerical Experiments	90
5.3.1	Monte Carlo Simulations	90
5.3.2	A Real Data Example	97
5.4	Conclusion	102
6	Conclusion	108
	Bibliography	112
	Curriculum Vitae	118

List of Tables

2.1	Empirical acceptance rates for testing change points in the mean of (4.12) under different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), results with 90% and 95% nominal levels are reported.	32
2.2	Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1) with different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	34
2.3	Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M2) with different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	35
2.4	Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M3) with different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	36
2.5	Empirical powers for testing change points in the mean of (M2) with $\rho = 0.3$ and the choice of trimming $\varepsilon \in \{0.02, 0.05, 0.08, 0.1, 0.2\}$ the significance level $\alpha = 0.05$ is used.	37

3.1	Relative computational time of the full-scale Alt test with respect to USN (baseline) for different sample sizes n . The results are based on average of five independent realizations.	41
3.2	Empirical acceptance rates for testing change points in the mean of (4.12) under different dependence strengths. For both the USN test and the alternative formulation Alt results with 90% and 95% nominal levels are reported. 41	
3.3	Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1)-(M3) with different dependence strengths. The proposed test is denoted by USN and it's alternative formulation version is denoted as Alt. For both tests the significance level $\alpha = 0.05$ is used.	43
3.4	Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1)-(M3) with different dependence strengths. The proposed test is denoted by USN and it's alternative formulation version is denoted as Alt. For both tests the significance level $\alpha = 0.05$ is used.	44
3.5	Empirical powers for testing change points in the mean of (M3) and (M3s), where the significance level is taken as $\alpha = 0.05$	45
4.1	Empirical acceptance rates for testing change points in the mean of (4.12) under different dependence strengths. For both the proposed UCbSN test and it's full-scale version results with 90% and 95% nominal levels are reported. 63	
4.2	Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1)-(M3) with different dependence strengths. The proposed test using the approximation scheme is denoted by UCbSN and it's full-scale version is denoted as UCbSN _{full} . For both tests the significance level $\alpha = 0.05$ is used.	64

4.3	Relative computational time of the full-scale UCbSN _{full} test with respect to UCbSN (baseline) for different sample sizes n . The results are based on average of five independent realizations.	65
4.4	Empirical power for testing change points in the mean of (M2) with $\rho = 0.3$ using grids with different densities. The significance level $\alpha = 0.05$ is used.	65
4.5	Empirical acceptance rates for testing change points in the median of under different dependence strengths. For the proposed unsupervised test, its contrast-based extension is denoted by UCbSN and the non-contrast-based extension is denoted by USN. For both test results with 90% and 95% nominal levels are reported.	68
4.6	Size-adjusted empirical powers for testing change points in the median of (4.12) under (M1)-(M3) with different dependence strengths. For the proposed unsupervised test, its contrast-based extension is denoted by UCbSN and the non-contrast-based extension is denoted by USN. For both test results with 90% and 95% nominal levels are reported.	69
4.7	Empirical acceptance rates for testing change points in the median of (4.12) under different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), results with 90% and 95% nominal levels are reported.	70
4.8	Size-adjusted empirical powers for testing change points in the median of (4.12) under (M1) with different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	71

4.9	Size-adjusted empirical powers for testing change points in the median of (4.12) under (M2) with different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	72
4.10	Size-adjusted empirical powers for testing change points in the median of (4.12) under (M3) with different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	73
4.11	Empirical powers for testing change points in the variance and first-order autocorrelation of (4.13) under different change-point alternatives. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	74
4.12	Empirical powers for testing change points in the first-order autocorrelation of (4.13) under different change-point alternatives. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.	75
4.13	Test statistics and their p -values for the tropical cyclone data.	76
4.14	Test statistics and their p -values for the tropical cyclone data.	76
5.1	Empirical coverage probabilities (in percentage) for confidence intervals constructed using different self-normalizers for the mean of the process (5.7) under different dependence strengths. The associated standard errors (also in percentage) are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	93

5.2	Average lengths of confidence intervals constructed using different self-normalizers for the mean of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	94
5.3	Empirical coverage probabilities (in percentage) for confidence intervals constructed using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors (also in percentage) are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	95
5.4	Average lengths of confidence intervals constructed using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	96
5.5	Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	98
5.6	Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	99
5.7	Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	100

5.8	Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	101
5.9	Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	102
5.10	Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	103
5.11	Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	104
5.12	Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.	105
5.13	Constructed 95% confidence intervals for the mean and the median using different methods for the relative daily change in the NMHC concentration during the period from March 10, 2004 to July 23, 2004.	107

List of Figures

2.1	The empirical rejection percentage of the naive SN-based test based on the alternative model 2.5	24
4.1	Satellite-derived lifetime-maximum wind speeds of tropical cyclones during 1981–2006.	76
5.1	Relative daily changes of NMHC concentrations from March 10, 2004 to July 23, 2004.	106

List of Abbreviations

$[k]$	Used for the integer part of k
$\mathbf{A} \preceq \mathbf{B}$	$\mathbf{B} - \mathbf{A}$ is positive semidefinite
H_0	The null hypothesis
H_1	The alternative hypothesis
\rightsquigarrow	The weak convergence on $(\{0\} \times [0, 1]) \cup ([0, 1] \times \{1\})$
$\mathbf{B}(\cdot)$	A standard multivariate Brownian motion

Chapter 1

Introduction

1.1 Self-Normalization for Independent Observations

The self-normalization approach in hypothesis testing and confidence interval construction has been studied and applied in statistics for many years. The work of William Gosset (who is widely known by his pseudonym “Student”) which was published in 1908 is identified as a prototypical example of a self-normalized approach in modern literature. Gosset introduced his renowned t -statistic based on a sample of normal independent identically distributed (later denoted as i.i.d.) observations X_1, \dots, X_n when he considered the problem of estimation of the mean μ in the framework of unknown variance σ^2 of the underlying distribution. For the sample mean $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$ and the sample variance $s_n^2 = (n-1)^{-1} \sum_{t=1}^n (X_t - \bar{X}_n)^2$ Gosset proposed the following form of the test statistic:

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{s_n}$$

to test the hypothesis that $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

Gosset (1908) studied and tabulated the distribution of T_n , later known in the literature as a t -distribution with $n-1$ degrees of freedom. He also showed that if certain assumptions are met, the distribution of T_n converges to the standard normal distribution as $n \rightarrow \infty$ even if assumption of normality of the X_i is eliminated; thus, T_n is asymptotically standard normal for the independent identically distributed

X_i . Before Gosset's work, the t -distribution was not known and standard normal distribution was used as an approximation for the distribution of T_n , which led to suboptimal performance in analyses of small samples.

To further illustrate the idea of self-normalization, we may consider the case of $\mu_0 = 0$ without loss of generality. For $\mu_0 = 0$, T_n can be written as:

$$T_n = \frac{\sqrt{n}(\bar{X}_n)}{s_n} = \frac{S_{1,n}}{V_n} \left\{ \frac{n-1}{n - (S_{1,n}/V_n)^2} \right\}^{\frac{1}{2}}, \quad (1.1)$$

where $S_{i,j} = \sum_{t=i}^j X_t$ and $V_n^2 = \sum_{t=1}^n X_t^2$. This form of T_n creates a new perspective which sheds light on relationship between limiting distribution of the self-normalized quantity $S_{1,n}/V_n$ and the T_n . It was studied in detail by Efron (1969) and Logat et al.(1973) who derived limiting distribution of the $S_{1,n}/V_n$ ratio and showed that it coincides with the limiting distribution of T_n .

Student's t -statistic became the foundation for a great number of works and studies in the late twentieth century. It has been generalized for statistical inference of many other quantities and functionals of distribution functions. The term "Studentized statistics" usually stands for the quantity $G_n(\theta) = (\hat{\theta}_n - \theta)/\hat{s}e_n$, where θ is a functional of the underlying distribution function F , $\hat{\theta}_n$ is usually defined as the corresponding sample version of θ and $\hat{s}e_n$ stands for a consistent estimator of the standard error of $\hat{\theta}_n$. Since limiting distribution of G_n is standard normal one can construct confidence intervals for θ based on the approximation through G_n . Multivariate generalization of $\mathbf{G}_n(\theta)$ can usually be obtained by replacing of consistent estimator for standard error with a consistent estimator of the covariance matrix of the $\hat{\theta}_n$. In order to find a consistent variance estimate, both univariate and multivariate generalizations require either derivation for the closed form for the asymptotic variance or use of resampling methods such as jackknife, bootstrapping and others.

1.2 Self-Normalization for Dependent Data

The primary focus of this dissertation is self-normalization as a statistical methodology for analysis of time series and dependent data. In this chapter, we will provide a detailed review of recent developments in self-normalization for dependent data. Readers who may be interested in a comprehensive summary of developments in self-normalization for independent data, refer to the book by Victor H. Pea and Shao (2009) and to the highly cited articles by Belloni et al. (2012), Belloni et al. (2014), Hansen et al. (2015). We will begin this chapter with an overview of the work by Lobato (2001), whose ideas became foundational for the novel studentization techniques for dependent data. Among others, one remarkable contribution by Lobato is the development of a framework which eliminates the need to choose tuning parameters. The latter are common in statistical methodologies for time series and can sometimes lead to conflicting conclusions. We will discuss this problem in detail in this dissertation. We will also provide a detailed overview of the works of Shao (2010a), Shao (2010b), Shao and Zhang (2010) and Shao (2015) which cover the latest developments in self-normalization methodologies in change-point detection and inference of time series. In addition, we will provide a brief overview of some extensions of the self-normalized approach to regression and other problems. The details on the application of self-normalization to other problems such as martingales estimation, long-range dependencies, multivariate regression and others can be found in Khan and Saleh (1997), Escanciano and Lobato (1997), Shao (2011) and Genay and Signori (2015).

Ideas published in the paper by Lobato (2001) are commonly cited as a major contribution to the recent developments in self-normalization. A major focus of his work is testing whether a possibly dependent stochastic process is uncorrelated up

to a certain order. Typical tests prior to this work were asymptotically correct for processes which are both uncorrelated and independent. Alternative tests described in the literature at that time have either required the selection of a user-chosen parameter or a bootstrap procedure. These bootstrap procedures involved the selection of some tuning parameters, such as block length. Thus, both approaches available at that time were dependent on some user-defined input. This is unfavorable due to the possibility of statistical inference being sensitive to this selection.

One of the key contributions of the work by Lobato (2001) was in proposing an alternative test statistic that does not require any user-defined numbers and whose asymptotic distribution does not contain any unknown parameters under the null hypothesis. In our work, we will focus on the same objective and will regard this property of a test statistic as indispensable.

The general case that Lobato (2001) has focused on in his work employs a key statistic T_K which is constructed through a vector of sample autocovariances $\mathbf{c} = (c_1, \dots, c_K)'$ normalized by a functional of the same vector \mathbf{c} .

In his work, Lobato makes an assumption that under the null hypothesis the observed process is uncorrelated up to the lag K , where K is unknown, but a fixed number. The vector of centralized partial sums of the underlying process X_t weakly converges to the Brownian motion process, so that $\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} \mathbf{z}_j$, where $\mathbf{z}_t = (z_{1,t}, \dots, z_{Kt})'$ and $z_{kt} = (X_t - \bar{X}_{1,n})(X_{t+k} - \bar{X}_{1,n})$ converges to the K -dimensional vector of independent Brownian motions. This invariance principle is not a trivial assumption and was later leveraged in multiple subsequent works involving self-normalization. This assumption is also crucial for many existing theorems in the field.

Lobato (2001) showed that under this invariance principle and the null hypothesis that the process is uncorrelated up to the lag K , the defined statistics T_K converges to a functional of a Brownian motion. Therefore, the asymptotic distribution of the

test statistic under the null hypothesis does not depend on any unknown parameters, which makes it pivotal.

In order to perform this test, one needs to calculate the value of T_n using the data observed and to compare it against the upper critical values of the limiting distribution. These critical values are provided by Lobato (2001). In order to study the behavior of the test under the alternative hypothesis, Lobato (2001) shows that statistic T_K diverges, which guarantees a nontrivial power of the test. The study of the test behavior under the alternative hypotheses will also be an important component of research presented in this dissertation. The work of Lobato (2001) demonstrated the importance of formulating a specific alternative hypothesis to guarantee some nontrivial power of the test. The alternative hypothesis of Lobato can be viewed as very general and non-restrictive because it implies a non-zero correlation in the process for some lag j , where j is unknown, but bounded by some fixed number K .

In addition to asymptotic properties of the proposed test statistic, Lobato (2001) provides a wide range of simulation studies and empirical examples to evaluate finite sample performance of the test and its potential for future analysis of real data.

The work of Lobato (2001) became influential for many researchers as it demonstrated a strong potential of self-normalization approach in analysis of dependent data. Analytical framework proposed by Lobato (2001) was later adopted for testing various statistical hypotheses and for the construction of confidence intervals.

The self-normalization approach formulated by Lobato (2001) was mainly developed for the stationary time series. Here we will review a work by Rho and Shao (2013) who proposed to generalize the approach of Lobato (2001) for the case of non-stationary process. Rho and Shao (2013) extended the application of the self-normalization idea to the time series regression model with non-stationary errors and unconditional heteroscedasticity. The framework developed by the authors is a gen-

eralization of the simple linear trend model that covers a wide class of non-stationary but weakly dependent models with unconditional heteroscedasticity.

The typical approach to inference of the model coefficients is through the ordinary least-square (OLS) estimators of the model coefficients. This approach is particularly difficult for the models with non-stationary errors due to the complex form of the asymptotic variance of the OLS estimator. One possible technique to address non-stationarity is to employ bootstrapping. A co-called Wild Bootstrap method of Wu (1986) can approximate the limiting distribution of the OLS estimators. Unfortunately, as the errors are not independent, the asymptotic variance cannot be captured properly without normalizing the OLS estimator. Appropriate normalization of the OLS estimator is done by Rho and Shao (2013) through the self-normalization. The main difference between the classical self-normalization method and the one proposed by Rho and Shao (2013) for the regression model is that the limiting distribution of the self-normalized quantity is not pivotal, but depends on the parameter which can be estimated through the bootstrap.

The method proposed by Rho and Shao (2013) offers a significant advantage over other approaches employed in the literature for estimation and inference of nonstationary time series regression models. The authors demonstrated that by employment of self-normalization, finite sample performance is improved in both simulated and real data. The improvements are also supported by theoretical results through analysis of limiting distributions of the OLS estimators and of the self-normalized quantity, which is defined through the function of the OLS estimators of the model coefficients normalized by a functional of the same OLS estimators.

This research work by Rho and Shao (2013) provides a new perspective on the possible applications of the self-normalization approach. It extends both the class of time series under consideration to the non-stationary and the class of studenti-

fied quantities of interest to the ones with non-pivotal limiting distribution. It also introduces a self-normalizer which has a trimming parameter ε , which is needed to guarantee the desired theoretical properties. Discussion on the trimming parameter proposed by Rho and Shao (2013) is also very important to our research. As discussed by the authors, trimming parameters in self-normalization are different from smoothing parameters, such as the truncation lag in long-run variance estimation (Newey and West, 1987; Andrews, 1991; Liu and Wu, 2010; Politis, 2011), the window size in dependent bootstrap (Künsch, 1989; Lahiri, 2003; Shao, 2010a; Zhou, 2013) or the subsampling width in subsampling methods (Hall and Jing, 1996; Politis et al., 1999; Zhang et al., 2013), as the effect of trimming is accounted for in the limiting distribution and its approximation. In our work, we will perform an additional sensitivity analysis on the choice of the trimming parameter ε .

A detailed analysis of times series regression models is out of scope of this dissertation. We suggest the works of Rho and Shao (2013), and Shao (2015) for further reading on this subject.

1.2.1 Self-Normalization in Change-Point Detection

Change-point detection is a wide field of research that has numerous applications. It was originally developed within the framework of quality control analysis, but has since evolved as a significant component of economics, environmental studies, financial analysis, etc. Over the last few decades, various statistical approaches have been studied in the context of structural stability. A review of the recent developments in this area can be found in Aue and Horváth (2013). Self-normalization technique for change-point detection has many recent developments and it is one of the major focuses of this dissertation. To demonstrate how self-normalization can be applied to change-point detection, we begin by testing for the presence of change-points in the

mean function.

The problem of testing for the presence of change-points in the mean function can naturally be formulated as a null hypothesis of the mean function having identical values for all observations in a given sample. Various alternatives to this hypothesis can be proposed and studied. One possible alternative to the null hypothesis of no change-points was studied in the work by Shao and Zhang (2010). Here we will provide a review of their results. Shao and Zhang (2010) have first considered a one-change point alternative hypothesis with an unknown location. They have also proposed extending their alternative hypothesis to account for multiple change points. This generalization requires the number of change-points to be specified.

As noted in Aue and Horváth (2013), a cumulative sum (CUSUM) process is foundational for many common statistical tests for the change point analysis. This process also forms the basis for the methodology proposed by Shao and Zhang (2010). CUSUM process tracks deviance between overall and partial mean of the underlying time series. Intuitively, one can expect the time series with no change-points to have a corresponding CUSUM process with values within a certain corridor. Thus, the CUSUM process that exceeds certain thresholds may indicate the presence of a change point in the underlying process.

Aue and Horváth (2013) state that much recent research in structural stability utilizes the invariance principle similar to the one proposed by Lobato (2001). Under this invariance principle, limiting distribution of the CUSUM process can be derived and is proportional to the one-dimensional Brownian motion. The exact form of the limiting distribution depends on the long-run variance of the underlying process, the estimation of which is in itself a nontrivial statistical problem.

As previously mentioned, this invariance principle is also crucially important for the theorems presented in this dissertation. It has been proven to hold for a wide range

of short-range dependent processes by using various dependence measures; see for example the mixing coefficients of Hannan (1979) and Herrndorf (1984), the functional dependence measure of Wu (2007) and Berkes et al. (2014), and references therein for other contributions.

Thus, testing for change-points through the CUSUM process leads to a problem of finding a consistent estimator for the long-run variance σ^2 , namely the spectral density at zero frequency. In order to find $\hat{\sigma}_n^2$, we either need to estimate an infinite number of parameters (γ_i for $i \in \mathbb{Z}$) or to use various approximations proposed in the literature. The methods available for estimation of σ^2 usually involve bandwidth selection; see for example Vogelsang (1999) and Crainiceanu and Vogelsang (2007). Thus, these methods depend on a user-defined tuning parameter, which needs to be avoided to derive consistent results.

In their paper, Shao and Zhang (2010) propose (i) to replace a consistent estimator of σ^2 with its inconsistent version and (ii) to find a quantity V_n , with the limiting distribution that is proportional to σ^2 . Therefore, under the null hypothesis of no change-point in the underlying data, the limiting distribution of the test statistic defined through the self-normalized CUSUM process would be pivotal. Unfortunately, even pivotal limiting distribution under the null hypothesis cannot guarantee that the test will demonstrate the desired power properties under the alternatives. Shao and Zhang (2010) show that for the naive choice of a self-normalizer, which does not take into account a specific form of the alternative hypotheses, a process can be constructed for which power of the test decreases to zero as the magnitude of change in the mean increases. This proves that a naive extension of the self-normalization idea fails if implemented without considering the alternative. Hence, the choice of the self-normalizer requires additional study.

In analyzing the reasons for the naive normalizer ultimate loss of power, Shao and

Zhang (2010) concluded that it is because it does not take into account the change-point alternative. A self-normalizer designed specifically for the one change point alternative can be used to resolve this problem. The authors introduce the following quantity:

$$V_n(k) = n^{-2} \left\{ \sum_{t=1}^k \left(S_{1,t} - \frac{t}{k} S_{1,k} \right)^2 + \sum_{t=k+1}^n \left(S_{t,n} - \frac{n-t+1}{n-k} S_{k+1,n} \right)^2 \right\}, k = 1, \dots, n-1 \quad (1.2)$$

where $S_{t_1, t_2} = \sum_{j=t_1}^{t_2} X_j$ if $t_1 \leq t_2$ and $S_{t_1, t_2} = 0$ otherwise. As one can see, $V_n(k)$ splits the sample in two parts and is designed for a one change-point alternative.

The asymptotic behavior of the test statistic defined with the normalizer $V_n(k)$ (1.2) was examined by Shao and Zhang (2010) who demonstrated that limiting distribution under the null hypothesis is pivotal for the mean function and thus can be tabulated. In addition, under the alternative hypothesis of one change point the test statistic diverges to ∞ in probability as the magnitude of the change in the mean increases at or above a rate of \sqrt{n} . This guarantees a nontrivial power of the test for one change-point alternative.

Finite sample performance of the proposed test statistic is examined in the paper using both simulated and real data. In all the presented examples, the proposed test demonstrated strong power and size results. Thus, the proposed test statistic for the one change-point alternative inherits the appealing features of the self-normalization approach. Namely, it does not require any user-defined parameters; therefore, it's asymptotic distribution under the null hypothesis is pivotal. In addition, this test is easy to implement.

Although Shao and Zhang (2010) focus their analysis on the one change-point alternative, they also discuss generalization of their test statistic to the multiple change-point alternative. The authors describe a procedure that accounts for a spe-

cific change-point alternative in the normalizer $V_n(k)$. They demonstrate how $V_n(k)$ can be modified for the two and three change points alternatives. Thus, they propose test statistics for alternatives with more than one change-point. However, these tests require the number of change-points being tested against to be pre-specified. The authors show that limiting distribution of their modified test statistic remains pivotal under the null hypothesis of no change-point in the mean function. Shao and Zhang (2010) do not provide the tabulated values of the limiting distribution because computational complexity of each multiple change-point alternative is significantly larger than the complexity of the test statistic for the one change point alternative. The alternative with multiple change points, for which the number of change points is not specified cannot be handled by the method of Shao and Zhang (2010). Since the actual number of change points is typically unknown, especially when one is at the stage of seeking a statistical test to determine their existence, it seems desirable if we can have an unsupervised counterpart that can be used for situations when there is no or ambiguous prior knowledge about the number of change points or when doubts are casted on this prior knowledge. In this dissertation, we will study change point detection problem for the case of multiple change points and will develop an unsupervised technique that does not require the number of change points in the alternatives to be pre-defined.

Change Point Detection for Quantities Other Than the Mean

The interest in change-point detection is not limited to the case of the mean function and has both theoretical and applied value in the more general settings. Change point detection in the variance, median, autocorrelation and other quantities of interest is of particular importance. In their work, Shao and Zhang (2010) extended the self-normalized test statistic for the case of the mean with conventional generalization.

Shao and Zhang (2010) use the following notation: \mathbf{F}^m is for the m -th marginal distribution of X_t , where the dimension m is fixed but arbitrary, $\mathbf{Y}_t = (X_t, \dots, X_{t+m-1})'$, $t = 1, \dots, N = n - m + 1$, and \mathbf{F}_t^m denotes the distribution of \mathbf{Y}_t . The authors focus on the change point detection in the quantity θ_t , where $\theta_t = \mathbf{T}(\mathbf{F}_t^m) \in \mathbb{R}^q$, $t = 1, \dots, N$ and \mathbf{T} is a functional that takes values in \mathbb{R}^q . $\hat{\mathbf{F}}_{j,k}$ denotes the empirical distribution of $(\mathbf{Y}_j, \dots, \mathbf{Y}_k)$, and $\mathbf{F}_{j,k}$ is for the uniform mixture of $\mathbf{F}_j, \dots, \mathbf{F}_k$, where $1 \leq j \leq k \leq N$. $\hat{\theta}_{j,k}$ denotes $\mathbf{T}(\hat{\mathbf{F}}_{j,k})$. Then the null and the alternative hypotheses can be formulated similarly to the case of the mean function discussed previously. Thus, the null hypothesis assumes no change-point in the quantity of interest, whereas the alternative hypothesis assumes the presence of one change-point with an unknown location.

The analysis by Shao and Zhang (2010) is restricted to the so-called approximately linear statistics. This class of statistics is fairly broad and includes, but is not limited to the cases of:

- marginal mean of X_t (the case discussed above in this chapter)
- marginal variance of X_t
- autocorrelation function at lags $(1, \dots, k)$
- quantiles of the distribution F^1 , including the median.

Natural generalization of the test statistic suggests the form for the normalizer which splits the underlying process in two parts, similar to the (1.2). This generalization mimics the idea of the CUSUM process with forward and backward recursive estimates so that the following representation holds:

$$t(\hat{\theta}_{1,t} - \hat{\theta}_{1,k}) = S_{1,t} - \frac{t}{k}S_{1,k}, \quad (1.3)$$

where $S_{1,t}$ is defined as in (1.2) This representation leads to the generalized analogue of the test statistic for the case of the mean.

In this dissertation we will propose a different generalization which will reduce to the form (1.2) in the case of the mean function, but will result in a different form of the normalizer in the general case.

Shao and Zhang (2010) provide tabulated values of the limiting distribution up to 10 dimensions. They also support their asymptotic results with a finite sample performance analysis using simulated and empirical data to show that their proposed test has nontrivial power even for small samples (sample size of $n = 200$ is studied). They showed that the test has a reasonable size and outperforms other tests previously proposed in the literature.

By focusing on the quantities that can be estimated through approximately linear statistics, the authors guarantee that Von Mises expansion holds for $\hat{\theta}_t$. (see Fernholz (1983), Frank R. Hampel (1986) for more details on Von Mises expansion for statistical functionals). Through this analysis, Shao and Zhang (2010) have demonstrated that self-normalization technique can be extended for the change-point detection in the cases other than the mean. They have also demonstrated how Von-Mises expansion can be leveraged in the theoretical analysis of the limiting distribution of self-normalized quantities.

It is important to note that similar to the case of the mean Shao and Zhang (2010) restrict their work to the alternative with one change point. A generalization to the multiple change-point alternatives is out of the scope of the work by Shao and Zhang (2010). In this dissertation, we will study multiple change-point alternatives which do not require pre-specification of the number of change points in the general framework.

1.2.2 Confidence Interval Construction for a Parameter in Time Series

As previously noted, change-point detection is not the only statistical problem with recent developments using self-normalization technique. Another problem with a long history of research interest is the construction of the confidence intervals for the unknown quantities. It has its origins in the central limit theorem, which was first formulated and studied for the independent, identically distributed (i.i.d.) data. In time series, the problem becomes significantly more challenging due to the presence of dependence. It is particularly difficult as it typically involves estimation of the asymptotic variances. As we have previously stated, many of the existing approaches depend on one or more user-defined parameters, the choice of which may lead to different results. The self-normalization approach can be employed to avoid user-dependent results because it uses asymptotically pivotal distributions. In this subsection, we will review several recent applications of self-normalization to the construction of confidence intervals.

Recent Variations of the Self-Normalization Approach to the Construction of Confidence Interval

Here we will review the research by Shao (2010b) and Shao (2015), who was the first to extend the concept of self-normalization to the construction of confidence intervals.

The framework in Shao (2010b) is similar to the one introduced in Shao and Zhang (2010). Shao has shown that the same technique of expansion through the influence functions can be applied not only to the change-point detection, but also to the problem of confidence interval construction.

By continuing to operate within the framework defined in (1.2.1), Shao (2010b) demonstrates that:

$$N(\theta_N - \theta)'W_N^{-1}(\theta_N - \theta) \Rightarrow U_q, \tag{1.4}$$

where U_q is a pivotal distribution discussed and tabulated by Lobato (2001) for $q = 1, \dots, 20$, where q is the number of dimensions and self-normalizer is defined as

$$W_N = N^{-2} \sum_{t=1}^N t^2 (\hat{\theta}_t - \hat{\theta}_N)(\hat{\theta}_t - \hat{\theta}_N)' \quad (1.5)$$

Thus, a $100(1 - \alpha)\%$ confidence interval for θ can be written as: $\{\theta : N(\theta_N - \theta)'W_N^{-1}(\theta_N - \theta) \leq U_{q,\alpha}\}$ In addition to the theoretical foundation, the Monte-Carlo simulations performed by Shao (2010b) showed strong finite sample performance.

To conclude, the test proposed in Shao (2010b) for the construction of confidence intervals inherits one of the most appealing properties of the self-normalization approach; namely, it does not involve any user-chosen parameters. The test is applicable to a wide class of statistics and is also fairly easy to implement as tabulated cut-off values are provided for up to 20-dimensional case.

According to Shao (2010b), the choice of self-normalizer W_N (1.5) is not unique, but the form of W_N specified by Shao (2010b) leads to the fairly simple form of the test statistic. An analysis of coverage rates and average interval lengths was out of the scope of his research. In his 2015 paper, Shao (2015) described several variants of the self-normalizers and analyzed the differences in their finite sample performance.

One possible self-normalizer can be constructed by replacing forward screening defined by W_N with its backward analogue $W_{1,N}$. Identical limiting distributions of $N(\theta_N - \theta)'W_{1,N}^{-1}(\theta_N - \theta)$ and $N(\theta_N - \theta)'W_N^{-1}(\theta_N - \theta)$ can be derived using the same assumptions as in Shao (2010b).

Going one step further, one can symmetrize the normalizer by taking an average of $W_{1,N}$ and W_N ; to construct a normalizer $W_{3,N} = (W_{1,N} + W_N)/2$. Symmetry is a rather natural property to require because in many practical cases it's desired to obtain the same confidence intervals by analyzing both the original time series and its

reversed version. That being said, there is more than one way to construct normalizers which fulfill the symmetry requirement. In the case of the mean function some of these normalizers might not coincide with the original self-normalizer, resulting in generally different confidence intervals. An example of this self-normalizer is also given in Shao (2015) and the difference in confidence intervals provided is noted to be an unfavourable property.

Small-scale simulation results accompany the discussion to compare performance of all of the normalizers introduced, though Shao (2015) suggests a more comprehensive analysis would be beneficial to further understand the difference in the finite sample performance of all the proposed tests.

In this dissertation we will construct a self-normalizer that will not only fulfill the symmetry requirement, but that also coincide with the original normalizer W_N in the case of the mean function, and in certain cases it might also yield the confidence intervals of smaller width than those demonstrated by Shao (2015). We will perform a more comprehensive study of coverage accuracy associated with all of the proposed tests for the case of the mean and other quantities of interest. In addition, we will provide theoretical support for the proposed self-normalizer's favourability over the other forms previously proposed.

1.3 Conclusion

In this chapter we have introduced the concept of self-normalization and provided a brief overview of its development since it was originally formulated in the early twentieth century. The latest research in self-normalization for dependent data was influenced significantly by the work of Lobato (2001), which we have covered in section 1.2. The detailed review of several recent developments in applications of self-

normalization approach for the dependent data is also given in this chapter.

In this dissertation, we will review our recent study of the self-normalization approach for the time series. We will particularly be interested in two applications of self-normalization: the change-point detection and the construction of confidence intervals. For both of these problems we have provided a review of recent developments and promised to extend the existing results as follows:

- We will propose a new test for the change-point detection which does not require the number of change-points in the alternative to be pre-specified. Therefore our test is unsupervised.
- We will propose a new approach to construction of confidence intervals which can be advantageous over the other forms proposed.

The rest of this dissertation is organized as follows. In Chapter 2, we will propose a new test for change-point detection for the case of the mean function. In Chapter 3, we will study an alternative formulation of the test statistic, which is designed to be potentially less sensitive to the magnitude of the first and the last change-points. In Chapter 4, we will demonstrate how our proposed test can be generalized to other quantities of interest, such as median, variance, etc. We will also propose a computational scheme to reduce the computational complexity of the test. In Chapter 5, we will address the problem of construction of confidence intervals.

Chapter 2

Change point detection in the case of the mean function

2.1 Introduction and Problem Formulation

The problem of assessing structural stability, briefly presented in Chapter 1, has been one of the major focuses of our research and will be covered in greater detail in this chapter and also in Chapters 3 and 4. The analysis that we are presenting here is based on the work of Zhang and Lavitas (2017).

Change-point detection problem has been traditionally phrased as a hypothesis testing problem with the null hypothesis indicating structural stability and the alternative containing one or multiple structural breaks, and one often relies heavily on statistical reasoning to find appropriate solutions. To avoid dependency on model specifications, one chooses to seek an approach that does not require the explicit definition of a certain time series model by studying the effect of dependence on the asymptotic distribution of given test statistics. This approach can be seen as a generalization of the methodology developed for independent observations with corrections for serial dependence. A cumulative sum (CUSUM) process is considered a foundational approach for this type of analysis. The CUSUM process that we have briefly introduced in (1.2.1) can be defined as following:

$$T_n(\lfloor nr \rfloor) = n^{-\frac{1}{2}} \sum_{t=1}^{\lfloor nr \rfloor} (X_t - \bar{X}_n), \quad r \in [0, 1], \quad (2.1)$$

Under certain regularity conditions, which we will discuss later, an invariance principle for the CUSUM process can be formulated as following:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} (X_t - E(X_t)) \rightarrow_D \sigma B(r), \quad (2.2)$$

where $\sigma^2 = \lim_{n \rightarrow \infty} n \text{var}(\bar{X}_{1,n}) = \sum_{k \in Z} \gamma(k) > 0$ is the long-run variance of the process $\{X_t\}$ and $B(r)$ is the one-dimensional Brownian motion.

Thus, to correct asymptotic distribution of the CUSUM process for serial dependency, a consistent estimator of the long-run variance, namely the spectral density at zero frequency, is required. The long-run variance involves autocovariances of all orders, and a user-defined bandwidth is typically needed for the estimator to be adaptive to the underlying dependence, which makes this method dependent on bandwidth specification. To avoid this issue, Shao and Zhang (2010) adopted the self-normalization idea of Lobato (2001) and extended it to the change-point problem as was briefly introduced above (1.2.1). Shao and Zhang (2010) exploited one of the main principles of self-normalization and instead of resorting to a consistent estimator of the long-run variance, they have leveraged a sequence of recursive estimators to form the normalizer, and in turn, pivotalize the asymptotic distribution of the test statistic.

As we have previously noted, the idea of self-normalization was originally developed in the context of a stationary time series, and its adaptation to change-point testing is not trivial. In particular, Shao and Zhang (2010) demonstrated that direct implementation of self-normalization can lead to inconsistent change-point tests. Specifically, the power of the resulting test may not necessarily increase to one as the alternative deviates from the null. The aforementioned paper took into account the change-point alternative by considering the situation with a single change-point and

devising a self-normalized test that has a monotonic power function. The extensions to multiple change-point alternatives, which will be discussed in detail later in this chapter, were also proposed in the same paper. The form of the test statistic, and thus its detailed implementation, depends on the number of actual change points, which is typically unknown in practice.

In our work we propose a new self-normalized test that does not require the *a priori* information on the number of actual change points yet still leads to a valid statistical procedure for testing change points. This is considered to be important because:

- the actual number of change points as required by the test of Shao and Zhang (2010) is typically unknown in practice
- when the number of change points is misspecified, the test of Shao and Zhang (2010) can suffer from power loss and in some situations the power can decrease to zero as the alternative deviates from the null; see the simulation study in Section 4.5

Shao and Zhang (2010) suggested estimating the number of change points before calling their testing procedure, for which they have mentioned several candidates including the information criterion of Yao (1988), the minimum description length criterion of Davis et al. (2006), and the sequential testing procedure of Bai and Perron (1998). Since different methods may yield different estimates for the number of potential change points, selecting the optimal one for this preliminary estimation step can be viewed as a tuning parameter selection problem for the test of Shao and Zhang (2010), and it can be a nontrivial task as the candidates mentioned by Shao and Zhang (2010) were mostly developed in different settings. Thus, ultimately, this approach is vulnerable to alternative misformulation. In addition, from a practical

point of view, it would be philosophically more natural to first consider a global test for the existence of change points, and if the existence is confirmed by the test, then further analysis may be called by the investigator to estimate the actual number and locations of change points. Therefore, it seems desirable if we can have a universal test as proposed in the current paper that does not require the *a priori* information on the number of actual change points. We also mention that the test of Shao and Zhang (2010) may become computationally prohibitive when the number of change points exceeds two, as it relies on simulation to obtain quantiles of the asymptotic distribution which is nonstandard and involves a maximization over a space whose dimension grows exponentially with the number of change points. In contrast, the current test, and thus its computational burden, do not depend on the number of change points. In Chapter 4 we will also propose an approximation scheme to further facilitate computations needed in the proposed testing procedure.

2.2 Proposed Test Statistics

The hypothesis of interest for this chapter is formulated in the following way. Given an observed time series X_1, \dots, X_n , the problem is to test the null hypothesis of no change point in the mean, namely

$$H_0 : E(X_1) = \dots = E(X_n) = \mu, \quad (2.3)$$

versus the alternative of having one or multiple change points, namely there exist $M \geq 1$ break points $k_0 = 1 < k_1 < \dots < k_M < n = k_{M+1}$ such that

$$E(X_i) \neq E(X_{i+1}), \quad i \in \{k_1, \dots, k_M\}, \quad (2.4)$$

and $E(X_i) = E(X_{i+1})$ otherwise. The actual number of change points M and their specific locations k_1, \dots, k_M are typically unknown in practice, especially when one is at the stage of seeking a global test for determining their existence. Let $\bar{X}_{j,k} = (k - j + 1)^{-1} \sum_{i=j}^k X_i$ be the sample mean of X_j, \dots, X_k and $S_{j,k} = \sum_{i=j}^k X_i$ be the associated partial sum, $1 \leq j \leq k \leq n$. In this notation the cumulative sum (CUSUM) process, which tracks if there is any significant deviation between the recursive mean $\bar{X}_{1,k} = S_{1,k}/k$ and the global mean $\bar{X}_{1,n} = S_{1,n}/n$ is defined as following:

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \bar{X}_{1,n}), \quad t \in [0, 1],$$

where $\lfloor nt \rfloor$ is the largest integer not exceeding nt . Under the null hypothesis of no change point, the process $\{Z_n(t)\}_{t \in [0,1]}$ is free of the unknown but constant mean parameter μ , and obtaining the appropriate cut-off value requires an asymptotic theory on the (centered) partial sum process

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \mu).$$

Asymptotic distribution of the CUSUM process under the null hypothesis is commonly analyzed through the Invariance Principle. We will formulate it here using the notation introduced above:

Let $B(\cdot)$ be a standard Brownian motion and \Rightarrow denote weak convergence in the Skorokhod space, then we assume the following:

(IP) There exists a $\sigma > 0$ such that $\{S_n(t), 0 \leq t \leq 1\} \Rightarrow \{\sigma B(t), 0 \leq t \leq 1\}$.

The invariance principle 2.2 has been proven to hold for a wide range of short-range dependent processes with the help of various dependence measures; see for example the mixing coefficients of Hannan (1979) and Herrndorf (1984), the functional dependence measure of Wu (2007) and Berkes et al. (2014), Billingsley (1999) and

Billingsley (2012) for other contributions. By Assumption (IP) and the continuous mapping theorem, one can obtain the weak convergence that

$$Z_n(t) \rightarrow_D \sigma\{B(t) - tB(1)\},$$

where the asymptotic distribution depends on the nuisance parameter σ . For time series data, the quantity σ^2 is typically the long-run variance which involves autocovariances of all orders, and a data-driven bandwidth is often needed for its estimator to be adaptive to the underlying dependence strength. However, as commented by Shao and Zhang (2010), both theoretical and empirical studies in the literature have found that the use of data-dependent bandwidth can lead to tests with nonmonotonic power. As we have shown in the previous chapter 1.2.1, to alleviate the problem, Shao and Zhang (2010) proposed to adopt the idea of self-normalization (Lobato, 2001 and Shao, 2010b) and have generalized it to the change-point setting. The main ingredient of self-normalization is to consider using a function of the CUSUM process as the normalizer to pivotalize the asymptotic distribution. In particular, we will here discuss an example of a naive normalizer, to illustrate the importance of incorporating alternative hypothesis into the form of the test statistic. Let

$$V_n^2 = \frac{1}{n^2} \sum_{i=1}^n \left\{ \sum_{j=1}^i (X_j - \bar{X}_{1,n}) \right\}^2,$$

then the continuous mapping theorem implies that

$$\frac{Z_n(t)^2}{V_n^2} \rightarrow_D \frac{\{B(t) - tB(1)\}^2}{\int_0^1 \{B(s) - sB(1)\}^2 ds},$$

where the asymptotic distribution no longer depends on the nuisance parameter σ and can thus be used to obtain the cut-off values. Nevertheless, this simple application of self-normalization unfortunately does not yield good performance for change-point

testing, and the main reason is that both $Z_n(t)$ and V_n can be affected by the change-point alternative, making the resulting test of less or no power. The simulations for the demonstration of this effect can be found in Shao and Zhang (2010) and we will summarize an experiment authors have constructed here. Let

$$X_t = \begin{cases} u_t, & \text{if } 1 \leq t \leq \frac{n}{2}; \\ \eta + u_t, & \text{if } \frac{n}{2} + 1 < t \leq n = 200. \end{cases} \quad (2.5)$$

where $u_t = 0.5u_{t-1} + \epsilon_t$, and $\epsilon_t \sim N(0, 1)$ and independent. From the plot below it is seen that when the magnitude of change η gets large, the power of the test constructed through $Z_n(t)^2/V_n^2$ decreases to zero.

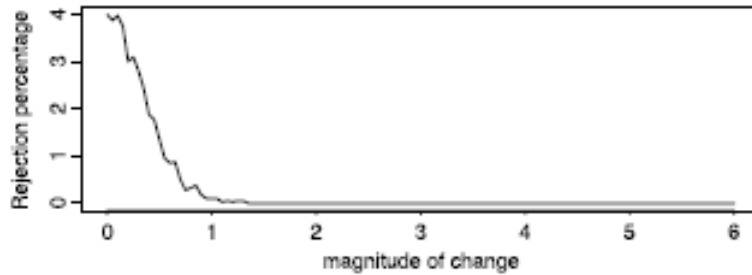


Figure 2.1: The empirical rejection percentage of the naive SN-based test based on the alternative model 2.5

To overcome this problem, Shao and Zhang (2010) proposed to split the data for calculating the self-normalizer according to the partition implied by hypothesized change-point locations (whose total number is assumed to be known), and after forming the self-normalized statistics for each possible partition with a prespecified number of segments, the maximum is then taken as the final test statistic. Such an approach is the most ideal for the situation with a single change-point alternative, on which the paper of Shao and Zhang (2010) was mainly focused. However, its application to situations with possibly more than one change points can pose certain challenges. To be more specific, the test of Shao and Zhang (2010) was developed under a supervised

learning setting where the total number of change points needs to be pre-specified. Thus, for each alternative a specific self-normalizer can be constructed. For example:

- form of the normalizer for the one change-point alternative:

$$\begin{aligned}
V_n(k) &= n^{-2} \left\{ \sum_{t=1}^k \left(S_{1,t} - \frac{t}{k} S_{1,k} \right)^2 \right. \\
&\quad \left. + \sum_{t=k+1}^n \left(S_{t,n} - \frac{n-t+1}{n-k} S_{k+1,n} \right)^2 \right\}, \quad k = 1, \dots, n-1.
\end{aligned} \tag{2.7}$$

where $S_{t_1, t_2} = \sum_{j=t_1}^t X_j$ if $t_1 \leq t_2$ and $S_{t_1, t_2} = 0$ otherwise.

- form of the normalizer for the two change-point alternative:

$$\begin{aligned}
V_{n_1+1, n_2}(k) &= (n_2 - n_1)^{-2} \left\{ \sum_{t=n_1+1}^k \left(S_{n_1+1, t} - \frac{t - n_1}{k - n_1} S_{n_1+1, k} \right)^2 \right. \\
&\quad \left. + \sum_{t=k+1}^{n_2} \left(S_{t, n_2} - \frac{n_2 - t + 1}{n_2 - k} S_{k+1, n_2} \right)^2 \right\},
\end{aligned} \tag{2.9}$$

where $n_1 + 1 \leq k \leq n_2 - 1$ and $S_{t_1, t_2} = \sum_{j=t_1}^{t_2} X_j$ if $t_1 \leq t_2$ and $S_{t_1, t_2} = 0$ otherwise.

However, the number of change-points is typically unknown in practice especially when one is at the preliminary stage of seeking a statistical test to determine the existence of change points. In addition, as the test based on the normalizers like 1.2 and 2.8 requires the calculation of self-normalized statistics for all possible partitions

with a prespecified number of segments, its computational burden will grow exponentially with the prespecified number of total change points. This, together with the large simulation needed to approximate the nonstandard asymptotic distribution with the same computational mechanism, makes the test of Shao and Zhang (2010) computationally prohibitive (or at least less favorable) for situations with more than two change points. Moreover, our simulation study in Section 4.5 indicates that, when the number of change points is misspecified, the test of Shao and Zhang (2010) can suffer from power loss and in some situations the power can decrease to zero as the alternative deviates from the null.

We shall here propose an alternative that does not require the *a priori* information on the number of change points yet still leads to a valid statistical procedure for change-point testing. The computational burden of the proposed test will remain at the same for situations with different numbers of change points. We will introduce the following notations:

$$D_{n,f}(j_1, j_2, j_3) = \frac{j_2 - j_1 + 1}{\sqrt{j_3 - j_1 + 1}} (\bar{X}_{j_1, j_2} - \bar{X}_{j_1, j_3}); \quad (2.10)$$

$$L_{n,f}(j_1, j_2, j_3) = \sum_{i=j_1}^{j_2} \left(\frac{i - j_1 + 1}{j_3 - j_1 + 1} \right)^2 (\bar{X}_{j_1, i} - \bar{X}_{j_1, j_2})^2; \quad (2.11)$$

$$R_{n,f}(j_1, j_2, j_3) = \sum_{i=j_2+1}^{j_3} \left(\frac{j_3 - i + 1}{j_3 - j_1 + 1} \right)^2 (\bar{X}_{i, j_3} - \bar{X}_{j_2+1, j_3})^2. \quad (2.12)$$

Since (2.10) has the equivalent form $D_{n,f}(j_1, j_2, j_3) = (j_3 - j_1 + 1)^{-1/2} \sum_{i=j_1}^{j_2} (X_i - \bar{X}_{j_1, j_3})$, it relates to the CUSUM process associated with X_{j_1}, \dots, X_{j_3} and recursively compares the forwarding partial mean \bar{X}_{j_1, j_2} , $j_2 \leq j_3$, with the overall mean \bar{X}_{j_1, j_3} to seek for potential change points. To make the normalization process not affected by the change-point alternative, one-sided self-normalizers are used and are given in

(2.11) and (2.12) respectively. Similarly, we introduce the backward version:

$$\begin{aligned}
D_{n,b}(j_1, j_2, j_3) &= \frac{j_3 - j_2 + 1}{\sqrt{j_3 - j_1 + 1}} (\bar{X}_{j_2, j_3} - \bar{X}_{j_1, j_3}); \\
L_{n,b}(j_1, j_2, j_3) &= \sum_{i=j_1}^{j_2-1} \left(\frac{i - j_1 + 1}{j_3 - j_1 + 1} \right)^2 (\bar{X}_{j_1, i} - \bar{X}_{j_1, j_2-1})^2; \\
R_{n,b}(j_1, j_2, j_3) &= \sum_{i=j_2}^{j_3} \left(\frac{j_3 - i + 1}{j_3 - j_1 + 1} \right)^2 (\bar{X}_{i, j_3} - \bar{X}_{j_2, j_3})^2.
\end{aligned}$$

Let $\Omega(\varepsilon) = \{(t_1, t_2) : \varepsilon \leq t_1 < t_2 \leq 1 - \varepsilon, t_2 - t_1 \geq \varepsilon\}$, $\Omega_n(\varepsilon) = \{(\lfloor nt_1 \rfloor, \lfloor nt_2 \rfloor) : (t_1, t_2) \in \Omega(\varepsilon)\}$, $\Xi_{n,f}(j_1, j_2, j_3) = L_{n,f}(j_1, j_2, j_3) + R_{n,f}(j_1, j_2, j_3)$ and $\Xi_{n,b}(j_1, j_2, j_3) = L_{n,b}(j_1, j_2, j_3) + R_{n,b}(j_1, j_2, j_3)$, our test statistic is then defined as

$$T_n = \max_{(l_1, l_2) \in \Omega_n(\varepsilon)} \frac{D_{n,f}(1, l_1, l_2)^2}{\Xi_{n,f}(1, l_1, l_2)} + \max_{(m_1, m_2) \in \Omega_n(\varepsilon)} \frac{D_{n,b}(m_1, m_2, n)^2}{\Xi_{n,b}(m_1, m_2, n)}. \quad (2.13)$$

The following theorem provides the asymptotic property of T_n under both the null and alternative.

Theorem 2.2.1. *Assume condition (IP). Then (i) under the null hypothesis (2.3), we have*

$$T_n \rightarrow_D T(\mathcal{B}) := \sup_{(r_1, r_2) \in \Omega(\varepsilon)} \frac{D(\mathcal{B}, 0, r_1, r_2)^2}{\Xi(\mathcal{B}, 0, r_1, r_2)} + \sup_{(s_1, s_2) \in \Omega(\varepsilon)} \frac{D(\mathcal{B}, s_1, s_2, 1)^2}{\Xi(\mathcal{B}, s_1, s_2, 1)}, \quad (2.14)$$

where

$$\begin{aligned}
D(\mathcal{B}, t_1, t_2, t_3) &= \frac{1}{\sqrt{t_3 - t_1}} \left[\mathcal{B}(t_2) - \mathcal{B}(t_1) - \frac{t_2 - t_1}{t_3 - t_1} \{ \mathcal{B}(t_3) - \mathcal{B}(t_1) \} \right], \\
\Xi(\mathcal{B}, t_1, t_2, t_3) &= \frac{1}{(t_3 - t_1)^2} \left(\int_{t_1}^{t_2} \left[\mathcal{B}(s) - \mathcal{B}(t_1) - \frac{s - t_1}{t_2 - t_1} \{ \mathcal{B}(t_2) - \mathcal{B}(t_1) \} \right]^2 ds \right. \\
&\quad \left. + \int_{t_2}^{t_3} \left[\mathcal{B}(t_3) - \mathcal{B}(s) - \frac{t_3 - s}{t_3 - t_2} \{ \mathcal{B}(t_3) - \mathcal{B}(t_2) \} \right]^2 ds \right);
\end{aligned}$$

and (ii) under the alternative (2.4) with $\min_{0 \leq i \leq M} |k_{i+1} - k_i|/n > \varepsilon$ and $E(X_{k_i+1}) - E(X_{k_i}) = n^{-1/2} L_i$ for some $L_i \neq 0$, $i = 1, \dots, M$, we have $T_n \rightarrow \infty$ in probability if

$\min_{1 \leq i \leq M} |L_i| \rightarrow \infty$.

Proof. Note that the partial sum $S_{j,k} = S_{1,k} - S_{1,j-1} = (k-j+1)\mu + \sqrt{n}[S_n(k/n) - S_n\{(j-1)/n\}]$, $1 \leq j \leq k \leq n$, we have

$$\begin{aligned} & \sqrt{[nr_2]/n}D_{n,f}(1, [nr_1], [nr_2]) - \sqrt{r_2}D(S_n, 0, r_1, r_2) \\ &= \{S_n([nr_1]/n) - ([nr_1]/[nr_2])S_n([nr_2]/n)\} - \{S_n(r_1) - (r_1/r_2)S_n(r_2)\} \\ &= \{(r_1/r_2) - ([nr_1]/[nr_2])\}S_n(r_2). \end{aligned}$$

Under condition (IP), both $\sup_{t \in [0,1]} |S_n(t)|$ and $\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |D_{n,f}(1, [nr_1], [nr_2])|$ are of order $O_p(1)$. Since $\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |(r_1/r_2) - ([nr_1]/[nr_2])| \leq 2/\{n(2\varepsilon - 1/n)^2\}$, we have

$$\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |\sqrt{[nr_2]/n}D_{n,f}(1, [nr_1], [nr_2]) - \sqrt{r_2}D(S_n, 0, r_1, r_2)| = O_p(n^{-1}),$$

and thus

$$\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |D_{n,f}(1, [nr_1], [nr_2]) - D(S_n, 0, r_1, r_2)| = O_p(n^{-1}).$$

Note that $\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |D_{n,f}(1, [nr_1], [nr_2])| = \max_{(l_1, l_2) \in \Omega_n(\varepsilon)} |D_{n,f}(1, l_1, l_2)|$ holds for all large n , by a similar argument and properties of the Brownian motion, it can be shown that

$$\left| T_n - \left\{ \sup_{(r_1, r_2) \in \Omega(\varepsilon)} \frac{D(S_n, 0, r_1, r_2)^2}{\Xi(S_n, 0, r_1, r_2)} + \sup_{(s_1, s_2) \in \Omega(\varepsilon)} \frac{D(S_n, s_1, s_2, 1)^2}{\Xi(S_n, s_1, s_2, 1)} \right\} \right| = o_p(1),$$

and (i) follows by the continuous mapping theorem. For (ii), under the alternative (2.4) with $E(X_{k_{i+1}}) - E(X_{k_i}) = n^{-1/2}L_i$, $i = 1, \dots, M$, we have

$$\begin{aligned} D_{n,f}(1, k_1, k_2) &= \frac{1}{\sqrt{k_2}} \left[\sum_{i=1}^{k_1} \{X_i - E(X_i)\} - \frac{k_1}{k_2} \sum_{j=1}^{k_2} \{X_j - E(X_j)\} \right] - \frac{k_1(k_2 - k_1)}{k_2\sqrt{nk_2}} L_1 \\ &:= D_{n,f,e}(1, k_1, k_2) - \frac{k_1(k_2 - k_1)}{k_2\sqrt{nk_2}} L_1. \end{aligned}$$

By the argument of (i) and the condition that $\min_{0 \leq i \leq M} |k_{i+1} - k_i|/n > \varepsilon$, one has

$$\frac{D_{n,f,e}(1, k_1, k_2)^2}{\Xi_{n,f}(1, k_1, k_2)} = O_p(1).$$

Note that $k_1(k_2 - k_1)/(k_2\sqrt{nk_2}) \geq \varepsilon^2$, the result follows by $\min_{1 \leq i \leq M} |L_i| \rightarrow \infty$. \square

To perform the test, by (2.14) we would reject the null hypothesis (2.3) at level $\alpha \in (0, 1)$ if $T_n > q_{1-\alpha}$ where $q_{1-\alpha}$ is the $(1-\alpha)$ -th quantile of $T(B)$, and its consistency is guaranteed by Theorem 2.2.1 (ii). The above testing procedure involves a trimming parameter ε , which controls the minimal length (in proportion) of partial sums. As commented by Zhou and Shao (2013), trimming parameters in self-normalization are different from smoothing parameters, such as the truncation lag in long-run variance estimation (Newey and West, 1987; Andrews, 1991; Liu and Wu, 2010; Politis, 2011), the window size in dependent bootstrap (Künsch, 1989; Lahiri, 2003; Shao, 2010a; Zhou, 2013) or the subsampling width in subsampling methods (Hall and Jing, 1996; Politis et al., 1999; Zhang et al., 2013), as the effect of trimming is accounted for in the limiting distribution and its approximation. Zhou and Shao (2013), along with Huang et al. (2015), conducted a series of numerical analyses and argued that the rule of thumb choice of $\varepsilon = 0.1$ seems to yield satisfactory performance for self-normalized methods. We have performed additional sensitivity analysis on the choice of the trimming parameter ε , and include it in the subsection (2.3). Our analysis confirms that performance of the proposed method is not very sensitive with respect to ε as long as ε is not too small or too large. Intuitively, if ε is too small, then the first few recursive estimates will be based on very few observations which can be unstable. On the other hand, the choice of ε being too large is also not desirable due to efficiency consideration. The proposed test has the potential of being generalized to quantities other than the mean, for example to quantiles, which we shall discuss in Chapter 4.

2.3 Simulation Results

We shall here carry out Monte Carlo simulations to examine the finite-sample performance of the proposed hypothesis testing procedure and make a comparison with the test of Shao and Zhang (2010). Due to the nonstandard limiting null distribution, critical values of both tests are obtained via simulation based on $n = 2,500$ and $10,000$ replications. Let $w_i, i \in \mathbb{Z}$, be independent standard normal random variables, and (e_i) be an autoregressive process satisfying the recursion $e_i = \rho e_{i-1} + w_i$. Let $1_{\{\cdot\}}$ denote the indicator function, we consider the model

$$X_i = \mu_i + e_i, \quad i = 1, \dots, n, \quad (2.15)$$

with the following change-point alternatives:

(M1) (one change-point alternative): $\mu_i = d1_{\{i/n > 2/3\}}$;

(M2) (two change-point alternative): $\mu_i = d1_{\{2/3 \geq i/n > 1/3\}} - d1_{\{i/n > 2/3\}}$;

(M3) (three change-point alternative): $\mu_i = d1_{\{i/n \leq 1/2\}} - d1_{\{i/n \leq 1/4\}} + d1_{\{i/n \geq 3/4\}}$,

where $d \in \mathbb{R}$ controls the amount of deviation from the null. For (M1)–(M3), the case with $d = 0$ corresponds to the null hypothesis of no change point. Let $n = 500$ and $\rho \in \{\pm 0.3, \pm 0.6, \pm 0.8\}$, we consider testing for change points in the mean of the process (4.12). For this, we consider the proposed new self-normalized test, namely the unsupervised self-normalized (USN) test, and compare it with the recent self-normalized test of Shao and Zhang (2010), which we will abbreviate as SZ10 hereafter. We shall here consider scenarios when prior information on the number of change-points is correctly specified and when it is misspecified. In particular, we consider SZ10 test statistics constructed under one, two and three hypothesized change points, which we denote by $SZ10_1$, $SZ10_2$ and $SZ10_3$ respectively. For the

proposed unsupervised test, it does not require this piece of prior information and the grid approximation scheme discussed in Section 4.4.1 is implemented to facilitate the computation. For the choice of the trimming parameter ε , Zhou and Shao (2013) and Huang et al. (2015) conducted a series of numerical analyses and argued that the rule of thumb choice of $\varepsilon = 0.1$ seems to yield satisfactory performance, which we shall use in our numerical analysis. Note that the effect of trimming, unlike the effect of smoothing as in nonparametric estimation, is accounted for in the limiting distribution and its approximation; see also the discussion in Zhou and Shao (2013). For each configuration, we generate 2,000 realizations, and the results are summarized in Sections 4.5.3 and 4.5.3 for comparisons under the null and under the alternative respectively.

Comparison: Empirical Acceptance Rate

We first examine the performance under the null, and Table 2.1 provides the empirical acceptance rates for 90% and 95% nominal levels. For the SZ10 test, it is supervised and it can be seen from Table 2.1 that using a larger prespecified number of change points generally leads to a larger size distortion for their test. In particular, the empirical acceptance rate of the SZ10₃ test is 0.913 for the mean case even with $\rho = 0.6$, which is far away from its nominal 95% level. In comparison, the proposed test is unsupervised and does not require prespecifying the number of change points. Its performance in size is comparable to the best of multiple change-point SZ10 tests in the mean case. Therefore, besides being unsupervised, the proposed test also seems to deliver more reasonable performance in size than its multiple change-point SZ10 competitors. The SZ10₁ test, though being designed only for cases with a single change point, seems to have the best performance in size due to its simpler form. However, it can be seen from Section 4.5.3 that the single change-point SZ10₁ test

ρ	USN		SZ10 ₁		SZ10 ₂		SZ10 ₃	
	90%	95%	90%	95%	90%	95%	90%	95%
<i>Results for the mean</i>								
0.3	0.886	0.944	0.896	0.950	0.882	0.942	0.890	0.938
0.6	0.872	0.930	0.897	0.944	0.868	0.929	0.854	0.913
0.8	0.794	0.872	0.889	0.936	0.795	0.881	0.724	0.807
-0.3	0.916	0.963	0.892	0.950	0.910	0.955	0.933	0.968
-0.6	0.938	0.975	0.909	0.958	0.941	0.972	0.963	0.983
-0.8	0.960	0.981	0.920	0.963	0.958	0.981	0.984	0.994

Table 2.1: Empirical acceptance rates for testing change points in the mean of (4.12) under different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), results with 90% and 95% nominal levels are reported.

can suffer from a serious power issue in the presence of multiple change points.

Comparison: Empirical Power

We shall now consider (M1)–(M3) and examine the performance under the alternative. Tables 2.2–4.10 provide the size-adjusted empirical powers under (M1)–(M3) respectively for $\rho \in \{0.3, 0.6\}$. For the SZ10 test, it requires prespecifying the number of change points and it can be seen from Tables 2.2–4.10 that, depending on whether this prior information is correctly specified or not, the power performance of their test can be largely affected. For example, when there is only a single change point as in (M1), then prespecifying two or more change points can lead to serious power losses for the SZ10 test, as in this case empirical powers of SZ10₂ and SZ10₃ are much lower than that of SZ10₁ from Table 2.2. On the other hand, when there are multiple change points as in (M2) and (M3), then serious power losses can be observed for the single change-point SZ10₁ test from Tables 2.3 and 4.10. In contrast, the proposed test does not require prespecifying the number of change points and seems to deliver a competitive power performance even when compared with the oracle SZ10 test. The

oracle SZ10 test is defined as the SZ10 test where the number of change points is assumed to be known and correctly prespecified, which corresponds to $SZ10_1$, $SZ10_2$ and $SZ10_3$ under (M1), (M2) and (M3) respectively. The oracle SZ10 test represents the power upper bound of SZ10 tests but is typically not obtainable in practice due to the unknown number of change points. It can be seen from Tables 2.2–4.10 that the power performance of the proposed test is mostly comparable to that of the oracle SZ10 test in the mean case. Therefore, in terms of the power performance, the proposed test seems to be more reliable in situations where the number of change points is unknown.

Sensitivity Analysis

In this subsection, we investigate the finite-sample performance of the proposed test with respect to the trimming parameter ε . For this, we consider model (M2) with $\rho = 0.3$, and would apply the proposed test with different choices of $\varepsilon \in \{0.02, 0.05, 0.08, 0.1, 0.2\}$ to test for change points in the mean. Let $n = 500$, the results are summarized in 2.5. It can be seen from Table 2.5 that the performance of the proposed test is not very sensitive with respect to ε as long as ε is not too small or too large. Intuitively, we can observe that if ε is too small, then the first few recursive estimates will only be based on very few first observations which can be unstable. On the other hand, the choice of ε being too large is also not desirable as the test's efficiency might be affected thought it.

2.4 Conclusion

The influential work of Shao and Zhang (2010) considered the situation with a single change-point alternative in time series and proposed a consistent self-normalized test for it. Its extension to multiple change-point alternatives was also given but

ρ	d	<i>Mean</i>			
		USN	SZ10 ₁	SZ10 ₂	SZ10 ₃
0.3	0.0	0.050	0.050	0.050	0.050
	0.2	0.147	0.198	0.110	0.074
	0.4	0.438	0.544	0.292	0.170
	0.6	0.792	0.845	0.598	0.405
	0.8	0.956	0.962	0.856	0.657
	1.0	0.994	0.991	0.971	0.872
	2.0	1.000	1.000	1.000	1.000
	3.0	1.000	1.000	1.000	1.000
	4.0	1.000	1.000	1.000	1.000
	5.0	1.000	1.000	1.000	1.000
0.6	0.0	0.050	0.050	0.050	0.050
	0.2	0.076	0.098	0.068	0.056
	0.4	0.161	0.216	0.108	0.074
	0.6	0.324	0.410	0.197	0.125
	0.8	0.546	0.623	0.354	0.216
	1.0	0.736	0.794	0.525	0.321
	2.0	0.998	0.998	0.984	0.936
	3.0	1.000	1.000	1.000	0.998
	4.0	1.000	1.000	1.000	1.000
	5.0	1.000	1.000	1.000	1.000

Table 2.2: Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1) with different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

requires prespecifying the actual number of change points under the alternative and is thus supervised. Since the actual number of change points is typically unknown especially when one is at the stage of seeking a statistical test to determine their existence, it seems desirable if we can have an unsupervised counterpart as considered in the current paper that can be used for situations when there is no or ambiguous prior

ρ	d	<i>Mean</i>			
		USN	SZ10 ₁	SZ10 ₂	SZ10 ₃
0.3	0.0	0.050	0.050	0.050	0.050
	0.2	0.299	0.132	0.316	0.243
	0.4	0.844	0.208	0.850	0.776
	0.6	0.992	0.166	0.992	0.988
	0.8	1.000	0.141	1.000	1.000
	1.0	1.000	0.110	1.000	1.000
	2.0	1.000	0.008	1.000	1.000
	3.0	1.000	0.000	1.000	1.000
	4.0	1.000	0.000	1.000	1.000
	5.0	1.000	0.000	1.000	1.000
0.6	0.0	0.050	0.050	0.050	0.050
	0.2	0.117	0.086	0.115	0.098
	0.4	0.344	0.154	0.346	0.251
	0.6	0.673	0.178	0.682	0.542
	0.8	0.893	0.195	0.902	0.824
	1.0	0.974	0.166	0.975	0.954
	2.0	1.000	0.058	1.000	1.000
	3.0	1.000	0.014	1.000	1.000
	4.0	1.000	0.002	1.000	1.000
	5.0	1.000	0.000	1.000	1.000

Table 2.3: Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M2) with different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

knowledge about the number of change points or when doubts are casted on this prior knowledge. As can be seen from our simulation study in Section 4.5, the supervised test of Shao and Zhang (2010) can suffer from serious power losses when the prespecified value on the number of change points is misspecified, and in certain situations the power of their test can even decrease to zero as the alternative deviates from the

ρ	d	<i>Mean</i>			
		USN	SZ10 ₁	SZ10 ₂	SZ10 ₃
0.3	0.0	0.050	0.050	0.050	0.050
	0.2	0.118	0.051	0.111	0.119
	0.4	0.292	0.040	0.225	0.346
	0.6	0.660	0.023	0.526	0.732
	0.8	0.903	0.006	0.792	0.937
	1.0	0.986	0.001	0.941	0.994
	2.0	1.000	0.000	1.000	1.000
	3.0	1.000	0.000	1.000	1.000
	4.0	1.000	0.000	1.000	1.000
	5.0	1.000	0.000	1.000	1.000
0.6	0.0	0.050	0.050	0.050	0.050
	0.2	0.069	0.048	0.070	0.070
	0.4	0.115	0.058	0.118	0.128
	0.6	0.209	0.042	0.184	0.260
	0.8	0.381	0.034	0.306	0.430
	1.0	0.560	0.020	0.455	0.636
	2.0	0.992	0.002	0.968	0.996
	3.0	1.000	0.000	1.000	1.000
	4.0	1.000	0.000	1.000	1.000
	5.0	1.000	0.000	1.000	1.000

Table 2.4: Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M3) with different dependence strengths. For both the proposed USN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

null. In contrast, the proposed unsupervised test does not require prespecifying the number of change points and seems to perform reasonably well for situations with different numbers of change points.

d	ε				
	0.02	0.05	0.08	0.1	0.2
0.0	0.108	0.070	0.059	0.050	0.048
0.2	0.354	0.346	0.320	0.358	0.349
0.4	0.851	0.853	0.890	0.890	0.876
0.6	0.990	0.991	0.996	0.996	0.995
0.8	1.000	1.000	1.000	1.000	1.000
1.0	1.000	1.000	1.000	1.000	1.000
2.0	1.000	1.000	1.000	1.000	1.000
3.0	1.000	1.000	1.000	1.000	1.000

Table 2.5: Empirical powers for testing change points in the mean of (M2) with $\rho = 0.3$ and the choice of trimming $\varepsilon \in \{0.02, 0.05, 0.08, 0.1, 0.2\}$ the significance level $\alpha = 0.05$ is used.

Chapter 3

Alternative formulation for the test statistic

3.1 Introduction

In chapter 2 we have propose a new self-normalized test that does not require the *a priori* information on the number of actual change points yet still leads to a valid statistical procedure for testing change points. This an important charasteristic of a test as (i) the actual number of change points as required by the test of Shao and Zhang (2010) is typically unknown in practice; (ii) when the number of change points is misspecified, the test of Shao and Zhang (2010) can suffer from power loss and in some situations the power can decrease to zero as the alternative deviates from the null

The proposed test is constructed based on the forward scanning process as well as the backward scanning process. The forward scanning process is inspecting the observed data based on the first k_2 observations $(X_1, X_2, \dots, X_{k_2})$ to test for the presence of a change-point within this subset. Since k_2 is unknown, the maximum over all possible k_2 will guarantee that change-point would be detected. The same idea also applies to the backward scanning process.

One might be interested in analyzing sensitivity of the test to the magnitude of the first and the last change points. Thus one would consider to design a test whose power performance is less sensitive to the magnitude of the first and the last change

points. In this chapter we will propose an alternative formulation of the test statistic and will demonstrate a numerical comparison of power properties of the proposed USN test (defined in 2.13) and the alternative test. In addition, we will examine computational comparison of the both approaches.

The work presented in this chapter was completed as a part of the research project by Zhang and Lavitas (2017).

3.2 Alternative formulation for the test statistic

We will remind that for the original test statistic T_n (see 2.13) both forward and backward scanning processes are examining the underlying data for all its subsets starting from the X_1 for the forward version and ending with X_n for the backward version. As commented above, one might be concerned that this may lead to the unpleasant sensitivity to the magnitude of the first and the last change-points. More specifically, if the magnitude of the first and the last change-points is significantly smaller than the magnitude of the rest change-points one might be concerned that this might negatively affect the power of the test.

To address these concerns we will not only analyze the empirical power of the USN test based on the Monte-Carlo simulations, but will also construct an alternative test to examine its power and size properties.

Let $\Omega(\varepsilon) = \{(t_1, t_2, t_3) : \varepsilon \leq t_1 < t_2 < t_3 \leq 1 - \varepsilon, t_2 - t_1 \geq \varepsilon, t_3 - t_2 \geq \varepsilon, \Omega_n(\varepsilon) = \{(\lfloor nt_1 \rfloor, \lfloor nt_2 \rfloor, \lfloor nt_3 \rfloor) : (t_1, t_2, t_3) \in \Omega(\varepsilon)\}$. We shall here consider an alternative formulation of the test statistic, which we further denote as Alt:

$$T_{nAlt} = \max_{(l_1, l_2, l_3) \in \Omega_n(\varepsilon)} \frac{D_{n,f}(l_1, l_2, l_3)^2}{\Xi_{n,f}(l_1, l_2, l_3)} + \max_{(m_1, m_2, m_3) \in \Omega_n(\varepsilon)} \frac{D_{n,b}(m_1, m_2, m_3)^2}{\Xi_{n,b}(m_1, m_2, m_3)}, \quad (3.1)$$

where $D_{n,f}(l_1, l_2, l_3)$, $D_{n,f}(l_1, l_2, l_3)$, $\Xi_{n,b}(l_1, l_2, l_3)$, and $\Xi_{n,f}(l_1, l_2, l_3)$ are defined as in 2.12. Following argument similar to the Chapter 2, $T_{n_{Alt}}$ is expected to have similar asymptotic properties as the test (2.14).

To perform the test, by (3.1) we would reject the null hypothesis (2.3) at level $\alpha \in (0, 1)$ if $T_{n_{Alt}} > q_{1-\alpha}$ where $q_{1-\alpha}$ is the $(1 - \alpha)$ -th quantile of $T_{Alt}(B)$. The above testing procedure involves same trimming parameter ε , which controls the minimal length (in proportion) of partial sums. As was noted before, the effect of trimming is accounted for in the limiting distribution and its approximation. Similary to the Chapter 2 we will be using the rule of thumb choice of $\varepsilon = 0.1$. Please, see discussion in Chapter 2 on the choice of the trimming parameter.

3.3 Simulation Results

In this section we will perform Monte-Carlo simulations to evaluate the alternative form of the test statistic, denoted by Alt. Our main goal is to compare the alternative formulation of the test with the USN test, proposed in Chapter 2

3.3.1 Computational complexity

First, as one can observe the alternative form of the test statistic requires more computation due to its involvement of maximizing over $O(n^3)$ terms. This computational complexity may possibly make this test computationally prohibitive in particular for the larger samples. To illustrate this we provide the computation time of this alternative test in terms of its ratio with respect to that of the proposed USN test (baseline) for different sample sizes $n = 100k$, $k \in \{1, 2, 3, 4, 5\}$ in Table 3.1. One can see from this comparison that the proposed USN test is as expected computationally much more friendly comparing with it's alternative formulation.

n	100	200	300	400	500
USN	1.00	1.00	1.00	1.00	1.00
Alt	13.33	25.63	38.51	47.89	61.10

Table 3.1: Relative computational time of the full-scale Alt test with respect to USN (baseline) for different sample sizes n . The results are based on average of five independent realizations.

ρ	USN		Alt	
	90%	95%	90%	95%
0.3	0.886	0.944	0.898	0.941
0.6	0.872	0.930	0.839	0.901
0.8	0.794	0.872	0.665	0.784
-0.3	0.916	0.963	0.940	0.972
-0.6	0.938	0.975	0.974	0.988
-0.8	0.960	0.981	0.988	0.998

Table 3.2: Empirical acceptance rates for testing change points in the mean of (4.12) under different dependence strengths. For both the USN test and the alternative formulation Alt results with 90% and 95% nominal levels are reported.

3.3.2 Finite sample size and power properties

As we have demonstrated with Table 3.1 the alternative formulation test has a much higher computational time due to its more complex form. Following the general rule of thumb one might expect the alternative formulated test to have also a larger size distortion than the proposed test due to its more complex form. We will further examine size properties of the alternative formulation of the test statistic for different dependency strengths $\rho \in \{\pm 0.3, \pm 0.6, \pm 0.8\}$ and the sample size of $n = 500$. As Table (3.2) suggests the Alt test generally has a larger size distortion than the proposed USN test. Please refer to the Table (3.2) below which compares the empirical acceptance rates for testing change points in the mean under different dependence strengths for the comparison.

We also provide a power comparison of the alternative formulation test and the proposed test in the Tables 4 and 5 of this chapter. We will simulate a time series process defined in a same way as in Chapter 2. Let w_i , $i \in \mathbb{Z}$, be independent standard normal random variables, and (e_i) be an autoregressive process satisfying the recursion $e_i = \rho e_{i-1} + w_i$. Let $1_{\{\cdot\}}$ denote the indicator function, we consider the model

$$X_i = \mu_i + e_i, \quad i = 1, \dots, n, \quad (3.2)$$

with the following change-point alternatives:

(M1) (one change-point alternative): $\mu_i = d1_{\{i/n > 2/3\}}$;

(M2) (two change-point alternative): $\mu_i = d1_{\{2/3 \geq i/n > 1/3\}} - d1_{\{i/n > 2/3\}}$;

(M3) (three change-point alternative): $\mu_i = d1_{\{i/n \leq 1/2\}} - d1_{\{i/n \leq 1/4\}} + d1_{\{i/n \geq 3/4\}}$,

where $d \in \mathbb{R}$ controls the amount of deviation from the null. For (M1)–(M3), the case with $d = 0$ corresponds to the null hypothesis of no change point. Let $n = 500$ and $\rho \in \{\pm 0.3, \pm 0.6\}$, we consider testing for change points in the mean of the process (4.12). Analyzing the tables 3.3 - 3.4 we can see that this alternative test, though being more complicated, does not seem to guarantee a superior performance in power when compared with the proposed USN test. For example, for $\rho = 0.3$ under the one-change point alternative with $d = 0.4$ the power of the alternative test is only 0.23, where is the power of the proposed USN test is 0.44, which is significantly higher.

Circling back to the original consideration for the development of the alternative formulation of the test we will also consider the case of the three change points. We will here remind that the alternative formulation of the test was originally designed in a way to be robust to the magnitude of the first and the last change-points, thus the Alt test is expected to have a better power performance than the proposed USN

rho	d	(M1)		(M2)		(M3)	
		Alt	USN	Alt	USN	Alt	USN
0.3	0.0	0.05	0.05	0.05	0.05	0.05	0.05
	0.2	0.09	0.15	0.24	0.30	0.10	0.12
	0.4	0.24	0.44	0.79	0.84	0.27	0.29
	0.6	0.56	0.79	0.98	0.99	0.61	0.66
	0.8	0.81	0.96	1.00	1.00	0.86	0.90
	1.0	0.95	0.99	1.00	1.00	0.98	0.99
	2.0	1.00	1.00	1.00	1.00	1.00	1.00
	3.0	1.00	1.00	1.00	1.00	1.00	1.00
0.6	0.0	0.05	0.05	0.05	0.05	0.05	0.05
	0.2	0.06	0.08	0.10	0.12	0.07	0.07
	0.4	0.09	0.16	0.27	0.34	0.11	0.12
	0.6	0.16	0.32	0.57	0.67	0.22	0.21
	0.8	0.28	0.55	0.83	0.89	0.33	0.38
	1.0	0.43	0.74	0.96	0.97	0.50	0.56
	2.0	0.97	1.00	1.00	1.00	0.98	0.99
	3.0	1.00	1.00	1.00	1.00	1.00	1.00

Table 3.3: Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1)-(M3) with different dependence strengths. The proposed test is denoted by USN and its alternative formulation version is denoted as Alt. For both tests the significance level $\alpha = 0.05$ is used.

test for the cases of smaller first and last change-points in comparison to the other change-points. Therefore here we will examine the sensitivity of the proposed USN test to the magnitude of the first and last change-points. In particular, we consider:

$$(M3) \text{ (three change-point alternative): } \mu_i = d1_{\{i/n \leq 1/2\}} - d1_{\{i/n \leq 1/4\}} + d1_{\{i/n \geq 3/4\}},$$

$$(M3s) \text{ (three change-point alternative): } \mu_i = d1_{\{i/n \leq 1/2\}} - 0.25d1_{\{i/n \leq 1/4\}} + 0.25d1_{\{i/n \geq 3/4\}},$$

The model (M3s) is constructed in such a way that the size of the first and last change points are made to be smaller than the second change-point, in particular

rho	d	(M1)		(M2)		(M3)	
		Alt	USN	Alt	USN	Alt	USN
-0.3	0.0	0.05	0.05	0.05	0.05	0.05	0.05
	0.2	0.26	0.41	0.80	0.83	0.29	0.29
	0.4	0.86	0.95	0.99	1.00	0.88	0.89
	0.6	0.94	1.00	1.00	1.00	0.99	1.00
	0.8	1.00	1.00	1.00	1.00	1.00	1.00
	1.0	1.00	1.00	1.00	1.00	1.00	1.00
	2.0	1.00	1.00	1.00	1.00	1.00	1.00
	3.0	1.00	1.00	1.00	1.00	1.00	1.00
-0.6	0.0	0.05	0.05	0.05	0.05	0.05	0.05
	0.2	0.94	0.63	0.95	0.94	0.47	0.44
	0.4	1.00	0.99	1.00	1.00	0.99	0.98
	0.6	1.00	1.00	1.00	1.00	1.00	1.00
	0.8	1.00	1.00	1.00	1.00	1.00	1.00
	1.0	1.00	1.00	1.00	1.00	1.00	1.00
	2.0	1.00	1.00	1.00	1.00	1.00	1.00
	3.0	1.00	1.00	1.00	1.00	1.00	1.00

Table 3.4: Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1)-(M3) with different dependence strengths. The proposed test is denoted by USN and its alternative formulation version is denoted as Alt. For both tests the significance level $\alpha = 0.05$ is used.

the one-fourth of that in model (M3). Let $n = 500$ and $\rho = 0.3$, the results are summarized in Table (3.5).

It can be seen from the Table (3.5) of this chapter that this alternative test is not guaranteed to significantly improve the robustness with respect to the magnitude of the first and last change points when compared with the proposed USN test. However, this might be a product of the limiting simulations results here. Additional theoretical power comparison of the two should be left as a future research topic.

d	(M3)		(M3s)	
	UCbSN	Alt	UCbSN	Alt
0.0	0.050	0.050	0.050	0.050
0.2	0.120	0.103	0.140	0.096
0.4	0.284	0.274	0.333	0.237
0.6	0.662	0.606	0.625	0.557
0.8	0.905	0.860	0.814	0.811
1.0	0.987	0.980	0.916	0.954
2.0	1.000	1.000	0.999	1.000
3.0	1.000	1.000	1.000	1.000

Table 3.5: Empirical powers for testing change points in the mean of (M3) and (M3s), where the significance level is taken as $\alpha = 0.05$.

3.4 Conclusion

In this chapter we have proposed an alternative formulation of the test statistic, whose power performance is designed to be less sensitive to the magnitude of the first and the last change points. The proposed test statistic has pivotal distribution under the null hypothesis and diverges under the general alternative hypothesis for the mean. We have also studied its final sample performance through the Monte-Carlo simulations and performed a numerical comparison of power properties on the proposed *USN* test and the alternative test and the computational comparison of both approaches. Through our analysis we were not able to confirm that the alternative formulation of the test statistic produces any significant benefit over the formulation of the *USN* test. Construction of an alternative hypothesis under which the alternative formulation of the test statistic would demonstrate better performance than the proposed *USN* method can be a topic for the future research.

Overall, we think the proposed *USN* test is more advantageous in the sense that:

- it generally has a less distortion in size

- its power performance is at least comparable to, and for some of the cases considered better than, that of the alternative test
- it requires significantly less computation.

Chapter 4

Change point detection in the case of other quantities of interest

4.1 Introduction

In our work we have aimed to propose a new approach in generalizing self-normalized statistics to handle quantities other than the mean. In the mean case, self-normalized statistics are generally functions of the partial sum process, as was discussed in Chapter 2. The CUSUM process after suitable scaling represents a sequence of recursive mean estimators. Hence, when one is interested in quantities other than the mean, the conventional approach is then to replace the role that the partial sum process plays in the mean case by recursive estimators of the quantity of interest. Such an approach was originally advocated by Shao (2010b) in the context of constructing confidence intervals for quantities associated with a stationary time series. However, due to its naturalness and ease of implementation, it gradually becomes a standard and has been adopted into various settings including change-point testing (Shao and Zhang, 2010). With this conventional approach, change-point testing is then conducted by sequentially comparing the recursive estimate with the overall estimate. Under the alternative, the overall estimate is computed by using observations both before and after the change point, and is thus expected to be different from the recursive estimate that is computed by using observations only before the hypothesized change point. However, when one is interested in robust quantities like the median,

the overall estimate in this case can be more resistant to the inhomogeneity caused by the change-point alternative, making the power of the resulting test to be suboptimal. To overcome the problem, we propose a contrast-based approach for generalizing self-normalized statistics to handle quantities other than the mean, which is specifically tailored for change-point testing problems. The consistency of the resulting change-point test is also established for cases with both the mean and quantities beyond the mean. Note that Shao and Zhang (2010) only studied the consistency of their test in the mean case.

We also propose an approximation scheme to further facilitate the computation needed in the proposed testing procedure. For the proposed unsupervised contrast-based method its computational burden would be the same for situations with different numbers of change points. As a comparison, the test of Shao and Zhang (2010) requires the prespecification of the total number of change points M^* , and their test involves the search of the maximum over $O(N^{M^*})$ partitions. Therefore, besides the improvements of (i) not requiring the *a priori* knowledge on M^* (which is typically unknown in practice); and (ii) being more specifically tailored for change-point testing by using the contrast-based approach in handling quantities other than the mean, the proposed test is also considered to be computationally more advantageous when there are more than two change points, in which case the computational burden of the test of Shao and Zhang (2010) would grow exponentially with respect to their hypothesized or prespecified value of M^* . We will also consider a grid approximation scheme to further reduce the computational burden of the current test and will demonstrate that we can dramatically improve speed of calculations with insignificant loss in accuracy.

The work presented in this chapter is based on the work by Zhang and Lavitas (2017). The rest of this chapter is organized as follows. We will demonstrate how

our test, proposed in Chapter 2 can be generalized to other quantities of interest, such as median, variance, etc. We will also propose a new approach: the contrast-based method for non-conventional generalization of the CUSUM process. We will also propose a computational scheme to reduce the computational complexity of the test. Our results will be supported with simulations and an analysis of real data.

4.2 The Conventional Approach

An attractive feature of self-normalization is that the resulting method can be readily generalized from the mean case to other quantities of interest, for example to quantiles. The key idea of achieving this is to exploit the role that the CUSUM process plays in the mean case, and replace it by an appropriate substitute in the general setting. To be more specific, the CUSUM process

$$\begin{aligned} Z_n(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \bar{X}_n) \\ &= \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{\lfloor nt \rfloor} X_i - \lfloor nt \rfloor \bar{X}_n \right) = \frac{\lfloor nt \rfloor}{\sqrt{n}} (\bar{X}_{\lfloor nt \rfloor} - \bar{X}_n), \quad t \in [0, 1], \end{aligned} \quad (4.1)$$

represents the difference between the recursive mean estimator $\bar{X}_{\lfloor nt \rfloor}$ and the overall mean estimator \bar{X}_n . Hence, if one is interested in a general parameter $\boldsymbol{\theta}$ that is possibly different from the mean, then it would be natural to consider replacing the $\bar{X}_{\lfloor nt \rfloor}$ and \bar{X}_n in (4.1) by the corresponding recursive estimator $\hat{\boldsymbol{\theta}}_{\lfloor nt \rfloor}$ and $\hat{\boldsymbol{\theta}}_n$, making the new statistic a function of the process $(\lfloor nt \rfloor / \sqrt{n})(\hat{\boldsymbol{\theta}}_{\lfloor nt \rfloor} - \hat{\boldsymbol{\theta}}_n)$, $t \in [0, 1]$. Such an approach was advocated by Shao (2010b) in the context of confidence interval construction, and gradually becomes a standard in generalizing self-normalized methods to quantities other than the mean; see for example the application in Shao and Zhang (2010) for change-point testing and Shao (2015) for a recent review. We

shall here consider a generalization of the test proposed in Section 2.2 by following this conventional approach.

To facilitate the mathematical formulation, suppose one is interested in the quantity $\boldsymbol{\theta}_i = \mathbf{Q}(\mathbf{F}_i) \in \mathbb{R}^d$, a function of \mathbf{F}_i , the distribution of $\mathbf{Y}_i = (X_i, X_{i+1}, \dots, X_{i+\ell-1})$. Then we consider the problem of testing for the null hypothesis

$$H_0^* : \boldsymbol{\theta}_1 = \dots = \boldsymbol{\theta}_N = \boldsymbol{\vartheta}, \quad N = n - \ell + 1, \quad (4.2)$$

versus the alternative that there exist $M^* \geq 1$ break points $k_0^* = 1 < k_1^* < \dots < k_{M^*}^* < N = k_{M^*+1}^*$ such that

$$\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_{i+1}, \quad i \in \{k_1^*, \dots, k_{M^*}^*\}, \quad (4.3)$$

and $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i+1}$ otherwise. Let $\hat{\mathbf{F}}_{j,k}$ be the empirical distribution of $(\mathbf{Y}_j, \dots, \mathbf{Y}_k)$, and denote by $\mathbf{F}_{j,k}$ the uniform mixture of $\mathbf{F}_j, \dots, \mathbf{F}_k$, $1 \leq j \leq k \leq N$. Let $\hat{\boldsymbol{\theta}}_{j,k} = \mathbf{Q}(\hat{\mathbf{F}}_{j,k})$, then counterparts of (2.10)–(2.12) are given by

$$\mathbf{D}_{N,f}^*(j_1, j_2, j_3) = \frac{j_2 - j_1 + 1}{\sqrt{j_3 - j_1 + 1}} (\hat{\boldsymbol{\theta}}_{j_1, j_2} - \hat{\boldsymbol{\theta}}_{j_1, j_3}), \quad (4.4)$$

$$\mathbf{L}_{N,f}^*(j_1, j_2, j_3) = \sum_{i=j_1}^{j_2} \left(\frac{i - j_1 + 1}{j_3 - j_1 + 1} \right)^2 (\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{j_1, j_2})(\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{j_1, j_2})^\top, \quad (4.5)$$

$$\mathbf{R}_{N,f}^*(j_1, j_2, j_3) = \sum_{i=j_2+1}^{j_3} \left(\frac{j_3 - i + 1}{j_3 - j_1 + 1} \right)^2 (\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2+1, j_3})(\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2+1, j_3})^\top. \quad (4.6)$$

Similarly, let

$$\begin{aligned} \mathbf{D}_{N,b}^*(j_1, j_2, j_3) &= \frac{j_3 - j_2 + 1}{\sqrt{j_3 - j_1 + 1}} (\hat{\boldsymbol{\theta}}_{j_2, j_3} - \hat{\boldsymbol{\theta}}_{j_1, j_3}), \\ \mathbf{L}_{N,b}^*(j_1, j_2, j_3) &= \sum_{i=j_1}^{j_2-1} \left(\frac{i - j_1 + 1}{j_3 - j_1 + 1} \right)^2 (\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{j_1, j_2-1}) (\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{j_1, j_2-1})^\top, \\ \mathbf{R}_{N,b}^*(j_1, j_2, j_3) &= \sum_{i=j_2}^{j_3} \left(\frac{j_3 - i + 1}{j_3 - j_1 + 1} \right)^2 (\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2, j_3}) (\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2, j_3})^\top, \end{aligned}$$

$\boldsymbol{\Xi}_{N,f}^*(j_1, j_2, j_3) = \mathbf{L}_{N,f}^*(j_1, j_2, j_3) + \mathbf{R}_{N,f}^*(j_1, j_2, j_3)$ and $\boldsymbol{\Xi}_{N,b}^*(j_1, j_2, j_3) = \mathbf{L}_{N,b}^*(j_1, j_2, j_3) + \mathbf{R}_{N,b}^*(j_1, j_2, j_3)$, then the generalized test statistic can be defined as

$$\begin{aligned} T_N^* &= \max_{(l_1, l_2) \in \Omega_N(\varepsilon)} \mathbf{D}_{N,f}^*(1, l_1, l_2)^\top \boldsymbol{\Xi}_{N,f}^*(1, l_1, l_2)^{-1} \mathbf{D}_{N,f}^*(1, l_1, l_2) \\ &+ \max_{(m_1, m_2) \in \Omega_N(\varepsilon)} \mathbf{D}_{N,b}^*(m_1, m_2, N)^\top \boldsymbol{\Xi}_{N,b}^*(m_1, m_2, N)^{-1} \mathbf{D}_{N,b}^*(m_1, m_2, N). \end{aligned} \quad (4.7)$$

Let $\mathbf{B}(\cdot)$ be a d -dimensional standard Brownian motion and $\mathcal{D} = \{(t_1, t_2) : 0 \leq t_1 \leq t_2 \leq 1\}$, we shall here make the following assumption on the recursive estimators $\hat{\boldsymbol{\theta}}_{j,k} = \mathbf{Q}(\hat{\mathbf{F}}_{j,k})$.

(IP*) There exists a positive definite matrix $\boldsymbol{\Sigma}$ such that

$$\begin{aligned} &\left[\frac{\lfloor Nt \rfloor - \lfloor Ns \rfloor \vee 1 + 1}{\sqrt{N}} \{ \mathbf{Q}(\hat{\mathbf{F}}_{\lfloor Ns \rfloor \vee 1, \lfloor Nt \rfloor}) - \mathbf{Q}(\mathbf{F}_{\lfloor Ns \rfloor \vee 1, \lfloor Nt \rfloor}) \}, (s, t) \in \mathcal{D} \right] \\ &\rightsquigarrow [\boldsymbol{\Sigma} \{ \mathbf{B}(t) - \mathbf{B}(s) \}, (s, t) \in \mathcal{D}], \end{aligned}$$

where $\lfloor Ns \rfloor \vee 1 = \max(\lfloor Ns \rfloor, 1)$ and \rightsquigarrow denotes the weak convergence in the sense of Hoffmann-Jørgensen (van der Vaart and Wellner, 1996).

Note that the partial sum process in Assumption (IP) relates to the empirical process, Assumption (IP*) provides a counterpart in the current general setting with the additional transform \mathbf{Q} , and can be verified by using the functional delta method of Volgushev and Shao (2014); see also the influence function approach of Shao (2010b).

Let

$$\Delta(\mathbf{B}, r, s, t, u) = \mathbf{B}(s) - \mathbf{B}(r) - \frac{s-r}{u-t} \{\mathbf{B}(u) - \mathbf{B}(t)\},$$

the following theorem provides an asymptotic theory on the test statistic T_N^* .

Theorem 4.2.1. *Assume condition (IP^{*}). Then (i) under the null hypothesis (4.2), we have*

$$\begin{aligned} T_N^* \rightarrow_D T^*(\mathbf{B}) &:= \sup_{(r_1, r_2) \in \Omega(\varepsilon)} \mathbf{D}^*(\mathbf{B}, 0, r_1, r_2)^\top \Xi^*(\mathbf{B}, 0, r_1, r_2)^{-1} \mathbf{D}^*(\mathbf{B}, 0, r_1, r_2) \\ &+ \sup_{(s_1, s_2) \in \Omega(\varepsilon)} \mathbf{D}^*(\mathbf{B}, s_1, s_2, 1)^\top \Xi^*(\mathbf{B}, s_1, s_2, 1)^{-1} \mathbf{D}^*(\mathbf{B}, s_1, s_2, 1), \end{aligned} \quad (4.8)$$

where

$$\begin{aligned} \mathbf{D}^*(\mathbf{B}, t_1, t_2, t_3) &= \frac{1}{\sqrt{t_3 - t_1}} \Delta(\mathbf{B}, t_1, t_2, t_1, t_3), \\ \Xi^*(\mathbf{B}, t_1, t_2, t_3) &= \frac{1}{(t_3 - t_1)^2} \left\{ \int_{t_1}^{t_2} \Delta(\mathbf{B}, t_1, s, t_1, t_2) \Delta(\mathbf{B}, t_1, s, t_1, t_2)^\top ds \right. \\ &\quad \left. + \int_{t_2}^{t_3} \Delta(\mathbf{B}, s, t_3, t_2, t_3) \Delta(\mathbf{B}, s, t_3, t_2, t_3)^\top ds \right\}; \end{aligned}$$

and (ii) under the alternative (4.3) with $\min_{0 \leq i \leq M^*} |k_{i+1}^* - k_i^*|/N > \varepsilon$, we have $T_N^* \rightarrow \infty$ in probability if $\sqrt{N} |\hat{\boldsymbol{\theta}}_{k_{i-1}^*+1, k_i^*} - \hat{\boldsymbol{\theta}}_{k_{i-1}^*+1, k_{i+1}^*}| \rightarrow \infty$ in probability, $i = 1, \dots, M^*$.

Proof. Under the general setting as framed in Section 4.2, $\hat{\boldsymbol{\theta}}_{j,k}$, $j \leq k$, can no longer be expressed as a linear combination of the recursive statistics $\hat{\boldsymbol{\theta}}_{1,i}$, $i = 1, \dots, n$. For this, we let

$$\begin{aligned} \Delta_{\mathbf{Q}, N}(\hat{\mathbf{F}}, r, s, t, u) &= \frac{1}{\sqrt{N}} \{ ([Ns] - [Nr] \vee 1 + 1) \mathbf{Q}(\hat{\mathbf{F}}_{[Nr] \vee 1, [Ns]}) \\ &\quad - (u-t)^{-1} (s-r) ([Nu] - [Nt] \vee 1 + 1) \mathbf{Q}(\hat{\mathbf{F}}_{[Nt] \vee 1, [Nu]}) \}, \end{aligned}$$

and define

$$\mathbf{D}_{\mathbf{Q}, N}^*(\hat{\mathbf{F}}, t_1, t_2, t_3) = \frac{1}{\sqrt{t_3 - t_1}} \Delta_{\mathbf{Q}, N}(\hat{\mathbf{F}}, t_1, t_2, t_1, t_3).$$

Note that

$$\begin{aligned} & \sqrt{[Nr_2]/N} \mathbf{D}_{N,f}^*(1, [Nr_1], [Nr_2]) - \sqrt{r_2} \mathbf{D}_{\mathbf{Q},N}^*(\hat{\mathbf{F}}, 0, r_1, r_2) \\ &= \left(\frac{r_1}{r_2} - \frac{[Nr_1]}{[Nr_2]} \right) \frac{[Nr_2]}{\sqrt{N}} \{ \mathbf{Q}(\hat{\mathbf{F}}_{1,[Nr_2]}) - \boldsymbol{\vartheta} \}, \end{aligned}$$

and thus by condition (IP^{*}), one has

$$\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |\sqrt{[Nr_2]/N} \mathbf{D}_{N,f}^*(1, [Nr_1], [Nr_2]) - \sqrt{r_2} \mathbf{D}_{\mathbf{Q},N}^*(\hat{\mathbf{F}}, 0, r_1, r_2)| = O_p(N^{-1}).$$

Thus, similarly to the proof for the case of the mean (2.2.1) one can write

$$\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |D_{n,f}(1, [nr_1], [nr_2]) - D(S_n, 0, r_1, r_2)| = O_p(n^{-1}).$$

Note that $\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |D_{n,f}(1, [nr_1], [nr_2])| = \max_{(l_1, l_2) \in \Omega_n(\varepsilon)} |D_{n,f}(1, l_1, l_2)|$ holds for all large n , by a similar argument and properties of the Brownian motion, it can be shown that

$$\left| T_n - \left\{ \sup_{(r_1, r_2) \in \Omega(\varepsilon)} \frac{D(S_n, 0, r_1, r_2)^2}{\Xi(S_n, 0, r_1, r_2)} + \sup_{(s_1, s_2) \in \Omega(\varepsilon)} \frac{D(S_n, s_1, s_2, 1)^2}{\Xi(S_n, s_1, s_2, 1)} \right\} \right| = o_p(1),$$

and (i) follows by the continuous mapping theorem. For (ii), under the alternative (2.4) with $E(X_{k_{i+1}}) - E(X_{k_i}) = n^{-1/2} L_i$, $i = 1, \dots, M$, we have

$$\begin{aligned} D_{n,f}(1, k_1, k_2) &= \frac{1}{\sqrt{k_2}} \left[\sum_{i=1}^{k_1} \{X_i - E(X_i)\} - \frac{k_1}{k_2} \sum_{j=1}^{k_2} \{X_j - E(X_j)\} \right] - \frac{k_1(k_2 - k_1)}{k_2 \sqrt{nk_2}} L_1 \\ &:= D_{n,f,e}(1, k_1, k_2) - \frac{k_1(k_2 - k_1)}{k_2 \sqrt{nk_2}} L_1. \end{aligned}$$

By the argument of (i) and the condition that $\min_{0 \leq i \leq M} |k_{i+1} - k_i|/n > \varepsilon$, one has

$$\frac{D_{n,f,e}(1, k_1, k_2)^2}{\Xi_{n,f}(1, k_1, k_2)} = O_p(1).$$

Note that $k_1(k_2 - k_1)/(k_2 \sqrt{nk_2}) \geq \varepsilon^2$, the result follows by $\min_{1 \leq i \leq M} |L_i| \rightarrow \infty$. \square

Therefore, similar to the mean case, one rejects the null hypothesis (4.2) at level

$\alpha \in (0, 1)$ if $T_N^* > q_{1-\alpha}^*$ where $q_{1-\alpha}^*$ is the $(1 - \alpha)$ -th quantile of $T^*(\mathbf{B})$ defined in (4.8). Nevertheless, in order to obtain the consistency of this test, unlike the mean case, one needs the stochastic condition that

$$\sqrt{N}|\hat{\boldsymbol{\theta}}_{k_{i-1}^*+1, k_i^*} - \hat{\boldsymbol{\theta}}_{k_{i-1}^*+1, k_{i+1}^*}| \rightarrow \infty \quad (4.9)$$

in probability as $n \rightarrow \infty$. Although Shao and Zhang (2010) did not study the power behavior of their test under this general setting, given the form of their test statistic and the argument used in the proof of Theorem 4.2.1, the same stochastic condition will also be required for their test to be consistent. Note that

$$\hat{\boldsymbol{\theta}}_{k_{i-1}^*+1, k_i^*} - \hat{\boldsymbol{\theta}}_{k_{i-1}^*+1, k_{i+1}^*} = \mathbf{Q}(\hat{\mathbf{F}}_{k_{i-1}^*+1, k_i^*}) - \mathbf{Q}(\hat{\mathbf{F}}_{k_{i-1}^*+1, k_{i+1}^*}),$$

where $\hat{\mathbf{F}}_{k_{i-1}^*+1, k_i^*}$ is the empirical distribution of $\mathbf{Y}_{k_{i-1}^*+1}, \dots, \mathbf{Y}_{k_i^*}, \mathbf{Y}_{k_i^*+1}, \dots, \mathbf{Y}_{k_{i+1}^*}$, a mixture of observations before and after the change point k_i^* , which under the change-point alternative is considered to be different from $\hat{\mathbf{F}}_{k_{i-1}^*+1, k_i^*}$, the empirical distribution based only on observations before the change point. Hence, condition (4.9) seems to be a reasonable requirement for change-point detection. In fact, the following proposition states that, in the case of the mean, it can be reduced to its deterministic counterpart.

Proposition 4.2.2. *If $\theta_i = \int_{\mathbb{R}} x dF_i(x)$ represents the mean, then assuming (IP), condition (4.9) will be satisfied if $\sqrt{N}|E(X_{k_i^*+1}) - E(X_{k_i^*})| \rightarrow \infty$.*

Proof. (Proposition 4.2.2) In the case of the mean, we have $\hat{\theta}_{j,k} = S_{j,k}/(k - j + 1) = (k - j + 1)^{-1} \sum_{i=j}^k \{X_i - E(X_i)\} + (k - j + 1)^{-1} \sum_{i=j}^k E(X_i)$. Let $S_{j,k}^\circ = S_{j,k} - E(S_{j,k})$ denote the centered partial sum and $d_i^* = E(X_{k_i^*+1}) - E(X_{k_i^*})$ represent the magnitude

of change, then

$$\begin{aligned}
\sqrt{N}|\hat{\theta}_{k_{i-1}^*+1, k_i^*} - \hat{\theta}_{k_{i-1}^*+1, k_{i+1}^*}| &= \sqrt{N} \left| \frac{S_{k_{i-1}^*+1, k_i^*}}{k_i^* - k_{i-1}^*} - \frac{S_{k_{i-1}^*+1, k_{i+1}^*}}{k_{i+1}^* - k_{i-1}^*} \right| \\
&= \sqrt{N} \left| \frac{S_{k_{i-1}^*+1, k_i^*}^\circ}{k_i^* - k_{i-1}^*} - \frac{S_{k_{i-1}^*+1, k_{i+1}^*}^\circ}{k_{i+1}^* - k_{i-1}^*} - \frac{k_{i+1}^* - k_i^*}{k_{i+1}^* - k_{i-1}^*} d^* \right| \\
&\geq \sqrt{N} \left(\frac{k_{i+1}^* - k_i^*}{k_{i+1}^* - k_{i-1}^*} |d^*| - \left| \frac{S_{k_{i-1}^*+1, k_i^*}^\circ}{k_i^* - k_{i-1}^*} - \frac{S_{k_{i-1}^*+1, k_{i+1}^*}^\circ}{k_{i+1}^* - k_{i-1}^*} \right| \right).
\end{aligned}$$

By condition (IP) and the proof of Theorem 2.2.1,

$$\sqrt{N} \left| \frac{S_{k_{i-1}^*+1, k_i^*}^\circ}{k_i^* - k_{i-1}^*} - \frac{S_{k_{i-1}^*+1, k_{i+1}^*}^\circ}{k_{i+1}^* - k_{i-1}^*} \right| = O_p(1),$$

and the result follows. \square

However, for general nonlinear statistics, it is not always the case that the stochastic condition (4.9) required to guarantee the consistency as in Theorem 4.2.1 will be implied by its more desirable deterministic counterpart $\sqrt{N}|\theta_{k_i^*+1} - \theta_{k_i^*}| \rightarrow \infty$. To illustrate the idea, we consider a simple example in which $X_i = d1_{\{i > pn\}}$, where $1_{\{\cdot\}}$ is the indicator function and $p \in (0, 1)$ represents the location (in proportion) of the change point. Hence, if one is interested in the mean, then the recursive mean estimate $\bar{X}_{[pn]} = 0$ and the overall mean estimate $\bar{X}_n = (1 - [pn]/n)d$, leading to the desirable result that $\sqrt{n}|\bar{X}_{[pn]} - \bar{X}_n| \rightarrow \infty$ if $\sqrt{nd} \rightarrow \infty$. However, if one is interested in nonlinear quantities like the median, then one can show that the recursive median estimate $\hat{\theta}_{[pn]} = 0$ and the overall median estimate $\hat{\theta}_n = 0$ when $p > 1/2 + 1/n$. Therefore, when the change point is located at the later half of the time series, the deterministic condition that $\sqrt{nd} \rightarrow \infty$ will no longer imply the needed stochastic condition that $\sqrt{n}|\hat{\theta}_{[pn]} - \hat{\theta}_n| \rightarrow \infty$, as $|\hat{\theta}_{[pn]} - \hat{\theta}_n| = 0$ in this case. Therefore, the CUSUM-type process $([nt]/\sqrt{n})(\hat{\theta}_{[nt]} - \hat{\theta}_n)$, $t \in [0, 1]$, may not be the most ideal for change-point detection in nonlinear quantities, especially when one is interested

in robust quantities like the median that are potentially resistant to the inhomogeneity under the change-point alternative. This motivates us to seek for a new scheme for generalizing self-normalized test statistics to quantities other than the mean that is specifically tailored for change-point testing problems, which we shall consider in Section 4.3.

4.3 A New Approach: The Contrast-Based Method

By viewing $\bar{X}_{[nt]}$ as a recursive mean estimator, the conventional approach generalizes self-normalized statistics developed for the mean to more general quantities by replacing $(\bar{X}_{[nt]} - \bar{X}_n)$ in the CUSUM process with its recursive counterpart $(\hat{\theta}_{[nt]} - \hat{\theta}_n)$. However, as discussed in Section 4.2, it may not be the most ideal for change-point testing problems as it was originally developed for constructing confidence intervals by assuming structural stability. One can notice that for the case of the mean we can write:

$$\frac{t}{\sqrt{n}}(\bar{X}_t - \bar{X}_n) = \frac{(n-t)t}{n\sqrt{n}}(\bar{X}_t - \bar{X}_{t+1,n}) \quad (4.10)$$

This form suggests that a generalization that is directly related to change-point testing can be suggested. We propose to focus on the contrast process $\{\mathbf{Q}(\hat{\mathbf{F}}_{1,[nt]}) - \mathbf{Q}(\hat{\mathbf{F}}_{[nt]+1,n})\}$, where $\mathbf{Q}(\hat{\mathbf{F}}_{1,[nt]})$ and $\mathbf{Q}(\hat{\mathbf{F}}_{[nt]+1,n})$ are parameter estimates based on $X_1, \dots, X_{[nt]}$ and $X_{[nt]+1}, \dots, X_n$ respectively. Hence, we are directly comparing estimates before and after the hypothesized change point, which is considered to be more relevant to change-point testing than the conventional $\{\mathbf{Q}(\hat{\mathbf{F}}_{1,[nt]}) - \mathbf{Q}(\hat{\mathbf{F}}_{1,n})\}$ for nonlinear quantities. This leads to the contrast-based counterparts of (4.4)–(4.6)

given by

$$\begin{aligned}
\mathbf{D}_{N,f}^\diamond(j_1, j_2, j_3) &= \frac{(j_2 - j_1 + 1)(j_3 - j_2)}{(j_3 - j_1 + 1)^{3/2}} (\hat{\boldsymbol{\theta}}_{j_1, j_2} - \hat{\boldsymbol{\theta}}_{j_2+1, j_3}), \\
\mathbf{L}_{N,f}^\diamond(j_1, j_2, j_3) &= \sum_{i=j_1}^{j_2} \frac{(i - j_1 + 1)^2 (j_2 - i)^2}{(j_3 - j_1 + 1)^2 (j_2 - j_1 + 1)^2} (\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{i+1, j_2}) (\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{i+1, j_2})^\top, \\
\mathbf{R}_{N,f}^\diamond(j_1, j_2, j_3) &= \sum_{i=j_2+1}^{j_3} \frac{(i - 1 - j_2)^2 (j_3 - i + 1)^2}{(j_3 - j_1 + 1)^2 (j_3 - j_2)^2} (\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2+1, i-1}) (\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2+1, i-1})^\top.
\end{aligned}$$

Similarly, let

$$\begin{aligned}
\mathbf{D}_{N,b}^\diamond(j_1, j_2, j_3) &= \frac{(j_2 - j_1)(j_3 - j_2 + 1)}{(j_3 - j_1 + 1)^{3/2}} (\hat{\boldsymbol{\theta}}_{j_2, j_3} - \hat{\boldsymbol{\theta}}_{j_1, j_2-1}), \\
\mathbf{L}_{N,b}^\diamond(j_1, j_2, j_3) &= \sum_{i=j_1}^{j_2-1} \frac{(i - j_1 + 1)^2 (j_2 - 1 - i)^2}{(j_3 - j_1 + 1)^2 (j_2 - j_1)^2} (\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{i+1, j_2-1}) (\hat{\boldsymbol{\theta}}_{j_1, i} - \hat{\boldsymbol{\theta}}_{i+1, j_2-1})^\top, \\
\mathbf{R}_{N,b}^\diamond(j_1, j_2, j_3) &= \sum_{i=j_2}^{j_3} \frac{(i - j_2)^2 (j_3 - i + 1)^2}{(j_3 - j_1 + 1)^2 (j_3 - j_2 + 1)^2} (\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2, i-1}) (\hat{\boldsymbol{\theta}}_{i, j_3} - \hat{\boldsymbol{\theta}}_{j_2, i-1})^\top,
\end{aligned}$$

$\boldsymbol{\Xi}_{N,f}^\diamond(j_1, j_2, j_3) = \mathbf{L}_{N,f}^\diamond(j_1, j_2, j_3) + \mathbf{R}_{N,f}^\diamond(j_1, j_2, j_3)$ and $\boldsymbol{\Xi}_{N,b}^\diamond(j_1, j_2, j_3) = \mathbf{L}_{N,b}^\diamond(j_1, j_2, j_3) + \mathbf{R}_{N,b}^\diamond(j_1, j_2, j_3)$, then our contrast-based generalization of (2.13) is given by

$$\begin{aligned}
T_N^\diamond &= \max_{(l_1, l_2) \in \Omega_N(\varepsilon)} \mathbf{D}_{N,f}^\diamond(1, l_1, l_2)^\top \boldsymbol{\Xi}_{N,f}^\diamond(1, l_1, l_2)^{-1} \mathbf{D}_{N,f}^\diamond(1, l_1, l_2) \\
&+ \max_{(m_1, m_2) \in \Omega_N(\varepsilon)} \mathbf{D}_{N,b}^\diamond(m_1, m_2, N)^\top \boldsymbol{\Xi}_{N,b}^\diamond(m_1, m_2, N)^{-1} \mathbf{D}_{N,b}^\diamond(m_1, m_2, N) \quad (4.11)
\end{aligned}$$

Note that in the case of the mean when $\theta_i = \int_{\mathbb{R}} x dF_i(x)$, we have $\hat{\theta}_{j,k} = S_{j,k}/(k-j+1)$ and one can show that both the conventional generalization (4.7) and the contrast-based generalization (4.11) reduce to the same form, namely the test statistic (2.13) proposed in Section 2.2. However, test statistics in (4.7) and (4.11) are generally different when parameter estimators are not additive, namely when $\hat{\boldsymbol{\theta}}_{j_1, j_2} + \hat{\boldsymbol{\theta}}_{j_2+1, j_3} \neq \hat{\boldsymbol{\theta}}_{j_1, j_3}$, $j_1 \leq j_2 < j_3$. Unlike the generalization in (4.7) which relies on the conventional method that was originally developed for inference of constant parameters of a sta-

tionary time series, the proposed contrast-based test statistic (4.11) is specifically tailored for change-point testing and can be more powerful in detecting change points in robust quantities like the median. The following theorem provides its asymptotic distribution and theoretical consistency.

Theorem 4.3.1. *Assume condition (IP^{*}). Then (i) under the null hypothesis (4.2), we have*

$$T_N^\diamond \rightarrow_D T^*(\mathbf{I}\mathbf{B}),$$

where $T^*(\mathbf{I}\mathbf{B})$ is given in (4.8); and (ii) under the alternative (4.3) with $\min_{0 \leq i \leq M^*} |k_{i+1}^* - k_i^*|/N > \varepsilon$ and $\boldsymbol{\theta}_{k_{i+1}^*} - \boldsymbol{\theta}_{k_i^*} = N^{-1/2}L_i^*$ for some $L_i^* \neq 0$, $i = 1, \dots, M^*$, we have $T_N^\diamond \rightarrow \infty$ in probability if $\min_{1 \leq i \leq M^*} |L_i^*| \rightarrow \infty$.

Proof. (Theorem 4.3.1) Let

$$\begin{aligned} \mathbf{D}_{\mathbf{Q},N}^\diamond(\hat{\mathbf{F}}, t_1, t_2, t_3) &= \frac{1}{\sqrt{N(t_3 - t_1)^3}} \{ (t_3 - t_2)(\lfloor Nt_2 \rfloor - \lfloor Nt_1 \rfloor \vee 1 + 1) \mathbf{Q}(\hat{\mathbf{F}}_{\lfloor Nt_1 \rfloor \vee 1, \lfloor Nt_2 \rfloor}) \\ &\quad - (t_2 - t_1)(\lfloor Nt_3 \rfloor - \lfloor Nt_2 \rfloor) \mathbf{Q}(\hat{\mathbf{F}}_{\lfloor Nt_2 \rfloor + 1, \lfloor Nt_3 \rfloor}) \}, \end{aligned}$$

then one can write

$$\begin{aligned} &\mathbf{D}_{N,f}^\diamond(1, \lfloor Nr_1 \rfloor, \lfloor Nr_2 \rfloor) - \mathbf{D}_{\mathbf{Q},N}^\diamond(\hat{\mathbf{F}}, 0, r_1, r_2) \\ &= \left\{ \frac{(\lfloor Nr_2 \rfloor - \lfloor Nr_1 \rfloor)/N}{(\lfloor Nr_2 \rfloor/N)^{3/2}} - \frac{r_2 - r_1}{r_2^{3/2}} \right\} \frac{\lfloor Nr_1 \rfloor}{\sqrt{N}} \{ \mathbf{Q}(\hat{\mathbf{F}}_{1, \lfloor Nr_1 \rfloor}) - \boldsymbol{\vartheta} \} \\ &\quad - \left(\frac{\lfloor Nr_1 \rfloor/N}{(\lfloor Nr_2 \rfloor/N)^{3/2}} - \frac{r_1}{r_2^{3/2}} \right) \frac{\lfloor Nr_2 \rfloor - \lfloor Nr_1 \rfloor}{\sqrt{N}} \{ \mathbf{Q}(\hat{\mathbf{F}}_{\lfloor Nr_1 \rfloor + 1, \lfloor Nr_2 \rfloor}) - \boldsymbol{\vartheta} \}. \end{aligned}$$

Note that for any $t \in [0, 1]$, $t - 1/N < \lfloor Nt \rfloor/N \leq t$, we have by condition (IP^{*}),

$$\sup_{(r_1, r_2) \in \Omega(\varepsilon)} |\mathbf{D}_{N,f}^\diamond(1, \lfloor Nr_1 \rfloor, \lfloor Nr_2 \rfloor) - \mathbf{D}_{\mathbf{Q},N}^\diamond(\hat{\mathbf{F}}, 0, r_1, r_2)| = O_p(N^{-1}),$$

and (i) follows by the continuous mapping theorem and a similar argument as in Theorem 2.2.1. For (ii), under the alternative (4.3) with $\boldsymbol{\theta}_{k_{i+1}^*} - \boldsymbol{\theta}_{k_i^*} = N^{-1/2}L_i^*$, $i = 1, \dots, M^*$, we have

$$\mathbf{D}_{N,f}^\diamond(1, k_1^*, k_2^*) = \frac{k_1^*(k_2^* - k_1^*)}{(k_2^*)^{3/2}} \{ (\hat{\boldsymbol{\theta}}_{1, k_1^*} - \boldsymbol{\theta}_{k_1^*}) - (\hat{\boldsymbol{\theta}}_{k_1^* + 1, k_2^*} - \boldsymbol{\theta}_{k_1^* + 1}) \} - \frac{k_1^*(k_2^* - k_1^*)}{k_2^* \sqrt{Nk_2^*}} \mathbf{L}_1^*.$$

Under the change-point alternative, we have $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_{k_1^*}$ and $\boldsymbol{\theta}_{k_1^*+1} = \cdots = \boldsymbol{\theta}_{k_2^*}$. Hence, by condition (IP *) and the argument of (i), one can obtain that

$$\frac{k_1^*(k_2^* - k_1^*)}{(k_2^*)^{3/2}} \{(\hat{\boldsymbol{\theta}}_{1, k_1^*} - \boldsymbol{\theta}_{k_1^*}) - (\hat{\boldsymbol{\theta}}_{k_1^*+1, k_2^*} - \boldsymbol{\theta}_{k_1^*+1})\} = O_p(1).$$

The result then follows by noticing that $\min_{1 \leq i \leq M^*} |L_i^*| \rightarrow \infty$. □

Compared with Theorem 4.2.1, the contrast-based method no longer requires the stochastic condition (4.9). By Theorem 4.3.1 (ii), the proposed test will have unit asymptotic power if the magnitude of change is parametrically distinguishable, namely when the more desirable deterministic condition $N^{1/2}|\boldsymbol{\theta}_{k_i^*+1} - \boldsymbol{\theta}_{k_i^*}| \rightarrow \infty$ is satisfied.

4.4 Implementation

4.4.1 An Approximation Scheme

Note that the test statistic (4.7) involves a maximum of $O(N^2)$ terms, and its specific form does not depend on M^* , the total number of change points. Therefore, the proposed testing procedure and its computational burden would be the same for situations with different numbers of change points. As a comparison, the test of Shao and Zhang (2010) requires the prespecification of the total number of change points M^* , and their test involves the search of the maximum over $O(N^{M^*})$ partitions. Therefore, besides the improvements of (i) not requiring the *a priori* knowledge on M^* (which is typically unknown in practice); and (ii) being more specifically tailored for change-point testing by using the contrast-based approach in handling quantities other than the mean, the proposed test is also considered to be computationally more advantageous when there are more than two change points, in which case the computational burden of the test of Shao and Zhang (2010) would grow exponentially

with respect to their hypothesized or prespecified value of M^* . We shall here in addition consider a grid approximation scheme to further reduce the computational burden of the current test, with which the maximum only needs to be searched over $O(N)$, instead of $O(N^2)$, partitions. To be more specific, let $\mathcal{G}_\varepsilon = \{(1 + k\varepsilon)/2, k \in \mathbb{Z}\} \cap [0, 1]$ be a grid on the unit interval, which is symmetric around the middle point. Let $\mathcal{G}_{\varepsilon, N, f} = \{(\lfloor Nt_1 \rfloor, \lfloor Nt_2 \rfloor) : (t_1, t_2) \in ([0, 1] \times \mathcal{G}_\varepsilon) \cap \Omega(\varepsilon)\}$ and $\mathcal{G}_{\varepsilon, N, b} = \{(\lfloor Nt_1 \rfloor, \lfloor Nt_2 \rfloor) : (t_1, t_2) \in (\mathcal{G}_\varepsilon \times [0, 1]) \cap \Omega(\varepsilon)\}$, we consider the test statistic

$$\begin{aligned} G_N^\diamond &= \max_{(l_1, l_2) \in \mathcal{G}_{\varepsilon, N, f}} \mathbf{D}_{N, f}^\diamond(1, l_1, l_2)^\top \boldsymbol{\Xi}_{N, f}^\diamond(1, l_1, l_2)^{-1} \mathbf{D}_{N, f}^\diamond(1, l_1, l_2) \\ &\quad + \max_{(m_1, m_2) \in \mathcal{G}_{\varepsilon, N, b}} \mathbf{D}_{N, b}^\diamond(m_1, m_2, N)^\top \boldsymbol{\Xi}_{N, b}^\diamond(m_1, m_2, N)^{-1} \mathbf{D}_{N, b}^\diamond(m_1, m_2, N), \end{aligned}$$

which provides a grid counterpart of the statistic T_n^\diamond introduced in (4.11). The asymptotic distribution of G_n^\diamond is given in Proposition 4.4.1 along with its consistency for testing the null hypothesis (4.2) against the change-point alternative.

Proposition 4.4.1. *Assume condition (IP *). Then (i) under the null hypothesis (4.2), we have*

$$G_N^\diamond \rightarrow_D G^\diamond(\mathbf{B}),$$

where

$$\begin{aligned} G^\diamond(\mathbf{B}) &= \sup_{(r_1, r_2) \in ([0, 1] \times \mathcal{G}_\varepsilon) \cap \Omega(\varepsilon)} \mathbf{D}^*(\mathbf{B}, 0, r_1, r_2)^\top \boldsymbol{\Xi}^*(\mathbf{B}, 0, r_1, r_2)^{-1} \mathbf{D}^*(\mathbf{B}, 0, r_1, r_2) \\ &\quad + \sup_{(s_1, s_2) \in (\mathcal{G}_\varepsilon \times [0, 1]) \cap \Omega(\varepsilon)} \mathbf{D}^*(\mathbf{B}, s_1, s_2, 1)^\top \boldsymbol{\Xi}^*(\mathbf{B}, s_1, s_2, 1)^{-1} \mathbf{D}^*(\mathbf{B}, s_1, s_2, 1); \end{aligned}$$

and (ii) under the alternative (4.3) with $\min_{0 \leq i \leq M^*} |k_{i+1}^* - k_i^*|/N > 3\varepsilon/2$ and $\boldsymbol{\theta}_{k_{i+1}^*} - \boldsymbol{\theta}_{k_i^*} = N^{-1/2} L_i^*$ for some $L_i^* \neq 0$, $i = 1, \dots, M^*$, we have $G_N^\diamond \rightarrow \infty$ in probability if $\min_{1 \leq i \leq M^*} |L_i^*| \rightarrow \infty$.

Therefore, the cut-off values can be obtained by simulating the distribution of $G^\diamond(\mathbf{B})$, which is often achieved in practice by simulating the test statistic G_N^\diamond for lengthy Gaussian white noises. Since a large number of simulated replications are

needed to accurately estimate the cut-off values, reducing the computational burden needed for computing the test statistic for each realization can be of significant importance. It can be seen from our simulation results in Section 5.3.1 that the above grid approximation scheme leads to tests with reasonably good performance in terms of both the size and power.

Proof. (Proposition 4.4.1) Note that $(l_1, l_2) \in \mathcal{G}_{\varepsilon, N, f}$ if and only if $(l_1, l_2) = (\lfloor Nr_1 \rfloor, \lfloor Nr_2 \rfloor)$ for some $(r_1, r_2) \in ([0, 1] \times \mathcal{G}_\varepsilon) \cap \Omega(\varepsilon)$, the first claim follows by the argument of Theorem 4.3.1 and the continuous mapping theorem. We shall now provide the proof for (ii). For this, note that for any $l_2 \in (k_1^*, k_2^*]$,

$$\mathbf{D}_{n, f}^\diamond(1, k_1^*, l_2) = \frac{k_1^*(l_2 - k_1^*)}{l_2^{3/2}} \{(\hat{\boldsymbol{\theta}}_{1, k_1^*} - \boldsymbol{\theta}_{k_1^*}) - (\hat{\boldsymbol{\theta}}_{k_1^*+1, l_2} - \boldsymbol{\theta}_{k_1^*+1})\} - \frac{k_1^*(l_2 - k_1^*)}{l_2 \sqrt{N} l_2} \mathbf{L}_1^*,$$

then by condition (IP^{*}), it suffices to prove that there exists a sequence $l_2(N) \in (k_1^*, k_2^*]$ such that $\{k_1^*, l_2(N)\} \in \mathcal{G}_{\varepsilon, N, f}$ and $\{l_2(N) - k_1^*\}/l_2(N)$ is bounded away from zero for all large N . For each $N = 1, 2, \dots$, by the construction of \mathcal{G}_ε , there exists an element, denoted by a_N , such that $a_N \in \mathcal{G}_\varepsilon \cap (k_2^*/N - \varepsilon/2, k_2^*/N]$. Note that $(Na_N - k_1^*)/(Na_N) \geq (k_2^*)^{-1}(N\varepsilon) \geq \varepsilon$ and $(Na_N - \lfloor Na_N \rfloor)/N \leq N^{-1}$, the result follows by letting $l_2(N) = \lfloor Na_N \rfloor \vee 1$. \square

4.5 Simulation Results

In this section, we will repeat our Monte Carlo simulations proposed in Chapter 2 to examine the finite-sample performance of the proposed hypothesis testing procedure and to make comparisons with the tests of Shao and Zhang (2010). This section is organized as follows. We will first provide a numerical comparison between the computationally efficient test which will use the proposed approximation scheme (UCbSN) and its full-scale counterpart (UCbSN_{full}). Later we will conduct a numerical comparison between the contrast-based method and the conventional method. Then, we will demonstrate finite sample performance of the proposed contrast-based method using the approximation scheme for the case of median, variance and autocorrelation

and it's comparison with the tests of Shao and Zhang (2010).

As previously defined, w_i , $i \in \mathbb{Z}$, are the independent standard normal random variables, and (e_i) is an autoregressive process satisfying the recursion $e_i = \rho e_{i-1} + w_i$. With $1_{\{\cdot\}}$ we denote the indicator function and the model of interest is:

$$X_i = \mu_i + e_i, \quad i = 1, \dots, n, \quad (4.12)$$

with the following change-point alternatives:

(M1) (one change-point alternative): $\mu_i = d1_{\{i/n > 2/3\}}$;

(M2) (two change-point alternative): $\mu_i = d1_{\{2/3 \geq i/n > 1/3\}} - d1_{\{i/n > 2/3\}}$;

(M3) (three change-point alternative): $\mu_i = d1_{\{i/n \leq 1/2\}} - d1_{\{i/n \leq 1/4\}} + d1_{\{i/n \geq 3/4\}}$,

where $d \in \mathbb{R}$ controls the amount of deviation from the null. For (M1)–(M3), the case with $d = 0$ corresponds to the null hypothesis of no change point. Let $n = 500$ and $\rho \in \{\pm 0.3, \pm 0.6, \pm 0.8\}$, we consider testing for change points in the median of the process (4.12). As before, we consider SZ10 test statistics constructed under one, two and three hypothesized change points, which we denote by $SZ10_1$, $SZ10_2$ and $SZ10_3$ respectively. For each configuration, we generate 2,000 realizations, and the results are summarized in Sections 4.5.3 and 4.5.3 for comparisons under the null and under the alternative respectively. Additional simulation results for the case of variance and the autocorrelation function will also be provided in Section 4.5.4.

4.5.1 Comparison: The computationally efficient test and its full-scale counterpart

In order to provide a comprehensive comparison between the approximation scheme and the full-scale test we will use both to examine the empirical acceptance rates and the power properties for the case of the mean function.

ρ	UCbSN		UCbSN _{full}	
	90%	95%	90%	95%
0.3	0.884	0.941	0.886	0.944
0.6	0.862	0.932	0.872	0.930
0.8	0.784	0.866	0.794	0.872
-0.3	0.910	0.958	0.916	0.963
-0.6	0.932	0.968	0.938	0.975
-0.8	0.953	0.976	0.960	0.981

Table 4.1: Empirical acceptance rates for testing change points in the mean of (4.12) under different dependence strengths. For both the proposed UCbSN test and its full-scale version results with 90% and 95% nominal levels are reported.

We first examine the empirical acceptance rates for the change points being tested in the case of the mean with different dependence strengths. Table 4.1 provides these empirical acceptance rates for 90% and 95% nominal levels in the case of the mean.

In table 4.2 below we also provide comparison of power properties of the two tests for the alternatives discussed and defined in 4.12. The proposed test that uses the approximation scheme is denoted by UCbSN and its full-scale version is denoted as UCbSN_{full}.

As shown in tables 4.1-4.2 the approximation scheme maintains a similar performance in size and power for all strengths of dependency. Thus, going forward by exploiting the grid version of the test, we should not expect any significant loss in size or power of the test. In addition, the proposed implementation scheme can significantly reduce the computational burden, especially if compared with the tests for the multiple change-point alternatives SZ10₂ and SZ10₃. In order to demonstrate the improvement in computational complexity of the test, we provide the computational time of the alternative test as a ratio with respect to the time of the proposed UCbSN test (baseline) for the different sample sizes $n = 100k$, $k \in \{1, 2, 3, 4, 5\}$. As shown

rho	d	(M1)		(M2)		(M3)	
		UCbSN	UCbSN _{full}	UCbSN	UCbSN _{full}	UCbSN	UCbSN _{full}
0.3	0.0	0.05	0.05	0.05	0.05	0.05	0.05
	0.2	0.16	0.15	0.31	0.30	0.12	0.12
	0.4	0.49	0.44	0.85	0.84	0.28	0.29
	0.6	0.83	0.79	0.99	0.99	0.66	0.66
	0.8	0.97	0.96	1.00	1.00	0.91	0.90
	1.0	1.00	0.99	1.00	1.00	0.99	0.99
	2.0	1.00	1.00	1.00	1.00	1.00	1.00
	3.0	1.00	1.00	1.00	1.00	1.00	1.00
0.6	0.0	0.05	0.05	0.05	0.05	0.05	0.05
	0.2	0.08	0.08	0.13	0.12	0.06	0.07
	0.4	0.18	0.16	0.38	0.34	0.11	0.12
	0.6	0.35	0.32	0.71	0.67	0.20	0.21
	0.8	0.59	0.55	0.91	0.89	0.38	0.38
	1.0	0.78	0.74	0.98	0.97	0.57	0.56
	2.0	1.00	1.00	1.00	1.00	0.99	0.99
	3.0	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.2: Size-adjusted empirical powers for testing change points in the mean of (4.12) under (M1)-(M3) with different dependence strengths. The proposed test using the approximation scheme is denoted by UCbSN and its full-scale version is denoted as UCbSN_{full}. For both tests the significance level $\alpha = 0.05$ is used.

in table 4.3 the proposed test is computationally much more efficient.

Additionally, as shown in table 4.3:

- computational times of the SZ10 test are highly dependent on the number of prespecified change points, and increase rapidly as the number of change points being tested against increases
- the single-change point test SZ10₁ has the simplest form and requires the least number of computations as expected

n	100	200	300	400	500
UCbSN	1.00	1.00	1.00	1.00	1.00
UCbSN _{full}	3.96	6.47	8.13	9.20	10.30
SZ10 ₁	0.26	0.40	0.50	0.55	0.63
SZ10 ₂	3.01	4.85	6.17	6.86	7.68
SZ10 ₃	54.28	169.88	315.37	461.30	652.88

Table 4.3: Relative computational time of the full-scale UCbSN_{full} test with respect to UCbSN (baseline) for different sample sizes n . The results are based on average of five independent realizations.

d	UCbSN ₂	UCbSN	UCbSN _{0.5}
0.0	0.057	0.059	0.063
0.2	0.352	0.352	0.357
0.4	0.873	0.877	0.874
0.6	0.994	0.995	0.994
0.8	1.000	1.000	1.000
1.0	1.000	1.000	1.000
2.0	1.000	1.000	1.000

Table 4.4: Empirical power for testing change points in the mean of (M2) with $\rho = 0.3$ using grids with different densities. The significance level $\alpha = 0.05$ is used.

- the proposed test can be computationally more efficient than multiple change-points tests, SZ10₂ and SZ10₃.

We also consider the case with a denser grid (denoted by UCbSN₂) which uses a grid with double density and a less dense grid (denoted by UCbSN_{0.5}) which uses a grid with half density. As shown in table 4.4 the results are not significantly affected by the density of the grid. The problem of defining and finding the optimal grid density requires quantifying the tradeoff between the approximation error and the corresponding computational burden, which we shall leave as a possible future research topic.

To summarize findings from this series of simulations studies we conclude that the approximation scheme we have proposed has similar performance as the version of the test with no approximation involved, but significantly improves the computational complexity of the test. Thus, we will perform all the subsequent simulations in this chapter using the implementation scheme proposed in 4.4.

4.5.2 Comparison: The proposed contrast-based method and the conventional method

To support our proposal for the contrast-based method we shall here provide a numerical comparison between the contrast-based method and the conventional method. We would remind that for the case of the mean the results of the conventional test are identical to the results of the contrast-based test, thus we will focus our attention on the case of the median function. More specifically, we consider the proposed unsupervised self-normalized test using both extensions for testing change point in the median. Our simulation results suggest that the contrast-based method in general leads to the less size distortions and better power performance. In addition, it enables us to study the asymptotic power of self-normalized test statistic under the more desirable deterministic condition $N^{1/2}|\boldsymbol{\theta}_{k_i^*+1} - \boldsymbol{\theta}_{k_i^*}| \rightarrow \infty$, while a less desirable stochastic condition is required for the conventional approach. The results are summarized in the tables 4.5 and 4.6 where size and power analysis is presented. Based on these results we will conclude that the contrast-based method has shown to be preferable over the conventional generalization and going forward we will focus on analyzing finite sample performance of the proposed UCbSN test.

ρ	UCbSN		USN	
	90%	95%	90%	95%
0.3	0.835	0.904	0.764	0.846
0.6	0.805	0.891	0.744	0.836
0.8	0.765	0.845	0.704	0.796
-0.3	0.825	0.900	0.732	0.811
-0.6	0.842	0.906	0.708	0.791
-0.8	0.842	0.910	0.708	0.779

Table 4.5: Empirical acceptance rates for testing change points in the median of under different dependence strengths. For the proposed unsupervised test, its contrast-based extension is denoted by UCbSN and the non-contrast-based extension is denoted by USN. For both test results with 90% and 95% nominal levels are reported.

4.5.3 Comparison with the tests of Shao and Zhang (2010)

Empirical Acceptance Rate

In this subsection we will compare the proposed contrast-based method using the approximation scheme with the tests of Shao and Zhang (2010).

We first examine the performance under the null, and Table 4.5 provides the empirical acceptance rates for 90% and 95% nominal levels. For the SZ10 test, it is supervised and it can be seen from Table 4.7 that the size distortions of SZ10₂ and SZ10₃, multiple change-point SZ10 tests hereby, deteriorates quickly for the median case. In particular, the empirical acceptance rate of the SZ10₃ test is 0.585 even with $\rho = 0.3$, which is far away from its nominal 95% level. The proposed test's performance in size can largely improve over both SZ10₂ and SZ10₃ in the median case. Therefore, besides being unsupervised, the proposed test also seems to deliver more reasonable performance in size than its multiple change-point SZ10 competitors.

rho	d	(M1)		(M2)		(M3)	
		UCbSN	USN	UCbSN	USN	UCbSN	USN
0.3	0.0	0.050	0.050	0.050	0.050	0.050	0.050
	0.2	0.109	0.096	0.219	0.180	0.092	0.083
	0.4	0.330	0.264	0.700	0.572	0.201	0.165
	0.6	0.644	0.545	0.953	0.866	0.444	0.336
	0.8	0.868	0.786	0.996	0.979	0.741	0.599
	1.0	0.972	0.932	1.000	0.998	0.916	0.792
	2.0	1.000	1.000	1.000	1.000	1.000	1.000
	3.0	1.000	1.000	1.000	1.000	1.000	1.000
0.6	0.0	0.050	0.050	0.050	0.050	0.050	0.050
	0.2	0.057	0.060	0.117	0.102	0.062	0.050
	0.4	0.121	0.099	0.316	0.244	0.082	0.070
	0.6	0.262	0.211	0.565	0.460	0.152	0.135
	0.8	0.444	0.360	0.830	0.718	0.287	0.212
	1.0	0.637	0.528	0.940	0.869	0.430	0.342
	2.0	0.991	0.973	1.000	0.999	0.968	0.913
	3.0	1.000	1.000	1.000	1.000	0.999	1.000

Table 4.6: Size-adjusted empirical powers for testing change points in the median of (4.12) under (M1)–(M3) with different dependence strengths. For the proposed unsupervised test, its contrast-based extension is denoted by UCbSN and the non-contrast-based extension is denoted by USN. For both test results with 90% and 95% nominal levels are reported.

Empirical Power

We shall now consider the same simulations settings for one and multiple change-point alternatives as in Chapter 2 to analyze the finite sample performance of the proposed test statistic. We will reuse the alternatives (M1)–(M3) and will examine the performance under these alternatives. Tables 4.8–4.10 provide the size-adjusted empirical powers under (M1)–(M3) respectively for $\rho \in \{0.3, 0.6\}$. For the SZ10 test, it requires prespecifying the number of change points and it can be seen from Tables 4.8–4.10 that, depending on whether this prior information is correctly specified or

ρ	UCbSN		SZ10 ₁		SZ10 ₂		SZ10 ₃	
	90%	95%	90%	95%	90%	95%	90%	95%
<i>Results for the median</i>								
0.3	0.835	0.904	0.867	0.920	0.743	0.825	0.456	0.572
0.6	0.805	0.891	0.880	0.936	0.723	0.817	0.493	0.602
0.8	0.765	0.845	0.872	0.927	0.697	0.792	0.434	0.534
-0.3	0.825	0.900	0.872	0.926	0.682	0.778	0.348	0.458
-0.6	0.842	0.906	0.869	0.922	0.655	0.740	0.269	0.357
-0.8	0.842	0.910	0.859	0.912	0.635	0.727	0.213	0.297

Table 4.7: Empirical acceptance rates for testing change points in the median of (4.12) under different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), results with 90% and 95% nominal levels are reported.

not, the power performance of their test can be largely affected. For example, when there is only a single change point as in (M1), then prespecifying two or more change points can lead to serious power losses for the SZ10 test, as in this case empirical powers of SZ10₂ and SZ10₃ are much lower than that of SZ10₁ from Table 4.8. On the other hand, when there are multiple change points as in (M2) and (M3), then serious power losses can be observed for the single change-point SZ10₁ test from Tables 4.9 and 4.10. In contrast, the proposed test does not require prespecifying the number of change points and seems to deliver a competitive power performance even when compared with the oracle SZ10 test. In terms of the power performance, the proposed test seems to be more reliable in situations where the number of change points is unknown.

4.5.4 Testing for Change Points in Variance and Autocorrelation

In this subsection, we consider testing for change points in the variance and autocorrelation, and similarly to the simulations study presented above we compare the proposed test with the test of Shao and Zhang (2010). Following Shao and Zhang

ρ	d	<i>Median</i>			
		UCbSN	SZ10 ₁	SZ10 ₂	SZ10 ₃
0.3	0.0	0.050	0.050	0.050	0.050
	0.2	0.109	0.112	0.070	0.052
	0.4	0.330	0.334	0.154	0.099
	0.6	0.644	0.613	0.328	0.164
	0.8	0.868	0.812	0.532	0.302
	1.0	0.972	0.935	0.741	0.478
	2.0	1.000	0.999	0.998	0.973
	3.0	1.000	1.000	1.000	1.000
	4.0	1.000	1.000	1.000	1.000
	5.0	1.000	1.000	1.000	1.000
0.6	0.0	0.050	0.050	0.050	0.050
	0.2	0.057	0.085	0.052	0.066
	0.4	0.121	0.163	0.068	0.070
	0.6	0.262	0.319	0.124	0.102
	0.8	0.444	0.493	0.199	0.161
	1.0	0.637	0.664	0.326	0.222
	2.0	0.991	0.979	0.882	0.743
	3.0	1.000	0.997	0.996	0.972
	4.0	1.000	0.999	1.000	0.999
	5.0	1.000	0.998	1.000	1.000

Table 4.8: Size-adjusted empirical powers for testing change points in the median of (4.12) under (M1) with different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

(2010), we consider first-order autoregressive models with shifts in its autoregressive coefficient. Let $w_i, i \in \mathbb{Z}$, be the independent standard normal random variables, and

$$X_i = \rho_i X_{i-1} + w_i, \quad i = 1, \dots, n \quad (4.13)$$

Similar to the mean and median case, we consider the following change-point alternatives:

ρ	d	<i>Median</i>			
		UCbSN	SZ10 ₁	SZ10 ₂	SZ10 ₃
0.3	0.0	0.050	0.050	0.050	0.050
	0.2	0.219	0.102	0.167	0.126
	0.4	0.700	0.151	0.522	0.340
	0.6	0.953	0.144	0.840	0.675
	0.8	0.996	0.100	0.971	0.904
	1.0	1.000	0.057	0.997	0.981
	2.0	1.000	0.000	1.000	1.000
	3.0	1.000	0.000	1.000	1.000
	4.0	1.000	0.000	1.000	1.000
	5.0	1.000	0.000	1.000	1.000
0.6	0.0	0.050	0.050	0.050	0.050
	0.2	0.117	0.074	0.095	0.081
	0.4	0.316	0.137	0.212	0.160
	0.6	0.565	0.157	0.412	0.321
	0.8	0.830	0.164	0.677	0.556
	1.0	0.940	0.130	0.839	0.739
	2.0	1.000	0.015	0.998	0.999
	3.0	1.000	0.000	1.000	1.000
	4.0	1.000	0.000	1.000	1.000
	5.0	1.000	0.000	1.000	1.000

Table 4.9: Size-adjusted empirical powers for testing change points in the median of (4.12) under (M2) with different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

(V1) (one change-point alternative): $\rho_i = r1_{\{i/n > 2/3\}}$;

(V2) (two change-point alternative): $\rho_i = r1_{\{2/3 \geq i/n > 1/3\}} - r1_{\{i/n > 2/3\}}$;

(V3) (three change-point alternative): $\rho_i = r1_{\{i/n > 1/4\}} - r1_{\{3/4 \geq i/n > 1/2\}}$,

where $r \in \{0, 0.3, 0.6, 0.8\}$ represents the amount of deviation from the null. Let $n = 500$, we consider testing change points in the variance and the first-order auto-

ρ	d	<i>Median</i>			
		UCbSN	SZ10 ₁	SZ10 ₂	SZ10 ₃
0.3	0.0	0.050	0.050	0.050	0.050
	0.2	0.092	0.046	0.142	0.075
	0.4	0.201	0.036	0.250	0.148
	0.6	0.444	0.021	0.455	0.284
	0.8	0.741	0.010	0.625	0.470
	1.0	0.916	0.000	0.991	0.696
	2.0	1.000	0.000	1.000	0.998
	3.0	1.000	0.000	1.000	1.000
	4.0	1.000	0.000	1.000	1.000
	5.0	1.000	0.000	1.000	1.000
0.6	0.0	0.050	0.050	0.050	0.050
	0.2	0.062	0.052	0.053	0.074
	0.4	0.082	0.050	0.074	0.138
	0.6	0.152	0.056	0.115	0.218
	0.8	0.287	0.047	0.183	0.320
	1.0	0.430	0.036	0.256	0.882
	2.0	0.968	0.006	0.780	0.997
	3.0	0.999	0.000	0.979	1.000
	4.0	1.000	0.001	1.000	1.000
	5.0	1.000	0.035	1.000	1.000

Table 4.10: Size-adjusted empirical powers for testing change points in the median of (4.12) under (M3) with different dependence strengths. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

correlation under the (V1) - (V3). The results summarized in the tables 4.11 - 4.12 are in concordance with the conclusions made previously for the case of the mean and the median. This supports our statement that the proposed test UCbSN is applicable to a wide range of statistics or interest in it demonstrates good power and size properties when applied even to a relatively small sample.

Model	r	Variance			
		SZ10 ₁	SZ10 ₂	SZ10 ₃	UCbSN
(V1)	0.0	0.052	0.048	0.046	0.062
	0.3	0.085	0.054	0.054	0.088
	0.6	0.542	0.255	0.137	0.480
	0.8	0.913	0.790	0.562	0.940
(V2)	0.0	0.052	0.048	0.046	0.062
	0.3	0.077	0.068	0.054	0.087
	0.6	0.329	0.281	0.147	0.416
	0.8	0.518	0.752	0.527	0.839
(V3)	0.0	0.052	0.048	0.046	0.062
	0.3	0.063	0.056	0.052	0.081
	0.6	0.084	0.206	0.264	0.292
	0.8	0.107	0.601	0.746	0.720

Table 4.11: Empirical powers for testing change points in the variance and first-order autocorrelation of (4.13) under different change-point alternatives. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

4.6 A Real Data Analysis

We shall here apply the results to a climate science data for assessing the temporal stability or stationarity of tropical cyclones. It has been argued and believed in the climate science literature that tropical cyclone winds should increase with increasing ocean temperature, one of the major consequences of global warming. As a result, researchers have been seeking empirical evidences on changes in tropical cyclone winds; see for example Elsner et al. (2008), Zhou (2010), Zhang and Wu (2011) and references therein. Also analysis using quantile regression methods here Koenker and Gilbert Bassett (1978) and Jagger and Elsner (2009) Analyses in the aforementioned papers require consistent estimation of the nuisance long-run variance, and we shall here use self-normalization to alleviate the difficult issue of estimating the long-run variance under the change-point alternative; see Shao (2010b) and Shao and Zhang

Model	r	<i>Autocorrelation</i>			
		SZ10 ₁	SZ10 ₂	SZ10 ₃	UCbSN
(V1)	0.0	0.059	0.091	0.126	0.086
	0.3	0.647	0.392	0.345	0.592
	0.6	0.992	0.924	0.903	0.989
	0.8	1.000	0.998	0.996	1.000
(V2)	0.0	0.059	0.091	0.126	0.086
	0.3	0.222	0.928	0.903	0.930
	0.6	0.184	1.000	1.000	1.000
	0.8	0.228	1.000	1.000	1.000
(V3)	0.0	0.059	0.091	0.126	0.086
	0.3	0.071	0.340	0.622	0.390
	0.6	0.118	0.893	0.997	0.958
	0.8	0.551	0.994	1.000	0.999

Table 4.12: Empirical powers for testing change points in the first-order autocorrelation of (4.13) under different change-point alternatives. For both the proposed UCbSN test and the SZ10 tests of Shao and Zhang (2010), the significance level $\alpha = 0.05$ is used.

(2010) for advantages of using self-normalization. A time series plot of the data is given in Figure 4.1, which contains satellite-derived lifetime-maximum wind speeds of 2098 tropical cyclones over the globe during 1981–2006, and we refer to Elsner et al. (2008) for a more detailed description. The results are summarized in Table 4.14, and a comparison with the supervised test of Shao and Zhang (2010) is also given. It can be seen from Table 4.14 that, at 5% significance level, the proposed unsupervised test finds statistical evidence against the hypotheses of mean constancy and variance constancy, indicating changes in tropical cyclone winds. As a comparison, results from the test of Shao and Zhang (2010) seem less conclusive and depend on what number of change points one would like to prespecify for their test. Since the number of change points is typically unknown in practice especially when one is at the stage of seeking statistical tests to determine their existence, the proposed unsupervised test

	<i>Mean</i>			
	UCbSN	SZ10 ₁	SZ10 ₂	SZ10 ₃
Test statistic	198.67	35.27	184.42	225.26
<i>p</i> -value	0.011	0.070	0.012	0.061

Table 4.13: Test statistics and their *p*-values for the tropical cyclone data.

	<i>Variance</i>			
	UCbSN	SZ10 ₁	SZ10 ₂	SZ10 ₃
Test statistic	210.84	97.42	133.65	135.47
<i>p</i> -value	0.008	0.003	0.052	0.329

Table 4.14: Test statistics and their *p*-values for the tropical cyclone data.

seems to provide a useful alternative for self-normalized change-point detection.

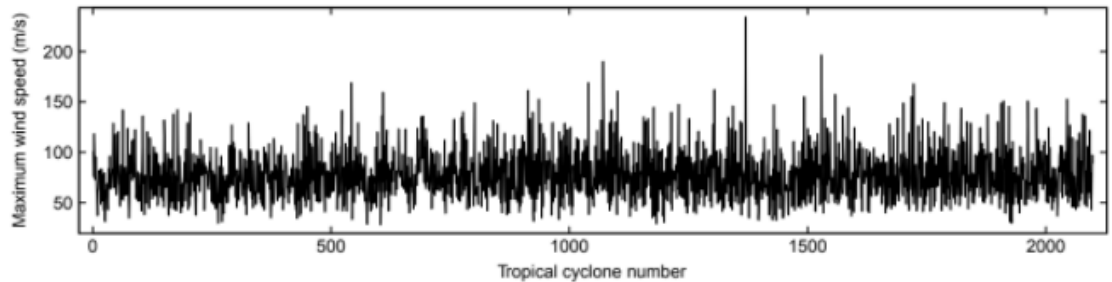


Figure 4.1: Satellite-derived lifetime-maximum wind speeds of tropical cyclones during 1981–2006.

4.7 Conclusion

We are proposing to use a new approach, called the contrast-based method, for generalizing self-normalized statistics to quantities other than the mean. Unlike the conventional approach which was mainly developed for constructing confidence intervals for quantities associated with a stationary time series, the proposed contrast-based approach is specifically tailored for change-point testing problems and can lead to tests with better power performance. Another useful feature of the proposed

contrast-based approach is that it allows one to study the asymptotic power behavior of self-normalized tests under a more desirable and interpretable deterministic condition as in Theorem 4.3.1 (ii); see condition (4.9) for a comparison.

Chapter 5

A Time-Symmetric Self-Normalization Approach for Inference of Time Series

5.1 Introduction and Problem Formulation

To make statistical inference about an unknown quantity, it is typical that the procedure may require the estimation of certain nuisance parameters, which are not of direct interest but need to be appropriately accounted for. For example, when one is interested in conducting hypothesis testing or constructing confidence interval for the mean of independent and identically distributed observations, the marginal variance then becomes the nuisance parameter which appears in the distribution of the sample mean. Same it true for statistical inference about unknown quantities in time series analysis. Self-normalization can be leveraged for this statistical analysis as well as for the problems discussed earlier to avoid direct estimation of the nuisance asymptotic variance. As we have presented earlier, in the mean case self-normalization is applied to the CUSUM process by exploiting the invariance principle of the partial sum process. In this case the normalizer will converge in distribution to a function of a Brownian motion on the unit interval. When the quantity of interest is beyond the mean, the role that the partial sum process played in the mean case is then replaced by a sequence of recursive estimators. In this case, however, the resulting self-normalized statistic can exhibit certain degrees of asymmetry resulting in an unbalanced use of the information. This problem was also discussed in Shao (2015) where unfavorability

of asymmetric self-normalizers was first brought up.

The issue of asymmetry can also lead to different p -values and confidence intervals depending on the direction of application. Thus, depending on the direction of application one can derive different conclusions if a asymmetric test is being applied. This is considered as an unpleasant issue, especially for time-reversible processes where the direction of application should not have any impact. Note that stationary Gaussian processes are widely used in time series analysis and they are all time-reversible. On the other hand, regardless of whether the underlying process is time-reversible, many existing inference procedures, including those based on Central Limit Theorem and Functional CLT and most existing long-run variance estimators, are all time-symmetric (T-symmetric). Given the aforementioned concerns and the discussion with Professor Steven Lalley and Professor Michael Stein at The University of Chicago, Shao (2015) considered two different approaches to symmetrizing the conventional self-normalizer.

The first approach considered by Shao (2015) is to compute recursive estimators from all possible blocks to form the self-normalizer, which makes the resulting self-normalizer T-symmetric but the associated computation is substantially more expensive than the conventional approach as observed by Shao (2015). In addition, such a self-normalizer does not unify with the conventional approach in the mean case, and small scale simulations provided by the author suggest that its performance in the mean case can be noticeably worse than the conventional self-normalizer of Shao (2010b). To address this issue, a second approach is the considered, based on the suggestion of Professor Michael Stein at The University of Chicago, whose most recent research in applied statistical analysis can be found in widely cited papers including Wang and Han (2016), Poppick et al. (2016) and Sun and M.L. Stein (2016). Professor Stein's suggestion was to compute the conventional self-normalizer for both

the original process and its reversed counterpart and take the average. A simulation comparison on constructing confidence intervals for the median of autoregressive processes was conducted in Shao (2015), and it was found that such a symmetrizing approach generally leads to wider confidence intervals, or tests with lower powers, when compared with the conventional self-normalizer of Shao (2010b). Thus, neither of existing approaches seems to be fully satisfactory, as none of them can exhibit symmetry together with strong performance.

In our work we seek a new approach in generalizing conventional self-normalizers to their T-symmetric counterpart. In particular, we propose to first augment the conventional self-normalizer to a wider class by exploiting mathematical properties of the cumulative sum (CUSUM) process, and then identify the T-symmetric subspace of the augmented class of self-normalizers. The proposed self-normalizer obtained through this approach enjoys the following desirable features. First, when one is interested in the mean where the conventional self-normalizer is already T-symmetric, then the proposed self-normalizer will automatically reduce to the conventional one to provide a unified inference procedure. Second, when the quantity of interest is beyond the mean where the conventional self-normalizer is not T-symmetric, then the proposed self-normalizer will serve as a T-symmetric generalization of the conventional approach to provide a T-symmetric inference procedure. In addition, it can be seen from our numerical results in Section 3 that the proposed self-normalizer can also lead to possibly narrower confidence intervals with improved empirical coverage probabilities for quantities such as the median when compared with the conventional self-normalizer of Shao (2010). In this chapter we will introduce the proposed class of self-normalizers and will study the asymptotic validity of the resulting statistical inference procedure. Numerical experiments are also conducted and described to investigate its finite-sample performance and make a comparison with the conventional

self-normalizer.

5.2 T-Symmetric Self-Normalizer

We shall here introduce the formal mathematical definition of T-symmetric statistics.

Definition 5.2.1. A statistic $G(X_1, \dots, X_n)$ is said to be T-symmetric if it is invariant to the direction of application, namely if $G(X_1, \dots, X_n) = G(X_n, \dots, X_1)$ for any time series data X_1, \dots, X_n .

In the time series literature, many commonly used inference procedures, including those based on central limit theorems such as including both traditional CLT and it's functional modification, and most existing long-run variance estimators, are all T-symmetric. The original self-normalizer:

$$U_n^2 = n^{-2} \sum_{k=1}^n (S_k - k\bar{X}_n)^2 \quad (5.1)$$

is also T-symmetric as was developed for the case of the mean function and is closely related to the CUSUM process.

Proposition 5.2.2. *For the statistic U_n^2 defined in 5.1, when being viewed as a function of X_1, \dots, X_n , it satisfies $U_n^2(X_1, \dots, X_n) = U_n^2(X_n, \dots, X_1)$*

Proof. (Proposition 5.2.2) Let $X_i^\bullet = X_{n-i+1}$, $i = 1, \dots, n$, then $(X_1^\bullet, \dots, X_n^\bullet)$ denotes the reversed sample (X_n, \dots, X_1) , and it can be seen that its partial sums satisfy

$$\sum_{i=1}^k X_i^\bullet = \sum_{i=1}^k X_{n-i+1} = S_n - S_{n-k}, \quad k = 1, \dots, n.$$

Therefore, by changing the index in the summation, one has

$$\begin{aligned}
U_n^2(X_1^\bullet, \dots, X_n^\bullet) &= n^{-2} \sum_{k=1}^{n-1} \left(S_n - S_{n-k} - \frac{k}{n} S_n \right)^2 \\
&= n^{-2} \sum_{k=1}^{n-1} \left(S_{n-k} - \frac{n-k}{n} S_n \right)^2 \\
&= n^{-2} \sum_{k=1}^{n-1} \left(S_k - \frac{k}{n} S_n \right)^2 = U_n^2(X_1, \dots, X_n),
\end{aligned}$$

and the result follows. \square

In the seminal work of Shao (2010b), the above self-normalizer (5.1) was generalized to handle quantities other than the mean. In particular, suppose one is interested in the parameter $\theta = T(F)$ where F denotes the marginal distribution of X_i , and let \hat{F}_k be the empirical distribution based on X_1, \dots, X_k . Note that (5.1), the self-normalizer for the mean, involves $\frac{S_k}{k}$, which can be viewed as a recursive mean estimator, and therefore a natural strategy as proposed by Shao (2010b) and advocated in its subsequent works is to replace $\frac{S_k}{k}$ in the mean self-normalizer (5.1) by the recursive estimator $T(\hat{F}_k)$, which leads to the self-normalizer:

$$V_n^2 = n^{-2} \sum_{k=1}^n k^2 \{T(\hat{F}_k) - T(\hat{F}_n)\}^2 \quad (5.2)$$

One can see that when being generalized to handle quantities other than the mean, for example the median, the conventional approach of Shao (2010) does not preserve the T-symmetric property, namely the self-normalizer 5.2 is generally speaking not T-symmetric. To illustrate, we consider the simple example where $n = 2m + 1$ for some positive integer $m \in \mathbb{Z}$ and $X_i = I_{(i)/2} \in \mathbb{Z}$, $i = 1, \dots, n$. Then it can be seen that for the median case the self-normalizer given in (5.2) as a statistic satisfies $V_n^2(X_1, \dots, X_n) = 0$ and $V_n^2(X_n, \dots, X_1) = (2m + 1)^{-2} \sum_{k=1}^m \{k^2 + (2k - 1)^2\} > 0$,

and is thus not T-symmetric.

We shall here consider a new approach in generalizing self-normalization methods to situations beyond the mean that is able to

- preserve the T-symmetric property
- reduce to the self-normalizer (5.1) in the mean case to yield a unified self-normalization protocol
- possibly lead to narrower confidence intervals with improved coverage probabilities when compared with the conventional approach (5.2) of Shao (2010b) for quantities such as the median or other quantiles

Although simple approaches may exist for making the self-normalizer (5.2) T-symmetric, finding a suitable generalization that can simultaneously achieve the aforementioned three merits can be nontrivial.

To introduce the proposed self-normalizer in its general form, suppose the parameter of interest is $\boldsymbol{\theta} = \mathbf{T}(\mathbf{F}^\ell)$, where \mathbf{F}^ℓ denotes the joint distribution of $\mathbf{Y}_k = (X_k, \dots, X_{k+\ell-1})^\top \in \mathbb{R}^\ell$ with \top being the transpose operator. Let $\hat{\mathbf{F}}_{i,j}^\ell$ be the empirical distribution function of $\mathbf{Y}_i, \dots, \mathbf{Y}_j$ and $\hat{\boldsymbol{\theta}}_{i,j} = \mathbf{T}(\hat{\mathbf{F}}_{i,j}^\ell)$, $1 \leq i \leq j \leq N = n - \ell + 1$. In this case, the self-normalizer (5.2) becomes a function of the recursive difference $\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,N} = \mathbf{T}(\hat{\mathbf{F}}_{1,k}^\ell) - \mathbf{T}(\hat{\mathbf{F}}_{1,N}^\ell)$, $k = 1, \dots, N$, and it is not T-symmetric mainly due to its involvement of only forward recursive estimators. As a result, a natural modification is to consider $(\hat{\boldsymbol{\theta}}_{1,k} + \hat{\boldsymbol{\theta}}_{N-k+1,N})/2 - \hat{\boldsymbol{\theta}}_{1,N}$, where $\hat{\boldsymbol{\theta}}_{1,k} = \mathbf{T}(\hat{\mathbf{F}}_{1,k}^\ell)$ and $\hat{\boldsymbol{\theta}}_{N-k+1,N} = \mathbf{T}(\hat{\mathbf{F}}_{N-k+1,N}^\ell)$ represent the forward and backward recursive estimator respectively. This simple approach, however, does not reduce to the self-normalizer (5.1) in the mean case. To be more specific, when $\hat{\boldsymbol{\theta}}_{i,j}$ is the sample mean of X_i, \dots, X_j , then $(\hat{\boldsymbol{\theta}}_{1,k} + \hat{\boldsymbol{\theta}}_{N-k+1,N})/2 - \hat{\boldsymbol{\theta}}_{1,N} = \{S_k/k + (S_n - S_{n-k})/k\}/2 - S_n/n$ which does

not correspond to $S_k/k - S_n/n$ as featured in (5.1). In addition, this discrepancy will also cause the resulting self-normalized statistic to have a different asymptotic distribution from that tabulated in Lobato (2001). We shall here consider a more sophisticated approach in generalizing the self-normalizer (5.2) of Shao (2010b).

For this, we propose to first augment the self-normalizer of Shao (2010b) to a wider class, within which we then aim at identifying its T-symmetric subspace. In particular, we propose to exploit the mathematical equivalency of $S_k - k\bar{X}_n$ that makes the self-normalizer (5.1) T-symmetric in the mean case to achieve the augmentation. To be more specific, by adding and subtracting $S_n = n\bar{X}_n$ in the mathematical expression, we have

$$\begin{aligned} S_k - k\bar{X}_n &= S_k - S_n + S_n - kS_n/n \\ &= (n - k)\bar{X}_n - (S_n - S_k) \\ &= (S_k - k\bar{X}_n)w - \{(S_n - S_k) - (n - k)\bar{X}_n\}(1 - w) \end{aligned}$$

holds for any $w \in [0, 1]$. Therefore, we propose to consider an augmented class of self-normalizers that has the form

$$\mathbf{\Lambda}_N(w) = N^{-1} \sum_{k=1}^{N-1} \boldsymbol{\lambda}_k(w) \boldsymbol{\lambda}_k(w)^\top, \quad (5.3)$$

where $w \in [0, 1]$ denotes the weight and

$$\boldsymbol{\lambda}_k(w) = N^{-1/2} \{k(\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,N})w - (N - k)(\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N})(1 - w)\}. \quad (5.4)$$

The augmented class of self-normalizers given in (5.3) generalizes the conventional one of Shao (2010b) by incorporating not only the forward but also backward recursive estimators. It can be seen from the construction in (5.4) that the augmentation mechanism is nontrivial as the terms $\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,N}$ and $\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N}$ need to be appropriately

weighted by k and $-(N-k)$ respectively; see also the illustration in the previous paragraph where the simple averaging mechanism $(\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,N})w + (\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N})(1-w)$ fails to possess certain desirable features. By setting $w = 1$ in (5.3), it includes the conventional self-normalizer of Shao (2010b) as a special case. The following theorem provides mathematical properties of this augmented class of self-normalizers, including the identification of its T-symmetric subspace. Its proof is provided in the Appendix.

Theorem 5.2.3. *For the augmented class of self-normalizers introduced in (5.3), when being viewed as a function of $\mathbf{Y}_1, \dots, \mathbf{Y}_N$, it satisfies*

- (i) *the conjugate property $\Lambda_N(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N) = \Lambda_N(1-w; \mathbf{Y}_N, \dots, \mathbf{Y}_1)$;*
- (ii) *the convexity*

$$\Lambda_N(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N) \preceq w\Lambda_N(1; \mathbf{Y}_1, \dots, \mathbf{Y}_N) + (1-w)\Lambda_N(0; \mathbf{Y}_1, \dots, \mathbf{Y}_N),$$

where for matrices $\mathbf{A} \preceq \mathbf{B}$ means $\mathbf{B} - \mathbf{A}$ is positive semidefinite;

- (iii) *in the case of the mean, $\Lambda_N(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N)$ is mathematically equivalent to the mean self-normalizer (5.1) for any $w \in [0, 1]$, which we call the mean-oracle property;*
- (iv) *$\Lambda_N(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N) \equiv \Lambda_N(w; \mathbf{Y}_N, \dots, \mathbf{Y}_1)$ for any \mathbf{T} and $(\mathbf{Y}_1, \dots, \mathbf{Y}_N)$ if and only if $w = 1 - w$.*

Proof. (Theorem 5.2.3) Following the proof of Proposition 5.2.2, let $\mathbf{Y}_i^\bullet = \mathbf{Y}_{N-i+1}$, $i = 1, \dots, N$, then $(\mathbf{Y}_1^\bullet, \dots, \mathbf{Y}_N^\bullet)$ denotes the reversed process $(\mathbf{Y}_N, \dots, \mathbf{Y}_1)$, and it can be seen that

$$\begin{aligned} N^{1/2}\boldsymbol{\lambda}_k(1-w; \mathbf{Y}_1^\bullet, \dots, \mathbf{Y}_N^\bullet) &= k(\hat{\boldsymbol{\theta}}_{N-k+1,N} - \hat{\boldsymbol{\theta}}_{1,N})(1-w) \\ &\quad - (N-k)(\hat{\boldsymbol{\theta}}_{1,N-k} - \hat{\boldsymbol{\theta}}_{1,N})w \\ &= -N^{1/2}\boldsymbol{\lambda}_{N-k}(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N). \end{aligned}$$

Therefore, by changing the index in the summation, we have

$$\begin{aligned}
\mathbf{\Lambda}_N(1-w; \mathbf{Y}_1^\bullet, \dots, \mathbf{Y}_N^\bullet) &= N^{-1} \sum_{k=1}^{N-1} \boldsymbol{\lambda}_k(1-w; \mathbf{Y}_1^\bullet, \dots, \mathbf{Y}_N^\bullet) \boldsymbol{\lambda}_k(1-w; \mathbf{Y}_1^\bullet, \dots, \mathbf{Y}_N^\bullet)^\top \\
&= N^{-1} \sum_{k=1}^{N-1} \boldsymbol{\lambda}_{N-k}(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N) \boldsymbol{\lambda}_{N-k}(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N)^\top \\
&= N^{-1} \sum_{k=1}^{N-1} \boldsymbol{\lambda}_k(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N) \boldsymbol{\lambda}_k(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N)^\top,
\end{aligned}$$

and (i) follows. For (ii), by (5.3) and properties of positive semidefinite matrices it suffices to show that

$$\boldsymbol{\lambda}_k(w) \boldsymbol{\lambda}_k(w)^\top \preceq w \boldsymbol{\lambda}_k(1) \boldsymbol{\lambda}_k(1)^\top + (1-w) \boldsymbol{\lambda}_k(0) \boldsymbol{\lambda}_k(0)^\top$$

holds for each $k \in \{1, \dots, N-1\}$. Let $\boldsymbol{\xi}_{k,N,f} = k(\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,N})$ and $\boldsymbol{\xi}_{k,N,b} = -(N-k)(\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N})$, and denote $\boldsymbol{\Xi}_k = w \boldsymbol{\lambda}_k(1) \boldsymbol{\lambda}_k(1)^\top + (1-w) \boldsymbol{\lambda}_k(0) \boldsymbol{\lambda}_k(0)^\top - \boldsymbol{\lambda}_k(w) \boldsymbol{\lambda}_k(w)^\top$, then by (5.4) we have

$$\boldsymbol{\Xi}_k = Nw(1-w)(\boldsymbol{\xi}_{k,N,f} - \boldsymbol{\xi}_{k,N,b})(\boldsymbol{\xi}_{k,N,f} - \boldsymbol{\xi}_{k,N,b})^\top.$$

Therefore, for any real-valued vector $\boldsymbol{\varphi}$ that has the same dimension as $\hat{\boldsymbol{\theta}}_{1,N}$, we have

$$\boldsymbol{\varphi}^\top \boldsymbol{\Xi}_k \boldsymbol{\varphi} = Nw(1-w) |\boldsymbol{\varphi}^\top (\boldsymbol{\xi}_{k,N,f} - \boldsymbol{\xi}_{k,N,b})|^2 \geq 0,$$

and (ii) follows. For (iii), note that in the mean case we have $\hat{\boldsymbol{\theta}}_{i,j} = (S_j - S_{i-1})/(j - i + 1)$, $1 \leq i \leq j \leq n$, with the convention that $S_0 = 0$. Therefore, one has

$$k(\hat{\boldsymbol{\theta}}_{1,k} - \hat{\boldsymbol{\theta}}_{1,N}) = k \left(\frac{S_k}{k} - \frac{S_n}{n} \right) = S_k - \frac{k}{n} S_n,$$

and

$$(N-k)(\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N}) = (n-k) \left(\frac{S_n - S_k}{n-k} - \frac{S_n}{n} \right) = \frac{k}{n} S_n - S_k.$$

Hence, we have $\boldsymbol{\lambda}_k(w) = n^{-1/2}(S_k - kS_n/n) = n^{-1/2}(S_k - k\bar{X}_n)$ in the mean case, and (iii) follows by the fact that $S_n - n\bar{X}_n = 0$. We shall now provide a proof for property (iv). Note that the self-normalizer (5.3) is T-symmetric when $w = 1 - w$ by the conjugate property (i), and therefore it suffices to prove that for any $w \neq 1 - w$ there

exists a mapping \mathbf{T} and $(\mathbf{Y}_1, \dots, \mathbf{Y}_N) \in \mathbb{R}^{\ell \times N}$ such that $\mathbf{\Lambda}_N(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N) \neq \mathbf{\Lambda}_N(w; \mathbf{Y}_N, \dots, \mathbf{Y}_1)$. For this, we construct the example where \mathbf{T} represents the marginal median and $\mathbf{Y}_i = \mathbf{I}(i > pN)$, $i = 1, \dots, N$, for some $p \in (1/2 + 1/N, 1)$. Then it can be seen that $\hat{\boldsymbol{\theta}}_{1,k} = 0$, $k = 1, \dots, N$, while

$$\hat{\boldsymbol{\theta}}_{k+1,N} = \begin{cases} 1, & \text{if } k > 2\lfloor pN \rfloor - N + 1; \\ 0.5, & \text{if } k = 2\lfloor pN \rfloor - N + 1; \\ 0, & \text{if } k < 2\lfloor pN \rfloor - N + 1. \end{cases}$$

Hence, we have

$$\mathbf{\Lambda}_N(w; \mathbf{Y}_1, \dots, \mathbf{Y}_N) = (1-w)^2 N^{-2} \sum_{k=1}^{N-1} (N-k)^2 (\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N}) (\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N})^\top,$$

and by the conjugate property (i),

$$\mathbf{\Lambda}_N(w; \mathbf{Y}_1^\bullet, \dots, \mathbf{Y}_N^\bullet) = w^2 N^{-2} \sum_{k=1}^{N-1} (N-k)^2 (\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N}) (\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N})^\top.$$

Since $\sum_{k=1}^{N-1} (N-k)^2 (\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N}) (\hat{\boldsymbol{\theta}}_{k+1,N} - \hat{\boldsymbol{\theta}}_{1,N})^\top$ is nonzero for all $N > 2$, the result follows. \square

By Theorem 5.2.3, the T-symmetric subset of the augmented class of self-normalizers (5.3) corresponds to the case where $w = 1 - w$, and therefore we propose to use $\mathbf{\Lambda}_N = \mathbf{\Lambda}_N(1/2)$ as a T-symmetric generalization of the conventional self-normalizer of Shao (2010b). Compared with the simple averaging scheme of Shao (2015), Theorem 5.2.3 (ii) provides a theoretical guarantee that the proposed T-symmetric self-normalizer can lead to confidence regions with smaller or at least the same volume. We will support this finding with numerical experiments which will be summarized in section (5.3). We shall in the following provide the associated inference procedure and its asymptotic property. In particular, for any significance level $\alpha \in (0, 1)$, if we can obtain a threshold ν_α such that the self-normalized quantity satisfies

$$\lim_{N \rightarrow \infty} \text{pr}\{N(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta})^\top \mathbf{\Lambda}_N^{-1}(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta}) > \nu_\alpha\} = \alpha,$$

then an asymptotic $(1 - \alpha)$ -th confidence region of $\boldsymbol{\theta}$ can be constructed as

$$\{\boldsymbol{\theta} : N(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta})^\top \boldsymbol{\Lambda}_N^{-1}(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta}) \leq \nu_\alpha\}, \quad (5.5)$$

and hypothesis tests on $\boldsymbol{\theta}$ can be performed similarly. In order to obtain the threshold q_α , one needs a distributional theory for the self-normalized quantity $N(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta})^\top \boldsymbol{\Lambda}_N^{-1}(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta})$, for which we need the following assumption.

(IP) There exists a positive definite matrix $\boldsymbol{\Sigma}$ such that

$$N^{-1/2}([\!|Nt|] \vee 1 - [\!|Ns|] \vee 1 + 1)\{\mathbf{T}(\hat{\mathbf{F}}_{[\!|Ns|] \vee 1, [\!|Nt|] \vee 1}^\ell) - \mathbf{T}(\mathbf{F}^\ell)\} \rightsquigarrow \boldsymbol{\Sigma}\{\mathbf{B}(t) - \mathbf{B}(s)\},$$

where \rightsquigarrow denotes the weak convergence on $(\{0\} \times [0, 1]) \cup ([0, 1] \times \{1\})$ in the sense of Hoffmann-Jørgensen (van der Vaart and Wellner, 1996), and $\mathbf{B}(\cdot)$ is a standard multivariate Brownian motion having the same dimension as $\boldsymbol{\theta}$.

Assumption (IP) provides a counterpart of the invariance principle (2.2) in the current general setting, and can be verified by using the functional delta method of Volgushev and Shao (2014); see also the influence function approach of Shao (2010b). The following theorem provides the asymptotic distribution of the self-normalized quantity, based on which one can obtain the threshold ν_α for inference of $\boldsymbol{\theta}$.

Theorem 5.2.4. *Assume condition (IP), and let*

$$\mathbf{W}(\mathbf{B}) = \int_0^1 \{\mathbf{B}(t) - t\mathbf{B}(1)\}\{\mathbf{B}(t) - t\mathbf{B}(1)\}^\top dt.$$

Then the self-normalized quantity satisfies

$$N(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta})^\top \boldsymbol{\Lambda}_N^{-1}(\hat{\boldsymbol{\theta}}_{1,N} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathbf{B}(1)^\top \mathbf{W}(\mathbf{B})^{-1} \mathbf{B}(1). \quad (5.6)$$

Proof. (Theorem 5.2.4) Recall that $\hat{\boldsymbol{\theta}}_{i,j} = \mathbf{T}(\hat{\mathbf{F}}_{i,j}^\ell)$ for $1 \leq i \leq j \leq N$, and we in addition define $\hat{\boldsymbol{\theta}}_{k,j} = \mathbf{T}(\hat{\mathbf{F}}_{j,j}^\ell)$ if $k > j$. Let $\boldsymbol{\Theta}_N(s, t) = N^{-1/2}([\!|Nt|] \vee 1 -$

$\lfloor Ns \rfloor)(\hat{\boldsymbol{\theta}}_{\lfloor Ns \rfloor + 1, \lfloor Nt \rfloor \vee 1} - \boldsymbol{\theta})$, then for any $t \geq N^{-1}$ we have

$$\begin{aligned} N^{-1/2} \lfloor Nt \rfloor (\hat{\boldsymbol{\theta}}_{1, \lfloor Nt \rfloor} - \hat{\boldsymbol{\theta}}_{1, N}) &= N^{-1/2} (\lfloor Nt \rfloor \vee 1) (\hat{\boldsymbol{\theta}}_{1, \lfloor Nt \rfloor \vee 1} - \hat{\boldsymbol{\theta}}_{1, N}) \\ &= \boldsymbol{\Theta}_N(0, t) - t \boldsymbol{\Theta}_N(0, 1) + N^{-1/2} (Nt - \lfloor Nt \rfloor) (\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta}), \end{aligned}$$

and similarly

$$(N - \lfloor Nt \rfloor) (\hat{\boldsymbol{\theta}}_{\lfloor Nt \rfloor + 1, N} - \hat{\boldsymbol{\theta}}_{1, N}) = N^{1/2} \{ \boldsymbol{\Theta}_N(t, 1) - (1-t) \boldsymbol{\Theta}_N(0, 1) \} - (Nt - \lfloor Nt \rfloor) (\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta}).$$

Therefore, $\boldsymbol{\lambda}_{\lfloor Nt \rfloor}(w)$ can be well approximated by

$$\boldsymbol{\lambda}(\boldsymbol{\Theta}, t, w) = w \{ \boldsymbol{\Theta}_N(0, t) - t \boldsymbol{\Theta}_N(0, 1) \} - (1-w) \{ \boldsymbol{\Theta}_N(t, 1) - (1-t) \boldsymbol{\Theta}_N(0, 1) \}$$

in the uniform sense that

$$\sup_{w \in [0, 1]} \sup_{t \in [N^{-1}, 1]} |\boldsymbol{\lambda}_{\lfloor Nt \rfloor}(w) - \boldsymbol{\lambda}(\boldsymbol{\Theta}, t, w)| \leq N^{-1/2} (Nt - \lfloor Nt \rfloor) |\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta}| = O_p(N^{-1}).$$

We shall now consider the region where $t \in [0, N^{-1})$, in which case $\lfloor Nt \rfloor = 0$ and thus by the definition in (5.4) we have $\boldsymbol{\lambda}_{\lfloor Nt \rfloor}(w) = 0$. On the other hand, for any $t \in [0, N^{-1})$, we have $\boldsymbol{\Theta}_N(0, t) = N^{-1/2} (\hat{\boldsymbol{\theta}}_{1, 1} - \boldsymbol{\theta})$ and $\boldsymbol{\Theta}_N(t, 1) = N^{1/2} (\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta}) = \boldsymbol{\Theta}_N(0, 1)$. As a result, $\boldsymbol{\lambda}(\boldsymbol{\Theta}, t, w) = w N^{-1/2} (\hat{\boldsymbol{\theta}}_{1, 1} - \boldsymbol{\theta}) - t N^{1/2} (\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta})$, and thus

$$\sup_{w \in [0, 1]} \sup_{t \in [0, N^{-1})} |\boldsymbol{\lambda}_{\lfloor Nt \rfloor}(w) - \boldsymbol{\lambda}(\boldsymbol{\Theta}, t, w)| \leq N^{-1/2} (|\hat{\boldsymbol{\theta}}_{1, 1} - \boldsymbol{\theta}| + |\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta}|) = O_p(N^{-1/2}).$$

Note that $\boldsymbol{\Lambda}_N(w) = \int_0^1 \boldsymbol{\lambda}_{\lfloor Nt \rfloor}(w) \boldsymbol{\lambda}_{\lfloor Nt \rfloor}(w)^\top dt$, it can be shown that

$$N (\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta})^\top \boldsymbol{\Lambda}_N(w)^{-1} (\hat{\boldsymbol{\theta}}_{1, N} - \boldsymbol{\theta}) = \boldsymbol{\Theta}_N(0, 1)^\top \boldsymbol{\Lambda}(\boldsymbol{\Theta}, w)^{-1} \boldsymbol{\Theta}_N(0, 1) + o_p(1),$$

where

$$\boldsymbol{\Lambda}(\boldsymbol{\Theta}, w) = \int_0^1 \boldsymbol{\lambda}(\boldsymbol{\Theta}, t, w) \boldsymbol{\lambda}(\boldsymbol{\Theta}, t, w)^\top dt.$$

Since $\sup_{t \in [0, 1]} |t - N^{-1} \lfloor Nt \rfloor| \leq N^{-1} \rightarrow 0$, we have $\boldsymbol{\Theta}_N(s, t) \rightsquigarrow \boldsymbol{\Sigma} \{ \boldsymbol{B}(t) - \boldsymbol{B}(s) \}$ on $(s, t) \in (\{0\} \times [0, 1]) \cup ([0, 1] \times \{1\})$ by condition (IP), and the result follows by the continuous mapping theorem. \square

By (5.6), one can set the threshold ν_α as the $(1 - \alpha)$ -th quantile of

$$\mathbf{B}(1)^\top \mathbf{W}(\mathbf{B})^{-1} \mathbf{B}(1),$$

which will make the confidence interval constructed in (5.5) possess the correct asymptotic size. Note that the asymptotic distribution in (5.6) is pivotal with respect to the transform $\mathbf{T}(\cdot)$, meaning that the same inference procedure can be applied to both the mean and quantities beyond the mean. This is considered as a convenient feature, as self-normalized statistics typically have non-standard asymptotic distributions which often require large scale Monte Carlo simulations to evaluate; see for example Lobato (2001), Shao (2010b), Shao (2015) and Zhang and Lavitas (2017) and references therein.

5.3 Numerical Experiments

5.3.1 Monte Carlo Simulations

We shall here conduct Monte Carlo simulations to examine the finite sample performance of the proposed T-symmetric self-normalized inference procedure, denoted by TSSN, and compare it with the conventional method of Shao (2010b) which is denoted by S10. Following Shao (2015), we consider constructing confidence intervals for the mean and median of the autoregressive model

$$X_i = \rho X_{i-1} + \epsilon_i, \quad i = 1, \dots, n, \tag{5.7}$$

where the autoregressive coefficient $\rho \in \{\pm 0.3, \pm 0.6, \pm 0.8\}$ is chosen to represent different strengths of dependence and (ϵ_i) is a sequence of independent standard normal random variables. The results are summarized in Tables 5.1–5.4, with Tables 5.1 and 5.2 containing results for the mean and Tables 5.3 and 5.4 containing results for

the median. We also include a comparison with the two symmetrizing approaches considered by Shao (2015). The first approach of Shao (2015), denoted by $S15_{W_2}$ to be consistent with their notation, is to involve recursive estimators from all possible blocks in the formation of the self-normalizer, and as a result the associated computation is substantially more than that for the remaining methods; see also the discussion in Shao (2015). Besides the computational disadvantage, in the case of the mean where no symmetrization is needed, it does not unify with the conventional self-normalizer of Shao (2010b), and it can be seen from Table 5.1 that its empirical coverage probabilities are generally worse than that of the remaining methods in the mean case; see also numerical results reported by Shao (2015) for a similar observation. Motivated by the suggestion of Professor Michael Stein at The University of Chicago, Shao (2015) then considered a second approach, denoted by $S15_{W_3}$ to be consistent with their notation, which takes an average of forward and backward self-normalizers. It can be seen from Tables 5.3 and 5.4 that such an approach generally leads to confidence intervals with better coverage probabilities but at the cost of having wider lengths (less test powers) when compared with the conventional S10 self-normalizer; see also the discussion in Shao (2015) for a similar finding. The proposed TSSN method provides an alternative approach to symmetrizing the self-normalizer for T-symmetric inference of time series. Unlike the $S15_{W_2}$ approach, its computation is in the same order as that of the conventional S10 self-normalizer, and in the mean case where the S10 is already T-symmetric it automatically reduces to the S10 to produce a unified inference procedure. In contrast to the $S15_{W_3}$ approach, it can be seen from Table 5.4 that the proposed TSSN method typically leads to confidence intervals with narrower lengths when compared with the conventional S10 self-normalizer. We shall here in addition mention the performance in challenging situations, namely when $n = 50$ and $\rho = \pm 0.8$. In particular, when $n = 50$ and $\rho = 0.8$ where there is

a strong positive dependence, then the $S15_{W_2}$ approach seems to yield the narrowest confidence intervals among the considered methods though at the cost of having the largest distortion in empirical coverage probabilities. On the other hand when $n = 50$ and $\rho = -0.8$ where there is a strong negative dependence, then it can be seen from Tables 5.3 and 5.4 that the $S15_{W_2}$ approach may lead to confidence intervals with larger size distortion and wider lengths (at the 90% confidence level) when compared with the conventional S10 approach. The proposed TSSN method seems to be the only one that generally improves over the conventional S10 approach in terms of both the size and the power, while maintaining the same order of computation.

Please note, that not all the methods tend to produce empirical coverage probabilities that are smaller than their nominal levels, and the direction of size distortion can depend on the underlying dependence structure. In particular, we can see from Table 3 that, when there is a negative dependence as in cases where $\rho \in \{-0.3, -0.6, -0.8\}$, methods $S15_{W_2}$ and $S15_{W_3}$ can produce empirical coverage probabilities that are higher than their nominal levels, and larger size distortions can be observed when the dependence gets stronger. For example, when $n = 50$ and $\rho = -0.8$ empirical coverage probabilities at the 90% nominal level for $S15_{W_2}$ and $S15_{W_3}$ are 95.71% and 94.2% respectively. In this case, having narrower confidence intervals can in fact help improve not only the power but also the size performance. For a comparison, the proposed TSSN method in this case has an empirical coverage probability of 91.3% for the 90% nominal, thus having less size distortion when compared with $S15_{W_2}$ (95.7%) and $S15_{W_3}$ (94.2%). In addition, it can be seen from Table 4 that the proposed method in this case can also lead to narrower confidence intervals with an average length of 0.740 when compared with $S15_{W_2}$ (0.803) and $S15_{W_3}$ (0.851). Therefore, we believe that the proposed method possesses certain merits when compared with $S15_{W_2}$ and $S15_{W_3}$ of Shao (2015). We would also clarify that $S15_{W_2}$ does not seem to exhibit superior

n	$1 - \alpha$	ρ	<i>Method</i>	
			S10/S15 $_{W_3}$ /TSSN	S15 $_{W_2}$
50	90%	0.3	88.6 (0.14)	88.7 (0.14)
		0.6	86.0 (0.16)	85.5 (0.16)
		0.8	81.1 (0.18)	78.6 (0.18)
		-0.3	91.1 (0.13)	91.9 (0.12)
		-0.6	93.0 (0.11)	94.2 (0.10)
		-0.8	95.7 (0.09)	96.8 (0.08)
	95%	0.3	93.9 (0.11)	94.0 (0.11)
		0.6	92.0 (0.12)	91.7 (0.12)
		0.8	87.7 (0.15)	85.9 (0.16)
		-0.3	95.8 (0.09)	96.3 (0.08)
		-0.6	97.1 (0.08)	97.6 (0.07)
		-0.8	98.4 (0.06)	99.0 (0.04)
200	90%	0.3	89.4 (0.14)	89.5 (0.14)
		0.6	89.2 (0.14)	88.9 (0.14)
		0.8	87.7 (0.15)	86.7 (0.15)
		-0.3	90.3 (0.13)	90.6 (0.13)
		-0.6	91.0 (0.13)	91.3 (0.13)
		-0.8	92.2 (0.12)	92.9 (0.12)
	95%	0.3	94.6 (0.10)	94.6 (0.10)
		0.6	94.3 (0.10)	94.1 (0.11)
		0.8	93.1 (0.11)	92.5 (0.12)
		-0.3	95.4 (0.09)	95.5 (0.09)
		-0.6	95.7 (0.09)	96.0 (0.09)
		-0.8	96.5 (0.08)	96.9 (0.08)

Table 5.1: Empirical coverage probabilities (in percentage) for confidence intervals constructed using different self-normalizers for the mean of the process (5.7) under different dependence strengths. The associated standard errors (also in percentage) are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

size properties over the other methods. In particular, it can be seen from Table 1 that empirical coverage probabilities of S15 $_{W_2}$ are generally worse than that of the remaining methods in the mean case; see also numerical results reported by Shao (2015) for

n	$1 - \alpha$	ρ	<i>Method</i>		
			S10/S15 $_{W_3}$ /TSSN	S15 $_{W_2}$	
50	90%	0.3	0.781 (0.0015)	0.721 (0.0010)	
		0.6	1.275 (0.0026)	1.151 (0.0018)	
		0.8	2.204 (0.0049)	1.906 (0.0034)	
		-0.3	0.455 (0.0008)	0.429 (0.0005)	
		-0.6	0.393 (0.0006)	0.378 (0.0004)	
		-0.8	0.389 (0.0006)	0.383 (0.0004)	
		95%	0.3	0.989 (0.0019)	0.895 (0.0013)
	0.6	1.614 (0.0033)	1.428 (0.0023)		
	0.8	2.791 (0.0063)	2.365 (0.0043)		
	-0.3	0.576 (0.0010)	0.532 (0.0007)		
	-0.6	0.498 (0.0008)	0.469 (0.0005)		
	-0.8	0.492 (0.0007)	0.475 (0.0005)		
	200	90%	0.3	0.402 (0.0007)	0.369 (0.0005)
			0.6	0.692 (0.0013)	0.631 (0.0009)
0.8			1.327 (0.0026)	1.198 (0.0018)	
-0.3			0.221 (0.0004)	0.204 (0.0003)	
-0.6			0.183 (0.0003)	0.170 (0.0002)	
-0.8			0.168 (0.0003)	0.158 (0.0002)	
95%			0.3	0.509 (0.0009)	0.458 (0.0006)
0.6		0.876 (0.0017)	0.783 (0.0011)		
0.8		1.681 (0.0033)	1.486 (0.0022)		
-0.3		0.280 (0.0005)	0.254 (0.0003)		
-0.6		0.232 (0.0004)	0.211 (0.0003)		
-0.8		0.213 (0.0004)	0.196 (0.0002)		

Table 5.2: Average lengths of confidence intervals constructed using different self-normalizers for the mean of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

a similar observation and the discussions therein. In the median case, it can be seen from Table 3 that S15 $_{W_2}$ tend to have the worst size performance in the presence of relatively strong dependence as in cases where $\rho \in \{-0.8, 0.8\}$ and the direction of

n	$1 - \alpha$	ρ	<i>Method</i>				
			S10	S15 $_{W_2}$	S15 $_{W_3}$	TSSN	
50	90%	0.3	86.4 (0.15)	88.9 (0.14)	89.7 (0.14)	87.5 (0.15)	
		0.6	84.9 (0.16)	86.0 (0.16)	87.7 (0.15)	85.7 (0.16)	
		0.8	80.1 (0.18)	79.3 (0.18)	83.1 (0.17)	80.9 (0.18)	
		-0.3	87.3 (0.15)	91.3 (0.13)	91.6 (0.12)	89.0 (0.14)	
		-0.6	87.5 (0.15)	92.8 (0.12)	92.5 (0.12)	89.8 (0.14)	
		-0.8	88.1 (0.14)	95.7 (0.09)	94.2 (0.10)	91.3 (0.13)	
		95%	0.3	91.7 (0.12)	94.0 (0.11)	94.6 (0.10)	93.0 (0.11)
	0.6	90.5 (0.13)	91.8 (0.12)	93.0 (0.11)	91.5 (0.12)		
	0.8	86.6 (0.15)	86.2 (0.15)	89.4 (0.14)	87.8 (0.15)		
	-0.3	92.0 (0.12)	95.6 (0.09)	95.7 (0.09)	93.9 (0.11)		
	-0.6	92.1 (0.12)	96.5 (0.08)	96.2 (0.09)	94.3 (0.10)		
	-0.8	92.4 (0.12)	98.3 (0.06)	97.3 (0.07)	95.5 (0.09)		
	200	90%	0.3	88.5 (0.14)	89.9 (0.13)	90.4 (0.13)	88.9 (0.14)
			0.6	88.2 (0.14)	89.1 (0.14)	89.9 (0.13)	88.6 (0.14)
0.8			86.8 (0.15)	87.0 (0.15)	88.2 (0.14)	87.1 (0.15)	
-0.3			88.2 (0.14)	90.2 (0.13)	90.9 (0.13)	88.8 (0.14)	
-0.6			88.0 (0.15)	91.0 (0.13)	91.3 (0.13)	89.1 (0.14)	
-0.8			88.0 (0.15)	92.1 (0.12)	92.1 (0.12)	89.5 (0.14)	
95%			0.3	93.4 (0.11)	94.7 (0.10)	95.0 (0.10)	94.0 (0.11)
0.6		93.3 (0.11)	94.1 (0.11)	94.7 (0.10)	93.8 (0.11)		
0.8		92.4 (0.12)	92.7 (0.12)	93.6 (0.11)	92.8 (0.12)		
-0.3		93.2 (0.11)	95.1 (0.10)	95.5 (0.09)	94.1 (0.11)		
-0.6		92.7 (0.12)	95.3 (0.09)	95.6 (0.09)	94.1 (0.11)		
-0.8		92.8 (0.12)	96.1 (0.09)	96.2 (0.09)	94.5 (0.10)		

Table 5.3: Empirical coverage probabilities (in percentage) for confidence intervals constructed using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors (also in percentage) are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

size distortion depends on the dependence structure. This can be a consequence of the complicated form of S15 $_{W_2}$. Note that S15 $_{W_2}$ involves recursive estimators from all possible blocks, and as a result the associated computation is substantially more

n	$1 - \alpha$	ρ	<i>Method</i>			
			S10	S15 $_{W_2}$	S15 $_{W_3}$	TSSN
50	90%	0.3	0.913 (0.0021)	0.849 (0.0014)	0.944 (0.0018)	0.878 (0.0018)
		0.6	1.411 (0.0034)	1.282 (0.0022)	1.452 (0.0030)	1.365 (0.0029)
		0.8	2.390 (0.0062)	2.081 (0.0042)	2.463 (0.0056)	2.324 (0.0054)
		-0.3	0.645 (0.0016)	0.611 (0.0010)	0.674 (0.0013)	0.613 (0.0012)
		-0.6	0.655 (0.0017)	0.633 (0.0010)	0.690 (0.0014)	0.618 (0.0012)
		-0.8	0.795 (0.0023)	0.803 (0.0015)	0.851 (0.0018)	0.740 (0.0016)
	95%	0.3	1.156 (0.0027)	1.053 (0.0017)	1.195 (0.0023)	1.112 (0.0022)
		0.6	1.787 (0.0043)	1.590 (0.0028)	1.839 (0.0038)	1.729 (0.0036)
		0.8	3.026 (0.0079)	2.581 (0.0052)	3.119 (0.0071)	2.943 (0.0068)
		-0.3	0.817 (0.0020)	0.757 (0.0012)	0.854 (0.0017)	0.777 (0.0015)
		-0.6	0.829 (0.0021)	0.785 (0.0013)	0.874 (0.0017)	0.783 (0.0016)
		-0.8	1.007 (0.0029)	0.996 (0.0019)	1.077 (0.0023)	0.937 (0.0020)
200	90%	0.3	0.469 (0.0010)	0.433 (0.0006)	0.479 (0.0009)	0.458 (0.0009)
		0.6	0.761 (0.0016)	0.696 (0.0010)	0.773 (0.0014)	0.745 (0.0014)
		0.8	1.417 (0.0030)	1.282 (0.0020)	1.439 (0.0028)	1.395 (0.0028)
		-0.3	0.321 (0.0007)	0.298 (0.0004)	0.331 (0.0006)	0.310 (0.0006)
		-0.6	0.319 (0.0007)	0.298 (0.0005)	0.330 (0.0006)	0.306 (0.0006)
		-0.8	0.357 (0.0009)	0.339 (0.0005)	0.373 (0.0007)	0.340 (0.0007)
	95%	0.3	0.594 (0.0012)	0.537 (0.0008)	0.607 (0.0011)	0.579 (0.0011)
		0.6	0.964 (0.0020)	0.863 (0.0013)	0.979 (0.0018)	0.944 (0.0018)
		0.8	1.795 (0.0038)	1.590 (0.0025)	1.823 (0.0036)	1.767 (0.0035)
		-0.3	0.407 (0.0009)	0.369 (0.0006)	0.419 (0.0008)	0.393 (0.0008)
		-0.6	0.404 (0.0009)	0.369 (0.0006)	0.418 (0.0008)	0.388 (0.0008)
		-0.8	0.452 (0.0011)	0.420 (0.0007)	0.472 (0.0009)	0.430 (0.0008)

Table 5.4: Average lengths of confidence intervals constructed using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

than that for the remaining methods; see also the discussion in Shao (2015). Besides the computational disadvantage, in the case of the mean where no symmetrization is needed, S15 $_{W_2}$ does not unify with the conventional self-normalizer. As a result,

$S15_{W_2}$ does not seem to be a superior choice over the remaining methods. We would also note that the proposed method is computational advantage over $S15_{W_2}$.

We will also provide comparison of power properties between $S15_{W_3}$ and the proposed TSSN test. It can be seen from Tables 5.6 - 5.11 that the TSSN has an advantage over $S15_{W_3}$ in terms of the raw power. We are focusing on raw power analysis only following the suggestions and arguments made in Horowitz and Savin (2000), Hansen (2005) and A. Canepa (2007) that the raw power is more related to empirical research and should be considered as a fair comparison. Based on the comparison of the raw power as in Tables 5.6 - 5.11, the proposed TSSN method is believed to have better power performance than $S15_{W_3}$.

5.3.2 A Real Data Example

We shall here use a real data example to further illustrate the proposed method. The data that we consider contains concentration measurements of atmospheric pollutants collected by an air quality chemical multisensor device in an Italian city, and we refer to De Vito et al. (2008), Mead and et all. (2013) and Marco and Gutierrez-Galvez (2012) for more details. We shall here focus on the daily average concentration of non-metanic hydrocarbons (NMHC) for the period from March 10, 2004 to July 23, 2004. The pollutant is believed to be related to respiratory diseases, and in order to assess if there is an increase or decrease pattern in its concentration, we consider its relative daily changes, for which a time series plot is given in Figure 5.1. We shall here apply different self-normalization methods for constructing confidence intervals for the mean and median of the relative daily change in the NMHC concentration. For this, we consider the conventional S10 approach of Shao (2010b), the two symmetrization methods $S15_{W_2}$ and $S15_{W_3}$ considered by Shao (2015), and the proposed

n	rho	d	S15 $_{W_3}$	TSSN
50	0.3	0	10.1 (0.43)	12.4 (0.47)
		0.2	19.3 (0.56)	22.6 (0.59)
		0.4	43.0 (0.7)	47.2 (0.71)
		0.6	68.1 (0.66)	71.6 (0.64)
		0.8	86.4 (0.49)	88.5 (0.45)
		1	94.9 (0.31)	95.9 (0.28)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)
		0.6	0	11.9 (0.46)
	0.2		14.9 (0.50)	17.7 (0.54)
	0.4		26.0 (0.62)	29.0 (0.64)
	0.6		41.4 (0.70)	45.1 (0.70)
	0.8		58.3 (0.70)	61.6 (0.69)
	1		72.3 (0.63)	75.5 (0.61)
	2		99.1 (0.14)	99.4 (0.11)
	3		100.0 (0.00)	100.0 (0.00)
	0.8		0	16.8 (0.53)
		0.2	17.7 (0.54)	19.7 (0.56)
		0.4	20.8 (0.57)	23.1 (0.60)
		0.6	26.7 (0.63)	29.2 (0.64)
		0.8	34.2 (0.67)	37.0 (0.68)
		1	42.5 (0.70)	45.1 (0.70)
		2	80.3 (0.56)	82.4 (0.54)
		3	95.8 (0.28)	96.6 (0.26)

Table 5.5: Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

TSSN method. For each method, we apply it to the daily relative change series from both directions to check if it yields a T-symmetric confidence interval or if the direction of application can actually affect the conclusion. The results are summarized in Table 5.13, from which we can observe the following. In the mean case, all the methods considered are T-symmetric, as confidence intervals constructed from forward

n	rho	d	S15 _{W₃}	TSSN
50	-0.3	0	8.0 (0.38)	10.4 (0.43)
		0.2	27.0 (0.63)	31.8 (0.66)
		0.4	64.8 (0.68)	69.6 (0.65)
		0.6	89.3 (0.44)	91.9 (0.39)
		0.8	97.7 (0.21)	98.6 (0.16)
		1	99.5 (0.10)	99.7 (0.07)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)
		-0.6	0	7.4 (0.37)
	0.2		26.1 (0.62)	32.3 (0.66)
	0.4		63.6 (0.68)	69.8 (0.65)
	0.6		88.5 (0.45)	91.6 (0.39)
	0.8		97.4 (0.22)	98.5 (0.17)
	1		99.5 (0.10)	99.8 (0.07)
	2		100.0 (0.00)	100.0 (0.00)
	3		100.0 (0.00)	100.0 (0.00)
	-0.8		0	5.5 (0.32)
		0.2	18.5 (0.55)	24.9 (0.61)
		0.4	51.3 (0.71)	59.7 (0.69)
		0.6	77.7 (0.59)	84.4 (0.51)
		0.8	91.8 (0.39)	95.4 (0.30)
		1	97.2 (0.23)	98.6 (0.16)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)

Table 5.6: Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

and backward applications match with each other. This is in line with Proposition 5.2.2, which states that the conventional S10 self-normalizer is already T-symmetric in the mean case. Note that all the methods, except for S15_{W₂}, find statistical evidence at the 5% significance level in suggesting a positive mean for the relative daily change. The S15_{W₂} method in this case produces a confidence interval that is wider

n	rho	d	S15 _{W₃}	TSSN
50	0.3	0	5.7 (0.33)	7.0 (0.36)
		0.2	11.7 (0.45)	14.1 (0.49)
		0.4	29.7 (0.65)	33.7 (0.67)
		0.6	54.1 (0.70)	58.6 (0.70)
		0.8	74.3 (0.62)	78.2 (0.58)
		1	87.9 (0.46)	90.4 (0.42)
		2	99.9 (0.04)	99.9 (0.03)
		3	100.0 (0.0)	100.0 (0.00)
		0.6	0	6.7 (0.35)
	0.2		8.4 (0.39)	10.4 (0.43)
	0.4		16.8 (0.53)	19.4 (0.56)
	0.6		29.0 (0.64)	32.7 (0.66)
	0.8		43.5 (0.70)	47.9 (0.71)
	1		58.5 (0.70)	62.4 (0.69)
	2		96.4 (0.26)	97.0 (0.24)
	3		99.9 (0.05)	99.9 (0.03)
	0.8		0	10.3 (0.43)
		0.2	11.2 (0.45)	12.7 (0.47)
		0.4	14.0 (0.49)	15.6 (0.51)
		0.6	17.9 (0.54)	20.1 (0.57)
		0.8	24.1 (0.60)	26.5 (0.62)
		1	31.1 (0.65)	34.5 (0.67)
		2	68.9 (0.65)	71.7 (0.64)
		3	90.6 (0.41)	92.2 (0.38)

Table 5.7: Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

than the remaining methods. In the median case, it can be seen from Table 5.13 that the conventional S10 approach yields different confidence intervals and different conclusions about the null hypothesis of a zero median, depending on the direction of application. Note that the median is considered to be irrelevant to the direction, and therefore it seems desirable to consider its T-symmetric generalizations as in the

n	rho	d	S15 _{W₃}	TSSN
50	-0.3	0	3.8 (0.27)	5.6 (0.33)
		0.2	16.5 (0.52)	20.7 (0.57)
		0.4	50.0 (0.71)	56.1 (0.70)
		0.6	78.8 (0.58)	83.1 (0.53)
		0.8	93.1 (0.36)	95.4 (0.30)
		1	98.1 (0.20)	98.9 (0.15)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)
		-0.6	0	3.8 (0.27)
	0.2		15.8 (0.52)	20.6 (0.57)
	0.4		48.4 (0.71)	55.5 (0.70)
	0.6		76.6 (0.60)	82.6 (0.54)
	0.8		92.3 (0.38)	95.1 (0.31)
	1		98.0 (0.20)	98.8 (0.15)
	2		100.0 (0.00)	100.0 (0.00)
	3		100.0 (0.00)	100.0 (0.00)
	-0.8		0	2.5 (0.22)
		0.2	10.2 (0.43)	14.7 (0.50)
		0.4	35.3 (0.68)	44.5 (0.70)
		0.6	64.4 (0.68)	72.7 (0.63)
		0.8	82.8 (0.53)	88.6 (0.45)
		1	92.9 (0.36)	96.1 (0.28)
		2	99.9 (0.05)	100.0 (0.02)
		3	100.0 (0.00)	100.0 (0.00)

Table 5.8: Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

current paper. Compared with the two symmetrization approaches of Shao (2015), the proposed TSSN method seems to produce a narrower confidence interval for the median from Table 5.13.

n	rho	d	S15 _{W₃}	TSSN
200	0.3	0	9.2 (0.41)	10.6 (0.44)
		0.2	42.1 (0.70)	45.0 (0.70)
		0.4	85.4 (0.50)	86.9 (0.48)
		0.6	98.7 (0.16)	99.0 (0.14)
		0.8	100.0 (0.02)	100.0 (0.02)
		1	100.0 (0.00)	100.0 (0.00)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)
		0.6	0	10.5 (0.43)
	0.2		23.2 (0.60)	25.1 (0.61)
	0.4		53.3 (0.71)	55.3 (0.70)
	0.6		81.1 (0.55)	82.5 (0.54)
	0.8		94.4 (0.33)	95.0 (0.31)
	1		98.6 (0.17)	98.8 (0.15)
	2		100.0 (0.00)	100.0 (0.00)
	3		100.0 (0.00)	100.0 (0.00)
	0.8		0	11.7 (0.45)
		0.2	15.7 (0.51)	16.8 (0.53)
		0.4	26.8 (0.63)	28.2 (0.64)
		0.6	42.4 (0.70)	44.3 (0.70)
		0.8	59.7 (0.69)	61.3 (0.69)
		1	74.0 (0.62)	75.6 (0.61)
		2	99.0 (0.14)	99.2 (0.12)
		3	100.0 (0.00)	100.0 (0.00)

Table 5.9: Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

5.4 Conclusion

Although self-normalization has been celebrated for its ability to avoid direct estimation of the nuisance asymptotic variance, its commonly adopted form due to the seminal work of Shao (2010b) is not T-symmetric when being applied to quantities

n	rho	d	S15 _{W₃}	TSSN
200	-0.3	0	9.4 (0.41)	11.5 (0.45)
		0.2	65.7 (0.67)	69.2 (0.65)
		0.4	97.8 (0.21)	98.3 (0.18)
		0.6	99.9 (0.04)	100.0 (0.03)
		0.8	100.0 (0.00)	100.0 (0.00)
		1	100.0 (0.00)	100.0 (0.00)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)
		-0.6	0	8.5 (0.39)
	0.2		64.4 (0.68)	68.9 (0.65)
	0.4		98.0 (0.20)	98.5 (0.17)
	0.6		100.0 (0.02)	100.0 (0.02)
	0.8		100.0 (0.00)	100.0 (0.00)
	1		100.0 (0.00)	100.0 (0.00)
	2		100.0 (0.00)	100.0 (0.00)
	3		100.0 (0.00)	100.0 (0.00)
	-0.8		0	7.7 (0.38)
		0.2	58 (0.70)	63.8 (0.68)
		0.4	95.9 (0.28)	97.3 (0.23)
		0.6	99.9 (0.05)	99.9 (0.04)
		0.8	100.0 (0.00)	100.0 (0.00)
		1	100.0 (0.00)	100.0 (0.00)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)

Table 5.10: Average raw power at the 90% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

other than the mean. The asymmetric issue can cause unnecessary confusions in practice, as different conclusions can be possibly obtained depending on the direction of application; see the discussion in Shao (2015) and also the real data illustration in Section 5.3.2. To address the issue, Shao (2015) considered two different approaches in symmetrizing the self-normalizer. The first approach is to involve parameter esti-

n	rho	d	S15 $_{W_3}$	TSSN
200	0.3	0	4.8 (0.30)	5.9 (0.33)
		0.2	28.0 (0.63)	30.8 (0.65)
		0.4	73.7 (0.62)	75.8 (0.61)
		0.6	94.8 (0.32)	95.8 (0.28)
		0.8	99.5 (0.10)	99.6 (0.09)
		1	100.0 (0.02)	100.0 (0.02)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)
		0.6	0	5.6 (0.32)
	0.2		13.9 (0.49)	15.8 (0.52)
	0.4		39.1 (0.69)	41.4 (0.70)
	0.6		67.5 (0.66)	69.9 (0.65)
	0.8		87.2 (0.47)	88.4 (0.45)
	1		95.9 (0.28)	96.4 (0.26)
	2		100.0 (0.00)	100.0 (0.00)
	3		100.0 (0.00)	100.0 (0.00)
	0.8		0	6.4 (0.35)
		0.2	9.0 (0.40)	10.0 (0.42)
		0.4	16.9 (0.53)	18.4 (0.55)
		0.6	29.3 (0.64)	31.2 (0.66)
		0.8	45.3 (0.70)	47.3 (0.71)
		1	60.1 (0.69)	62.2 (0.69)
		2	96.7 (0.25)	96.9 (0.24)
		3	99.9 (0.04)	99.9 (0.03)

Table 5.11: Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

mators of all possible blocks into the formulation of the self-normalizer, which at a cost makes the required computation substantially more than that of the conventional approach; see also the discussion in Shao (2015). In addition, such an approach does not unify with the mean self-normalizer of (Lobato, 2001), which has been celebrated for its application in the mean due to its connection with the CUSUM process and is

n	rho	d	S15 _{W₃}	TSSN
200	-0.3	0	4.8 (0.30)	6.2 (0.34)
		0.2	51.0 (0.71)	55.2 (0.70)
		0.4	93.9 (0.34)	94.9 (0.31)
		0.6	99.6 (0.09)	99.7 (0.08)
		0.8	100.0 (0.00)	100.0 (0.00)
		1	100.0 (0.00)	100.0 (0.00)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)
		-0.6	0	4.1 (0.28)
	0.2		49.1 (0.71)	53.9 (0.71)
	0.4		93.3 (0.35)	95.1 (0.30)
	0.6		99.7 (0.08)	99.8 (0.06)
	0.8		100.0 (0.00)	100.0 (0.00)
	1		100.0 (0.00)	100.0 (0.00)
	2		100.0 (0.00)	100.0 (0.00)
	3		100.0 (0.00)	100.0 (0.00)
	-0.8		0	3.7 (0.27)
		0.2	42.2 (0.70)	48.6 (0.71)
		0.4	88.9 (0.44)	91.9 (0.39)
		0.6	99.1 (0.13)	99.5 (0.10)
		0.8	100.0 (0.02)	100.0 (0.00)
		1	100.0 (0.00)	100.0 (0.00)
		2	100.0 (0.00)	100.0 (0.00)
		3	100.0 (0.00)	100.0 (0.00)

Table 5.12: Average raw power at the 95% nominal level using different self-normalizers for the median of the process (5.7) under different dependence strengths. The associated standard errors are reported in the parentheses. The results are based on 50,000 realizations for each configuration.

already T-symmetric. Due to the suggestion of Professor Michael Stein at The University of Chicago, Shao (2015) then considered a second approach, which is to compute the conventional self-normalizer of Shao (2010b) for both the original process and its reversed counterpart and take the average. Although such an approach makes the self-normalizer and its associated inference procedure T-symmetric, it generally leads

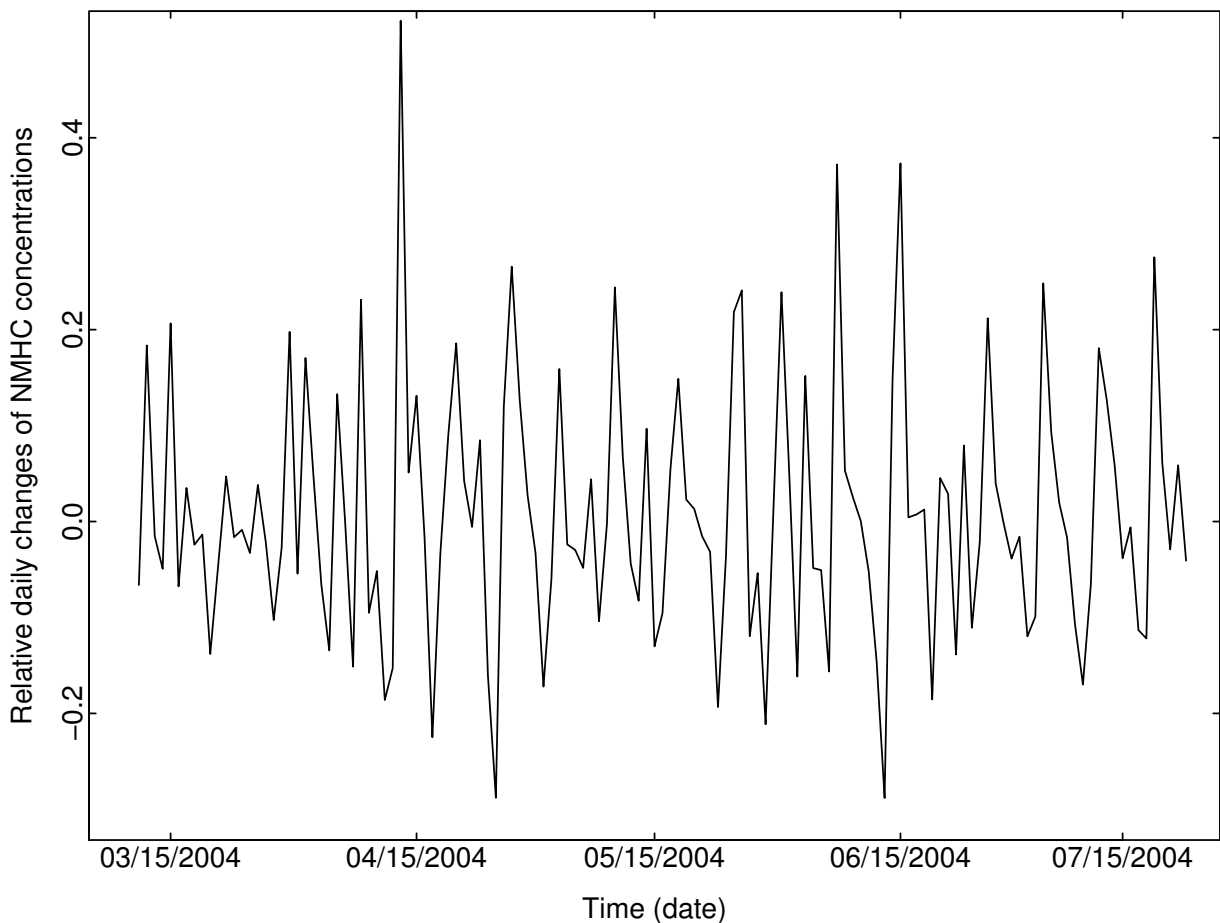


Figure 5.1: Relative daily changes of NMHC concentrations from March 10, 2004 to July 23, 2004.

to confidence intervals with wider lengths or tests with less power when compared with the conventional self-normalizer; see the simulation results in Section 5.3 and also a similar observation reported by Shao (2015). The current paper seeks a new approach in symmetrizing the self-normalizer by exploiting mathematical properties of the CUSUM process. In particular, instead of interpreting the CUSUM process as the difference between forward recursive estimators and the overall estimator as in Shao (2010b), we propose to reorganize the CUSUM process in a mathematically equivalent but T-symmetric form before replacing block mean estimators by their

95% Confidence Intervals		
	<i>Mean</i>	<i>Median</i>
<i>Forward Application</i>		
S10	(0.0014, 0.0158)	(-0.0196, -0.0103)
S15 _{W₂}	(-0.0003, 0.0175)	(-0.0395, 0.0096)
S15 _{W₃}	(0.0014, 0.0158)	(-0.0405, 0.0106)
TSSN	(0.0014, 0.0158)	(-0.0345, 0.0045)
<i>Backward Application</i>		
S10	(0.0014, 0.0158)	(-0.0508, 0.0209)
S15 _{W₂}	(-0.0003, 0.0175)	(-0.0395, 0.0096)
S15 _{W₃}	(0.0014, 0.0158)	(-0.0405, 0.0106)
TSSN	(0.0014, 0.0158)	(-0.0345, 0.0045)

Table 5.13: Constructed 95% confidence intervals for the mean and the median using different methods for the relative daily change in the NMHC concentration during the period from March 10, 2004 to July 23, 2004.

generalized counterparts. This new approach provides us with an augmented class of self-normalizers, which includes the conventional self-normalizer as a special case. Algebraic properties of this augmented class of self-normalizers are studied, and a sufficient and necessary condition for identifying its T-symmetric subspace is also provided in Section 5.2. In addition to its pleasant theoretical properties, it can be seen from our numerical experiments in Section 5.3, and the discussions therein, that the proposed TSSN method seems to also possess certain advantages in terms of the finite-sample performance, when compared with the conventional self-normalizer of Shao (2010b) and with the two symmetrization methods considered by Shao (2015).

Chapter 6

Conclusion

Since the influential work of Shao (2010b), self-normalization has experienced deep and exciting developments in various statistical inference problems. In this dissertation we gave a brief overview of a field and pointed the diversity of existing research in the area. One of the major focuses of our work was in testing for the presence of change-points in the time series process. The majority of the results existed prior were focused on making inference of quantities associated with a known stationary time series, and as commented by Shao and Zhang (2010), its adaptation to change-point testing problems can be nontrivial as direct implementation can lead to inconsistent tests. The aforementioned paper considered the situation with a single change-point alternative and proposed a consistent self-normalized test. Its extension to multiple change-point alternatives was also given but required prespecifying the actual number of change points under the alternative and is thus supervised. Since the actual number of change points is typically unknown especially when one is at the stage of seeking a statistical test to determine their existence, it seems desirable if we can have an unsupervised counterpart as considered in the current paper that can be used for situations when there is no or ambiguous prior knowledge about the number of change points or when doubts are casted on this prior knowledge. As can be seen from our simulation study, the supervised test of Shao and Zhang (2010) can suffer

from serious power losses when the prespecified value on the number of change points is misspecified, and in certain situations the power of their test can even decrease to zero as the alternative deviates from the null. In contrast, the proposed unsupervised test does not require prespecifying the number of change points and seems to perform reasonably well for situations with different numbers of change points. We have also proposed an alternative formulation of the test statistic, which is designed to reduce sensitivity to the magnitude of the first and the last change-points, but requires further investigation and we see further development of this approach as a potential topic for future research. In addition, we have proposed a new approach, called the contrast-based method, for generalizing self-normalized statistics to quantities other than the mean. Unlike the conventional approach which was mainly developed for constructing confidence intervals for quantities associated with a stationary time series, the proposed contrast-based approach is specifically tailored for change-point testing problems and can lead to tests with better power performance. Another useful feature of the proposed contrast-based approach is that it allows one to study the asymptotic power behavior of self-normalized tests under a more desirable and interpretable deterministic condition. This also complements the result of Shao and Zhang (2010) who only studied the asymptotic power of their test for the mean case.

Our second major focus of interest was in studying the self-normalization approach for statistical inference about the unknown quantities. Although self-normalization has been celebrated for its ability to avoid direct estimation of the nuisance asymptotic variance, its commonly adopted form due to the seminal work of Shao (2010b) is not T -symmetric when being applied to quantities other than the mean. The asymmetric issue can cause unnecessary confusions in practice, as different conclusions can be possibly obtained depending on the direction of application; see the discussion in Shao (2015) and also the real data illustration Chapter 5. To address the issue,

Shao (2015) considered two different approaches in symmetrizing the self-normalizer. The first approach is to involve parameter estimators of all possible blocks into the formulation of the self-normalizer, which at a cost makes the required computation substantially more than that of the conventional approach. In addition, such an approach does not unify with the mean self-normalizer, which has been celebrated for its application in the mean due to its connection with the CUSUM process and is already T -symmetric. Due to the suggestion of Professor Michael Stein at The University of Chicago, Shao (2015) then considered a second approach, which is to compute the conventional self-normalizer of Shao (2010b) for both the original process and its reversed counterpart and take the average. Although such an approach makes the self-normalizer and its associated inference procedure T -symmetric, it generally leads to confidence intervals with wider lengths or tests with less power when compared with the conventional self-normalizer and we have demonstrated in Chapter 5. In our work we have proposed a new approach in symmetrizing the self-normalizer by exploiting mathematical properties of the CUSUM process. In particular, instead of interpreting the CUSUM process as the difference between forward recursive estimators and the overall estimator as in Shao (2010b), we propose to reorganize the CUSUM process in a mathematically equivalent but T -symmetric form before replacing block mean estimators by their generalized counterparts. This new approach provides us with an augmented class of self-normalizers, which includes the conventional self-normalizer as a special case. Algebraic properties of this augmented class of self-normalizers are studied, and a sufficient and necessary condition for identifying its T -symmetric subspace is also provided. In addition to its pleasant theoretical properties, it can be seen from our numerical experiments and the discussions therein, that the proposed TSSN method seems to also possess certain advantages in terms of the finite-sample performance, when compared with the conventional self-normalizer of Shao (2010b)

and with the two symmetrization methods considered by Shao (2015).

Bibliography

- A. Canepa, L. G. G. (2007). Improvement of the quasi-likelihood ratio test in arma models: some results for bootstrap methods. *Journal of Time Series Analysis*, 28:434–453.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.
- Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34:1–16.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.
- Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80:2369–2429.
- Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81:608–650.
- Berkes, I., Liu, W., and Wu, W. B. (2014). Komlós-major-tusnády approximation under dependence. *The Annals of Probability*, 42(2):794–817.
- Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley, 2 edition.

- Billingsley, P. (2012). *Probability and Measure*. Wiley, 4 edition.
- Crainiceanu, C. M. and Vogelsang, T. J. (2007). Nonmonotonic power for tests of a mean shift in a time series. *Journal of Statistical Computation and Simulation*, 77:454–476.
- Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239.
- De Vito, S., Massera, E., Piga, M., Martinotto, L., and Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129:750–757.
- Elsner, J. B., Kossin, J. P., and Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, 455:92–95.
- Escanciano, J. C. and Lobato, I. N. (1997). Testing the martingale hypothesis. *Palgrave Handbook of Econometrics*, 39:972–1003.
- Fernholz, L. T. (1983). *Von Mises Calculus For Statistical Functionals*. Springer-Verlag New York, 1 edition.
- Frank R. Hampel, Elvezio M. Ronchetti, P. J. R. W. A. S. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, 1 edition.
- Genay, R. and Signori, D. (2015). Multi-scale tests for serial correlation. *Journal of Econometrics*, 184:62–80.
- Hall, P. and Jing, B.-Y. (1996). On sample reuse methods for dependent data. *Journal of the Royal Statistical Society: Series B*, 58(4):727–737.

- Hannan, E. (1979). The central limit theorem for time series regression. *Stochastic Processes and their Applications*, 9:281–289.
- Hansen, N. R., Reynaud-Bouret, P., and Rivoirard, V. (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21:83–143.
- Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business Economic Statistics*, 23:365–380.
- Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 12(1):141–153.
- Horowitz and Savin (2000). Empirically relevant critical values for hypothesis tests: A bootstrap approach. *Journal of Econometrics*, 95:375–389.
- Huang, Y., Volgushev, S., and Shao, X. (2015). On self-normalization for censored dependent data. *Journal of Time Series Analysis*, 36:109–124.
- Jagger, T. H. and Elsner, J. B. (2009). Modeling tropical cyclone intensity with quantile regression. *International journal of climatology*, 29:1351–1361.
- Khan, S. and Saleh, A. K. M. E. (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate student-t errors. *Biometrical Journal*, 39:131–147.
- Koenker, R. and Gilbert Bassett, J. (1978). Regression quantiles. *Econometrica*, 46:33–50.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241.

- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer-Verlag, New York.
- Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory*, 26:1218–1245.
- Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96:1066–1076.
- Marco, S. and Gutierrez-Galvez, A. (2012). Signal and data processing for machine olfaction and chemical sensing: A review. *IEEE Sensors Journal*, 12:3189–3214.
- Mead, M. I. and et all., O. P. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.
- Politis, D. N. (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, 27:703–744.
- Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer, New York.
- Poppick, A., McInerney, D., and et all (2016). Temperatures in transient climates: improved methods for simulations with evolving temporal covariances. *Annals of Applied Statistics*, 10:477–505.
- Rho, Y. and Shao, X. (2013). Improving the bandwidth-free inference methods by prewhitening. *Journal of Statistical Planning and Inference*, 143:1912–1922.

- Shao, X. (2010a). The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235.
- Shao, X. (2010b). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72:343–366.
- Shao, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis*, 32:598–606.
- Shao, X. (2015). Self-normalization for time series: a review of recent developments. *Journal of the American Statistical Association*, page to appear.
- Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105:1228–1240.
- Sun, Y. and M.L. Stein, c. a. (2016). A stochastic space-time model for intermittent precipitation occurrences. *Annals of Applied Statistics*, 9:2110–2132.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag, New York.
- Victor H. Pea, T. L. L. and Shao, Q.-M. (2009). *Self-Normalized Processes*. Springer-Verlag Berlin Heidelberg, 1 edition.
- Vogelsang, T. J. (1999). Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88:283–299.
- Volgushev, S. and Shao, X. (2014). A general approach to the joint asymptotic analysis of statistics from sub-samples. *Electronic Journal of Statistics*, 8(1):390–431.

- Wang, J. and Han, Y. (2016). Evaluation of dynamically downscaled extreme temperature using a spatially-aggregated generalized extreme value (gev) model. *Climate Dynamics*, 47:2833–2849.
- Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussions). *Annals of Statistics*, 14:1261–1350.
- Wu, W. B. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability*, 35(6):2294–2320.
- Yao, Y.-C. (1988). Estimating the number of change-points via schwarz' criterion. *Statistics & Probability Letters*, 6:181–189.
- Zhang, T., Ho, H.-C., Wendler, M., and Wu, W. B. (2013). Block sampling under strong dependence. *Stochastic Processes and their Applications*, 123:2323–2339.
- Zhang, T. and Lavitas, L. (2017). Unsupervised self-normalized change-point testing for time series. *Journal of the American Statistical Association*, page forthcoming.
- Zhang, T. and Wu, W. B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika*, 98:599–614.
- Zhou, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *The Annals of Statistics*, 38(4):2187–2217.
- Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108:726–740.
- Zhou, Z. and Shao, X. (2013). Inference for linear models with dependent errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75:323–343.

CURRICULUM VITAE

