INFERENCE OF TIME SERIES REGRESSION MODELS WITH WEAKLY DEPENDENT ERRORS

BY

YEONWOO RHO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Statistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Xiaofeng Shao, Chair
Professor Douglas G. Simpson
Professor Annie Qu
Professor John I. Marden

Abstract

In this thesis we develop inferential methods for time series models with weakly dependent errors in the following three aspects. The first aspect concerns the issue of the size-distortion in the presence of strong temporal dependence, which is well-known in the literature. There are recently proposed bandwidth-free methods, which generally reduces the size-distortion compared to the traditional method. However, these methods still suffer from severe size distortion when the temporal dependence in the error process is strong. We propose to use the prewhitening to handle the strong temporal dependence so that the size distortion is greatly reduced in the presence of strong temporal dependence in the error. This work is presented as Chapter 2, in the context of time series regression with dynamic regressors and stationary and weakly dependent errors. The second and third aspects are motivated by the recent surge of awareness that the stationarity assumption for the error is often too restrictive for real data. Some macroeconomic series are often observed to have heteroscedastic behavior. In Chapter 3, we introduce short-memory nonstationary error framework that can accommodate a wide range of nonstationary linear processes or modulated stationary processes in the context of trend assessment setting. We propose a method that can handle both heteroscedastic behavior and the temporal dependence in the error process. In Chapter 4, we further introduce a piecewise locally stationary framework for the error process that can cover a wide range of linear and nonlinear processes that are short-memory nonstationary in the unit root setting. A bootstrap-based method is proposed and its consistency is proved.

To my family.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Xiaofeng Shao. His enthusiasm, motivation, patience, and immense knowledge have guided me through my PhD study and research. I cannot imagine having a better advisor, mentor, and a role model for my Ph.D study and academic career.

Besides my advisor, my completion of thesis could not be accomplished without the support of rest of my thesis committee: Dr. Douglas Simpson, Dr. Annie Qu, and Dr. John Marden. I cannot thank enough for their continuous support, encouragement, insightful comments, and hard questions, which made my thesis more solid. In addition, I appreciate the teaching and consulting opportunities given by Dr. Annie Qu and Dr. Douglas Simpson, which helped me better understand and practice my major.

Last but not least, I would like to thank my husband Seokwoo Choi for the support and love throughout my PhD study. It was a great comfort and pleasure to have a lifetime partner who is also an expert in the same field.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis we consider inference for regression models in time series analysis. In time series literature, error processes of regression models are often assumed to be stationary and weakly dependent. The limiting distribution of the ordinary least squares (OLS) estimators under stationary error assumption follows normal distribution, but the temporal dependence in the error makes the asymptotic variance more complicated than the case with independent and identically distributed (iid) errors. We call this asymptotic variance the long-run variance, because it contains autocovariances of the error process with lags up to infinity, whereas in the iid case the asymptotic variance contains marginal variances only.

In the traditional inference approach, the heteroscedasticity and autocorrelation consistent (HAC) estimation is often involved to consistently estimate the asymptotic covariance matrix of regression parameter estimator, which is basically the kernel-based nonparametric estimator. Since the bandwidth parameter in the HAC estimation is difficult to choose in practice, there has been a recent surge of interest in developing bandwidth-free inference methods. There are two such new approaches. One is the KVB method Kiefer et al. (2000) and the other is the self-normalized (SN) method (Lobato, 2001; Shao, 2010b). To avoid the choice of the bandwidth parameters, the KVB and SN methods inconsistently estimate the long-run variance. The limiting distribution is nonstandard but pivotal, and in their simulations, the finite sample coverages or sizes are observed to be closer to the nominal level compared to the HAC method. However, existing simulation studies show that these new methods still suffer from severe size distortion in the presence of strong temporal dependence for a medium sample size. To remedy the problem, in Chapter 2[1], we propose to apply the prewhitening to the inconsistent long-run variance estimator in these methods to reduce the size distortion in the testing problem of regression parameters in time series regression models with dynamic regressors and stationary errors. The asymptotic distribution of the prewhitened

[1] Reprinted from Journal of Statistical Planning and Inference, Vol 143, Yeonwoo Rho and Xiaofeng Shao, "Improving the bandwidth-free inference methods by prewhitening," Pages 1912–1922, Copyright (2013), with permission from Elsevier.

Wald statistic is obtained and the general effectiveness of prewhitening is shown through simulations.

However, stationary assumption for the error process is often too ideal for the real data. For example, it has been recently pointed out that some macroeconomic series may exhibit short-memory nonstationary behavior such as nonconstant (unconditional) variances. Motivated by the need to assess the significance of the trend in some macroeconomic series, in Chapter 3[2], we consider inference of parameters in parametric trend functions when the errors exhibit certain degrees of nonstationarity with changing unconditional variances. We adopt the SN approach to avoid the difficulty involved in the estimation of the asymptotic variance of the ordinary least squares estimator. The limiting distribution of the SN quantity is non-pivotal due to the nonstationarity, unlike the usual SN method in stationary setting. However, the self-normalization does remove the part due to the temporal dependence in the limiting distribution so that the limiting distribution can be consistently approximated by using the wild bootstrap. Numerical simulation demonstrates favorable coverage properties of the proposed coupled method of the SN and wild bootstrap in comparison with alternative ones. The U.S. nominal wages series is analyzed to illustrate the finite sample performance.

Chapter 4[3] further generalizes the nonstationary framework in Chapter 3 in the context of unit root testing. Unit root testing is a well-studied topic in Econometrics and Statistics. In most of the existing works for unit root testing, the errors are assumed to be stationary. However, there has been a recent surge of awareness that the stationary error assumption is too restrictive for real data, and new unit root tests are needed to deal with nonstationary errors. Existing studies often assume similar error processes as we used in Chapter 3, but these type of error process is still somewhat restrictive in the sense that, for example, it basically only allows for heteroscedastic behavior in the error variance. In Chapter 4, we adopt the piecewise locally stationary process as our error process, which allows for both smooth and abrupt changes in second or higher order properties and accommodates the unconditional heteroscedasticity and weak dependence. This kind of process covers a wide range of linear and nonlinear time series that are short-memory nonstationary. Under this new framework, we derive the limiting null distributions of the conventional unit root test statistics, which contain a number of unknown parameters that are difficult to estimate. To facilitate the inference, we propose to use the dependent wild bootstrap to approximate the non-pivotal limiting null distributions, and the bootstrap consistency is theoretically justified under

[2]Accepted subject to minor changes for publication in Journal of Business and Economic Statistics.
[3]Submitted for publication in Econometric Theory.

both the null and alternative. Through finite sample simulations, we demonstrate the size accuracy of our procedure as compared to the block bootstrap-based counterpart.

Chapter 2

Improving the bandwidth-free inference methods by prewhitening

2.1 Introduction

Inference for a parameter in a stationary time series or a time series regression model is a well-studied problem in econometrics and statistics. The traditional approach typically requires consistent estimation of the asymptotic covariance matrix using the so-called HAC (heteroskedasticity autocorrelation consistent) estimator. The HAC estimator involves a choice of the bandwidth or truncation lag which is a smoothing parameter that balances bias and variance in estimation or size distortion and power loss in testing. See Andrews (1991) and Newey and West (1987) for important early developments. Lately, there has been some new developments on inference without using a bandwidth parameter. Kiefer et al. (2000) (KVB, hereafter) can be considered to be the first one to propose an inconsistent estimator of the asymptotic covariance matrix for the ordinary least squares (OLS) estimator in dynamic linear regression models. The limiting distribution of their Wald statistic is nonstandard but pivotal and the critical values can be obtained by simulations. In a similar spirit, Lobato (2001) developed a new bandwidth-free method to test the nullity of the first k-lag autocorrelations for a weakly dependent stationary time series. Shao (2010b) further generalized Lobato's method to allow approximately linear statistic that has a non-differentiable influence function in the context of confidence interval construction, and called it the "self-normalized" (SN, hereafter) approach. The above-mentioned approaches share a common feature: to circumvent the problem of choosing bandwidth parameters in the consistent covariance matrix estimation, inconsistent estimators of asymptotic covariance matrix are used and the resulting test statistics have non-standard but pivotal limiting distributions.

In finite samples the HAC-based tests can exhibit a large size distortion especially in the presence of strong dependence; see Andrews (1991). This is mainly due to the difficulty involved with the spectral density estimation. In the stationary time series setting, the long-run variance corresponds to the spectral density at zero frequency up to a multiplicative constant. The HAC-based tests take advantage of this fact and utilize the usual lag-window type spectral density estimators to consistently estimate the long-run variance. If the data is close to the white noise, spectral density is flat so that the smoothing strategy easily gives unbiased estimations. The difficulty rises when the temporal dependence is strong. The spectral density becomes spiky, and the estimation using the smoothing technique, which is basically taking averages on local neighborhoods, would yield heavily biased estimation. Prewhitening, or prefiltering, is originally developed in the literature of spectral density estimation. Prewhitening helps reduce the bias in the estimation by first flattening spectral density with a linear filter; the flattened density can now be estimated with less bias, and transforming it back completes the estimation. See Priestley (1981) for more details. Andrews and Monahan (1992) applied prewhitening to the HAC estimator using the vector autoregression (VAR) filter to help reduce the bias, and in their Monte Carlo experiment, the bias and size distortion become much less severe especially in the strongly dependent case.

The bandwidth-free approaches by KVB, Lobato, and Shao also aim to improve the size distortion of the traditional HAC-based approach, but as shown in their simulation studies the size distortion is still apparent when temporal dependence is strong for a given small/medium sample size. In this chapter we consider the hypothesis testing problem in the time series regression setting. Our main goal is to investigate if we can improve the size of the recently developed bandwidth-free inference approaches by employing prewhitening. It is worth noting that the prewhitening method has been implemented in the simulations of Kiefer and Vogelsang (2005) for the studentized mean under the fixed-b asymptotics, but no rigorous theory seems provided. As Kiefer et al. (2000) and Shao (2010b) developed their methods under different frameworks, we first extend the SN method to the regression setting to make them comparable. We further demonstrate that the KVB and SN approaches are analytically different in finite samples and apply the idea of prewhitening to both KVB's and SN methods and prove their theoretical validity.

The rest of this chapter is organized as follows. Section 2.2 introduces the KVB method and its prewhitened version with a theoretical justification. In Section 2.3 we extend the SN method to the regression setting and make an analytical comparison with the KVB method. The prewhitened SN method is also presented in Section 2.3. The finite sample size and power results are presented in Section 2.4. Section 2.5 concludes and figures and tables are in Section 2.6. Technical details are gathered in Appendix

A.

Throughout the chapter, we use "$\longrightarrow_p$" for convergence in probability, "$\longrightarrow_D$" for convergence in distribution, and "$\Rightarrow$" for weak convergence in $\mathcal{D}[0,1]$, the space of functions on [0,1] which are right continuous and have left limits, endowed with the Skorokhod topology (Billingsley, 1968). The symbols $O_p(1)$ and $o_p(1)$ signify being bounded in probability and convergence to zero in probability, respectively. If $O_p(1)$ and $o_p(1)$ are used for matrices, they mean elementwise boundedness and convergence to zero in probability. We use $\lfloor a \rfloor$ to denote the integer part of $a \in \mathbb{R}$ and $B_k(\cdot)$ to denote a k-dimensional vector of independent Brownian motions.

2.2 Prewhitening of the KVB method

In this section we present the KVB approach and its prewhitened version. Following Kiefer et al. (2000) we consider the regression model

$$y_t = X_t'\beta + u_t, \quad t = 1, 2, \cdots, T, \tag{2.1}$$

where β is a $k \times 1$ vector of regression parameters, X_t is a $k \times 1$ vector of stationary regressors, and u_t is a stationary error that satisfies $E(u_t|X_t) = 0$. Define $v_t = X_t u_t$ and let $\Omega = \sum_{j=-\infty}^{\infty} E(v_t v_{t-j}')$ be the long-run variance of v_t. Note that v_t is stationary so that Ω does not depend on t. Let Ψ be the lower triangular matrix from the Cholesky decomposition of Ω, that is, $\Psi\Psi' = \Omega$. Throughout the chapter, we consider testing the null hypothesis

$$H_0: \ R\beta = a, \tag{2.2}$$

where R is an $m \times k$ matrix, a is an $m \times 1$ vector, and the number of hypotheses is m. We assume $m \leq k$ and rank$(R) = m$. The extension of our prewhitening idea to nonlinear models with nonlinear hypotheses can be done following the development in Bunzel et al. (2001), but we shall not pursue this generality here for the sake of simplicity.

Consider the OLS estimator $\widehat{\beta}_T$ of β. To obtain the limiting distribution of $T^{1/2}(\widehat{\beta}_T - \beta)$, we introduce the following two assumptions from Kiefer et al. (2000).

K1. $T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} X_t u_t \Rightarrow \Psi B_k(r)$ for all $r \in (0, 1]$.

K2. $T^{-1}\sum_{t=1}^{\lfloor rT\rfloor} X_t X_t' \longrightarrow_p rQ$ for all $r \in (0,1]$ and Q^{-1} exists.

These two conditions hold when $\{(X_t', u_t)\}$ is a stationary vector time series satisfying suitable weak dependence and moment conditions. For example, Assumption K1 is satisfied if $\{X_t u_t\}$ is weakly stationary, strong mixing with mixing coefficients decaying at certain rates, and has finite moments of order greater than two; see Assumption 2.1 of Phillips (1987a). Assumption K2 holds by an application of law of large numbers for stationary and weakly dependent sequences; see White (1984). Under K1 and K2, as shown in Kiefer et al. (2000), we have

$$T^{1/2}(\widehat{\beta}_T - \beta) \longrightarrow_D Q^{-1}\Psi B_k(1) \sim N(0, V), \tag{2.3}$$

where $V = Q^{-1}\Omega Q^{-1} = Q^{-1}\Psi\Psi' Q^{-1}$. Denote

$$U_m = B_m(1)' \left[\int_0^1 \big(B_m(r) - rB_m(1)\big)\big(B_m(r) - rB_m(1)\big)' dr \right]^{-1} B_m(1)$$

for any $m \in \mathbb{N}$. Kiefer et al. (2000) proposed the following Wald-type test statistic

$$G_T^{KVB} = T(R\widehat{\beta}_T - a)' \left[R\left\{ \widehat{Q}_T^{-1} \left(\frac{1}{T^2} \sum_{t=1}^{T} \widehat{S}_t^{KVB} \widehat{S}_t^{KVB'} \right) \widehat{Q}_T^{-1} \right\} R' \right]^{-1} (R\widehat{\beta}_T - a)$$

where $\widehat{Q}_T^{-1} = \big(\frac{1}{T}\sum_{j=1}^{T} X_j X_j'\big)^{-1}$, $\widehat{S}_t^{KVB} = \sum_{j=1}^{t} X_j \widehat{u}_j$ for $t = 1, \cdots, T$, and $\widehat{u}_j = y_j - X_j'\hat{\beta}_T$ are the OLS residuals. Under the null hypothesis, they showed that $G_T^{KVB} \longrightarrow_D U_m$.

While the traditional approach estimates $\Omega = \sum_{j=-\infty}^{\infty} E(v_t v_{t-j}')$ consistently using the HAC estimator, the KVB method estimates Ω by its inconsistent sample counterpart $\widehat{\Omega} = T^{-2}\sum_{t=1}^{T} \widehat{S}_t^{KVB} \widehat{S}_t^{KVB'}$, which corresponds to the usual HAC estimator with bandwidth equal to sample size and the Bartlett kernel (Kiefer and Vogelsang, 2002). As seen from the simulation results in Kiefer et al. (2000), the size distortion is still large when the dependence of v_t is strong for a medium sample size. To alleviate the problem, we propose to estimate Ω by a prewhitened version of $\widehat{\Omega}$. Following Andrews and Monahan (1992), we fit a VAR(p) model to $\widehat{v}_t = X_t \widehat{u}_t$ with a fixed p

$$\widehat{v}_t = \sum_{l=1}^{p} \widehat{A}_l \widehat{v}_{t-l} + \widehat{\epsilon}_t, \qquad t = p+1, \cdots, T, \tag{2.4}$$

and calculate the fitted residuals

$$\widehat{\epsilon}_t = \widehat{v}_t - \sum_{l=1}^{p} \widehat{A}_l \widehat{v}_{t-l}, \qquad t = p+1, \cdots, T.$$

Here we do not require the underlying true model for v_t or $\widehat{v}_t$ to be VAR(p). The parametric VAR(p) model is expected to capture the second order dependence in $\hat{v}_t$ and the autocorrelations for the residuals $\hat{\epsilon}_t$ are expected to be weaker than those in $\hat{v}_t$. Here the use of VAR(p) prewhitening filter is out of technical/practical convenience. In practice, a suitable parametric model that captures the main autocorrelation feature in $\{\hat{v}_t\}$ can be used and some model selection procedures can be adopted to achieve a balance of goodness-of-fit and model parsimony.

Define

$$V_{T,PW}^{KVB} = \widehat{Q}_T^{-1} \widehat{D}_p \left(\frac{1}{(T-p)^2} \sum_{t=p+1}^{T} \widetilde{S}_t^{KVB} \widetilde{S}_t^{KVB\prime} \right) \widehat{D}_p' \widehat{Q}_T^{-1},$$

where $\widetilde{S}_t^{KVB} = \sum_{j=p+1}^{t} \widehat{\epsilon}_j$ for $t = p+1, \cdots, T$, and $\widehat{D}_p = (I_k - \sum_{l=1}^{p} \widehat{A}_l)^{-1}$. Here I_k is the $k \times k$ identity matrix. To show the theoretical validity of the prewhitened KVB method, we need the following two assumptions in addition to K1 and K2.

PW1. $\widehat{A}_l - A_l = o_p(1)$ for some $A_l \in \mathbb{R}^{k \times k}$ for all $l = 1, \cdots, p$.

PW2. $I_k - \sum_{l=1}^{p} A_l$ is non-singular.

Note that if we replace PW1 with a stronger assumption $\sqrt{T}(\widehat{A}_l - A_l) = O_p(1)$, then PW1 and PW2 are the same as Assumption D (i) and (ii) in Andrews and Monahan (1992). In general, the assumptions PW1 and PW2 are not primitive. In the appendix, we present an example that demonstrates PW1 and PW2 indeed hold provided that the OLS estimate is used in (2.4) and some other mild assumptions are satisfied. Note that PW1 can easily be satisfied if MLE or Yule-Walker estimates are used for $\widehat{A}_l$ as shown in Taniguchi and Kakizawa (2000) and Lütkepohl (2005), although we need a separate proof.

Theorem **2.2.1**. *If assumptions K1, K2, PW1, and PW2 are satisfied, then under the null hypothesis (2.2),* $G_{T,PW}^{KVB} = T(R\widehat{\beta}_T - a)'(RV_{T,PW}^{KVB}R')^{-1}(R\widehat{\beta}_T - a) \longrightarrow_D U_m$.

Thus the prewhitened Wald statistic admits the same limiting null distribution as its unprewhitened counterpart, which is consistent with the results in Andrews and Monahan (1992). In the latter paper, the bias reduction property of the prewhitened (consistent) long-run variance estimator has been shown to be effective in reducing the size distortion associated with the unprewhitened counterpart. Here our long-run variance estimators before or after prewhitening are both inconsistent, so it would be interesting to see if prewhitening still works for the KVB method.

2.3 Extension of the SN method to the regression model and its prewhitened version

As seen from Kiefer et al. (2000) and Shao (2010b), the KVB method and the SN method seem very similar. They both use inconsistent estimates of asymptotic variance matrix without using a bandwidth parameter and they have the same asymptotic distribution that only depends on the number of hypotheses for a testing problem. However, since the SN method is formulated in the setting of univariate stationary time series and the KVB method for regression models, it is unclear if the two methods deliver the same finite sample performance. In this section we extend the SN method to the regression setting, compare the KVB and SN methods, and investigate the prewhitening of the SN method.

We consider the same regression model (2.1) and the null hypothesis (2.2) as in Section 2.2. Let $\widehat{\theta}_t$ be the OLS estimate of $\theta = \beta$ based on the first $t+k-1$ observations, $\{(y_1, X_1'), \cdots, (y_{t+k-1}, X_{t+k-1}')\}$ for $t = 1, \cdots, N$. These $\widehat{\theta}_t$'s are the same as $\widehat{\beta}_{t+k-1}$'s in Section 2.2. Let $N = T - k + 1$. Define the SN-based Wald statistic as

$$G_N^{SN} = N(R\widehat{\theta}_N - a)'(RW_N R')^{-1}(R\widehat{\theta}_N - a),$$

where $W_N = N^{-2}\sum_{t=1}^N t^2(\widehat{\theta}_t - \widehat{\theta}_N)(\widehat{\theta}_t - \widehat{\theta}_N)'$. Define $\widehat{S}_t^{SN} = t(\widehat{\theta}_t - \widehat{\theta}_N)$ and $\widehat{Q}_t^{-1} = \left(\frac{1}{t}\sum_{j=1}^t X_j X_j'\right)^{-1}$ for $t \geq k$. We rewrite the partial sum process $\widehat{S}_t^{KVB}$ from the KVB method as

$$\widehat{S}_t^{KVB} = \sum_{j=1}^t X_j \widehat{u}_j = \sum_{j=1}^t X_j(y_j - X_j'\widehat{\theta}_N) = \sum_{j=1}^t X_j y_j - \left(\sum_{j=1}^t X_j X_j'\right)\widehat{\theta}_N$$

and $\widehat{S}_t^{SN}$ from the SN method as

$$\widehat{S}_t^{SN} = t\widehat{\theta}_t - t\widehat{\theta}_N = t\left(\sum_{j=1}^{t+k-1} X_j X_j'\right)^{-1}\left(\sum_{j=1}^{t+k-1} X_j y_j\right) - t\widehat{\theta}_N = \frac{t}{t+k-1}\widehat{Q}_{t+k-1}^{-1}\widehat{S}_{t+k-1}^{KVB}, \qquad (2.5)$$

for $t = 1, \cdots, N$. Since K2 implies

$$\widehat{Q}_{\lfloor rN \rfloor+k-1} = \frac{T}{\lfloor rN \rfloor + k - 1} \frac{1}{T} \sum_{j=1}^{\lfloor rN \rfloor+k-1} X_j X_j' \longrightarrow_p \frac{1}{r} rQ = Q \quad \text{for all } r \in (0,1]$$

and K1 implies

$$T^{-1/2} \widehat{S}^{KVB}_{\lfloor rN \rfloor+k-1} \Rightarrow \Psi(B_k(r) - rB_k(1)),$$

we have

$$N^{-1/2} \widehat{S}^{SN}_{\lfloor rN \rfloor} = \frac{N^{-1/2}}{T^{-1/2}} \frac{\lfloor rN \rfloor}{\lfloor rN \rfloor + k - 1} \widehat{Q}^{-1}_{\lfloor rN \rfloor+k-1} T^{-1/2} \widehat{S}^{KVB}_{\lfloor rN \rfloor+k-1} \Rightarrow Q^{-1}\Psi(B_k(r) - rB_k(1)).$$

The following theorem presents the limiting null distribution of the SN-based Wald statistic, which is a direct consequence of the continuous mapping theorem.

Theorem **2.3.1**. *Assume K1 and K2. Under the null hypothesis (2.2), we have* $G_N^{SN} \longrightarrow_D U_m$.

Remark **2.3.1**. Therefore the KVB and the SN methods are asymptotically equivalent at the first order. From our discussion preceding Theorem 2.3.1, we can see that the two Wald statistics (i.e. KVB versus SN) differ in their forms, which leads to different finite sample performance. We can rewrite KVB's and SN statistics without prewhitening as

$$\text{(KVB)} \quad T(R\widehat{\theta}_N - a)' \left[R \left(\frac{1}{T^2} \sum_{t=1}^{T} \widehat{Q}_T^{-1} \widehat{S}_t^{KVB} \widehat{S}_t^{KVB'} \widehat{Q}_T^{-1} \right) R' \right]^{-1} (R\widehat{\theta}_N - a)$$

and

$$\text{(SN)} \quad N(R\widehat{\theta}_N - a)' \left[R \left(\frac{1}{N^2} \sum_{t=k}^{T} \alpha_t^2 \widehat{Q}_t^{-1} \widehat{S}_t^{KVB} \widehat{S}_t^{KVB'} \widehat{Q}_t^{-1} \right) R' \right]^{-1} (R\widehat{\theta}_N - a),$$

where $\alpha_t = (t-k+1)/t$ for $t = k, \cdots, T$. The main difference between the SN method and the KVB method is that in the SN method the recursive estimate $\widehat{Q}_t$ is used in place of $\widehat{Q}_T$ in the KVB formulation. In other words, the KVB method applies recursive estimation not to the asymptotic variance $V = Q^{-1}\Omega Q^{-1}$ itself but only to the 'meat' part Ω of the sandwich expression. Simulation results in Section 2.4 show that the SN method delivers much less size distortion in finite samples. It is not fully clear why the SN method performs better, but it suggests that recursive estimate $\widehat{Q}_t$ helps with the finite sample size.

Similar observations were made in Lee (2006), who proposed an SN-type test statistic in the M-estimation context.

To apply the prewhitening to the SN-based Wald statistic, we note that the SN method introduced above implicitly employs the setting of strictly stationary time series in Shao (2010b). Let $Z_t = \{(y_t, X_t'), \cdots, (y_{t+k-1}, X_{t+k-1}')\}'$, F_k be the marginal distribution function of Z_t, and $\beta = H(F_k) \in \mathbb{R}^k$ where H is a functional that corresponds to OLS estimation. Let $\widehat{\theta}_t = \widehat{\beta}_{t+k-1}$ be the OLS estimator using the first $t+k-1$ observations $\{(y_1, X_1'), \cdots, (y_{t+k-1}, X_{t+k-1}')\}$ as before. As in Shao (2010b), suppose we can do the following expansion around the true coefficient $\theta = \beta$,

$$\widehat{\theta}_N = \theta + \frac{1}{N}\sum_{t=1}^{N} IF(Z_t; F_k) + R_N, \tag{2.6}$$

where $IF(z; F_k)$ is the influence function of the functional H and R_N is the remainder term. The asymptotic variance V in (2.3) is the long-run variance of $IF(Z_t; F_k)$, which may involve unknown nuisance parameters. Prewhitening can be applied to an estimate of the influence function, denoted as $\widehat{IF}(Z_t; F_k)$. To motivate our estimate, we note that

$$\widehat{\theta}_t = \theta + \frac{1}{t}\sum_{j=1}^{t} IF(Z_j; F_k) + R_t \quad \text{and} \quad \widehat{\theta}_{t-1} = \theta + \frac{1}{t-1}\sum_{j=1}^{t-1} IF(Z_j; F_k) + R_{t-1},$$

which imply that

$$IF_t = IF(Z_t; F_k) = t\widehat{\theta}_t - (t-1)\widehat{\theta}_{t-1} - \theta + (t-1)R_{t-1} - tR_t.$$

In general, we expect the remainder term $(t-1)R_{t-1} - tR_t$ to be negligible, so the influence function can be naturally estimated by

$$\widehat{IF}_t = \widehat{IF}(Z_t; F_k) = t\widehat{\theta}_t - (t-1)\widehat{\theta}_{t-1} - \widehat{\theta}_N.$$

Then we have $\sum_{j=1}^{t} \widehat{IF}_j = t(\widehat{\theta}_t - \widehat{\theta}_N)$, which coincides with $\widehat{S}_t^{SN}$ in (2.5), suggesting that our choice of the estimated influence function is consistent with the self-normalizer used in the SN method. It is worth noting that there exist other choices of $\widehat{IF}(Z_t; F_k)$. For example, one may use

$$\widehat{IF}(Z_t; F_k) = (N-1)(\widehat{\theta}_N - \widehat{\theta}_{-t}),$$

where $\widehat{\theta}_{-t}$ is the estimate based on the leave-one-out subsample $(Z_1, \cdots, Z_{t-1}, Z_{t+1}, \cdots, Z_N)$. It can be motivated from the expression $\widehat{\theta}_{-t} = \theta + \frac{1}{N-1} \sum_{j \neq t} IF(Z_j; F_k) + R_{-t}$ and (2.6). However, this Jackknife-based estimator did not work well in our simulation, so we shall not pursue further investigation along this line.

Next we consider the prewhitening of the SN-based Wald statistic based on $\widehat{IF}_t$. Again we fit a VAR(p) model to $\{\widehat{IF}_t\}$ and get the fitted residuals

$$\widehat{\epsilon}_t = \widehat{IF}_t - \sum_{l=1}^{p} \widehat{A}_l \widehat{IF}_{t-l}, \quad t = p+1, \cdots, T.$$

Here $\widehat{IF}_t$ may not be from a VAR(p) model. The choice of p can be based on an inspection of partial autocorrelation function or a particular model selection algorithm like AIC. We can also choose p by testing for zero coefficients or doing GLR tests for nested models as recommended in Lütkepohl (2005).

Define

$$V_{N,PW}^{SN} = \widehat{D}_p \left(\frac{1}{(N-p)^2} \sum_{t=p+1}^{N} \widetilde{S}_t^{SN} \widetilde{S}_t^{SN\prime} \right) \widehat{D}_p',$$

where $\widetilde{S}_t^{SN} = \sum_{j=p+1}^{t} \widehat{\epsilon}_j$ for $t = p+1, \cdots, T$, and $\widehat{D}_p = (I_k - \sum_{l=1}^{p} \widehat{A}_l)^{-1}$. The prewhitened SN-based Wald test statistic is then

$$T_{N,PW}^{SN} = N(R\widehat{\theta}_N - a)'(RV_{N,PW}^{SN} R')^{-1}(R\widehat{\theta}_N - a).$$

To establish its asymptotic distribution, we further assume

S1. $N^{-1/2} \lfloor rN \rfloor (\widehat{\theta}_{\lfloor rN \rfloor} - \theta) \Rightarrow \Delta B_k(r)$ for $r \in (0, 1]$, where Δ is the lower triangular matrix from the Cholesky decomposition of V in (2.3).

Remark **2.3.2**. Assumption S1 follows from the functional central limit theorem for $IF(Z_t; F_k)$, which holds under suitable moment and mixing conditions on Z_t, and the uniform negligibility of the remainder terms $\{R_t\}_{t=1}^{N}$. In the regression setting (2.1), if the OLS estimator is used, then K1 and K2 imply S1. Specifically, since $\widehat{\theta}_{\lfloor rN \rfloor} = (\sum_{t=1}^{\lfloor rN \rfloor} X_t X_t')^{-1} (\sum_{t=1}^{\lfloor rN \rfloor} X_t u_t) + \theta$, we have that

$$N^{-1/2} \lfloor rN \rfloor (\widehat{\theta}_{\lfloor rN \rfloor} - \theta) \Rightarrow Q^{-1} \Psi B_k(r)$$

under Assumptions K1 and K2, i.e., S1 holds in view of the fact that $V = Q^{-1}\Psi\Psi' Q^{-1} = \Delta\Delta'$.

Theorem **2.3.2**. *Suppose $\{(y_t, X_t')\}$ is stationary. If assumptions S1, PW1, and PW2 are satisfied, then, under the null hypothesis (2.2), $T_{N,PW}^{SN} \longrightarrow_D U_m$.*

Remark **2.3.3**. Existence of $\widehat{A}_l$'s and A_l's that satisfy PW1 and PW2 can be shown following the argument for the KVB approach in the appendix. PW1 holds with the same choices of $\widehat{A}$ and A as provided in the appendix by replacing $\widehat{v}_t$ with $\widehat{IF}_t$ and v_t with IF_t. To derive the local asymptotic power, we can follow exactly the same argument as in Kiefer et al. (2000). Since the limiting distributions (under the null and alternative) are all the same for the KVB and SN methods as well as their prewhitened versions, there is no difference in the asymptotic local power for the four methods. The details are omitted.

Remark **2.3.4**. In general, Lobato's, KVB's, and the SN methods yield slightly different test statistics, but for the inference of the mean of a stationary time series, the three methods coincide and so do their prewhitening versions. Note that testing the mean of a stationary time series can be recast as a special case of the testing problem in the regression models. If we let X_t in the regression model (2.1) be a constant $\mathbf{1}_m$, where m is the dimension of the time series, $R = \mathbf{1}_m$, and $a = \mu_0$, where μ_0 is the hypothesized mean under the null for the hypothesis (2.2), testing the regression parameter becomes the testing mean problem. In this case, all three methods result in the following test statistic

$$n(\overline{X}_n - \mu_0)' {V_n}^{-1} (\overline{X}_n - \mu_0) \longrightarrow_D U_m,$$

where n is the sample size, $\overline{X}_n$ denotes the sample mean, $V_n = n^{-2}\sum_{t=1}^{n} \widehat{S}_t \widehat{S}_t'$, and $\widehat{S}_t = \sum_{j=1}^{t}(X_j - \overline{X}_n)$.

2.4 Finite Sample Performance

In this section we compare finite sample size and power of the KVB and SN-based Wald test in the regression setting. We also compare both tests with and without prewhitening to see if prewhitening can help to reduce the size distortion. Following the simulation design in Kiefer et al. (2000), we replicate the 6 data-generating processes in their paper. We consider the regression model (2.1) with $k = 5$, one constant regressor and four stochastic regressors. The four stochastic regressors and the errors are drawn independently from AR(1) homoskedastic processes (AR(1)-HOMO), AR(1) heteroskedastic processes (AR(1)-HET1 and AR(1)-HET2), MA(1) homoskedastic processes (MA(1)-HOMO), and MA(1) heteroskedastic processes (MA(1)-HET1 and MA(1)-HET2). Then the regressors are transformed so that $X'X = TI_5$. For details about this transformation, we refer to Andrews and Monahan (1992). Specifically, errors and four stochastic regressors are drawn from $\eta_t = \rho\eta_{t-1} + e_t$ where $e_t \sim N(0, 1-\rho^2)$ for AR(1) models. MA(1) models use $\eta_t = e_t + \theta e_{t-1}$ with $e_t \sim N\big(0, (1+\theta^2)^{-1}\big)$. HET1 and HET2 models are constructed with the same AR(1) or MA(1) process with e_t's replaced by $|X_{t,2}|e_t$ and $|\frac{1}{2}\sum_{i=2}^{5} X_{t,i}|e_t$, respectively, where $X_{t,i}$ denotes the ith regressor at time t. For AR(1) models, we use ρ=-0.95,-0.9,-0.7,-0.5,-0.3,0,0.3,0.5,0.7,0.9,0.95 and for MA(1) models, θ=-0.99,-0.7,-0.5,-0.3,0,0.3,0.5,0.7,0.99. Our results are based on 5,000 replications, sample size $T = 128$, and nominal level is set to be 0.05. Table II in Kiefer et al. (2000) is used for asymptotic critical values. Note that we need to divide our test statistics G_T^{KVB}, $G_{T,PW}^{KVB}$, G_N^{SN}, and $G_{N,PW}^{SN}$ by the number of hypotheses m to use their critical values. Following Andrews and Monahan (1992), we use VAR(1) as the prewhitening filter for all the cases. However, we do not follow their singularity treatment for $\widehat{D}_1^{-1}$ because it does not affect the result much in our simulation. We denote the KVB method without prewhitening by "K", KVB with prewhitening by "K-PW", SN without prewhitening by "SN", and SN with prewhitening by "SN-PW".

Tables 2.1, 2.2, and 2.3 present empirical sizes for AR(1)-HOMO, AR(1)-HET1, and AR(1)-HET2, respectively. It can be seen that prewhitening generally improves both KVB's and SN methods in the presence of strong temporal dependence. The stronger temporal dependence is associated with a better correction of size distortion. However, when temporal dependence is weak and $m = 4$, prewhitening becomes ineffective, falsely rejecting more true null hypotheses than the original methods. This may be due to the relatively large m, weak dependence, and ineffectiveness of VAR(1) filter in this case. When $m = 4$, the VAR(1) model contains too many parameters relative to the sample size $T = 128$ and a more parsimonious model might be able to fit the data better and make the prewhitening effective. Similar findings hold for MA(1) models, which are not reported here to save space. A comparison of the KVB method and the SN method shows that the SN method usually does noticeably better than the KVB method. In particular, size distortion of the SN method is almost as half as that of the KVB method when temporal dependence is very strong, say $\rho \geq 0.9$. On the other hand, when temporal dependence is weak or strongly negative, and when $m = 3$ or 4, the SN method performs slightly worse than the KVB method after prewhitening. The prewhitened SN-based test still outperforms the prewhitened KVB test in most cases.

Figure 2.1 presents size-adjusted power curves for the AR(1)-HOMO case. Size-adjusted powers are calculated for $m = 1$, where the 95% quantile of 5,000 test statistics calculated with data generated under the null hypothesis is used as the size-adjusted critical value. Percentages of true rejection (power) are calculated with 1,000 replications for each of 19 β's at 0.1,0.2,$\cdots$,2. We present power curves for only six ρ's, $\rho = -0.95, -0.9, -0.7, 0, 0.9, 0.95$. Graphs for other values of ρ look very similar to that of $\rho = 0$, in which all of the four methods have almost the same size-adjusted powers. When dependence is strongly negative, power of the KVB method is slightly better than that of the SN method, and when dependence is strongly positive, the SN method has slightly better power, but the difference is not too large in either case. Overall, it seems that prewhitening does not have much impact on size-adjusted power for both methods.

On the basis of limited simulation results we conclude that prewhitening filter, if properly chosen, makes both KVB's and SN methods less size-distorted without losing much power, especially in the strongly dependent case. The SN method can substantially outperform the KVB method in size with little power loss or even gain some power in some cases.

To explain why prewhitening helps to reduce the size distortion for both KVB's and SN methods theoretically, we need to derive the Edgeworth expansion of the distribution of the studentized estimator in both cases. For studentized sample mean, Zhang and Shao (2013) recently provided such an expansion result under the fixed-b asymptotics in the framework of the Gaussian location model. However, it seems very challenging to generalize their result to the time series regression model adopted here. Nevertheless, we expect the (high order) bias of the studentizer as an estimator of asymptotic variance will show up in the leading error term of the expansion. The fact that prewhitening is capable of reducing the bias for both consistent HAC (Andrews, 1991) estimator and inconsistent variance estimator used in the KVB and SN methods explains the improvement in size. Similarly, the inferiority of the KVB method and its prewhitening version relative to their SN counterparts, as seen from the simulation results in most cases, can also be explained by deriving the Edgeworth expansions for these studentized estimators, although this is beyond the scope of this article.

2.5 Conclusion

In this chapter we investigate the effect of prewhitening when applied to the nontraditional bandwidth-free inference methods proposed by Kiefer et al. (2000) and Shao (2010b) in the time series regression models. The nontraditional methods differ from the traditional HAC-based tests by using inconsistent estimators of the asymptotic covariance matrix, which lead to nonstandard but pivotal limiting distributions. All the nontraditional methods with or without prewhitening have the same limiting distribution that only depends on the number of hypotheses we are testing. Our simulation results demonstrate that prewhitening often helps to reduce the size distortion. Improvement in size distortion by prewhitening using a VAR(1) filter is seen to be substantial when the number of tests is smaller and temporal dependence is (positively) strong. The VAR(1) prewhitening filter used in our simulation studies is a convenient one. In practice, we recommend the practitioner to carefully examine the pattern of the autocorrelations/partial autocorrelations of $\widehat{v}_t$ ($\widehat{IF}_t$) and select an appropriate prewhitening filter or use some well-justified model selection algorithm to find a good parametric model to capture the dependence in $\widehat{v}_t$ ($\widehat{IF}_t$). As an important part of this chapter, we present an extension of the SN method to the regression setup and make an analytical comparison with the KVB method. Simulation studies show the SN-based test and its prewhitened version tend to have better size than the KVB counterparts especially when temporal dependence is strong.

2.6 Tables and Figures

Table 2.1: Finite sample rejection probabilities for OLS estimation of AR(1) models with homoskedasticity, T=128, 5,000 replications, nominal level 0.05; Asymptotic critical values from Kiefer et al. (2000) are used.

AR(1)-HOMO								
	$m=1$				$m=2$			
ρ	K	K-PW	SN	SN-PW	K	K-PW	SN	SN-PW
-0.95	0.281	0.241	0.136	0.119	0.386	0.315	0.225	0.199
-0.9	0.163	0.139	0.088	0.078	0.234	0.189	0.155	0.131
-0.7	0.087	0.081	0.071	0.068	0.103	0.092	0.084	0.076
-0.5	0.070	0.065	0.056	0.053	0.073	0.068	0.070	0.066
-0.3	0.059	0.058	0.049	0.048	0.065	0.062	0.061	0.064
0	0.054	0.053	0.046	0.048	0.060	0.061	0.052	0.056
0.3	0.064	0.063	0.049	0.047	0.070	0.065	0.055	0.055
0.5	0.063	0.059	0.048	0.043	0.084	0.074	0.060	0.061
0.7	0.087	0.078	0.055	0.048	0.118	0.099	0.070	0.057
0.9	0.184	0.156	0.064	0.051	0.246	0.198	0.088	0.072
0.95	0.287	0.239	0.070	0.058	0.395	0.315	0.114	0.094
	$m=3$				$m=4$			
ρ	K	K-PW	SN	SN-PW	K	K-PW	SN	SN-PW
-0.95	0.469	0.379	0.350	0.311	0.537	0.442	0.478	0.437
-0.9	0.298	0.234	0.240	0.208	0.355	0.278	0.318	0.280
-0.7	0.131	0.105	0.115	0.103	0.145	0.118	0.143	0.128
-0.5	0.086	0.081	0.083	0.082	0.098	0.086	0.103	0.106
-0.3	0.072	0.069	0.064	0.072	0.072	0.070	0.073	0.078
0	0.058	0.061	0.053	0.058	0.066	0.068	0.063	0.068
0.3	0.073	0.067	0.061	0.069	0.075	0.074	0.066	0.073
0.5	0.091	0.082	0.063	0.062	0.097	0.083	0.076	0.074
0.7	0.129	0.105	0.080	0.069	0.147	0.118	0.101	0.091
0.9	0.335	0.266	0.137	0.113	0.362	0.280	0.177	0.148
0.95	0.471	0.378	0.156	0.137	0.528	0.425	0.209	0.186

Table 2.2: Finite sample rejection probabilities for OLS estimation of AR(1) models with heteroskedasticity 1, T=128, 5,000 replications, nominal level 0.05; Asymptotic critical values from Kiefer et al. (2000) are used.

AR(1)-HET1								
	$m=1$				$m=2$			
ρ	K	K-PW	SN	SN-PW	K	K-PW	SN	SN-PW
-0.95	0.310	0.268	0.165	0.151	0.390	0.328	0.249	0.229
-0.9	0.188	0.162	0.124	0.111	0.242	0.197	0.175	0.158
-0.7	0.094	0.090	0.084	0.080	0.122	0.107	0.105	0.101
-0.5	0.068	0.066	0.066	0.063	0.082	0.077	0.079	0.078
-0.3	0.069	0.067	0.066	0.067	0.066	0.063	0.069	0.067
0	0.062	0.060	0.062	0.062	0.064	0.062	0.064	0.067
0.3	0.069	0.066	0.062	0.064	0.071	0.071	0.071	0.070
0.5	0.077	0.073	0.068	0.067	0.085	0.078	0.075	0.071
0.7	0.109	0.099	0.079	0.075	0.132	0.115	0.096	0.088
0.9	0.208	0.187	0.109	0.100	0.256	0.211	0.119	0.107
0.95	0.302	0.266	0.114	0.104	0.380	0.321	0.139	0.123
	$m=3$				$m=4$			
ρ	K	K-PW	SN	SN-PW	K	K-PW	SN	SN-PW
-0.95	0.454	0.362	0.336	0.308	0.499	0.394	0.440	0.416
-0.9	0.282	0.228	0.236	0.212	0.312	0.239	0.303	0.274
-0.7	0.119	0.106	0.124	0.118	0.138	0.118	0.151	0.142
-0.5	0.090	0.081	0.090	0.092	0.084	0.077	0.104	0.107
-0.3	0.074	0.072	0.075	0.081	0.074	0.074	0.080	0.084
0	0.065	0.067	0.069	0.074	0.068	0.070	0.077	0.082
0.3	0.070	0.067	0.065	0.069	0.074	0.073	0.077	0.083
0.5	0.092	0.084	0.081	0.079	0.090	0.082	0.090	0.094
0.7	0.126	0.111	0.096	0.087	0.142	0.124	0.121	0.119
0.9	0.316	0.259	0.152	0.135	0.348	0.274	0.190	0.172
0.95	0.447	0.359	0.184	0.164	0.488	0.380	0.214	0.189

Table 2.3: Finite sample rejection probabilities for OLS estimation of AR(1) models with heteroskedasticity 2, T=128, 5,000 replications, nominal level 0.05; Asymptotic critical values from Kiefer et al. (2000) are used.

AR(1)-HET2								
	$m=1$				$m=2$			
ρ	K	K-PW	SN	SN-PW	K	K-PW	SN	SN-PW
-0.95	0.283	0.255	0.135	0.119	0.387	0.332	0.226	0.209
-0.9	0.178	0.155	0.109	0.097	0.228	0.193	0.152	0.139
-0.7	0.095	0.089	0.079	0.075	0.109	0.095	0.097	0.089
-0.5	0.065	0.065	0.063	0.059	0.083	0.074	0.080	0.077
-0.3	0.067	0.064	0.064	0.066	0.070	0.066	0.068	0.069
0	0.055	0.058	0.054	0.055	0.055	0.055	0.057	0.056
0.3	0.064	0.063	0.052	0.052	0.072	0.070	0.062	0.063
0.5	0.064	0.060	0.049	0.047	0.079	0.074	0.065	0.069
0.7	0.093	0.084	0.065	0.061	0.118	0.107	0.084	0.080
0.9	0.193	0.175	0.079	0.067	0.248	0.208	0.099	0.090
0.95	0.290	0.256	0.078	0.067	0.386	0.328	0.119	0.102
	$m=3$				$m=4$			
ρ	K	K-PW	SN	SN-PW	K	K-PW	SN	SN-PW
-0.95	0.430	0.352	0.331	0.304	0.482	0.387	0.447	0.420
-0.9	0.281	0.225	0.234	0.214	0.312	0.254	0.302	0.279
-0.7	0.122	0.105	0.130	0.122	0.136	0.120	0.150	0.142
-0.5	0.080	0.074	0.087	0.087	0.085	0.082	0.105	0.107
-0.3	0.069	0.069	0.075	0.073	0.073	0.072	0.083	0.089
0	0.061	0.063	0.064	0.068	0.069	0.069	0.075	0.080
0.3	0.073	0.070	0.072	0.072	0.079	0.075	0.078	0.083
0.5	0.080	0.074	0.069	0.071	0.087	0.084	0.090	0.093
0.7	0.132	0.115	0.096	0.088	0.143	0.123	0.116	0.110
0.9	0.305	0.252	0.148	0.131	0.338	0.270	0.180	0.160
0.95	0.420	0.336	0.154	0.134	0.477	0.386	0.215	0.193

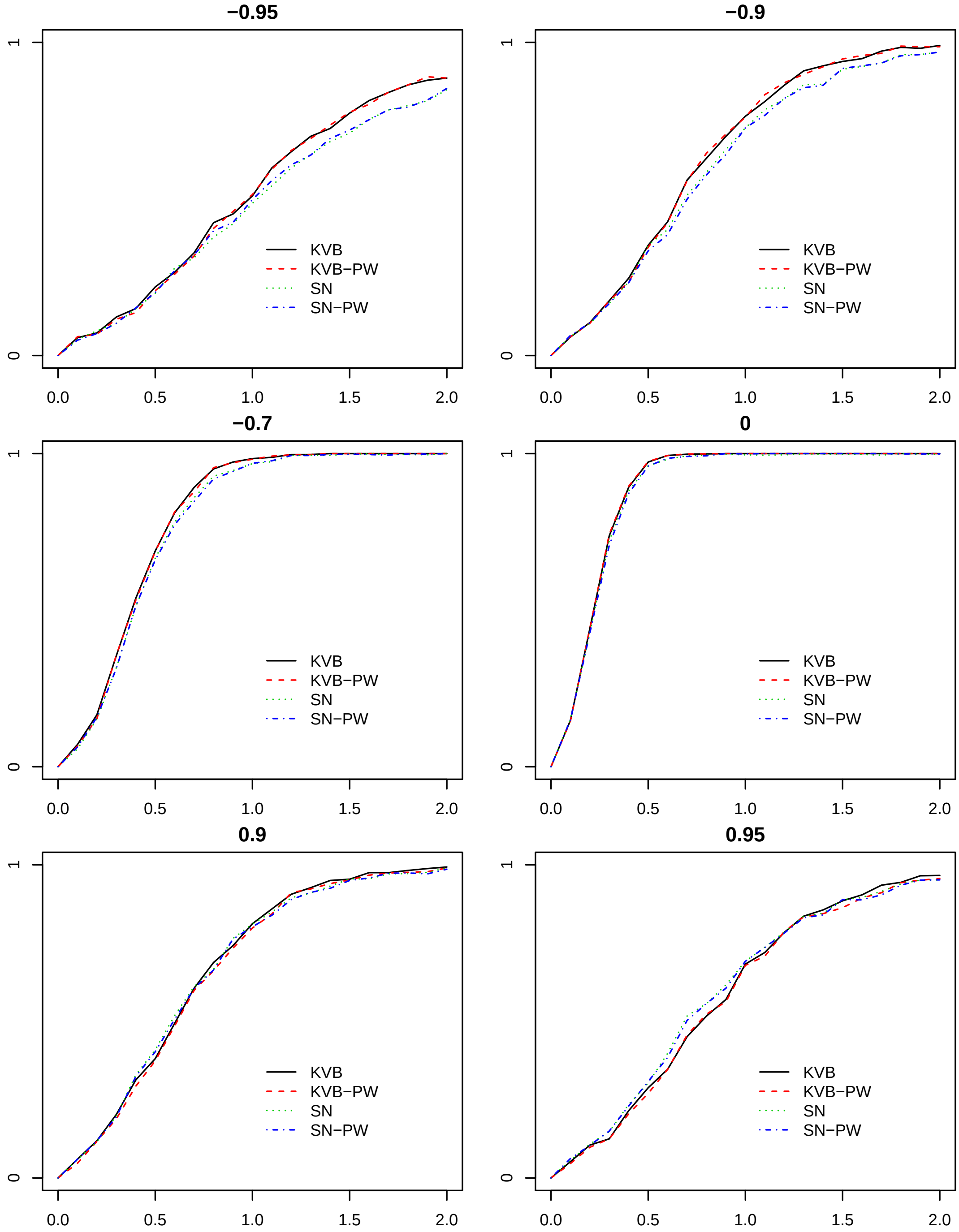

Figure 2.1: Size-adjusted power for AR(1)-HOMO for $m = 1$, $T = 128$, and $\rho = -0.95, -0.9, -0.7, 0, 0.9, 0.95$, with 1,000 replications. For other values of ρ, the size-adjusted power curves look very similar to the case of $\rho = 0$.

Chapter 3

Inference for Time Series Regression Models with Nonstationary Heteroscedastic Errors

3.1 Introduction

Consider the simple linear trend model,

$$X_{t,n} = b_1 + b_2(t/n) + u_{t,n}, \qquad t = 1, \ldots, n. \tag{3.1}$$

Inference of the linear trend coefficient b_2 is an important problem in many fields such as econometrics, statistics, and environmental sciences, and there is a rich literature. (Hamilton, 1994, Chapter 16) derived the limiting distribution of the ordinary least squares (OLS) estimators of b_1 and b_2 when the errors are independent and identically distributed (iid) from a normal distribution. To account for possible dependence in the error, $\{u_{t,n}\}$ has often been assumed to be a stationary weakly dependent process. For example, Sherman (1997) applied the subsampling approach when the errors are from a stationary mixing process, and Zhou and Shao (2013) considered the M-estimation with stationary errors whose weak dependence was characterized by physical dependence measures (Wu, 2005). Another popular way in econometrics is to model the dependence in the error as an autoregressive (AR) process, and testing procedures that are robust to serial correlation with possible unit root have been developed; see Canjels and Watson (1997), Vogelsang (1998), Bunzel and Vogelsang (2005), and Harvey et al. (2007), among others.

However, there is increasing evidence that many macroeconomic series exhibit heteroscedastic behavior. For example, the U.S. gross domestic product series has been observed to have less variability since 1980s; see Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Busetti and Taylor (2003), and references therein. In addition, Sensier and van Dijk (2004) argued that majority of the macroeconomic data in Stock and Watson (1999) had abrupt changes in unconditional variances. The empirical evidence of nonstationary heteroscedastic errors can also be found from environmental time series; see Rao (2004), Zhou and Wu (2009), and Zhang and Wu (2011) for their data illustrations using global temperature series. There have been a number of papers on the inference of time series models with nonstationary heteroscedastic errors. For example, Phillips and Xu (2006) and Xu and Phillips (2008) studied the inference on the AR coefficients when the innovations are modulated stationary or heteroscedastic linear processes. Xu (2008) focused on inference of polynomial trends with errors exhibiting nonstationary volatility. Zhao (2011) and Zhao and Li (2013) explored inferences on the constant mean in the presence of modulated stationary errors. Xu (2012) presented a statistic with a pivotal limiting distribution for the multivariate trend model when the error process is generated from a stable vector AR process with heteroscedastic innovations.

Our framework generalizes the model (3.1) in two nontrivial aspects. For one, our trend function is a finite linear combination of deterministic trends, each of which can contain finite number of breaks at known points. Most studies in the literature are restricted to the linear or polynomial trend, but allowing for break in intercept can be useful in practice. For example, if $X_{t,n} = b_1 + b_2\mathbf{1}(t > t_B) + b_3(t/n) + u_{t,n}$, i.e., a linear trend model with a jump at a known intervention time t_B, then testing the significance of b_2 is of interest in assessing the significance of the jump at the intervention in the presence of a linear trend. This kind of one break model was considered in the analysis of nominal wage series by Perron (1989), and its inference is reinvestigated in Section 3.5. For the other, we allow two types of nonstationary processes for the errors $u_{t,n}$. One type of nonstationary process is adapted from Cavaliere and Taylor (2007, 2008a,b) and Xu and Phillips (2008). For this class of models, the error process is represented as a linear process of independent but heteroscedastic innovations. This includes autoregressive moving average (ARMA) models with independent but heteroscedastic innovations as a special case. The other type of nonstationary process is the modulated stationary processes (Zhao, 2011; Zhao and Li, 2013), where a stationary process is amplified by a (possibly periodic) deterministic function that only depends on the relative location of an observation. The latter is a type of locally stationary processes (Priestley, 1965; Dahlhaus, 1997), as pointed out in Zhao and Li (2013). Together, they cover a wide class of nonstationary models with unconditional heteroscedasticity.

When the errors are nonstationary, the difficulty rises due to the complex form of the asymptotic variance of the OLS estimator. In this chapter, we apply the wild bootstrap (WB) (Wu, 1986) to approximate the limiting behavior of the OLS estimator. However, simply applying the wild bootstrap to the unstudentized OLS estimator does not work because it cannot properly capture the terms in the asymptotic variance that are due to temporal dependence in the error. To overcome this difficulty, we propose to adopt the self-normalized (SN) method (Lobato, 2001; Shao, 2010b), which was mainly developed for stationary time series. Due to the heteroscedasticity and the non-constant regressor, the limiting distribution of the SN-based quantity is non-pivotal, unlike the existing SN-based methods. The advantage of using the SN method, though, is that the limiting distribution of the studentized OLS estimator only depends on the unknown heteroscedasticity, which can be captured by the wild bootstrap. Compared to the conventional heteroscedasticity and autocorrelation consistent (HAC) method or block-based bootstrap methods that are known to handle mildly nonstationary errors, the use of the SN method and the wild bootstrap are practically convenient and are shown to lead to more accurate inference.

To summarize, we provide an inference procedure of trends in time series that is robust to smooth and abrupt changes in unconditional variances of the temporally dependent error process, in a quite general framework. Our assumptions on deterministic trends and nonstationary heteroscedastic errors are less restrictive than those from most earlier works in the trend assessment literature. Besides, our inference procedure is convenient to implement. Although our method is not entirely bandwidth-free, the finite-sample performance is not overly sensitive to the choice of the trimming parameter ϵ, and its choice is accounted for in the first order limiting distribution and the bootstrap approximation. The work can be regarded as the first extension of the SN method to the regression model with nonstationary errors. One of the key theoretical contributions of the present chapter is to derive the functional central limit theorem of the recursive OLS estimator, which leads to the weak convergence of the SN-based quantity. At a methodological level, the self-normalization coupled with the wild bootstrap eliminates the need to directly estimate the temporal dependence in the error process. Although some earlier works such as Cavaliere and Taylor (2008b) and Zhao and Li (2013) applied the wild bootstrap for similar error processes, it seems that they need to involve the choice of a block size or consistent estimation of a nuisance parameter to make the wild bootstrap work. In our simulation, our method is demonstrated to deliver more accurate finite-sample coverage compared to alternative methods, and thus inherits a key property of the SN approach which was known to hold only for stationary time series (Shao, 2010b). Recently, Zhou and Shao (2013) extended the SN method to the regression setting with fixed parametric regressors, where the error is stationary and the response variable is only nonstationary in mean. The framework here is considerably more general in that the error is allowed to exhibit local stationarity or unconditional heteroscedasticity and the response variable is nonstationary at all orders. On the other hand, we only consider the inference based on the ordinary least squares (OLS) estimator, whereas the M-estimation was considered in Zhou and Shao (2013).

The rest of the chapter is organized as follows. Section 3.2 presents the model, the OLS estimation, and the limiting distribution of the OLS estimator. Section 3.3 contains a functional central limit theorem for the recursive OLS estimators, which leads to the limiting distribution of the SN-based quantity. The wild bootstrap is described and its consistency for the SN quantity is justified in Section 3.3. Section 4.4 presents some simulation results and Section 3.5 contains an application of our method to the U.S. nominal wage data. Section 4.5 concludes and Section 3.7 presents tables and figures. Technical details are relegated to Appendix B.

Throughout the chapter, we use $\xrightarrow{\mathcal{D}}$ for convergence in distribution and $\Rightarrow$ weak convergence in $D[\epsilon, 1]$ for some $\epsilon \in (0, 1)$, the space of functions on $[\epsilon,1]$ which are right continuous and have left limits, endowed with Skorohod metric Billingsley (1968). The symbols $O_p(1)$ and $o_p(1)$ signify being bounded in probability and convergence to zero in probability, respectively. If $O_p(1)$ and $o_p(1)$ are used for matrices, they mean elementwise boundedness and convergence to zero in probability. We use $\lfloor a \rfloor$ to denote the integer part of $a \in \mathbb{R}$, $B(\cdot)$ a standard Brownian motion, and $N(\mu, \Sigma)$ the (multivariate) normal distribution with mean μ and covariance matrix Σ. Denote by $||X||_p = (E|X|^p)^{1/p}$. For a $p \times q$ matrix $A = (a_{ij})_{i \le p, j \le q}$, let $||A||_F = (\sum_{i=1}^p \sum_{j=1}^q a_{ij}^2)^{1/2}$ be the Frobenius norm. Let $\mathcal{X}_n = (X_{1,n}, \ldots, X_{n,n})$ denote the data.

3.2 The Model and Estimation

Consider the model

$$X_{t,n} = F_{t,n}'\beta + u_{t,n} = b_1 f_1(t/n) + \cdots + b_p f_p(t/n) + u_{t,n}, \quad t = 1, \ldots, n,$$

where $X_{t,n}$ is a univariate time series, n is the sample size, $\beta = (b_1, \ldots, b_p)'$ is the parameter of interest, and the regressors $F_{t,n} = \{f_1(t/n), \ldots, f_p(t/n)\}'$ are non-random. The regressors $\{f_j(t/n)\}_{j=1}^p$ are rescaled, thus increasing the sample size n means we have more data in a local window. We use $F(s) = \{f_1(s), \ldots, f_p(s)\}'$, $s \in [0, 1]$ to denote the regressors with relative location s, and we observe $F_{t,n} = F(t/n)$. Let $N = n - p + 1$, where p is fixed.

The error process $\{u_{t,n}\}$ is assumed to be either one of the followings;

(A1) [Generalized Linear Process] The error process is defined as

$$u_{t,n} = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j,n} = \mathbf{C}(L)\varepsilon_{t,n},$$

where $\varepsilon_{t,n} = \omega_{t,n} e_t$ with e_t iid (0,1), $\mathbf{C}(L) = \sum_{j=0}^{\infty} c_j L^j$, and L is the lag operator. We further assume

- (i) Let $\omega(s) = \lim_{n\to\infty} \omega_{\lfloor ns\rfloor,n}$ be some deterministic, positive (cadlag) function on $s \in [0,1]$. Let $\omega(s)$ be piecewise Lipschitz continuous with at most finite number of breaks. If $t < 0$, let $\omega_{t,n} < \omega^*$ for some $0 < \omega^* < \infty$.
- (ii) $0 < |\mathbf{C}(1)| < \infty$, $\sum_{j=0}^{\infty} j|c_j| < \infty$.
- (iii) $E|e_t|^4 < \infty$.

(A2) [Modulated Stationary Process] The error process is defined as

$$u_{t,n} = \nu_{t,n}\eta_t,$$

where $\{\eta_t\}$ is a mean zero strictly stationary process that can be expressed as $\eta_t = G(\mathcal{F}_t)$ for some measurable function G and $\mathcal{F}_t = (\ldots, \epsilon_{t-1}, \epsilon_t)$ where ϵ_t are iid (0,1). We further assume

- (i) Let $\nu(s) = \lim_{n\to\infty} \nu_{\lfloor ns\rfloor,n}$ be some deterministic, positive (cadlag) function on $s \in [0,1]$. Let $\nu(s)$ be piecewise Lipschitz continuous with at most finite number of breaks.
- (ii) $E(\eta_t) = 0$, $E(\eta_t^2) = 1$, and $E(\eta_t^{4+\delta}) < \infty$ for some $\delta > 0$.
- (iii) For some $\chi \in (0,1)$, $||G(\mathcal{F}_k) - G(\{\mathcal{F}_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_k\})||_4 = O(\chi^k)$ if $k \geq 0$ and 0 otherwise. Here, $\{\epsilon_t'\}$ is an iid copy of $\{\epsilon_t\}$.
- (iv) $\Gamma^2 > 0$, where $\Gamma^2 = \sum_{h=-\infty}^{\infty} \mathrm{cov}(\eta_t, \eta_{t+h})$ is the long run variance of $\{\eta_t\}$.

The key feature of the two settings above is that they allow both smooth and abrupt changes in the unconditional variance and second order properties of $u_{t,n}$ through $\omega(s)$ in (A1) (i) and $\nu(s)$ in (A2) (i), which depend only on the relative location $s = t/n$ in a deterministic fashion. If this part should be constant, then the two models correspond to the popular stationary models; the linear process for (A1) and stationary causal process [see Wu (2005)] for (A2). The model (A1) is considered in, for example, Cavaliere

(2005), Cavaliere and Taylor (2007, 2008a,b), and Xu and Phillips (2008). The assumptions (A1) (ii) is popular in the linear process literature to ensure the central limit theorem and the invariance principle, and (A1) (iii) implies the existence of the fourth moment of $\{u_{t,n}\}$. The model (A2) is adapted from Zhao (2011) and Zhao and Li (2013) and is called the modulated stationary process, which were originally developed to account for the seasonal change, or periodicity, observed in financial or environmental data. The condition (A2) (ii) is slightly stronger than the existence of the fourth moment, which is required in the proof of the bootstrap consistency. (A2) (iii) implies that $\{u_{t,n}\}$ is short-range dependent and the dependence decays exponentially fast. These two classes of models are both nonstationary and nonparametric, which represent extensions of stationary linear/nonlinear processes to nonstationary processes with unconditional heteroscedasticity.

Remark **3.2.1**. Note that the model (A1) is a special case of Xu and Phillips (2008), but it can be made as general as the framework of Xu and Phillips (2008), by letting e_t be a martingale difference sequence with its natural filtration $\mathcal{E}_t$ satisfying $n^{-1}\sum_{t=1}^{n} E(e_t^2|\mathcal{E}_{t-1}) \to C < \infty$ for some positive constant C, rather than an iid sequence. See Remark 1 of Cavaliere and Taylor (2008b). In this chapter we do not pursue this generalization only for the simplicity in the proofs. For (A2), our framework for the modulation, $\nu(s)$ is more general than that of Zhao (2011). For example, the so-called "k-block asymptotically equal cumulative variance condition" in Definition 1 in Zhao (2011) rules out a linear trend in the unconditional variance. The model (A2) can be replaced with some mixing conditions for the stationary part, but the details are omitted for simplicity.

In this article, we are interested in the inference of $\beta = (b_1, ..., b_p)'$, based on the OLS estimator $\widehat{\beta}_N = (\sum_{t=1}^{n} F_{t,n}F_{t,n}')^{-1}(\sum_{t=1}^{n} F_{t,n}X_{t,n})$. If the errors satisfy (A1) or (A2) and the fixed regressors $F(s)$ are piecewise Lipschitz continuous, then under certain regularity conditions, the OLS estimator is approximately normally distributed, i.e.,

$$n^{1/2}(\widehat{\beta}_N - \beta) \xrightarrow{\mathcal{D}} N(\mathbf{0}_p, \Psi), \tag{3.2}$$

where the covariance matrix Ψ has the sandwich form $Q_1^{-1}VQ_1^{-1}$, $Q_1 = \int_0^1 F(s)F(s)'ds$, and

$$V = \begin{cases} \mathbf{C}^2(1)\int_0^1 \omega^2(s)F(s)F(s)'ds & \text{if (A1)} \\ \Gamma^2 \int_0^1 \nu^2(s)F(s)F(s)'ds & \text{if (A2)} \end{cases}$$

The statement (3.2) is a direct consequence of Theorem 3.3.1 presented below. Notice that Q_1 can be consistently estimated by $n^{-1}\sum_{t=1}^{n} F_{t,n}F'_{t,n}$, but V depends on the nuisance parameters $\mathbf{C}(1)$, Γ, and $\{\omega(s), \nu(s); s \in [0,1]\}$. To perform hypothesis testing or construct a confidence region for β, the conventional approach is to consistently estimate the unknown matrix V using, e.g., the HAC estimator in Andrews (1991). The HAC estimator involves a bandwidth parameter, and the finite sample performance critically depends on the choice of this bandwidth parameter. Moreover, the consistency of the HAC estimator is shown under the assumption that the errors are stationary or approximately stationary (Newey and West, 1987; Andrews, 1991). Alternatively, block-based resampling approaches [see Lahiri (2003)] such as the moving block bootstrap and subsampling [see Politis et al. (1999)] are quite popular to deal with the dependence in the error process, see Fitzenberger (1998), Sherman (1997), and Romano and Wolf (2006) among others. However, these methods are designed for stationary processes with different blocks having the same or approximately the same stochastic property. Note that Paparoditis and Politis (2002) proposed the so-called local block bootstrap for the inference of the mean with locally stationary errors. Their method involves two tuning parameters and no guidance on their choice seems provided.

There have been recently proposed alternative methods to the HAC-based inference for dynamic linear regression models such as Kiefer et al. (2000) [KVB, hereafter]. KVB's approach is closely related to the SN method by Lobato (2001) and Shao (2010b). The KVB and SN methods share the same idea that by using an inconsistent estimator of asymptotic variance, a bandwidth parameter can be avoided. In finite samples this strategy is shown to achieve better coverage and size compared to the conventional HAC-based inference. In a recent paper by Rho and Shao (2013) or in Chapter 2, the KVB method is shown to differ from the SN method in the regression setting: the KVB method applies an inconsistent recursive estimation of the "meat" part V of the covariance matrix Ψ in (3.2) and consistently estimates the "bread" part Q_1, whereas the SN method involves the inconsistent recursive estimation for the whole covariance matrix Ψ. Although the KVB and SN methods have the same limiting distribution, Rho and Shao (2013) have shown in their simulations that the SN method tends to have better finite sample performance than the KVB method. For this reason, we adopt the SN method for our problem.

To apply the SN method, we need to define an inconsistent estimate of Ψ based upon recursive estimates of β. Consider the OLS estimator $\widehat{\beta}_t = \widehat{\beta}_{t,N}$ of β using the first $t+p-1$ observations, $\widehat{\beta}_t = \left(\sum_{i=1}^{t+p-1} F_{i,n}F'_{i,n}\right)^{-1}\left(\sum_{i=1}^{t+p-1} F_{i,n}X_{i,n}\right)$, for all $t = 1,\dots,N$. We only have N estimates of β because $\sum_{t=1}^{k} F_{t,n}F'_{t,n}$ is not invertible if $k < p$. Then, following Shao (2010b), we construct the self-normalizer $\Omega_N(\epsilon) = N^{-2}\sum_{t=\lfloor N\epsilon\rfloor}^{N} t^2(\widehat{\beta}_t - \widehat{\beta}_N)(\widehat{\beta}_t - \widehat{\beta}_N)'$ using the recursive OLS estimators, where $\epsilon \in (0,1)$ is a trimming parameter. Define the SN quantity

$$T_N = N(\widehat{\beta}_N - \beta)'\{\Omega_N(\epsilon)\}^{-1}(\widehat{\beta}_N - \beta). \tag{3.3}$$

Here, $\Omega_N(\epsilon)$ is an estimate of Ψ, but unlike the HAC method, the major role of this self-normalizer is to remove nuisance parameters in the limiting distribution of the SN quantity T_N, rather than as a consistent estimator of the true asymptotic variance Ψ. With stationary errors, the same self-normalizer does this job well, providing pivotal limiting distributions for T_N. However, under our framework of nonstationary errors, the self-normalizer $\Omega_N(\epsilon)$ cannot completely get rid of the nuisance parameter in the limiting distribution of T_N, as seen from the next section.

3.3 Inference

For the regressor $F_{t,n}$, we impose the following assumptions.

(R1) For all $r \in [\epsilon, 1]$, $Q_r = r^{-1}\int_0^r F(s)F(s)'ds$ is well-defined and Q_r^{-1} exists, and further, $\inf_{r\in[\epsilon,1]}\det(Q_r) > 0$ and $\sup_{r\in[\epsilon,1]}||Q_r||_F < \infty$.

(R2) For each $j = 1, \ldots, p$, $f_j(s)$ is piecewise Lipschitz continuous with at most finite number of breaks.

The assumptions (R1)-(R2) are satisfied by commonly used trend functions such as $f_j(s) = s^\tau$ for some nonnegative integer τ, $f_j(s) = \sin(2\pi s)$ or $\cos(2\pi s)$. The assumption $\inf_{r\in[\epsilon,1]}\det(Q_r) > 0$ in (R1) excludes the collinearity of fixed regressors and is basically equivalent to the assumption (C4) in Zhou and Shao (2013), and the trimming parameter $\epsilon \in (0,1)$ has to be chosen appropriately. The assumption (R1) appears slightly weaker than Kiefer et al. (2000)'s Assumption 2, $\lfloor nr\rfloor^{-1}\sum_{t=1}^{\lfloor nr\rfloor} F_{t,n}F_{t,n}' = Q + o_p(1)$ for some $p \times p$ matrix Q, which is identical for all $r \in (0,1]$. Although, in general, Kiefer et al. (2000) allows for dynamic stationary regressors whereas our framework targets for fixed regressors, our method can be extended to allow for dynamic regressors by conditioning on the regressors. Since our work is motivated by the trend assessment of macroeconomic series, it seems natural to assume the regressor is fixed. The assumption (R2) allows for structural breaks in trend functions.

Theorem **3.3.1**. *Assume (R1)-(R2).*

(i) Let the error process $\{u_{t,n}\}$ be generated from (A1). The recursive OLS estimator $\widehat{\beta}_{\lfloor Nr \rfloor}$ of β converges weakly, i.e.,

$$N^{-1/2}\lfloor Nr \rfloor(\widehat{\beta}_{\lfloor Nr \rfloor} - \beta) \Rightarrow \mathbf{C}(1)Q_r^{-1}B_{F,\omega}(r), \tag{3.4}$$

where $B_{F,\omega}(r) = \int_0^r F(s)\omega(s)dB(s)$.

(ii) Assume (A2) for the error process $\{u_{t,n}\}$. Then

$$N^{-1/2}\lfloor Nr \rfloor(\widehat{\beta}_{\lfloor Nr \rfloor} - \beta) \Rightarrow \Gamma Q_r^{-1}B_{F,\nu}(r),$$

where $B_{F,\nu}(r) = \int_0^r \nu(s)F(s)dB(s)$.

(iii) Define T_N as in (3.3) and $\widetilde{B}_F(r) = B_{F,\omega}(r)$ if (A1) and $\widetilde{B}_F(r) = B_{F,\nu}(r)$ if (A2). It follows from the continuous mapping theorem that

$$T_N \xrightarrow{\mathcal{D}} \mathcal{L}_F = \widetilde{B}_F(1)'Q_1^{-1}\{\Omega(\epsilon)\}^{-1}Q_1^{-1}\widetilde{B}_F(1),$$

where $\Omega(\epsilon) = \int_\epsilon^1 \left\{Q_r^{-1}\widetilde{B}_F(r) - rQ_1^{-1}\widetilde{B}_F(1)\right\}\left\{Q_r^{-1}\widetilde{B}_F(r) - rQ_1^{-1}\widetilde{B}_F(1)\right\}'dr.$

Remark **3.3.1**. For a technical reason, the above functional central limit theorem has been proved on $D[\epsilon, 1]$ for some $\epsilon \in (0, 1)$ instead of on $D[0, 1]$. For most of the trend functions that include polynomial s^τ with $\tau > 1/2$ or $\sin(2\pi s)$, this ϵ has to be strictly greater than 0. However, for some trend functions, ϵ can be set as 0. In particular, when the trend function is a constant mean function, then $Q_r = 1$ for $r \in (0, 1]$, $B_F(r) = \int_0^r dB(s) = B(r)$, and $B_{F,\nu}(r) = \int_0^r \nu(s)dB(s)$. Our proof goes through when $\epsilon = 0$. In any case, the choice of trimming parameter ϵ is captured by the first order limiting distribution in the same spirit of the fixed-b approach (Kiefer and Vogelsang, 2005).

Remark **3.3.2**. If the trend function has breaks, the trimming parameter ϵ has to be rather carefully chosen so that the condition (R1) can be satisfied. For example, if we use a linear trend function with a jump in the mean, i.e., $X_{t,n} = b_1 + b_2\mathbf{1}(t > t_B) + b_3(t/n) + u_{t,n}$, with a break point at t_B, then $Q_r = r^{-1}\int_0^r F(s)F(s)'ds$ is singular for $r \in [0, t_B/n]$. However, as long as we choose $\epsilon > t_B/n$ so that Q_r is invertible, the choice of trimming parameter does not seem to affect the finite sample performance much, as we can see in the third simulation in Section 4.4.

The limiting distribution $\mathcal{L}_F$ in Theorem 3.3.1 is not pivotal due to the heteroscedasticity of the error process and non-constant nature of Q_r. If the error process is stationary, then $\omega(s) \equiv \omega$ in (A1) and $\nu(s) \equiv \nu$ in (A2) would be cancelled out, and the only unknown part in the limiting distribution is Q_r. If

$Q_r = Q$, $r \in [\epsilon, 1]$ for some $p \times p$ matrix Q, then Q would be cancelled out in the limiting distribution, which occurs in the use of the SN method for stationary time series with a constant regressor; see Lobato (2001) and Shao (2010b). In our case, the part that reflects the contribution from the temporal dependence, $\mathbf{C}(1)$ or Γ, does cancel out. However, the heteroscedasticity remains even if $Q_r = Q$, and the limiting distribution still depends on the unknown nuisance parameter $\{\omega(s), \nu(s); s \in [0,1]\}$.

Estimating unknown parameters in $\mathcal{L}_F$ seems quite challenging due to the estimation of $\omega(s)$, $\nu(s)$, and integral of them over a Brownian motion. Instead of directly estimating the unknown parameters, we approximate the limiting distribution $\mathcal{L}_F$ using the wild bootstrap introduced in Wu (1986). Although there has been some recent attempts to use the wild bootstrap to the trend assessment with nonstationary errors (for example, Xu (2012) for the linear trend model and Zhao and Li (2013) for the constant mean model), this article seems to be the first to apply the wild bootstrap to the SN-based quantity. The main reason the wild bootstrap works for the SN method in such settings is that the part due to the temporal dependence, which is difficult to capture with the wild bootstrap, no longer exist in the limiting distribution $\mathcal{L}_F$ of the SN-based quantity so that $\mathcal{L}_F$ can be consistently approximated by the wild bootstrap. In contrast, without the self-normalization, the unknown nuisance parameter $\mathbf{C}(1)$ or Γ still remains in the limiting distribution, and thus, the wild bootstrap is not consistent in this case.

Let $\widehat{u}_{t,n} = X_{t,n} - F_{t,n}'\widehat{\beta}_N$ denote the residuals. We first generate the resampled residuals $u_{t,n}^* = \widehat{u}_{t,n} W_t$, where W_t is a sequence of external variables. Then we generate the bootstrap responses as

$$X_{t,n}^* = F_{t,n}'\widehat{\beta}_N + u_{t,n}^* \qquad t = 1, \ldots, n.$$

Based on the bootstrap sample $(F_{t,n}, X_{t,n}^*)_{t=1}^n$, we obtain the OLS estimator $\widehat{\beta}_N^* = \left(\sum_{t=1}^n F_{t,n}F_{t,n}'\right)^{-1} \left(\sum_{t=1}^n F_{t,n}X_{t,n}^*\right)$. Then the sampling distribution of $N^{1/2}(\widehat{\beta}_N - \beta)$ is approximated by the conditional distribution of $N^{1/2}(\widehat{\beta}_N^* - \widehat{\beta}_N)$ given the data $\mathcal{X}_n$, and $\mathcal{L}_F$ can be approximated by its bootstrap counterpart. We assume

(B1) $\{W_t\}_{t=1}^n$ are iid with $E(W_t) = 0$ and $\text{var}(W_t) = 1$, $\{W_t\}_{t=1}^n$ are independent of the data, and $E(W_1^4) < \infty$.

In practice, we can sample $\{W_t\}_{t=1}^n$ from the standard normal distribution with mean zero or use the distribution recommended by Mammen (1993). Notice that there is no user-determined parameters, which makes the wild bootstrap convenient to implement in practice.

Let $\widehat{\beta}^*_{\lfloor Nr \rfloor} = \left(\sum_{t=1}^{\lfloor Nr \rfloor + p-1} F_{t,n}F'_{t,n}\right)^{-1} \sum_{t=1}^{\lfloor Nr \rfloor + p-1} F_{t,n}X^*_{t,n}$ be the OLS estimate using the first $\lfloor Nr \rfloor + p - 1$ of the bootstrapped sample $(F_{t,n}, X^*_{t,n})_{t=1}^n$. The following theorem states the consistency of the wild bootstrap.

THEOREM **3.3.2**. *Assume (R1)-(R2) and (B1).*

(i) If $\{u_{t,n}\}$ is generated from (A1), we have

$$N^{-1/2}\lfloor Nr \rfloor(\widehat{\beta}^*_{\lfloor Nr \rfloor} - \widehat{\beta}_N) \Rightarrow Q_r^{-1} B_{F,\mathbf{D},\omega}(r) \qquad \textit{in probability,} \tag{3.5}$$

where $B_{F,\mathbf{D},\omega}(r) = \{\mathbf{D}(1)\}^{1/2} \int_0^r \omega(s)dB(s)$ and $\mathbf{D}(1) = \sum_{j=0}^{\infty} c_j^2$.

(ii) If $\{u_{t,n}\}$ is generated from (A2), we have

$$N^{-1/2}\lfloor Nr \rfloor(\widehat{\beta}^*_{\lfloor Nr \rfloor} - \widehat{\beta}_N) \Rightarrow Q_r^{-1} B_{F,\nu}(r) \qquad \textit{in probability,} \tag{3.6}$$

where $B_{F,\nu}(r) = \int_0^r \nu(s)dB(s)$.

(iii) It follows from the continuous mapping theorem that

$$T_N^* = N(\widehat{\beta}^*_N - \widehat{\beta}_N)'\{\Omega^*_N(\epsilon)\}^{-1}(\widehat{\beta}^*_N - \widehat{\beta}_N) \xrightarrow{\mathcal{D}} \mathcal{L}_F$$

in probability, where $\Omega^*_N(\epsilon) = N^{-2}\sum_{t=\lfloor N\epsilon \rfloor}^N t^2(\widehat{\beta}^*_t - \widehat{\beta}^*_N)(\widehat{\beta}^*_t - \widehat{\beta}^*_N)'$ and $\mathcal{L}_F$ as defined in Theorem 3.3.1.

REMARK **3.3.3**. Notice that the wild bootstrap cannot successfully replicate the original distribution without the normalization, unless $\mathbf{C}^2(1) = \mathbf{D}(1)$ or $\Gamma = 1$, because the wild bootstrap cannot capture the temporal dependence of the original error series. However, in the limiting distribution of $N^{-1/2}\lfloor Nr \rfloor(\widehat{\beta}_{\lfloor Nr \rfloor} - \beta)$, the part that reflects the long run dependence, $\mathbf{C}(1)$ or Γ, can be separated from the rest, for both data generating processes (A1) and (A2). This property makes it possible to construct a self-normalized quantity, whose limiting distribution depends only on the heteroscedastic part (captured by the wild bootstrap), not on the temporal dependence part.

REMARK **3.3.4**. Hypothesis testing can be conducted following the argument in Zhou and Shao (2013). Let the null hypothesis be $H_0: \ R\beta = \lambda$, where R is a $k \times p$ ($k \leq p$) matrix with rank k and λ is a $k \times 1$ vector. Under the null hypothesis H_0, we have

$$T_{N,R} = N(\widehat{\beta}_N - \beta)'R'\{R\Omega_N(\epsilon)R'\}^{-1}R(\widehat{\beta}_N - \beta) \xrightarrow{\mathcal{D}} \mathcal{L}_{F,R},$$

where $\mathcal{L}_{F,R} = \widetilde{B}_F(1)'Q_1^{-1}R'\{R\Omega(\epsilon)R'\}^{-1}RQ_1^{-1}\widetilde{B}_F(1)$, and the test can be formed by using the bootstrapped critical values, owing to the fact that

$$T^*_{N,R} = N(\widehat{\beta}^*_N - \widehat{\beta}_N)'R'\{R\Omega^*_N(\epsilon)R'\}^{-1}R(\widehat{\beta}^*_N - \widehat{\beta}_N) \xrightarrow{\mathcal{D}} \mathcal{L}_{F,R}$$

in probability under the assumptions of Theorem 3.3.2. For example, if we let $k = 1$, $\lambda = 0$, and R be a vector of 0's except for the jth element being 1, we can construct a confidence interval or test the significance of individual regression coefficient b_j. Let $\widehat{\beta}_{t,j}$ be the jth element of $\widehat{\beta}_t$. The $100(1-\alpha)\%$ confidence interval for b_j is constructed as $\widehat{\beta}_{N,j} \pm \left\{\mathcal{C}_{F,j,1-\alpha}N^{-3}\sum_{t=\lfloor N\epsilon \rfloor}^{N} t^2(\widehat{\beta}_{t,j} - \widehat{\beta}_{N,j})^2\right\}^{1/2}$, where $\mathcal{C}_{F,j,1-\alpha}$ is the $(1-\alpha)$th quantile of $\{T^{*(i)}_{N,j}\}_{i=1}^{B}$, $T^{*(i)}_{N,j} = N^3(\widehat{\beta}^{*(i)}_{N,j} - \widehat{\beta}_{N,j})^2/\sum_{t=\lfloor N\epsilon \rfloor}^{N} t^2(\widehat{\beta}^{*(i)}_{t,j} - \widehat{\beta}^{*(i)}_{N,j})^2$, $\widehat{\beta}^{*(i)}_{N,j}$ is the jth element of the OLS estimate of β using the ith bootstrapped sample, and B is the number of bootstrap replications.

Remark **3.3.5**. Our method is developed in the same spirit of Xu (2012), who focused on the multivariate trend inference under the assumption that the trend is linear and errors follow an VAR(p) model with heteroscedastic innovations. In particular, Xu's class 4 test used the studentizer first proposed in Kiefer et al. (2000) and the wild bootstrap to capture the heteroscedasticity in the innovations. Since the KVB method and the SN method are closely related (see Rho and Shao (2013) for a detailed comparison), it seems that there are some overlap between our work and Xu (2012). However, our work differs from Xu's in at least three important aspects.

(1) The settings in these two papers are different. Xu (2012) allows for multivariate time series and are interested in the inference of linear trends, whereas ours is restricted to univariate time series. Our assumption on the form of the trend function and the error structure are considerably more general. In particular, we allow for more complex trend functions as long as it is known up to a finite dimensional parameter. In addition, the AR(p) model with heteroscedastic innovation used in Xu (2012) is a special case of our (A1), i.e., linear process with heteroscedastic innovations. Also Xu's method does not seem directly applicable to the modulated stationary process (A2) we assumed for the errors.

(2) The wild bootstrap is applied to different residuals. In Xu's work, he assumed VAR(p) structure for the error process with known p. He needs to estimate the VAR(p) model to obtain the residuals, which are approximately independent but heteroscedastic. His wild bootstrap is applied to the residuals from the VAR(p) model, and is expected to work due to the well-known ability of wild bootstrap to capture heteroscedasticity. By contrast, we apply the wild bootstrap to OLS residuals

from the regression model and our OLS residuals are both heteroscedastic and temporally dependent. The wild bootstrap is not expected to capture/mimic temporal dependence, but since the part that is due to temporal dependence is cancelled out in the limiting distribution of our self-normalized quantity, the wild bootstrap successfully captures the remaining heteroscedasticity and provides a consistent approximation of the limiting distribution. Note that our method works for errors of both types, but it seems that Xu's method would not work for the modulated stationary process as assumed in (A2) without a nontrivial modification.

(3) Xu's class 4 test without prewhitening is in a sense an extension of Kiefer et al. (2000) to the linear trend model, whereas our method is regarded as an extension of SN method to the regression setting. In time series regression framework, Rho and Shao (2013) showed that there is a difference between the SN method and the KVB method; see Remark 3.3.1 therein and also Section 3.2 for a detailed explanation. It is indeed possible to follow the KVB method in our framework and we would expect the wild bootstrap to be consistent to approximate the limiting distribution of the studentized quantity owing to the separate structure of the errors.

To save space, we do not present the details for the KVB studentizer in our setting. Finite sample comparison between our method and Xu's class 3 and 4 tests are provided in Section 3.4.2.

3.4 Simulations

In this section, we compare the finite sample performance of our method to other comparable methods in Zhao (2011) and Xu (2012) for (i) constant mean models, (ii) linear trend models, and (iii) linear trend models with a break in the intercept. For the constant mean models, Zhao's method is applicable only when the errors have periodic heteroscedasticity. On the other hand, Xu's method can only be used for the linear trend models when the errors are from an AR process. In the cases where neither Zhao's nor Xu's methods can be directly applied, we compare our method to the conventional HAC method. Although it has not been rigorously proven that the traditional HAC method works in our framework, we can construct a HAC-type estimator that consistently estimates V in (3.2). Define

$$\widehat{V}_{n,l_n} = n^{-1} \sum_{t_1=1}^{n} \sum_{t_2=(t_1-l_n+1)\wedge 1}^{(t_1+l_n-1)\vee n} k\left(\frac{t_1-t_2}{l_n}\right) \widehat{u}_{t_1,n}\widehat{u}_{t_2,n} F\left(\frac{t_1}{n}\right) F\left(\frac{t_2}{n}\right)',$$

where $k(s)$ is a kernel function that is defined as 0 if $|s| \geq 1$ and 1 if $s = 0$, l_n is the bandwidth parameter, and $\widehat{u}_{t,n}$ is the OLS residual, i.e., $\widehat{u}_{t,n} = X_{t,n} - F'_{t,n}\widehat{\beta}_N$. This HAC-type estimate $\widehat{V}_{n,l_n}$ can be shown to be consistent for appropriately chosen l_n, and the inference is based on

$$T_N^{HAC} = n(\widehat{\beta}_N - \beta)'(\widehat{Q}_N^{-1}\widehat{V}_{n,l_n}\widehat{Q}_N^{-1})^{-1}(\widehat{\beta}_N - \beta), \tag{3.7}$$

where $\widehat{Q}_N = n^{-1}\sum_{t=1}^{n} F_{t,n}F'_{t,n}$, and the χ^2 approximation. In fact, in the linear trend testing context, this statistic is almost identical to the Class 3 test without prewhitening in Xu (2012).

3.4.1 Constant Mean Models

Consider the linear model

$$X_{t,n} = \mu + u_{t,n}, \quad t = 1, \ldots, n, \tag{3.8}$$

where μ is an unknown constant. Without loss of generality, let $\mu = 0$. Note that for this constant mean models our method is completely bandwidth free, since the trimming parameter can be set to be 0, as mentioned in Remark 3.3.1.

We conduct two sets of simulations for different kinds of error processes. The first simulation follows the data generating process in Zhao (2011), where the error process $\{u_{t,n}\}$ is modulated stationary and has periodic heteroscedasticity. Consider the model (A2) with $v(t/n) = \cos(3\pi t/n)$ and

(M1) $\eta_t = \xi_t - E(\xi_t)$, $\xi_t = \theta|\xi_{t-1}| + (1-\theta^2)^{1/2}\epsilon_t$, $|\theta| < 1$,

(M2) $\eta_t = \sum_{j=0}^{\infty} a_j\epsilon_{t-j}$, $a_j = (j+1)^{-\zeta}/10$, $\zeta > 1/2$,

where, ϵ_t are independent standard normal variables in M1. We consider M2 with two different kinds of innovations: M2-1 denotes M2 with standard normal innovations, and M2-2 stands for M2 with t-distributed innovations with $k = 3, 4, 5$ degrees of freedom. We used 10000 replications to get the coverage and interval length with nominal level 95% and sample size $n = 150$. In each replication, 1000 pseudoseries of iid $N(0,1)$ W_t are generated. The coverage rates and interval lengths of Zhao's method (Zhao) and our method (SN) are reported in Table 3.1. Note that our method does not require any user-chosen number in this particular example and m in the table is only for Zhao's method. With our choice of $v(t/n)$, Zhao's "k-block asymptotically equal cumulative variance condition" holds for $m = 25, 50, 75, \ldots$, which means that when $m = 15$ and 30 in M1 and M2-1, the conditions in Zhao (2011) are not satisfied. For the model M2-2, $m = 25$ is used for Zhao's method.

Zhao (2011) provided a comparison with some existing block-based bootstrap methods and showed that his method tends to have more accurate coverage rates. Our method is comparable to Zhao's based on our simulation. For the nonlinear threshold autoregressive models M1, it seems that our method delivers slightly less coverage especially when the dependence is strong and generally slightly wider confidence intervals compared to Zhao's method with correctly specified m. However, for the linear process models M2, our method has better coverage and slightly narrower intervals when ζ is small, i.e., when dependence is too strong so that the short-range dependence condition (A2) (iii) does not hold.

In the second simulation, the errors $u_{t,n}$ are generated from (A1) with an AR(1) model. Let e_t be generated from iid $N(0,1)$. For (A1), $u_{t,n} = \rho u_{t-1,n} + \omega(t/n)e_t$, $t = 1, \ldots, n$, which satisfies the condition (A1) (ii) and (iii), by letting $c_j = \rho^j$. For $\omega(s)$, we consider the single volatility shift model

$$\omega^2(s) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbf{1}(s \geq \tau), \qquad s \in [0,1], \ \tau \in (0,1). \tag{3.9}$$

We let the location of the break $\tau \in \{0.2, 0.8\}$, the initial variance $\sigma_0^2 = 1$, and the ratio $\delta = \sigma_0/\sigma_1 \in \{1/3, 3\}$. This kind of break model and the choice of parameters were used in Cavaliere and Taylor (2007, 2008a,b). Here, sample size $n \in \{100, 250\}$ and the AR coefficient $\rho \in \{0, 0.5, 0.8\}$, and 2000 replications are used to get the coverage percentages and average interval lengths. For the wild bootstrap, $\{W_t\}_{t=1}^n$ are generated from iid $N(0,1)$, and we use 1000 bootstrap replications. We also try the two-point distribution in Mammen (1993) for the bootstrap and get similar simulation results (not reported).

Since this error structure do not show any periodicity, Zhao's method is no longer applicable. We use the HAC-type statistic introduced in (3.7) to make a comparison with our method. However, the performance of the HAC-type quantity is sensitive to the user-chosen parameter l_n, the choice of which is often difficult. For the purpose of our simulation, we choose the oracle optimal l_n, favoring the HAC method. Specifically, since we know the true V, we choose l_n such that $\hat{l} = \text{argmin}_l ||\widehat{V}_{n,l_n} - V||_F^2$, where $|| \cdot ||_F$ stands for the Frobenius norm. Note that this choice of l_n is not feasible in practice. For the choice of the kernel, we use the Bartlett kernel, i.e., $a(s) = (1 - |s|)\mathbf{1}(|s| \leq 1)$, which guarantees the positive-definiteness of $\widehat{V}_{n,l_n}$.

The results are presented in Table 3.2. The columns under the HAC (l) represent the result using the HAC-type quantity with $\hat{l}$, $\hat{l}/2$, and $2\hat{l}$. When $\rho = 0$, the optimal bandwidths chosen for the HAC estimators are $\widehat{l} = 1$, so we do not have any values for the column corresponding to $\hat{l}/2$.

If $\rho = 0$, all methods perform very well, which is expected because the error process is simply iid with some heteroscedasticity. The wild bootstrap without the self-normalization is also consistent because $\mathbf{C}(1) = 1$. However, as the dependence increases, the performance of all methods gets worse. If $\rho \neq 0$, the wild bootstrap without the self-normalization is no longer consistent and thus produce the worst coverage. Of the three types of methods considered (SN-WB, WB, HAC), the SN method with the wild bootstrap delivers the most accurate coverage probability. For the case of moderate dependence $\rho = 0.5$, the SN-WB method still performs quite well, while the HAC-type method shows some undercoverage. When the temporal dependence strengthens, i.e., $\rho = 0.8$, the SN-WB method delivers the best coverage. For the HAC-type methods, note that the optimal bandwidth is chosen to minimize the distance from the true covariance matrix. However, this choice of the bandwidth may not coincide with the bandwidth that produces the best coverage; as can be seen for the cases $(\tau, \delta) = (0.2, 3)$ or $(\tau, \delta) = (0.8, 0.33)$, $2\hat{l}$ tends to produce better coverage than $\hat{l}$. The interval lengths for the SN-WB method are somewhat longer than those of HAC methods, but considering the gain in coverage accuracy, this increase of interval length seems reasonable.

3.4.2 Linear Trend Models

In this subsection, we are interested in inference on the linear trend coefficient from a linear trend model

$$X_{t,n} = b_1 + b_2(t/n) + u_{t,n}, \qquad t = 1, \ldots, n. \tag{3.10}$$

Without loss of generality, let $b_1 = 0$ and $b_2 = 5$. The error process $\{u_{t,n}\}$ follows the same setting as in the second simulation of Section 3.4.1, except that $\{u_{t,n}\}$ is either generated from AR(1) or MA(1), i.e.,

$$\begin{aligned} AR(1): &\qquad u_{t,n} = \rho u_{t-1,n} + \varepsilon_{t,n}, \\ MA(1): &\qquad u_{t,n} = \theta \varepsilon_{t-1,n} + \varepsilon_{t,n}, \end{aligned}$$

where $\varepsilon_{t,n} = \omega(t/n)e_t$, e_t iid $N(0,1)$, and $\omega(s)$ as in (3.9).

For this setting, Xu (2012)'s methods can be used. In fact, Xu's Class 3 test is almost identical to the HAC-type statistic in (3.7). Xu's Class 4 test is a fixed-b version of his Class 3 test, and it is closely related to our method, as mentioned in Remark 3.3.5. For this reason, we compare our method with Xu's Class 3 and 4 tests. In addition to his original statistics, Xu further applied the prewhitening. In our simulation, we examine finite sample performance of the following methods;

"SN-WB(ϵ)	Our method with various values for the trimming parameter ϵ,
"WB"	Wild bootstrap without self-normalization,
"F3"	Xu's Class 3 test without prewhitening with χ^2 limiting distribution,
"PW3"	Xu's prewhitened Class 3 test with χ^2 limiting distribution,
"iid"	Xu's prewhitened Class 3 test with the iid bootstrap,
"WB"	Xu's prewhitened Class 3 test with the wild bootstrap,
"F4"	Xu's Class 4 test without prewhitening with the wild bootstrap,
"PW4"	Xu's prewhitened Class 4 test with the wild bootstrap.

For Xu's Class 3 tests we need to choose the truncation parameter. Following Xu, we used Andrews (1991)'s bandwidth selection for AR(1) or MA(1) model with Bartlett kernel. In the implementation of the wild bootstrap for Xu's method, we assumed the true order of the AR structure of the error process is known.

Remark **3.4.1**. It should be noted that a direct comparison of our method and Xu's methods is not fair. We gave some advantages to Xu's methods in two aspects. The first aspect is that we assume the knowledge of the true error model, VAR(1), following Xu (2012), which may be unrealistic in practice. Xu's methods can take advantage of this knowledge of the true model; one advantage is their bootstrap samples are generated following the correctly-specified data generating process, and the other advantage is that his Class 3 tests can be based on the most (theoretically) efficient choice of the bandwidth parameters. In contrast, our method does not rely on parametric assumption on the error structure, so knowing the true data generating process does not provide useful information in the implementation of our method. The second aspect is that some of Xu's methods use another layer of correction, prewhitening. The prewhitening can be highly effective when the true order is known. The only tests that our method is directly comparable with are Xu's original Class 3 test (F3) and Class 4 test (F4), for which the advantage from knowing the true model may still exist.

Table 3.3 presents the finite sample coverage rates of 95% confidence intervals of the linear trend coefficient b_2 of (3.10), along with the mean interval lengths in parentheses, for the AR(1) models. Comparing our method with the HAC method (Xu's Class 3 test without prewhitening), our method always have more accurate coverage rates regardless of the choice of the trimming parameter ϵ. Not surprisingly, prewhitening is very effective in bringing the coverage level closer to the nominal level. Xu's Class 4 test seems to provide the best coverage rates, for both prewhitened and nonprewhitened versions.

However, Xu's methods and his prewhitening are based on the assumption that the errors are from an AR model with a known lag. When the errors are not from an AR model, then the performance may deteriorate. As can be seen from Table 3.4, where the errors are generated from an MA model, in some cases Xu's tests, except for his test 3, tend to provide over-coverage, especially when $\delta = 3$. On the other hand, our SN-WB method tends to provide more stable results. Furthermore, in our unreported simulations, when the order of the VAR model is misspecified, Xu's methods can be sensitive to the form of misspecification. Thus when the true model of the error process is not VAR or true order of the VAR model is not known, Xu's methods may not be advantageous.

In this linear trend assessment setting, we need to use $\epsilon > 0$ so that our method is theoretically valid. As can be seen in Tables 3.3, if the dependence is weak or moderate ($\rho = 0$ or 0.5), the choice of ϵ does not affect the coverage accuracy much, and $\epsilon = 0.1$ or above seems appropriate. However, the effect of ϵ is more apparent when the dependence is stronger, i.e., when $\rho = 0.8$. It seems that $\epsilon \in \{0.2, 0.3, 0.4\}$ are all reasonable choices here, considering the amount of improvement in coverage percentage. Notice that even when our choice of ϵ is not optimal, the SN-WB method delivers more accurate coverage than the HAC method. In terms of the interval length, there is a mild monotonic relationship between ϵ and interval length. The larger ϵ is, the longer the interval tends to become. As a rule of thumb in practice, we can use $\epsilon \in [0.2, 0.4]$ if the temporal dependence in the data is likely to be strong, otherwise, we can also use $\epsilon = 0.1$, which was also recommended by Zhou and Shao (2013) under a different setting.

3.4.3 Linear trend models with a break in intercept

This section concerns the linear trend models with a break in its mean level at a known point,

$$X_{t,n} = b_1 + b_2\mathbf{1}(t/n > s_b) + b_3(t/n) + u_{t,n}, \quad t = 1, \ldots, n, \tag{3.11}$$

where $s_b \in (0, 1)$ indicates the relative location for the intercept break, which is assumed to be known. Without loss of generality, let $s_b = 0.3$, $b_1 = 0$, $b_2 = 2$, and $b_3 = 5$, which was modified from the model we fit in the data illustration in Section 3.5. Our interest is on constructing confidence intervals for the the break parameter b_2. For the error process $\{u_{t,n}\}$, we use the AR(1) model, as introduced in the second simulation of Section 3.4.1. Because of the general form of the trend function, neither Xu's nor Zhao's methods work in this setting. Here we compare the HAC-type statistic defined in (3.7) with our SN-WB(ϵ) method, and the wild bootstrap alone. For the HAC-type method, the bandwidth parameter is chosen following Andrews (1991)'s suggestion for the AR(1) model, so in a sense, this HAC-type method takes some advantage of the knowledge of the true model.

Table 3.5 presents the empirical coverage rates for 95% confidence intervals of b_2. The mean interval lengths are in parentheses. When there is no temporal dependence, i.e., $\rho = 0$, all methods work well. When there is moderate ($\rho = 0.5$) or strong ($\rho = 0.8$) dependence, the wild bootstrap is no longer consistent. In this case, the HAC method with Andrews' bandwidth parameter choice also shows some departures from the nominal level, converging to its limit much slower than our method. In this setting, our SN-WB method provides the most accurate coverage rates. Even the worst coverage of our SN-WB method for a range of ϵ's under examination seems to be more accurate than the HAC method in all simulation settings. Notice that this kind of linear trend model with a break in the intercept, the trimming parameter ϵ in our method should be chosen so that $\epsilon > s_b$. See Remark 3.3.2 for more details. In this simulation, since $s_b = 0.3$, we choose ϵ to be 0.35, 0.4, 0.5, 0.6, and 0.7. The results does not seem to be sensitive to the choice of this trimming parameter.

3.5 Data Analysis

In this section, we illustrate our method using the logarithm of the U.S. nominal wage data, which was originally from Nelson and Plosser (1982) for the period 1900-1970 and has been intensively analyzed in econometrics literature, including Perron (1989). The latter author argued that this series has a fairly stable linear trend with a sudden decrease in level between 1929 and 1930 (the Great Crash). In this section we examine his assertion for an extended dataset for the period 1900-1988, provided by Koop and Steel (1994). The left panel of Figure 3.1 presents the data fitted with the linear trend model with one break in the intercept

$$X_{t,n} = b_1 + b_2 \mathbf{1}(t > 30) + b_3(t/n) + u_{t,n}, \qquad t = 1, \ldots, n = 89. \tag{3.12}$$

Here the second regressor $\mathbf{1}(t > 30)$ represents the break at year 1929. From the residual plot on the right panel of Figure 3.1, the error appears to exhibit some degree of heteroscedasticity. Our inference method, which is robust to this kind of heteroscedasticity and weak temporal dependence, is expected to be more appropriate than other existing methods, the validity of which hinges on the stationarity of the errors. Table 3.6 reports the 100(1-α)% confidence intervals with $\alpha = 0.01, 0.05, 0.1$, for the three coefficients using our method. For the wild bootstrap, we used 1000 pseudo series $\{W_t\}_{t=1}^n$ drawn from iid $N(0,1)$. For the choice of the trimming parameter ϵ, We use $\epsilon = 0.4, 0.5, 0.6$. Note that since there is a break in the regressor at $30/89 = 0.34$, ϵ cannot be less than 0.34, see Remark 3.3.2. As can be seen in Table 3.6, the resulting confidence intervals are not very sensitive to the choice of the trimming parameter ϵ. In fact, the confidence intervals unanimously suggest that the linear trend is significantly positive, i.e., the nominal wage is linearly increasing at significance level 1% with a 99% confidence interval of about [3.6,6], and the decrease at the Great Crash is significant at 5% level but not at 1% level.

3.6 Conclusion

In this chapter we study the estimation and inference of nonstationary time series regression models by extending the SN method (Lobato, 2001; Shao, 2010b) to the regression setting with fixed parametric regressors and nonstationary errors with unconditional heteroscedasticity. Due to the heteroscedasticity and non-constant regressor, the limiting distribution of the SN quantity is no longer pivotal and contains a number of nuisance parameters. To approximate the limiting distribution, we apply the wild bootstrap (Wu, 1986) and rigorously justify its consistency. Simulation comparison demonstrates the advantage of our method in comparison with HAC-based alternatives. The necessity of self-normalization is also demonstrated in theory and finite samples.

The two kinds of nonstationary models we consider in this article are natural generalizations of the classical linear process and the stationary causal process to capture unconditional heteroscedasticity often seen on the data. However, it is restricted only to the short-range dependence as seen from our condition (A1) (ii) and (A2) (iii). When there is long-range dependence, it is not clear if the SN method is still applicable. This can be an interesting topic for future research. Furthermore, the OLS is a convenient but not an efficient estimator of the regression parameter when the errors exhibit autocorrelation and heteroscedasticity. How to come up with a more efficient estimator and perform the SN-based inference is worthy of a careful investigation. Finally, our method assumes a parametric form for the mean function so it is easy to interpret and the prediction is straightforward. However, a potential drawback is that the result may be biased if the parametric trend function is misspecified. We can avoid this misspecification problem by pre-applying Zhang and Wu (2011)'s test to see if a parametric form fits the data well. The inferential procedure developed here can be readily used provided that a suitable parametric form is specified.

3.7 Tables and Figures

Table 3.1: Percentages of coverage of the 95% confidence intervals (mean lengths of the interval in the parentheses) of the mean μ in (3.8), where μ is set to be zero. The error processes are from Zhao (2011)'s setting with periodic heteroscedasticity. The method in Zhao (2011) and our method are compared. The interval lengths of the confidence intervals are in the parentheses. The number of replications is 10,000, and 1,000 pseudoseries with iid $N(0,1)$ are used for the wild bootstrap.

M1	m=15				m=25				m=30			
θ	0	0.2	0.5	0.8	0	0.2	0.5	0.8	0	0.2	0.5	0.8
SN					95.0	94.7	94.9	93.7				
					(0.29)	(0.29)	(0.32)	(0.40)				
Zhao	95.2	95.3	95.7	94.6	95.2	94.8	95.1	94.7	95.2	95.0	95.3	95.5
	(0.25)	(0.26)	(0.28)	(0.34)	(0.28)	(0.29)	(0.31)	(0.40)	(0.30)	(0.30)	(0.33)	(0.42)
M2-1	m=15				m=25				m=30			
ζ	2.5	2.01	1.01	0.7	2.5	2.01	1.01	0.7	2.5	2.01	1.01	0.7
SN					94.5	93.9	94.2	95.7				
					(0.04)	(0.04)	(0.09)	(0.15)				
Zhao	95.2	94.3	94.1	95.5	94.7	94.2	95.9	97.2	95.1	95.2	96.4	97.8
	(0.03)	(0.04)	(0.07)	(0.13)	(0.04)	(0.04)	(0.09)	(0.17)	(0.04)	(0.05)	(0.10)	(0.19)
M2-2	k=5				k=4				k=3			
ζ	2.5	2.01	1.01	0.7	2.5	2.01	1.01	0.7	2.5	2.01	1.01	0.7
SN	94.3	94.0	94.0	96.1	94.6	93.9	94.1	96.2	94.2	93.4	94.2	96.0
	(0.048)	(0.06)	(0.11)	(0.19)	(0.05)	(0.06)	(0.12)	(0.20)	(0.06)	(0.07)	(0.14)	(0.24)
Zhao	95.0	94.9	95.6	97.3	95.1	94.5	95.9	97.4	94.8	94.7	95.7	97.6
	(0.05)	(0.06)	(0.12)	(0.22)	(0.05)	(0.06)	(0.13)	(0.24)	(0.06)	(0.07)	(0.15)	(0.29)

Table 3.2: Percentages of coverage of the 95% confidence intervals (mean lengths of the interval in the parentheses) of the mean μ in (3.8), where μ is set to be zero. The error processes exhibit nonperiodic heteroscedasticity; (A1)-AR(1) with single volatility shift (3.9). The number of replications is 2000, and 1000 pseudoseries with iid $N(0, 1)$ are used for the wild bootstrap.

Model				SN-WB	WB	HAC (l)			$\hat{l}$
n	τ	δ	ρ			$\hat{l}$	$\hat{l}/2$	$2\hat{l}$	
100	0.2	0.33	0	94.5 (1.3)	94.5(1.1)	94.8(1.1)	-	94.5(1.1)	1
100	0.2	3	0	94.8 (0.2)	94.8(0.2)	94.9(0.2)	-	94.8(0.2)	1
100	0.8	0.33	0	94.8 (0.7)	93.5(0.6)	93.6(0.6)	-	93.9(0.6)	1
100	0.8	3	0	94.3 (0.4)	94.8(0.4)	95.0(0.4)	-	94.5(0.3)	1
250	0.2	0.33	0	95.2 (0.8)	94.7(0.7)	94.8(0.7)	-	94.8(0.7)	1
250	0.2	3	0	95.0 (0.1)	94.8(0.1)	94.8(0.1)	-	95.0(0.1)	1
250	0.8	0.33	0	95.0 (0.4)	94.9(0.4)	94.9(0.4)	-	95.0(0.4)	1
250	0.8	3	0	94.3 (0.3)	95.0(0.2)	95.2(0.2)	-	94.8(0.2)	1
100	0.2	0.33	0.5	93.2 (2.4)	70.8(1.2)	87.5(1.8)	84.5(1.6)	87.4(1.8)	6
100	0.2	3	0.5	94.8 (0.5)	73.6(0.2)	89.0(0.3)	83.2(0.3)	90.5(0.4)	5
100	0.8	0.33	0.5	93.7 (1.4)	72.5(0.7)	88.8(1.0)	81.7(0.9)	90.0(1.0)	5
100	0.8	3	0.5	93.5 (0.8)	71.9(0.4)	88.7(0.6)	85.1(0.5)	88.4(0.6)	6
250	0.2	0.33	0.5	94.7 (1.6)	74.2(0.8)	91.6(1.2)	88.7(1.1)	91.5(1.2)	8
250	0.2	3	0.5	95.0 (0.3)	73.5(0.2)	91.1(0.2)	86.1(0.2)	92.3(0.2)	6
250	0.8	0.33	0.5	94.3 (0.9)	74.0(0.5)	91.6(0.7)	87.6(0.6)	93.3(0.7)	6
250	0.8	3	0.5	93.5 (0.5)	73.5(0.3)	91.6(0.4)	89.0(0.4)	91.8(0.4)	8
100	0.2	0.33	0.8	90.8 (5.4)	45.1(1.7)	80.5(3.7)	75.1(3.2)	79.1(3.7)	13
100	0.2	3	0.8	92.2 (1.1)	45.6(0.3)	82.6(0.7)	74.7(0.6)	84.1(0.8)	11
100	0.8	0.33	0.8	90.8 (3.0)	47.9(1.0)	83.8(2.1)	76.0(1.7)	84.2(2.1)	11
100	0.8	3	0.8	89.8 (1.9)	45.6(0.6)	80.2(1.3)	75.1(1.1)	79.0(1.3)	13
250	0.2	0.33	0.8	93.2 (3.7)	46.9(1.1)	87.3(2.7)	80.8(2.3)	87.9(2.8)	15
250	0.2	3	0.8	93.8 (0.7)	48.8(0.2)	87.4(0.5)	80.6(0.4)	90.0(0.5)	15
250	0.8	0.33	0.8	93.0 (2.1)	47.7(0.6)	89.3(1.5)	82.0(1.3)	90.5(1.6)	15
250	0.8	3	0.8	92.5 (1.3)	47.9(0.4)	87.0(0.9)	80.8(0.8)	87.6(0.9)	15

Table 3.3: Percentages of coverage of the 95% confidence intervals (mean lengths of the interval in the parentheses) of the linear trend model with $(b_1, b_2) = (0, 5)$, (A1)-AR(1), single volatility shifts. The number of replications is 2000, and 1000 pseudoseries with iid $N(0, 1)$ are used for the wild bootstrap.

Model				SN-WB (ϵ)						WB	Xu-Class3				Xu-Class4	
n	τ	δ	ρ	0	0.1	0.2	0.3	0.5	0.7		F3	PW3	iid	WB	F4	PW4
100	0.2	0.33	0	94.8 (3.4)	94.7 (3.4)	94.7 (3.5)	95.0 (3.7)	94.9 (4.2)	95.3 (4.6)	94.0 (3.2)	94.2 (3.2)	95.2 (3.4)	95.6 (3.4)	94.8 (3.3)	94.5 (3.9)	94.9 (4.0)
100	0.2	3	0	96.5 (1.0)	96.2 (1.0)	95.2 (1.1)	95.2 (1.1)	95.5 (1.2)	95.7 (1.2)	94.3 (0.9)	94.0 (0.9)	94.8 (0.9)	94.1 (0.9)	94.0 (0.9)	96.0 (1.0)	96.5 (1.1)
100	0.8	0.33	0	95.0 (2.7)	95.0 (2.7)	94.8 (2.8)	95.2 (2.8)	95.1 (3.0)	96.0 (3.0)	93.7 (2.7)	92.7 (2.6)	94.0 (2.8)	94.2 (2.8)	94.5 (2.8)	94.8 (2.9)	95.3 (3.0)
100	0.8	3	0	96.0 (1.2)	94.3 (1.3)	94.0 (1.3)	93.9 (1.4)	94.8 (1.5)	95.5 (1.6)	94.8 (1.1)	94.3 (1.1)	96.0 (1.1)	95.0 (1.1)	94.1 (1.1)	94.0 (1.3)	94.8 (1.3)
250	0.2	0.33	0	95.4 (2.1)	94.8 (2.2)	95.0 (2.2)	94.8 (2.3)	95.2 (2.6)	95.2 (2.9)	95.0 (2.1)	94.7 (2.1)	95.3 (2.1)	95.5 (2.1)	95.2 (2.1)	94.5 (2.5)	94.7 (2.5)
250	0.2	3	0	96.3 (0.6)	94.9 (0.6)	95.3 (0.7)	95.0 (0.7)	95.2 (0.7)	95.3 (0.8)	95.0 (0.6)	94.7 (0.6)	95.3 (0.6)	94.5 (0.6)	94.8 (0.6)	95.8 (0.6)	96.0 (0.6)
250	0.8	0.33	0	95.9 (1.7)	94.5 (1.7)	94.2 (1.8)	94.5 (1.8)	94.8 (1.8)	94.8 (1.9)	95.0 (1.7)	94.0 (1.7)	94.8 (1.8)	95.0 (1.8)	95.4 (1.8)	94.2 (1.8)	94.8 (1.8)
250	0.8	3	0	96.0 (0.8)	95.0 (0.8)	95.8 (0.8)	95.5 (0.9)	95.9 (0.9)	95.6 (1.0)	95.6 (0.7)	95.0 (0.7)	95.3 (0.7)	94.8 (0.7)	94.9 (0.7)	95.7 (0.8)	96.2 (0.8)
100	0.2	0.33	0.5	91.8 (5.9)	92.9 (6.3)	93.4 (6.6)	94.0 (7.0)	93.7 (7.9)	93.7 (8.6)	73.7 (3.6)	87.6 (5.1)	93.8 (6.6)	95.1 (6.9)	94.5 (6.7)	94.5 (8.0)	94.7 (8.1)
100	0.2	3	0.5	91.0 (1.6)	93.5 (1.9)	94.3 (2.1)	94.2 (2.1)	94.6 (2.2)	93.9 (2.3)	71.7 (1.0)	84.2 (1.3)	95.0 (1.8)	94.2 (1.8)	94.2 (1.8)	95.9 (2.1)	96.6 (2.1)
100	0.8	0.33	0.5	90.9 (4.5)	92.7 (4.9)	93.3 (5.2)	93.3 (5.3)	93.8 (5.6)	94.3 (5.7)	72.5 (3.0)	84.2 (3.9)	92.2 (5.4)	94.7 (5.6)	94.5 (5.6)	94.9 (5.9)	95.5 (6.4)
100	0.8	3	0.5	90.5 (2.0)	92.5 (2.3)	92.8 (2.5)	93.1 (2.6)	93.7 (2.8)	94.5 (3.0)	74.0 (1.2)	87.4 (1.7)	95.2 (2.2)	94.7 (2.3)	94.0 (2.2)	94.0 (2.7)	94.9 (2.6)
250	0.2	0.33	0.5	93.5 (3.8)	94.0 (4.2)	93.9 (4.3)	94.2 (4.5)	94.5 (5.0)	94.8 (5.6)	75.3 (2.4)	91.1 (3.6)	95.2 (4.2)	95.5 (4.3)	95.3 (4.2)	94.2 (4.9)	95.0 (5.0)
250	0.2	3	0.5	92.4 (1.1)	94.4 (1.3)	95.7 (1.4)	94.5 (1.4)	94.8 (1.4)	95.0 (1.5)	73.2 (0.6)	89.3 (1.0)	95.6 (1.2)	94.7 (1.1)	94.8 (1.1)	95.8 (1.3)	96.2 (1.3)
250	0.8	0.33	0.5	92.8 (3.0)	93.6 (3.3)	93.8 (3.4)	93.7 (3.5)	94.0 (3.6)	94.2 (3.6)	74.9 (2.0)	89.2 (2.9)	94.2 (3.5)	94.9 (3.5)	94.8 (3.5)	93.8 (3.7)	94.7 (3.8)
250	0.8	3	0.5	93.2 (1.3)	94.7 (1.5)	95.2 (1.6)	95.1 (1.7)	95.5 (1.8)	95.0 (1.9)	74.3 (0.8)	91.2 (1.2)	95.0 (1.4)	95.0 (1.4)	94.9 (1.4)	95.7 (1.7)	96.4 (1.6)
100	0.2	0.33	0.8	83.4 (11.5)	88.3 (13.1)	90.1 (14.3)	91.6 (15.7)	92.0 (17.8)	90.6 (18.8)	48.1 (4.9)	77.4 (9.8)	90.5 (15.4)	94.0 (17.7)	92.9 (16.9)	94.0 (19.9)	94.2 (21.7)
100	0.2	3	0.8	76.6 (2.7)	85.5 (3.7)	89.7 (4.4)	91.2 (4.7)	90.8 (5.0)	89.5 (5.1)	42.6 (1.3)	69.0 (2.4)	93.9 (12.2)	91.8 (4.3)	91.6 (4.4)	93.6 (5.3)	95.0 (5.4)
100	0.8	0.33	0.8	80.7 (8.1)	87.2 (9.7)	89.2 (10.7)	90.0 (11.3)	90.5 (12.0)	90.6 (12.2)	47.1 (3.8)	70.2 (6.6)	88.6 (13.1)	94.8 (14.1)	94.3 (13.4)	93.7 (13.8)	95.3 (18.3)
100	0.8	3	0.8	80.4 (3.8)	88.0 (4.7)	89.6 (5.4)	91.1 (5.8)	90.7 (6.5)	90.9 (6.9)	44.8 (1.7)	76.3 (3.4)	92.7 (5.5)	92.9 (5.7)	92.4 (5.5)	93.1 (6.8)	94.0 (6.6)
250	0.2	0.33	0.8	88.0 (8.4)	91.6 (9.6)	92.6 (10.1)	93.1 (10.7)	93.1 (12.0)	92.3 (13.1)	49.2 (3.3)	85.9 (7.7)	93.3 (10.1)	95.2 (10.9)	95.0 (10.6)	94.8 (12.4)	95.5 (12.7)
250	0.2	3	0.8	83.7 (2.0)	92.5 (2.8)	94.2 (3.2)	94.1 (3.3)	94.3 (3.4)	93.3 (3.6)	46.6 (0.9)	81.7 (2.0)	95.6 (2.9)	94.2 (2.8)	94.3 (2.9)	95.6 (3.2)	96.1 (3.3)
250	0.8	0.33	0.8	85.5 (6.1)	91.6 (7.4)	92.4 (7.8)	92.2 (8.1)	92.5 (8.3)	92.8 (8.5)	47.9 (2.7)	82.0 (5.8)	91.8 (8.3)	94.8 (8.9)	94.8 (8.8)	93.7 (9.1)	94.5 (10.2)
250	0.8	3	0.8	87.3 (2.7)	93.2 (3.5)	93.7 (3.8)	94.1 (4.0)	94.3 (4.3)	94.1 (4.6)	48.4 (1.1)	86.2 (2.6)	94.9 (3.5)	94.6 (3.5)	94.8 (3.5)	95.3 (4.2)	96.2 (4.1)

Table 3.4: Percentages of coverage of the 95% confidence intervals (mean lengths of the interval in the parentheses) of the linear trend model with $(b_1, b_2) = (0, 5)$, (A1)-MA(1), single volatility shifts. The number of replications is 2000, and 1000 pseudoseries with iid $N(0,1)$ are used for the wild bootstrap.

Model				SN-WB (ϵ)						WB	Xu-Class3				Xu-Class4	
n	τ	δ	θ	0	0.1	0.2	0.3	0.5	0.7		F3	PW3	iid	WB	F4	PW4
100	0.2	0.33	0	94.8 (3.4)	94.7 (3.4)	94.7 (3.5)	95.0 (3.7)	94.9 (4.2)	95.3 (4.6)	94.0 (3.2)	94.2 (3.2)	95.2 (3.4)	95.3 (3.5)	95.0 (3.4)	94.5 (3.9)	95.0 (4.0)
100	0.2	3	0	96.5 (1.0)	96.2 (1.0)	95.2 (1.1)	95.2 (1.1)	95.5 (1.2)	95.7 (1.2)	94.3 (0.9)	94.0 (0.9)	94.8 (0.9)	95.2 (0.9)	94.7 (0.9)	96.1 (1.0)	97.2 (1.1)
100	0.8	0.33	0	95.0 (2.7)	95.0 (2.7)	94.8 (2.8)	95.2 (2.8)	95.1 (3.0)	96.0 (3.0)	93.7 (2.7)	92.7 (2.6)	94.0 (2.8)	94.2 (2.8)	94.6 (2.9)	94.7 (2.9)	95.3 (3.1)
100	0.8	3	0	96.0 (1.2)	94.3 (1.3)	94.0 (1.3)	93.9 (1.4)	94.8 (1.5)	95.5 (1.6)	94.8 (1.1)	94.3 (1.1)	96.0 (1.1)	96.0 (1.1)	95.1 (1.1)	94.0 (1.3)	95.0 (1.3)
250	0.2	0.33	0	95.4 (2.1)	94.8 (2.2)	95.0 (2.2)	94.8 (2.3)	95.2 (2.6)	95.2 (2.9)	95.0 (2.1)	94.7 (2.1)	95.3 (2.1)	95.5 (2.1)	95.4 (2.1)	94.5 (2.5)	94.8 (2.5)
250	0.2	3	0	96.3 (0.6)	94.9 (0.6)	95.3 (0.7)	95.0 (0.7)	95.2 (0.7)	95.3 (0.8)	95.0 (0.6)	94.7 (0.6)	95.3 (0.6)	95.3 (0.6)	95.1 (0.6)	95.8 (0.6)	96.1 (0.6)
250	0.8	0.33	0	95.9 (1.7)	94.5 (1.7)	94.2 (1.8)	94.5 (1.8)	94.8 (1.8)	94.8 (1.9)	95.0 (1.7)	94.0 (1.7)	94.8 (1.8)	94.8 (1.8)	95.2 (1.8)	94.2 (1.8)	94.3 (1.9)
250	0.8	3	0	96.0 (0.8)	95.0 (0.8)	95.8 (0.8)	95.5 (0.9)	95.9 (0.9)	95.6 (1.0)	95.6 (0.7)	95.0 (0.7)	95.3 (0.7)	95.2 (0.7)	95.0 (0.7)	95.8 (0.8)	96.0 (0.8)
100	0.2	0.33	0.5	93.2 (4.8)	94.0 (5.0)	94.2 (5.2)	94.5 (5.4)	94.5 (6.1)	94.7 (6.7)	83.9 (3.6)	91.4 (4.3)	96.9 (5.8)	97.6 (6.1)	97.1 (6.0)	95.2 (6.1)	95.5 (6.5)
100	0.2	3	0.5	94.8 (1.4)	95.0 (1.5)	95.2 (1.6)	94.5 (1.7)	95.0 (1.7)	94.9 (1.8)	84.2 (1.0)	89.4 (1.1)	97.6 (1.6)	98.4 (1.6)	97.8 (1.5)	96.6 (1.6)	97.9 (1.7)
100	0.8	0.33	0.5	93.8 (3.7)	94.5 (4.0)	94.2 (4.1)	94.3 (4.2)	94.8 (4.4)	95.2 (4.5)	83.5 (2.9)	89.0 (3.5)	96.2 (4.7)	96.3 (5.0)	96.7 (5.2)	95.2 (4.5)	96.3 (5.2)
100	0.8	3	0.5	93.8 (1.7)	93.8 (1.8)	93.2 (1.9)	93.4 (2.0)	94.2 (2.2)	95.2 (2.3)	85.2 (1.2)	90.6 (1.4)	97.7 (1.9)	98.4 (2.0)	97.7 (1.9)	94.2 (2.1)	96.0 (2.1)
250	0.2	0.33	0.5	94.5 (3.0)	94.6 (3.2)	94.6 (3.3)	94.8 (3.4)	95.0 (3.8)	94.9 (4.3)	85.7 (2.3)	93.5 (2.9)	97.5 (3.6)	97.8 (3.7)	98.0 (3.7)	94.7 (3.7)	95.5 (3.8)
250	0.2	3	0.5	95.2 (0.9)	95.0 (1.0)	96.1 (1.0)	95.2 (1.1)	94.7 (1.1)	95.5 (1.2)	85.5 (0.6)	92.0 (0.8)	97.8 (1.0)	97.8 (1.0)	97.5 (1.0)	96.1 (0.9)	96.8 (1.0)
250	0.8	0.33	0.5	94.8 (2.4)	94.0 (2.6)	94.0 (2.6)	93.8 (2.6)	94.7 (2.7)	94.8 (2.8)	85.4 (1.9)	92.1 (2.4)	96.8 (3.0)	97.2 (3.1)	97.5 (3.1)	94.2 (2.8)	94.7 (2.9)
250	0.8	3	0.5	94.8 (1.1)	95.0 (1.2)	95.5 (1.2)	95.2 (1.3)	95.8 (1.3)	95.5 (1.4)	85.2 (0.8)	93.0 (1.0)	97.8 (1.2)	98.2 (1.2)	97.5 (1.2)	95.9 (1.2)	96.5 (1.3)
100	0.2	0.33	0.8	93.3 (5.7)	94.3 (6.0)	94.2 (6.2)	94.3 (6.5)	94.8 (7.3)	94.8 (8.0)	82.0 (4.1)	90.8 (5.1)	97.6 (7.4)	98.4 (7.9)	98.0 (7.8)	95.2 (7.4)	95.9 (8.0)
100	0.2	3	0.8	94.5 (1.6)	94.7 (1.8)	95.1 (1.9)	94.6 (2.0)	94.7 (2.1)	95.0 (2.2)	81.6 (1.1)	88.4 (1.3)	98.3 (2.0)	98.8 (2.2)	98.2 (2.0)	97.2 (2.0)	98.2 (2.2)
100	0.8	0.33	0.8	93.5 (4.4)	94.2 (4.7)	94.1 (4.8)	94.2 (5.0)	94.5 (5.2)	95.3 (5.3)	82.1 (3.3)	87.8 (4.0)	96.7 (6.1)	97.5 (6.6)	98.1 (6.8)	95.4 (5.5)	96.5 (6.8)
100	0.8	3	0.8	93.2 (2.0)	94.0 (2.2)	93.4 (2.3)	93.4 (2.4)	94.2 (2.6)	95.1 (2.8)	83.2 (1.4)	90.3 (1.7)	98.4 (2.4)	99.0 (2.6)	98.4 (2.4)	94.8 (2.5)	96.5 (2.6)
250	0.2	0.33	0.8	94.7 (3.6)	94.7 (3.8)	94.3 (3.9)	94.8 (4.1)	94.8 (4.6)	95.0 (5.1)	83.5 (2.6)	93.3 (3.5)	98.4 (4.6)	98.4 (4.7)	98.4 (4.7)	94.8 (4.5)	95.8 (4.7)
250	0.2	3	0.8	95.0 (1.0)	95.3 (1.2)	96.0 (1.2)	95.0 (1.3)	94.8 (1.3)	95.2 (1.4)	83.4 (0.7)	91.7 (0.9)	98.2 (1.3)	98.4 (1.3)	98.2 (1.3)	96.1 (1.1)	97.2 (1.2)
250	0.8	0.33	0.8	94.3 (2.9)	94.2 (3.1)	93.8 (3.1)	94.0 (3.2)	94.3 (3.3)	95.0 (3.3)	83.4 (2.2)	91.6 (2.8)	97.8 (3.8)	98.0 (4.0)	98.0 (4.0)	94.4 (3.4)	95.8 (3.7)
250	0.8	3	0.8	94.7 (1.3)	95.0 (1.4)	95.5 (1.5)	95.1 (1.5)	95.7 (1.6)	95.3 (1.7)	83.0 (0.9)	93.0 (1.1)	98.6 (1.5)	98.7 (1.6)	98.7 (1.5)	96.2 (1.5)	96.9 (1.5)

Table 3.5: Percentages of coverage of the 95% confidence intervals of b_2 (mean lengths of the interval in the parentheses) of the linear trend model with $(b_1, b_2, b_3) = (0, 2, 5)$, (A1)-AR(1), single volatility shifts. The number of replications is 2000, and 1000 pseudoseries with iid $N(0, 1)$ are used for the wild bootstrap.

Model				SN-WB (ϵ)					WB	HAC
n	τ	δ	ρ	0.35	0.4	0.5	0.6	0.7		
100	0.2	0.33	0	94.5 (5.5)	94.5 (5.5)	94.3 (5.6)	94.8 (5.7)	94.8 (5.9)	94.0 (3.9)	93.0 (3.8)
100	0.2	3	0	94.6 (1.1)	94.7 (1.1)	95.0 (1.1)	95.0 (1.0)	95.4 (1.0)	94.4 (0.6)	94.5 (0.6)
100	0.8	0.33	0	95.0 (2.4)	95.0 (2.4)	95.1 (2.4)	95.3 (2.5)	95.5 (2.5)	94.5 (1.9)	93.6 (1.9)
100	0.8	3	0	94.2 (2.0)	94.2 (2.0)	94.3 (2.0)	94.7 (2.0)	94.7 (2.2)	93.3 (1.3)	93.2 (1.3)
250	0.2	0.33	0	95.6 (3.6)	95.6 (3.6)	95.5 (3.6)	95.5 (3.6)	95.1 (3.7)	94.2 (2.5)	94.2 (2.5)
250	0.2	3	0	95.0 (0.7)	95.0 (0.7)	95.0 (0.7)	94.8 (0.6)	94.8 (0.6)	94.5 (0.4)	94.6 (0.4)
250	0.8	0.33	0	95.8 (1.5)	95.8 (1.5)	95.8 (1.5)	95.9 (1.5)	95.8 (1.5)	94.3 (1.2)	94.0 (1.2)
250	0.8	3	0	95.0 (1.3)	95.0 (1.3)	94.8 (1.3)	95.1 (1.3)	95.6 (1.4)	94.2 (0.8)	94.2 (0.8)
100	0.2	0.33	0.5	93.7 (10.1)	93.8 (10.1)	93.8 (10.1)	93.3 (10.3)	93.0 (10.6)	73.1 (4.3)	84.6 (5.7)
100	0.2	3	0.5	93.9 (2.0)	93.8 (2.0)	93.8 (2.0)	94.0 (2.0)	93.5 (1.9)	71.9 (0.7)	87.4 (1.0)
100	0.8	0.33	0.5	93.7 (4.5)	93.7 (4.5)	93.7 (4.5)	93.5 (4.5)	93.5 (4.5)	71.9 (2.1)	84.9 (2.8)
100	0.8	3	0.5	93.2 (3.6)	93.2 (3.6)	93.2 (3.6)	93.1 (3.7)	93.0 (4.0)	72.1 (1.4)	84.3 (1.9)
250	0.2	0.33	0.5	95.1 (6.8)	95.2 (6.8)	95.2 (6.9)	94.9 (7.0)	94.7 (7.2)	73.4 (2.8)	88.5 (4.2)
250	0.2	3	0.5	94.0 (1.3)	94.0 (1.3)	93.8 (1.3)	94.0 (1.3)	94.3 (1.2)	72.0 (0.5)	89.5 (0.7)
250	0.8	0.33	0.5	95.4 (2.9)	95.4 (2.9)	95.2 (2.9)	95.0 (2.9)	95.0 (2.9)	74.2 (1.4)	88.8 (2.1)
250	0.8	3	0.5	94.5 (2.4)	94.5 (2.4)	94.5 (2.4)	94.7 (2.5)	95.2 (2.7)	72.5 (0.9)	88.5 (1.4)
100	0.2	0.33	0.8	91.0 (20.2)	91.0 (20.2)	91.0 (20.3)	90.8 (20.8)	90.1 (21.3)	47.9 (5.6)	72.5 (9.8)
100	0.2	3	0.8	90.8 (4.4)	90.9 (4.4)	91.0 (4.3)	91.0 (4.2)	89.8 (4.1)	42.9 (1.0)	71.4 (1.8)
100	0.8	0.33	0.8	93.2 (9.3)	93.2 (9.3)	93.0 (9.3)	92.6 (9.3)	92.0 (9.3)	46.0 (2.8)	74.9 (4.9)
100	0.8	3	0.8	92.5 (7.3)	92.4 (7.3)	92.2 (7.3)	92.4 (7.6)	92.5 (8.0)	47.4 (1.9)	72.9 (3.3)
250	0.2	0.33	0.8	93.7 (15.4)	93.7 (15.4)	93.9 (15.5)	94.0 (15.8)	94.1 (16.2)	48.9 (3.9)	81.9 (8.4)
250	0.2	3	0.8	91.8 (3.1)	92.0 (3.1)	92.0 (3.0)	92.0 (3.0)	92.4 (2.9)	45.2 (0.7)	82.1 (1.5)
250	0.8	0.33	0.8	93.5 (6.7)	93.5 (6.7)	93.4 (6.7)	93.8 (6.7)	93.3 (6.7)	48.6 (1.9)	82.8 (4.2)
250	0.8	3	0.8	93.5 (5.5)	93.5 (5.5)	93.4 (5.5)	93.6 (5.7)	94.2 (6.1)	48.0 (1.3)	81.2 (2.8)

Figure 3.1: The nominal wage series fitted with a break at 1929 (the Great Crash) and the corresponding residuals.

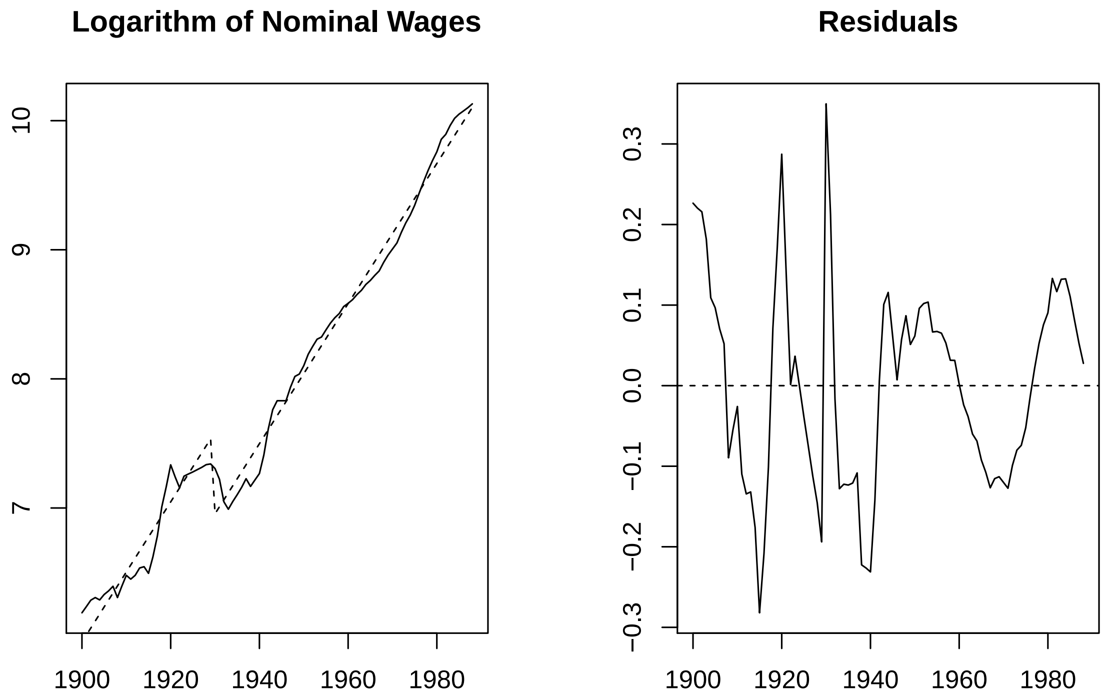

Table 3.6: The 100(1-α)% confidence intervals for the parameters in the model (3.12) for the nominal wage series. The estimated coefficients are shown in the parenthesis.

		b_1(5.91)		b_2(-0.63)		b_3(4.83)	
α	ϵ	Lower	Upper	Lower	Upper	Lower	Upper
	0.4	5.45	6.37	-1.45	0.18	3.64	6.02
0.01	0.5	5.50	6.31	-1.48	0.21	3.62	6.04
	0.6	5.57	6.25	-1.49	0.22	3.72	5.93
	0.4	5.63	6.18	-1.10	-0.17	4.13	5.53
0.05	0.5	5.65	6.17	-1.14	-0.13	4.16	5.49
	0.6	5.67	6.15	-1.12	-0.15	4.19	5.47
	0.4	5.70	6.12	-1.03	-0.24	4.29	5.37
0.10	0.5	5.71	6.11	-1.04	-0.23	4.35	5.31
	0.6	5.73	6.09	-1.04	-0.23	4.35	5.30

Chapter 4

Bootstrap-Assisted Unit Root Testing With Piecewise Locally Stationary Errors

4.1 Introduction

Unit root testing has received intensive attention in econometrics since the seminal work by Dickey and Fuller (1979, 1981). In their papers, the unit root tests were developed assuming independent and identically distributed (iid) error processes. When the error processes are stationary and weakly dependent, many variants of Dickey-Fuller test have been proposed, but unit root tests in the literature more or less rely on the two fundamental ways of approximating the weak dependence in the error. One is the Phillips-Perron test (Phillips, 1987a; Phillips and Perron, 1988), where the longrun variance of the error process is consistently estimated in a nonparametric way using the heteroscedasticity and autocorrelation consistent estimators (Newey and West, 1987; Andrews, 1991). The other is the augmented Dickey-Fuller test (Said and Dickey, 1984) which approximates the dependence structure in the error processes with an AR(p) model, where p can grow with respect to the sample size. In addition to these two conventional methods and their variants, bootstrap-based methods were also proposed. See Paparoditis and Politis (2002, 2003), Chang and Park (2003), Cavaliere and Taylor (2009a), among others. For reviews and comparisons of some of these bootstrap based methods, we refer to Paparoditis and Politis (2005) and Palm et al. (2008).

Recently, it has been argued in the literature that many macroeconomic series exhibit heteroscedastic behavior in the error. For instance, the U.S. gross domestic product series is observed to have less variability since 1980s; see Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Busetti and Taylor (2003), and references therein. Also, the majority of macroeconomic data in Stock and Watson (1999) exhibit heteroscedasticity in unconditional variances as pointed out by Sensier and van Dijk (2004). If there are breaks in the error structure, it is known that the traditional unit root tests such as Dickey and Fuller (1979) are biased towards rejecting the stationarity assumption (Busetti and Taylor, 2003). For this reason, a number of unit root tests that are robust to the heteroscedasticity have been developed in the literature, such as Busetti and Taylor (2003) and Cavaliere and Taylor (2007, 2008a,b, 2009a), which allow for smooth and abrupt changes in the unconditional or conditional variance in the error processes.

Most of the existing methods handle heteroscedasticity by assuming their error processes to be linear processes with heteroscedastic innovations, where the error process u_t has the form

$$u_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \qquad e_t = \omega_t \varepsilon_t. \tag{4.1}$$

Here, $\{\varepsilon_t\}$ is assumed to be independent and identically distributed (iid) or martingale difference sequence, and ω_ts are a sequence of deterministic numbers that account for the heteroscedasticity. This process can be considered as a generalization of the popular linear processes. Although this generalization allows for some departures from stationarity, it is still restrictive in the following three aspects. First, this kind of linear process cannot accommodate some popular nonlinear models in time series analysis, such as threshold, bilinear models. Second, this error structure is somewhat special, and it seems that most of existing methods developed to account for the heteroscedasticity in the error take advantage of this special error structure. See Section 4.3 for more details. Third, this kind of heteroscedastic linear process only allows heteroscedasticity in the error variance and and may not be appropriate in the case of changes in other second order properties such as autocorrelations.

In this chapter we adapt a more general framework of nonstationarity to capture smooth and abrupt changes in second or higher order properties of the error processes. The piecewise locally stationary (PLS) process was recently proposed by Zhou (2013), as a generalization of the locally stationary process. The locally stationary processes have received a lot of attention since the seminal work of Priestley (1965) and Dahlhaus (1997). Local stationarity is a concept that can naturally expand the notion of stationarity by allowing for smoothly changing second order properties of a time series; see Dahlhaus (1997), Mallat et al. (1998), Giurcanu and Spokoiny (2004), and Zhou and Wu (2009), among others, for more related work. However, locally stationary processes exclude abrupt changes in the second or higher order properties, which is often observed in real data. To accommodate the abrupt changes, the PLS processes were proposed to allow for a finite number of breaks in addition to the smooth changes. For example, Adak (1998) proposed a PLS model in the frequency domain, generalizing Dahlhaus (1997)'s local stationary model. On the other hand, Zhou (2013) proposed another PLS model in the time domain as an extension of the framework of Zhou and Wu (2009) and Draghicescu et al. (2009). This type of model allows for both nonlinearity and local stationarity and covers a wide range of processes; see Wu (2005), Zhou and Wu (2009), and Zhou (2013) for more discussions.

Under the general PLS framework for the errors, the limiting null distributions of the conventional unit root test statistics are not pivotal; they depend on the local long-run variance of the PLS error and some other nuisance parameters. A direct estimation of the unknown parameters in the limiting null distributions is very involved, unlike the case of stationary errors (Phillips, 1987a). To overcome this difficulty, we apply the dependent wild bootstrap (DWB) proposed in Shao (2010a) to approximate the limiting null distributions and provide a rigorous theoretical justification by establishing the functional central limit theorem for the partial sum process of the bootstrapped residuals. This seems to be the first time the DWB is justified for the PLS processes and in the unit root setting, which suggests the ability of the DWB to accommodate both piecewise local stationarity and weak dependence, and it can be potentially used for other inference problems related to locally stationary processes.

The rest of the chapter is organized as follows. Section 4.2 presents the model, the statistics and their limiting distributions under the null and local alternatives. In Section 4.3 the DWB is described and its consistency is justified. The power behavior under the local alternatives of the DWB is also presented. Section 4.4 presents some simulation results. Section 4.5 concludes. Section 4.6 presents tables and figures. Technical details are relegated to Appendix C.

Throughout the chapter, we use $\xrightarrow{\mathcal{D}}$ for convergence in distribution and $\Rightarrow$ weak convergence in $D[0,1]$, the space of functions on [0,1] which are right continuous and have left limits, endowed with Skorohod metric (Billingsley, 1968). Let $a_n \asymp c_n$ indicates $a_n/c_n \to 1$ as $n \to \infty$. The symbols $O_p(1)$ and $o_p(1)$ signify being bounded in probability and convergence to zero in probability, respectively. We use $\lfloor a \rfloor$ to denote the integer part of $a \in \mathbb{R}$, $B(\cdot)$ a standard Brownian motion, and $N(\mu, \Sigma)$ the (multivariate) normal distribution with mean μ and covariance matrix Σ. Denote by $||X||_p = (E|X|^p)^{1/p}$. For a $p \times q$ matrix $A = (a_{ij})_{i \le p, j \le q}$, let $||A||_F = (\sum_{i=1}^p \sum_{j=1}^q a_{ij}^2)^{1/2}$ be the Frobenius norm. Let $\mathbf{1}(\mathcal{E})$ be the indicator function, being 1 if the event $\mathcal{E}$ occurs and 0 otherwise.

4.2 The unit root test under piecewise locally stationary errors

Following the framework in Phillips and Xiao (1998), we assume that the data $\{y_{1,n}, \ldots, y_{n,n}\}$ is generated from

$$\begin{aligned} y_{t,n} &= X_{t,n} + \beta' z_{t,n} \quad t = 0, 1, \ldots, n, \\ X_{t,n} &= \rho X_{t-1,n} + u_{t,n}, \quad t = 1, 2, \ldots, n. \end{aligned} \tag{4.2}$$

Here, $z_{t,n}$ is a $p \times 1$ vector of deterministic trend functions, and β is a $p \times 1$ vector of corresponding coefficients, which satisfies the following conditions.

(Z1) There exists a scaling matrix D_n and a piecewise continuous function $Z(r)$ such that $D_n^{-1} z_{\lfloor nr \rfloor, n} \Rightarrow Z(r)$ as $n \to \infty$.

(Z2) $\int_0^1 Z(r) Z(r)' dr$ is positive definite.

These assumptions allow some popular choices for trend functions such as $(p-1)$th order polynomial trends and are quite standard in the literature; see Section 2.1 of Phillips and Xiao (1998) and Section 2 in Cavaliere and Taylor (2007). For initial conditions, we assume $X_{0,n} = 0$ to simplify the argument. This assumption can be relaxed to, for example, $X_{0,n} = O_p(1)$, which does not alter our asymptotic results.

Following Zhou (2013)'s framework, the error process $\{u_{t,n}\}_{t=1}^n$ is assumed to be mean zero piecewise locally stationary with τ break points, i.e., there exist constants $0 = b_0 < b_1 < \ldots < b_\tau < b_{\tau+1} = 1$ and some functions $G_0, G_1, \ldots, G_\tau$ such that

$$u_{t,n} = G_j(s_t, \mathcal{F}_t), \quad \text{if } b_j \leq s_t < b_{j+1},$$

where $s_t = t/n$, $\mathcal{F}_t = (\ldots, \varepsilon_0, \ldots, \varepsilon_{t-1}, \varepsilon_t)$, and ε_t's are i.i.d. random variables with mean 0 and variance 1. We assume

(A1) The process $\{u_t\}$ is piecewise stochastic Lipschitz continuous, i.e., for all $j \in \{0, 1, \ldots, \tau\}$ and $s_1, s_2 \in [b_j, b_{j+1}]$, $s_1 \neq s_2$,

$$||G_j(s_1, \mathcal{F}_0) - G_j(s_2, \mathcal{F}_0)||_2 / |s_1 - s_2| \leq C$$

for some finite constant C.

(A2) $\max_{j\in\{0,1,\ldots,\tau\}} \sup_{s\in[b_j,b_{j+1}]} ||G_j(s,\mathcal{F}_0)||_4 < \infty$.

(A3) $\delta_4(k) = O(\chi^k)$ for some $\chi \in (0,1)$ and $\delta_p(k)$ is the physical dependence measure defined as

$$\delta_p(k) = \max_{0\le j\le \tau} \sup_{b_j\le s\le b_{j+1}} ||G_j(s,\mathcal{F}_k) - G_j\{s,(\mathcal{F}_{-1},\varepsilon_0',\varepsilon_1,\ldots,\varepsilon_k)\}||_p,$$

if $k \ge 0$, and $\delta_p(k) = 0$ if $k < 0$.

(A4) $\inf_{s\in[0,1]} \sigma^2(s) > 0$, where $\sigma^2(s) = \sum_{h=-\infty}^{\infty} c_j(s;h)$ is the long-run variance function, $c_j(s;h) = \mathrm{cov}\{G_j(s,\mathcal{F}_0), G_j(s,\mathcal{F}_h)\}$ for $s \in [b_j, b_{j+1})$, and $\sigma^2(1) = \lim_{s\uparrow 1} \sigma^2(s)$.

The assumption (A1) states that if t/n and t'/n are close and there is no break point in between, then $u_{t,n}$ and $u_{t',n}$ are expected to be stochastically close. In other words, the second or higher order property of $u_{t,n}$ should be smoothly changing, except for a finite number of break points. The physical dependence measure in (A3) was introduced in Zhou and Wu (2009) as an extension of its stationary counterpart first introduced by Wu (2005). The assumption (A3) implies that $u_{t,n}$ is locally short-range dependent and the dependence decays exponentially fast. When location $s \in [0,1]$ is fixed, the process $\{G_j(s,\mathcal{F}_t)\}_{t\in\mathbb{Z}}$ is stationary for each j, and (A4) introduces the time varying longrun variance parameter $\sigma^2(s)$. Note that in the stationary error case, $\sigma^2(s)$ is a constant. The assumptions (A1)-(A4) are similar to those of Zhou and Wu (2009), Wu and Zhou (2011), and Zhou (2013). See (Zhou, 2013, Section 2) for more specific examples. These assumptions are not the weakest possible but are satisfied by a wide class of time series models.

Given the observations $\{y_{t,n}, z_{t,n}\}_{t=1}^n$, we are interested in testing the unit root hypothesis

$$H_0: \ \rho = 1 \qquad \text{vs} \qquad H_1: \ |\rho| < 1.$$

Consider the OLS estimator $\widehat{\rho}_n = (\sum_{t=1}^n \widehat{X}_{t,n}\widehat{X}_{t-1,n})/(\sum_{t=1}^n \widehat{X}_{t-1,n}^2)$ of ρ, where $\widehat{X}_{t,n} = y_{t,n} - \widehat{\beta}_n' z_{t,n}$ are the OLS residuals of $y_{t,n}$ regressed on $z_{t,n}$. Theorem 4.2.1 below states the limiting null distributions of two test statistics that are popular in the literature.

Theorem **4.2.1**. *Assume (A1)-(A4) and (Z1)-(Z2). Under the null hypothesis* $\rho = 1$*,*

$$\mathbf{T}_n = n(\widehat{\rho}_n - 1) = \frac{n^{-1}\sum_{t=1}^n \widehat{X}_{t-1,n}u_{t,n}}{n^{-2}\sum_{t=1}^n \widehat{X}_{t-1,n}^2} \xrightarrow{\mathcal{D}} \mathcal{L}_{\mathbf{T}} := \frac{\{B_{\sigma|Z}(1)^2 - \sigma_u^2\}}{2\int_0^1 B_{\sigma|Z}(r)^2 dr} \tag{4.3}$$

and the t*-statistic*

$$\mathbf{t}_n = \frac{(\sum_{t=1}^n \widehat{X}_{t-1,n}^2)^{-1/2}(\widehat{\rho}_n - 1)}{(s_n^2)^{1/2}} \xrightarrow{\mathcal{D}} \mathcal{L}_{\mathbf{t}} := \frac{\{B_{\sigma|Z}(1)^2 - \sigma_u^2\}}{2\{\sigma_u^2\int_0^1 B_{\sigma|Z}(r)^2 dr\}^{1/2}}, \tag{4.4}$$

where

$s_n^2 = (n-2)^{-1}\sum_{t=1}^n(\widehat{X}_{t,n} - \widehat{\rho}_n\widehat{X}_{t-1,n})^2$.

$\sigma_u^2 = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n Eu_{t,n}^2$,

$B_\sigma(r) = \int_0^r \sigma(s)dB(s)$*, and*

$B_{\sigma|Z}(r) = B_\sigma(r) - \{\int_0^1 B_\sigma(s)Z(s)'ds\}\{\int_0^1 Z(s)Z(s)'ds\}^{-1}Z(r)$ *is the Hilbert projection of* $B_\sigma(\cdot)$ *onto the space orthogonal to* $Z(\cdot)$*.*

As justified by Lemma C.0.12 (ii) in Appendix, σ_u^2, the limit of average marginal variance, can also be written as $\int_0^1 c(s;0)ds$, where $c(s;h) = c_j(s;h)$ for $s \in [b_j, b_{j+1})$ or $s \in [b_\tau, 1]$ with $c_j(s;h)$ defined in (A4).

If the error process is stationary, we have $\sigma^2(s) = \sigma^2$ for some constant $\sigma > 0$ and $\sigma_u^2 = E(u_t^2)$. Thus if there is no deterministic trend functions, i.e., $\beta \equiv 0$, the limiting null distributions, $\mathcal{L}_{\mathbf{T}}$ and $\mathcal{L}_{\mathbf{t}}$, reduce to

$$\mathcal{L}_{\mathbf{T},\sigma} = \frac{\{B(1)^2 - \sigma_u^2/\sigma^2\}}{2\int_0^1 B(r)^2dr}, \quad \mathcal{L}_{\mathbf{t},\sigma} = \frac{\sigma/\sigma_u\{B(1)^2 - \sigma_u^2/\sigma^2\}}{2\{\int_0^1 B(r)^2dr\}^{1/2}},$$

respectively. These limiting null distributions contain only a couple of unknown parameters and coincide with those in Phillips (1987a). To make the inference possible in this stationary error case, as Phillips (1987a) and Phillips and Perron (1998) suggested, we can estimate the longrun variance σ^2 of the error process consistently using the heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West, 1987; Andrews, 1991). The new statistics [see page 287 on Phillips (1987a)] adjusted using consistent estimates of σ and σ_u have pivotal limiting null distributions. However, in the piecewise locally stationary error case, the usual Phillips-Perron adjustment would not lead to pivotal limiting null distributions. Specifically, the HAC-based estimator of the nuisance parameter $\sigma^2(s)$, $s \in [0,1]$, is not known to be consistent in the PLS framework. The $\sigma^2(s)$ is unknown at infinitely many points, and we not only need to estimate $\sigma^2(s)$ but also the integral of $\sigma(s)$ over a Brownian motion, which makes the direct estimation of the unknown parameters in the limiting null distributions difficult.

In the unit root literature, local alternatives, $\rho_n = 1 + c/n$, $c < 0$, are often considered to examine the behavior of the test when the true ρ is close to the unity. The Ornstein-Uhlenbeck process $J_c(r) = \int_0^r e^{(r-s)c} dB(s)$ is usually involved in the limiting distributions of $\mathbf{T}_n$ and $\mathbf{t}_n$ for this near-integrated case. Under our error assumptions, we define a similar process $J_{c,\sigma}(r) := \int_0^r e^{(r-s)c}\sigma(s)dB(s)$, which is generated by the stochastic differential equation $dJ_{c,\sigma}(r) = cJ_{c,\sigma}(r)dr + \sigma(s)dB(r)$ with initial condition $J_{c,\sigma}(0) = 0$. Then the limiting distributions of the two test statistics under the local alternative can be expressed as in the following theorem.

Theorem **4.2.2**. *Assume (A1)-(A4) and (Z1)-(Z2). When* $\rho = \rho_n = 1 + c/n$, $c < 0$,

$$n^{-1/2}\widehat{X}_{\lfloor nr \rfloor, n} \Rightarrow J_{c,\sigma|Z}(r), \tag{4.5}$$

where $J_{c,\sigma|Z}(r) = J_{c,\sigma}(r) - \{\int_0^1 J_{c,\sigma}(s)Z(s)'ds\}\{\int_0^1 Z(s)Z(s)'ds\}^{-1}Z(r)$ *is the Hilbert projection of* $J_{c,\sigma}(\cdot)$ *onto the space orthogonal to* $Z(\cdot)$. *Then limiting distributions of the two statistics are*

$$\mathbf{T}_n \xrightarrow{\mathcal{D}} \mathcal{L}_{\mathbf{T},c} = \frac{\int_0^1 J_{c,\sigma|Z}(r)\sigma(r)dB(r) + 2^{-1}\{\int_0^1 \sigma^2(r)dr - \sigma_u^2\}}{\int_0^1 J^2_{c,\sigma|Z}(r)dr} + c \tag{4.6}$$

and

$$\mathbf{t}_n \xrightarrow{\mathcal{D}} \mathcal{L}_{\mathbf{t},c} = \frac{\int_0^1 J_{c,\sigma|Z}(r)\sigma(r)dB(r) + 2^{-1}\{\int_0^1 \sigma^2(r)dr - \sigma_u^2\}}{\{\sigma_u^2 \int_0^1 J^2_{c,\sigma|Z}(r)dr\}^{1/2}} + \frac{c\{\int_0^1 J^2_{c,\sigma|Z}(r)dr\}^{1/2}}{\sigma_u}. \tag{4.7}$$

Notice that when $c = 0$, i.e., under the null hypothesis, we have $J_{c,\sigma|Z}(r) = B_{\sigma|Z}(r)$ so that $\int_0^1 B_{\sigma|Z}(r)\sigma(r) dB(r) = 2^{-1}\{B^2_{\sigma|Z}(1) - \int_0^1 \sigma^2(r)dr\}$, which is due to the Itô's formula.[1] so that the limiting distributions $\mathcal{L}_{\mathbf{T},c}$ and $\mathcal{L}_{\mathbf{t},c}$ are identical to the limiting null distributions $\mathcal{L}_{\mathbf{T}}$ and $\mathcal{L}_{\mathbf{t}}$, respectively. On the other hand, when c is extremely negative, the limiting distributions take very negative values, in which case the unit root null hypothesis would be rejected with high probability, implying nontrivial power under the (local)

[1] It can be written as $dB_\sigma(r) = \sigma(r)dB(r)$. Using the Itô's formula, we derive $B^2_\sigma(r) = B^2_\sigma(0) - \int_0^r 2\sigma(s)B_\sigma(s)dB(s) + 2^{-1}\int_0^r 2\sigma^2(s)ds$, which leads to $\int_0^r B_\sigma(s)\sigma(s)dB(s) = 2^{-1}\{B^2_\sigma(r) - \int_0^r \sigma^2(s)ds\}$.

alternative. In the special case $\beta \equiv 0$ and $\sigma(s) = \sigma$, i.e., when there is no deterministic trend and the error is stationary, the two limiting distributions $\mathcal{L}_{\mathbf{T},c}$ and $\mathcal{L}_{\mathbf{t},c}$ reduce to those in Theorem 1 from Phillips (1987b).

4.3 Bootstrap-assisted unit root test

To implement the (asymptotic) level α test in practice, we need to know the α-th quantiles of the limiting null distributions $\mathcal{L}_{\mathbf{T}}$ and $\mathcal{L}_{\mathbf{t}}$. However, as mentioned after Theorem 4.2.1, it is difficult to consistently estimate the unknown parameters $\sigma(s)$ for all $s \in [0, 1]$. As a way out, we shall use a bootstrap method to approximate the limiting null distributions. When the errors are stationary, it is well-known that the Phillips-Perron test or augmented Dickey-Fuller tests have size distortions in finite samples, even though they are justified to work asymptotically. Bootstrap-based methods have been proposed to improve the finite sample performance. Psaradakis (2001), Chang and Park (2003), and Palm et al. (2008) used the sieve bootstrap (Kreiss, 1988) assuming an infinite order AR structure for the error process. Paparoditis and Politis (2003) applied the block bootstrap (Künsch, 1989), which randomly samples from overlapping blocks of residuals. Swensen (2003) extended the stationary bootstrap (Politis and Romano, 1994) to the unit root testing, where not only the overlapping blocks are randomly chosen, but also the block size is chosen from a geometric distribution. Cavaliere and Taylor (2009a) applied the wild bootstrap (Wu, 1986) for the unit root M tests (Perron and Ng, 1996), which are modifications of the Phillips-Perron test.

To further accommodate the heteroscedasticity in the error as well as the temporal dependence, bootstrap-based methods have been developed by Cavaliere and Taylor (2008b, 2009b) and Smeekes and Taylor (2012). In their papers, the error is assumed to be a linear process with heteroscedastic innovations [see (4.1)], and the wild bootstrap (Wu, 1986) was used. However, these wild bootstrap-based tests would not work for the PLS errors because the wild bootstrap can account for heteroscedasticity but not the temporal dependence structure. The wild bootstrap works in the framework of (4.1) since the parts from the limiting null distribution that are due to the temporal dependence in the error were removed using Phillips-Perron adjustment (Cavaliere and Taylor, 2008b) or augmented Dickey-Fuller adjustment (Smeekes and Taylor, 2012). This type of removal is only possible under the assumption that the error is a heteroscedastic linear process, in which case the longrun variance $\sigma^2(s)$ can be factored into two parts; one part is due to the heteroscedastic nature of the innovations ω_t, and the other part is due to the temporal dependence $\sum_{j=0}^{\infty} c_j$ in the error, as shown in Chapter 3. As a result, the limiting null distributions after the two popular adjustments for the temporal dependence in the unit root test literature depend only on the heteroscedasticity, which can be handled by the wild bootstrap. However, for the PLS error processes, even after the Phillips-Perron or augmented Dickey-Fuller adjustment, the limiting null distributions still depend on the part that is due to the temporal dependence in the error. Therefore the wild bootstrap is not expected to work in our setting, both in theory and finite sample simulation; see Section 4.4 for some numerical evidence.

To accommodate the nonstationarity and temporal dependence in the error, we propose to adopt the so-called dependent wild bootstrap (DWB), which was first introduced by Shao (2010a) in the context of stationary time series. It turns out that the DWB is capable of mimicking the local weak dependence in the error process and provide a consistent approximation of the limiting null distributions of $\mathbf{T}_n$ and $\mathbf{t}_n$. Note that the DWB was developed for stationary time series and its applicability was only proved for the smooth function model. This chapter seems to be the first to show the validity of the DWB in the PLS framework and for the unit root testing. We expect that the local block bootstrap (Paparoditis and Politis, 2002), which involves two tuning parameters, also works here in terms of providing consistent distribution approximation. However, since the DWB is capable of capturing weak dependence and piecewise local stationarity simultaneously using only one tuning parameter, it may be preferred for practical reasons.

To implement the DWB, we generate pseudo-residuals by perturbing the original (OLS) residuals using a set of external variables $\{W_{t,n}\}_{t=1}^{n}$. The difference between the DWB and the original wild bootstrap is that $\{W_{t,n}\}_{t=1}^{n}$ is made to be dependent in the DWB, whereas $\{W_{t,n}\}_{t=1}^{n}$ is assumed to be independent in the usual wild bootstrap. The following assumption on $\{W_{t,n}\}_{t=1}^{n}$ are from Shao (2010a).

(B1) $\{W_{t,n}\}_{t=1}^{n}$ is a realization from a stationary time series with $E(W_{t,n}) = 0$ and $\text{var}(W_{t,n}) = 1$, $\{W_{t,n}\}_{t=1}^{n}$ are independent of the data, $\text{cov}(W_{t,n}, W_{t',n}) = a\{(t-t')/l\}$, $a(\cdot)$ is a kernel function and $l = l_n$ is a bandwidth parameter that satisfies $l \asymp Cn^{\kappa}$ for some $0 < \kappa < 1/3$. Assume that $W_{t,n}$ is l-dependent and $E(W_1^4) < \infty$.

(B2) Assume that $a : \mathbb{R} \to [0,1]$ is symmetric and has compact support on [-1,1], $a(0) = 1$, $\lim_{x\to 0}\{1 - a(x)\}/|x|^q = k_q \neq 0$ for some $q \in (0,2]$, and $\int_{-\infty}^{\infty} a(u)e^{-iux}du \geq 0$ for $x \in \mathbb{R}$.

In practice, we can sample $\{W_{t,n}\}_{t=1}^n$ from the multivariate normal distribution with mean zero and the covariance function $\text{cov}(W_{t,n}, W_{t',n}) = a\{(t-t')/l\}$. There are two user-determined parameters: the kernel function $a(\cdot)$ and the bandwidth parameter l. The kernel function affects the performance to a less degree than the bandwidth parameter l, and the choice of l will be discussed in Section 4.4. For the kernel function, some commonly used kernels such as the Bartlett kernel satisfy (B2).

The DWB in the unit root testing can be performed as follows.

Step 1. Calculate the OLS estimates $\widehat{\beta}_n$ of β by fitting $y_{t,n}$ on $z_{t,n}$, and let $\widehat{X}_{t,n} = y_{t,n} - \widehat{\beta}_n' z_{t,n}$.

Step 2. Let $\widehat{\rho}_n$ be the OLS estimate of $\widehat{X}_{t,n}$ on $\widehat{X}_{t-1,n}$. Calculate the statistics $\mathbf{T}_n = n(\widehat{\rho}_n - 1)$ and $\mathbf{t}_n = (s_n^2 \sum_{t=1}^n \widehat{X}_{t-1,n}^2)^{-1/2}(\widehat{\rho}_n - 1)$.

Step 3. Calculate the residuals $\widehat{u}_{t,n} = \widehat{X}_{t,n} - \widehat{\rho}_n \widehat{X}_{t-1,n}$ for all $t = 1, \ldots, n$.

Step 4. Randomly generate the l-dependent mean-zero stationary series $\{W_{t,n}\}_{t=1}^n$ satisfying condition (B1)-(B2) and generate the perturbed residuals $u_{t,n}^* = \widehat{u}_{t,n} W_{t,n}$.

Step 5. Construct the bootstrapped sample $y_{t,n}^*$ using $\{u_{t,n}^*\}$ as if $\rho = 1$ is true;

$$(y_{t,n}^* - \widehat{\beta}_n' z_{t,n}) = (y_{t-1,n}^* - \widehat{\beta}_n' z_{t-1,n}) + u_{t,n}^*,$$

$t = 2, \ldots, n$, and $y_{1,n}^* = \widehat{\beta}_n' z_{1,n} + u_{1,n}^*$.

Step 6. Calculate $\widehat{\beta}_n^*$ by refitting $y_{t,n}^*$ on $z_{t,n}$ and $\widehat{X}_{t,n}^* = y_{t,n}^* - (\widehat{\beta}_n^*)' z_{t,n}$.

Step 7. Calculate bootstrapped versions of $\widehat{\rho}_n$ and s_n^2, i.e., $\widehat{\rho}_n^*$ and ${s_n^*}^2$, based on $\{\widehat{X}_{t,n}^*\}_{t=1}^n$, and the bootstrapped test statistics $\mathbf{T}_n^* = n(\widehat{\rho}_n^* - 1)$ and $\mathbf{t}_n^* = \{{s_n^*}^2 \sum_{t=1}^n (\widehat{X}_{t-1,n}^*)^2\}^{-1/2}(\widehat{\rho}_n^* - 1)$.

Step 8. Repeat steps 2-7 B times, and record the bootstrapped test statistics $\{\mathbf{T}_n^{*(1)}, \ldots, \mathbf{T}_n^{*(B)}\}$ and $\{\mathbf{t}_n^{*(1)}, \ldots, \mathbf{t}_n^{*(B)}\}$. The p-values are

$$\frac{\sum_{b=1}^B \mathbf{1}\{\mathbf{T}_n^{*(b)} < \mathbf{T}_n\}}{B} \quad \text{and} \quad \frac{\sum_{b=1}^B \mathbf{1}\{\mathbf{t}_n^{*(b)} < \mathbf{t}_n\}}{B}.$$

Notice that the null hypothesis is not enforced in Step 3, i.e., we are using the unrestricted residuals. There is another way of constructing the bootstrap sample as discussed in Paparoditis and Politis (2003), where the null hypothesis is imposed in Step 3 and the corresponding residuals are referred as the restricted residuals, i.e., Step 3 is replaced by the following;

Step 3'. Calculate the restricted residuals $\widehat{u}^{+}_{t,n} = \widehat{X}_{t,n} - \widehat{X}_{t-1,n}$ for all $t = 1, \dots, n$.

Both procedures are consistent under the null hypothesis, but the DWB with unrestricted residuals delivers higher power than the DWB with restricted residuals, as seen from our simulations. Similar observations were made in Paparoditis and Politis (2003) for their residual block bootstrap.

The following theorem provides the core result in the proof of the consistency of the DWB in Theorem 4.3.2 and may be of independent interest.

THEOREM **4.3.1**. *Assume (A1)-(A4) and (B1)-(B2). Suppose there is no deterministic trends, i.e., $\beta \equiv 0$. Under the null and the alternative hypotheses, the partial sum process of the unrestricted bootstrap residual converges, i.e.,*

$$n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} u^{*}_{t,n} \Rightarrow \int_0^r \sigma(s) dB(s) \qquad \textit{in probability.}$$

Note that Theorem 4.3.1 holds not only under the null $\rho = 1$ but also under the alternative $|\rho| < 1$. This property makes the DWB method (with unrestricted residual) more powerful because the bootstrapped distributions correctly mimic the limiting null distributions under both the null and alternative. Although the null hypothesis was not forced in Step 3, the DWB can still correctly approximate the limiting null distribution, mainly because $\widehat{X}^{*}_{t,n}$ are constructed assuming $\rho = 1$ in Step 5.

THEOREM **4.3.2** (Bootstrap consistency). *Assume (A1)-(A4), (Z1)-(Z2), and (B1)-(B2). Under both the null and the alternative hypotheses,*

$$\mathbf{T}^{*}_{n} \xrightarrow{\mathcal{D}} \mathcal{L}_{\mathbf{T}} \qquad \textit{in probability,} \qquad \mathbf{t}^{*}_{n} \xrightarrow{\mathcal{D}} \mathcal{L}_{\mathbf{t}} \qquad \textit{in probability.}$$

Then Theorem 4.3.3 below follows immediately from Theorems 4.2.2 and 4.3.2, and the argument in the proof of Theorem 5.1 in Paparoditis and Politis (2003).

THEOREM **4.3.3** (Power of the DWB statistics for Near-integrated processes). *Assume (A1)-(A4), (Z1)-(Z2), and (B1)-(B2). For any $c < 0$,*

$$P(\mathbf{T}_n \le \mathbf{T}^{*}_{n,\alpha} | \rho = 1 + c/n) \xrightarrow{\mathcal{P}} P(L_{T,c} \le \mathcal{L}^{\alpha}_{\mathbf{T}}),$$

$$P(\mathbf{t}_n \le \mathbf{t}^*_{n,\alpha} | \rho = 1 + c/n) \xrightarrow{\mathcal{P}} P(L_{t,c} \le \mathcal{L}^\alpha_{\mathbf{t}}),$$

where $L_{T,c}$ and $L_{t,c}$ are random variables with distribution $\mathcal{L}_{\mathbf{T},c}$ and $\mathcal{L}_{\mathbf{t},c}$, respectively, which are defined in Theorem 4.2.2. $\mathcal{L}^\alpha_{\mathbf{T}}$ and $\mathcal{L}^\alpha_{\mathbf{t}}$ are the α-quantiles of the limiting null distributions, $\mathcal{L}_{\mathbf{T}}$ and $\mathcal{L}_{\mathbf{t}}$, respectively. We define $\mathbf{T}^*_{n,\alpha}$ and $\mathbf{t}^*_{n,\alpha}$ be the α-quantiles of $\{\mathbf{T}^{*(1)}_n, \ldots, \mathbf{T}^{*(B)}_n\}$ and $\{\mathbf{t}^{*(1)}_n, \ldots, \mathbf{t}^{*(B)}_n\}$ in Step 8, respectively.

As noted after Theorem 4.2.2, when $c = 0$, $\mathcal{L}_{\mathbf{T},c}$ and $\mathcal{L}_{\mathbf{t},c}$ are identical to the limiting null distributions, which makes the probability of rejecting the null hypothesis, or the (asymptotic) size of the test, be exactly the same as the level of the test. On the other hand, if c is negative and very far from 0, the probability of rejecting the null, or the asymptotic power of the test, will be close to 1, so the test is powerful if the true ρ is far below the unity. If c is not 0 but not too far from 0, Theorem 4.3.3 states that the probability of rejecting the null is somewhere in the middle of the level of the test and 1. That is, the DWB unit root tests have nontrivial power for the local alternatives.

Remark **4.3.1**. Notice that the DWB was originally developed for stationary time series (Shao, 2010). In the construction of the DWB samples, $\{W_{t,n}\}_{t=1}^n$ is generated as l-dependent stationary time series, so it is natural to expect that DWB would work for stationary time series. However, it seems not straightforward that this simple form of stationary time series would work in the case of locally stationary process with unknown breaks. What Theorem 4.3.1 suggests is that the DWB is capable of capturing nonstationary behaviors, without the need to specify any parametric forms of the error structures or to know the specific form of nonstationarity such as location of breaks.

4.4 Simulations

In this section we compare our DWB method with the residual block bootstrap (RBB) (Paparoditis and Politis, 2003). and investigate how the choice of the bandwidth parameter affect the size and power of the two tests. we set $\beta \equiv 0$ for simplicity so that we have $\widehat{X}_{t,n} = X_{t,n}$. Consider

$$X_{t,n} = \rho X_{t-1,n} + u_{t,n}, \quad t = 1, 2, \ldots, n$$

where $u_{t,n}$ is generated from time-varying MA(1) model, $u_{t,n} = e_{t,n} + \phi(t/n)e_{t-1,n}$, $t = 1, \ldots, n$, and $e_{t,n} = \omega(t/n)\varepsilon_t$, ε_t are iid $N(0,1)$. We consider the following models for the MA(1) coefficient $\phi(s)$ and the standard deviation of the innovations $\omega(s)$, which are piecewise smooth functions.

(M1) Stationary $u_{t,n}$:

$$\phi(s) = 0.5, \quad \omega(s) = 0.5, \quad s \in [0,1].$$

(M2) Increasing $\phi_{t,n}$ and $\omega(s)$:

$$\phi(s) = 1.6s - 0.8, \quad \omega(s) = 0.5s + 0.1, \quad s \in [0,1]$$

(M3) Jump in $\phi_{t,n}$ and increasing $\omega(s)$ 1:

$$\phi(s) = 0.2 + 0.6\mathbf{1}(s > 0.5), \quad \omega(s) = 0.5s + 0.1, \quad s \in [0,1]$$

(M4) Jump in $\phi_{t,n}$ and increasing $\omega(s)$ 2:

$$\phi(s) = 0.2 + 0.6\mathbf{1}(s > 0.5), \quad \omega(s) = 0.5s + 0.5, \quad s \in [0,1]$$

(M1) represents stationary case with moderate temporal dependence in which case, both the RBB and DWB are consistent. (M2)-(M4) satisfy the piecewise local stationarity assumption (A1)-(A4), but are not stationary. The underlying assumption for the RBB to work is that the candidate blocks need to have the same or roughly the same distribution, which is violated in the PLS case, so in theory the RBB is expected to be inconsistent for the models (M2)-(M4).

To implement the DWB, we need to choose the bandwidth parameter l. We propose to use the minimum volatility (MV) method from Politis et al. (1999). The rationale behind the MV method is that the approximation of the limiting distribution is stable when the bandwidth parameter l is in a certain range. We can find an optimal bandwidth l, which depends on the data $\{X_{t,n}\}$, as follows:

1. Choose some candidate l's; $l_1, \ldots, l_k$.

2. For each l_i $(i = 1, \ldots, k)$, generate the bootstrap sample $y_{t,n}^{*(i)}$ $(t = 1, \ldots, n)$ and calculate $\mathbf{T}_n^{(1,i)}$

3. Repeat B times so that we have $(\mathbf{T}_n^{*(1,i)}, \ldots, \mathbf{T}_n^{*(B,i)})$ for each l_i.

4. Let D_i be the empirical distribution function of $(\mathbf{T}_n^{*(1,i)}, \ldots, \mathbf{T}_n^{*(B,i)})$, i.e., $D_i(x) = B^{-1} \sum_{j=1}^{B} \mathbf{1}(\mathbf{T}_n^{*(j,i)} \leq x)$. For $i = 1, \ldots, k-1$, calculate the Kolmogorov-Smirnov distance between D_i and D_{i+1}, $H_i = \sup_{x \in \mathbb{R}} |D_i(x) - D_{i+1}(x)|$.

5. The optimal l is $l_{\widehat{i}}$, where $\widehat{i} = \text{argmin}_{i=1,\ldots,k-1} H_i$.

Although Paparoditis and Politis (2003) provided some guidance of choosing the block size in the RBB, it seems that it still depends on some subjective choices. To make a fair comparison between the DWB and RBB, we use the MV method for both bootstrap-based tests.

In our simulation, the sample size is $n = 100$ and the number of Monte-Carlo replications is 2000. For the DWB, we generate $B = 1000$ bootstrap replications; in each replication, pseudoseries $(W_1, \ldots, W_n)'$ are generated from iid $N(\mathbf{0}_n, \Sigma)$, where Σ is n by n matrix with its (i,j)th element being $a\{(t-t')/l\}$. Here the Bartlett kernel is used, i.e., $a(s) = (1 - |s|)\mathbf{1}(|s| \leq 1)$.

Tables 4.1 presents the empirical rejection percentage of the DWB and RBB unit root tests with unrestricted residuals and restricted residuals at nominal level 5% using various l's. The second to last line of each part of the table represents the rejection rates when we choose l from the set $\{1, 4, 7, \ldots, 49\}$ using the MV method in each replication, and the last line presents the average of the optimal ls over all Monte Carlo replications. To save space, the results are only shown for l in $\{1, 7, 13, \ldots, 49\}$, although our simulation is based on a finer grid of $\{1, 4, 7, 10, 13, \ldots, 49\}$. Notice that the wild bootstrap, i.e., when $l = 1$, is not consistent when serial correlation is present, which is expected in theory. In contrast, the DWB and RBB with the bandwidth l chosen by the MV method, however, seem to work reasonably well. This justifies the use of the MV method in real data practice, not only for the DWB unit root test but also for the RBB unit root test. In addition, Table 4.2 further investigates the case for $n = 400$ with unrestricted residuals. As expected, the DWB generally produces better size than the case for $n = 100$, but the RBB does not do any better or sometimes worse than that for $n = 100$.

For the stationary case (M1), all eight tests DWB-$\mathbf{T}_n$, DWB-$\mathbf{t}_n$, RBB-$\mathbf{T}_n$, RBB-$\mathbf{t}_n$, using unrestricted and restricted residuals perform similarly well, all having the empirical rejection rates close to the nominal level. The difference between the DWB and RBB shows up for the models (M2)-(M4). The RBB is not consistent, rejecting substantially more frequently than the nominal level, while the DWB is consistent under (M2)-(M4) with finite sample rejection rates being pretty close to its nominal level in general. In particular, for (M2) and (M3) when the innovation variance is relatively small, the empirical sizes of the RBB test are about 10%, but the DWB still delivers an accurate size. The distinctions between $\mathbf{T}_n$ and $\mathbf{t}_n$ or using the restricted and unrestricted residuals are not as apparent, though, at least in terms of empirical size.

Tables 4.3 presents the size adjusted powers when $n = 100$ using unrestricted and restricted residuals. The local alternative with $\rho = 1 - 0.15/n = 0.85$ is considered. To make a fair power comparison, we calculated size adjusted power following Domínguez and Lobato (2001). When the error process is stationary, (M1), the DWB and RBB have almost identical power. For other nonstationary models (M2)-(M4), the RBB tends to have slightly better power than the DWB. However, considering the size distortion of the RBB test, the power loss is not too serious. In terms of power comparison of the unrestricted and restricted residuals, using the restricted residuals result in substantial power loss, regardless of the model or whether using $\mathbf{T}_n$ and $\mathbf{t}_n$. Thus using unrestricted residuals is strongly preferred in practice. In terms of the choice between $\mathbf{T}_n$ and $\mathbf{t}_n$, the former tends to be slightly more powerful for all models (M1)-(M4) and for both unrestricted and restricted residuals. Thus in practice, $\mathbf{T}_n$ may be preferred.

To summarize, the RBB is not consistent in PLS setting, resulting in the large size distortion. By contrast, the DWB delivers quite accurate size in all situations without much power loss. Between two statistics $\mathbf{T}_n$ and $\mathbf{t}_n$, the former is slightly more powerful and thus may be preferred. Lastly, algorithm using the unrestricted residuals in Section 4.2 is preferred to the one using the restricted residuals, due to the significant increase in power, corresponding to the unrestricted residuals.

4.5 Conclusion

In this chapter we presented a new bootstrap-based unit root testing procedure that is robust to changing second and higher order properties in the error process. The error process we adapted is the piecewise locally stationary (PLS) framework, which is general enough to allow for not only the unconditional error variance changes but also autocorrelation changes, and it seems that it is the first time this framework is introduced to the unit root testing literature. Under the PLS framework, we derived the limiting null distributions of two popular test statistics and proposed to approximate them using the dependent wild bootstrap (DWB). As an important theoretical contribution, we established the functional central limiting theorem of the partial sum process of the DWB residuals and justified the bootstrap consistency under both the null and the alternative hypotheses. Consequently, our test was shown to have nontrivial power under the local alternatives both asymptotically and in finite samples. In the simulation, our DWB-based method provided more accurate size with only slight power loss compared to the RBB-based counterpart (Paparoditis and Politis, 2003). It is also worth noting that the DWB was originally proposed for stationary time series. Thus this chapter has broadened the applicability of the dependent wild bootstrap, whose use in the locally stationary context is worth further exploring.

4.6 Tables and Figures

Table 4.1: Empirical rejection rates of DWB (D) and RBB (R) for models (M1)-(M4) using unrestricted and restricted residuals (Step 3 and Step 3', respectively, in Section 4.2). The last row in each part ($\hat{l}_{\text{opt}}$) reports the mean optimal bandwidths over 2000 Monte Carlo replications. The sample size is $n = 100$, and the nominal level is 5%.

	Unrestricted Residuals (Step 3)															
	$\mathbf{T}_n$								$\mathbf{t}_n$							
	(M1)		(M2)		(M3)		(M4)		(M1)		(M2)		(M3)		(M4)	
l	D	R	D	R	D	R	D	R	D	R	D	R	D	R	D	R
1	0.8	0.1	4.0	1.8	0.6	0.3	0.7	0.2	1.1	0.1	4.3	1.6	0.9	0.1	0.9	0.2
7	3.4	3.4	6.0	12.3	4.0	7.8	4.4	5.1	3.8	3.2	6.2	10.8	4.2	6.8	4.8	5.4
13	4.0	4.0	6.9	11.8	4.3	8.3	5.1	5.9	3.9	3.8	6.8	10.5	4.3	7.2	4.6	5.3
19	4.4	5.0	7.8	12.3	5.1	9.2	5.2	6.2	4.2	4.2	7.1	10.8	4.6	8.0	4.8	5.3
25	4.5	5.3	6.8	12.2	4.6	10.2	4.9	7.3	4.3	4.7	6.3	10.6	4.4	8.8	4.8	6.2
31	5.0	5.5	7.7	13.2	5.3	11.0	6.2	7.6	4.8	4.8	7.1	11.1	5.2	9.4	5.6	6.0
37	5.6	6.9	8.1	14.4	5.4	11.9	6.4	8.9	5.1	6.2	7.8	13.2	5.5	10.4	5.9	7.6
43	4.8	6.8	8.8	13.8	5.6	12.0	5.9	9.6	4.6	5.8	8.1	12.4	5.1	10.0	5.3	7.6
49	5.6	7.4	8.3	14.9	5.1	12.8	6.8	9.2	5.2	6.2	7.6	14.2	4.9	11.6	6.0	8.1
MV	**5.2**	**5.7**	**6.9**	**12.0**	**5.4**	**10.0**	**6.0**	**7.0**	**5.5**	**5.5**	**6.6**	**12.0**	**5.2**	**8.6**	**5.6**	**7.0**
$\hat{l}_{\text{opt}}$	28	19	28	20	31	20	28	18	27	19	27	20	31	20	28	19
	Restricted Residuals (Step 3')															
	$\mathbf{T}_n$								$\mathbf{t}_n$							
	(M1)		(M2)		(M3)		(M4)		(M1)		(M2)		(M3)		(M4)	
l	D	R	D	R	D	R	D	R	D	R	D	R	D	R	D	R
1	0.8	0.1	4.0	1.9	0.6	0.3	0.7	0.2	1.1	0.1	4.5	1.6	0.9	0.1	1.0	0.2
7	3.5	3.1	6.2	13.0	5.0	8.3	4.8	5.3	3.9	3.4	6.0	11.3	5.0	7.5	4.8	5.5
13	4.5	3.9	7.3	12.8	6.1	9.2	5.5	6.2	4.8	3.7	7.1	12.0	5.8	7.8	5.2	5.3
19	4.8	4.9	8.5	13.9	6.7	10.3	5.9	6.7	4.7	4.2	7.5	12.0	6.3	8.8	5.1	5.5
25	5.1	5.4	7.2	13.5	6.8	11.2	5.8	8.0	4.6	4.5	7.2	12.5	6.2	9.4	5.4	6.6
31	5.7	5.3	8.6	13.5	8.4	11.6	7.3	8.3	5.1	4.8	8.0	12.2	7.3	9.8	6.4	7.0
37	6.6	6.6	9.0	16.0	8.8	13.6	7.9	9.6	6.2	6.2	9.0	14.5	8.1	11.3	7.5	8.2
43	6.2	6.6	9.8	15.6	8.8	12.8	7.6	8.8	5.6	5.3	9.3	14.1	8.6	11.2	7.1	7.3
49	6.9	7.3	10.1	16.1	9.5	13.9	8.1	9.5	6.4	5.9	8.9	14.8	9.0	12.0	7.6	8.2
MV	**6.3**	**5.5**	**8.2**	**13.2**	**7.8**	**10.8**	**6.8**	**7.2**	**6.6**	**4.8**	**7.9**	**12.6**	**7.7**	**9.5**	**6.4**	**6.8**
$\hat{l}_{\text{opt}}$	28	19	29	21	33	20	28	19	27	19	29	21	32	20	28	19

Table 4.2: Empirical rejection rates of DWB (D) and RBB (R) for models (M1)-(M4) using the unrestricted residuals (Step 3 in Section 4.2). The last row reports the mean optimal bandwidths over 2000 Monte Carlo replications. The sample size is $n = 400$, the nominal level is 5%.

	$\mathbf{T}_n$								$\mathbf{t}_n$							
	(M1)		(M2)		(M3)		(M4)		(M1)		(M2)		(M3)		(M4)	
l	D	R	D	R	D	R	D	R	D	R	D	R	D	R	D	R
1	0.8	0.0	3.2	1.5	0.4	0.0	0.9	0.0	1.1	0.0	3.5	1.2	0.6	0.0	0.9	0.0
11	3.2	5.1	4.6	12.5	4.9	9.4	2.8	6.3	3.1	5.1	5.1	11.6	5.2	8.6	3.1	6.2
21	4.3	5.3	4.7	12.0	4.8	9.4	3.5	6.6	3.9	4.8	5.0	10.8	5.4	9.2	3.6	5.8
31	4.2	4.8	4.8	11.3	5.5	8.9	4.2	6.2	4.5	4.8	4.9	10.2	5.6	8.5	4.1	6.3
41	4.6	6.0	5.1	11.6	5.0	9.6	4.5	7.0	4.6	5.5	4.8	10.7	5.1	9.1	4.8	6.6
51	4.4	5.8	4.9	11.7	5.3	8.8	4.8	6.6	4.6	5.4	4.8	10.2	4.8	8.2	4.6	5.8
61	5.0	5.9	5.5	12.6	4.5	9.6	5.2	7.0	4.8	5.2	5.4	10.4	4.8	9.0	4.8	6.3
71	4.9	5.5	4.7	12.7	4.4	10.2	5.1	7.4	5.1	5.3	5.1	11.2	4.8	9.2	5.2	7.2
81	5.0	5.9	5.1	10.7	4.3	9.2	4.5	7.2	5.0	5.3	4.8	9.8	4.6	8.3	4.7	5.9
91	5.0	7.1	4.3	12.4	4.8	10.2	5.1	7.3	5.2	6.4	4.6	11.4	4.5	8.4	5.0	6.4
101	4.8	6.7	5.2	12.9	4.4	10.8	4.5	7.5	4.9	5.6	5.2	10.9	5.0	9.0	4.4	7.1
111	5.0	6.6	5.3	12.3	4.8	10.6	5.3	7.3	5.1	5.9	5.3	11.3	5.0	9.3	5.1	6.7
121	4.8	6.9	4.8	12.2	4.9	10.8	4.9	8.3	5.1	6.2	5.2	10.8	5.2	9.1	4.5	7.1
131	5.1	7.8	5.1	13.1	4.6	11.2	5.6	9.2	4.8	6.2	5.1	11.3	5.1	9.8	5.3	8.0
141	5.0	8.8	5.3	14.6	5.0	11.9	5.3	9.6	5.1	7.3	5.3	13.0	5.1	10.6	5.3	8.3
151	5.6	8.1	4.8	15.2	4.8	12.8	5.5	9.6	5.4	6.5	5.1	13.3	4.8	10.9	5.5	8.3
161	5.5	8.5	5.1	14.9	4.8	13.0	5.4	9.8	5.4	6.9	5.2	13.5	5.0	11.2	5.5	8.3
171	5.2	8.5	5.3	15.3	5.1	12.2	5.1	9.6	4.9	7.1	4.7	12.8	4.8	10.3	5.1	8.1
181	5.3	8.8	5.2	14.8	4.8	12.3	5.0	8.7	5.1	7.8	5.2	13.1	4.6	9.9	4.8	7.8
191	5.2	8.4	5.2	15.4	4.2	13.1	5.1	10.2	5.3	7.2	5.1	13.6	4.8	11.8	5.1	8.9
MV	**5.0**	**7.0**	**4.6**	**12.6**	**4.6**	**10.4**	**4.8**	**7.8**	**5.1**	**6.0**	**5.1**	**11.2**	**4.6**	**9.3**	**5.0**	**7.0**
$\hat{l}_{\text{opt}}$	121	81	120	89	127	84	120	82	122	82	120	87	127	86	121	80

Table 4.3: The size adjusted powers for DWB (D) and RBB (R) for models (M1)-(M4) using the unrestricted and restricted residuals (Step 3 and Step 3', respectively, in Section 4.2). The last row of each part ($\hat{l}_{\text{opt}}$) reports the mean optimal bandwidths over 2000 Monte Carlo replications. The sample size is $n = 100$, and the local alternative $\rho = 0.85$ is considered.

	Unrestricted Residuals (Step 3)															
	$\mathbf{T}_n$								$\mathbf{t}_n$							
	(M1)		(M2)		(M3)		(M4)		(M1)		(M2)		(M3)		(M4)	
l	D	R	D	R	D	R	D	R	D	R	D	R	D	R	D	R
1	96.6	96.7	60.5	59.8	71.9	72.0	87.1	86.4	93.4	93.8	59.1	57.7	68.8	68.2	83.5	83.2
7	89.5	89.2	55.5	58.9	68.3	65.2	77.8	78.9	85.8	86.1	53.5	56.1	64.6	61.8	76.1	76.8
13	88.1	87.5	53.4	58.8	63.8	63.0	75.9	77.1	83.5	83.6	51.0	56.0	61.3	59.2	73.8	74.0
19	86.6	85.4	53.2	60.0	62.3	62.6	73.8	75.9	81.5	81.2	51.4	57.5	59.7	59.2	71.4	73.0
25	85.7	85.2	54.1	59.2	62.5	64.1	73.5	75.8	81.0	81.0	52.0	57.8	59.2	60.0	71.2	72.8
31	84.4	82.8	53.9	61.0	59.7	64.0	72.6	75.2	79.2	79.1	52.4	58.6	57.7	60.1	70.7	72.0
37	83.7	85.0	52.9	63.7	59.7	66.4	73.2	76.8	79.7	80.5	50.8	59.9	57.0	62.4	70.5	74.4
43	83.0	83.0	54.1	64.6	58.4	66.7	72.8	76.6	78.9	79.6	52.1	61.8	57.0	63.2	71.0	73.8
49	83.8	81.2	52.8	65.5	58.6	66.8	72.7	76.1	78.6	78.6	50.6	63.2	55.4	62.9	69.3	73.0
MV	**83.8**	**84.5**	**52.0**	**60.2**	**61.3**	**63.5**	**72.6**	**76.0**	**80.7**	**82.7**	**51.0**	**57.5**	**58.4**	**58.7**	**71.2**	**73.4**
$\hat{l}_{\text{opt}}$	32	19	32	21	30	20	32	19	32	19	32	21	29	19	32	19
	Restricted Residuals (Step 3')															
	$\mathbf{T}_n$								$\mathbf{t}_n$							
	(M1)		(M2)		(M3)		(M4)		(M1)		(M2)		(M3)		(M4)	
l	D	R	D	R	D	R	D	R	D	R	D	R	D	R	D	R
1	96.7	96.5	61.7	60.4	72.4	71.9	86.9	86.6	93.3	93.8	59.7	57.9	68.8	68.2	83.3	83.2
7	80.7	83.2	40.4	44.2	51.0	55.0	67.8	70.5	77.0	80.0	40.7	43.4	48.4	52.2	66.6	69.2
13	67.3	74.5	33.2	37.4	40.1	47.1	53.8	62.0	64.3	71.4	31.9	35.6	39.1	44.0	54.4	59.5
19	57.4	66.5	30.5	33.4	35.9	42.3	48.2	55.4	54.4	63.9	29.0	32.2	33.4	39.6	47.2	54.5
25	52.7	67.7	27.1	33.0	33.1	43.7	44.0	55.0	50.0	63.1	26.3	32.6	30.9	40.6	42.4	53.5
31	48.1	59.6	28.2	32.8	32.6	41.3	41.3	50.5	45.4	56.8	26.9	31.5	30.4	38.1	39.8	49.5
37	45.6	62.9	25.4	34.0	30.0	43.0	37.5	52.4	42.7	59.4	24.3	32.8	27.4	38.6	35.7	49.1
43	43.5	58.2	24.0	34.8	29.3	42.8	36.9	49.3	40.4	54.9	23.2	33.1	27.3	39.0	34.5	47.5
49	41.9	54.2	24.1	34.9	29.0	40.3	34.3	47.4	38.2	52.4	22.9	34.4	26.2	38.6	33.1	45.4
MV	**47.5**	**66.1**	**24.3**	**32.6**	**30.2**	**44.0**	**40.1**	**55.0**	**44.1**	**63.9**	**24.9**	**32.9**	**28.9**	**41.7**	**39.0**	**53.8**
$\hat{l}_{\text{opt}}$	33	20	34	22	33	20	33	20	32	20	33	21	32	19	32	19

Appendix A

Technical Details for Chapter 2

Proof of Theorem 2.2.1. For $r \in (0,1]$, we rewrite $(T-p)^{-1/2}\widetilde{S}^{KVB}_{\lfloor rT \rfloor}$ as

$$\begin{aligned}
(T-p)^{-1/2}\widetilde{S}^{KVB}_{\lfloor rT \rfloor} &= (T-p)^{-1/2}\sum_{j=p+1}^{\lfloor rT \rfloor}(\widehat{v}_j - \sum_{l=1}^{p}\widehat{A}_l\widehat{v}_{j-l}) \\
&= (T-p)^{-1/2}\big(\sum_{j=p+1}^{\lfloor rT \rfloor}\widehat{v}_j - \sum_{j=p+1}^{\lfloor rT \rfloor}\sum_{l=1}^{p}A_l\widehat{v}_{j-l} \\
&\quad - \sum_{j=p+1}^{\lfloor rT \rfloor}\sum_{l=1}^{p}(\widehat{A}_l - A_l)\widehat{v}_{j-l}\big).
\end{aligned} \tag{A.1}$$

Since $T^{-1/2}\sum_{j=1}^{\lfloor rT \rfloor}\widehat{v}_j \Rightarrow \Psi(B_k(r) - rB_k(1))$ as proved in Kiefer et al. (2000) under K1 and K2, $(T-p)^{-1/2}\sum_{j=p+1}^{\lfloor rT \rfloor}\widehat{v}_{j-l} \Rightarrow \Psi(B_k(r) - rB_k(1))$ for each l, which implies that the last term in (A.1) is $o_p(1)$ under the assumption PW1. The first two terms can be rewritten as $(T-p)^{-1/2}\big(\sum_{j=1}^{\lfloor rT \rfloor}\widehat{v}_j - \sum_{l=1}^{p}A_l\sum_{j=1}^{\lfloor rT \rfloor}\widehat{v}_j + O_p(1)\big) = (T-p)^{-1/2}\big(I_k - \sum_{l=1}^{p}A_l\big)\sum_{j=1}^{\lfloor rT \rfloor}\widehat{v}_j + o_p(1)$, so we have

$$(T-p)^{-1/2}\widetilde{S}^{KVB}_{\lfloor rT \rfloor} \Rightarrow D_p^{-1}\Psi(B_k(r) - rB_k(1)),$$

where $D_p = (I_k - \sum_{l=1}^{p}A_l)^{-1}$. By K2, PW1, and PW2, $\widehat{Q}_T^{-1} \longrightarrow_p Q^{-1}$ and $\widehat{D}_p \longrightarrow_p D_p$, so we have

$$(T-p)^{-1/2}R\widehat{Q}_T^{-1}\widehat{D}_p\widetilde{S}^{KVB}_{\lfloor rT \rfloor} \Rightarrow \widetilde{\Psi}(B_k(r) - rB_k(1)),$$

where $\widetilde{\Psi} = RQ^{-1}\Psi$ is an $m \times k$ matrix. Using the continuous mapping theorem, we have

$$RV^{KVB}_{T,PW}R' \longrightarrow_D \widetilde{\Psi}\Big(\int_0^1 \big(B_k(r) - rB_k(1)\big)\big(B_k(r) - rB_k(1)\big)' dr\Big)\widetilde{\Psi}'.$$

From (2.3), $T^{1/2}R(\widehat{\beta}_T - \beta) \longrightarrow_D \widetilde{\Psi}B_k(1)$, and the continuous mapping theorem leads to

$$G_{T,PW}^{KVB} \longrightarrow_D B_k(1)'\widetilde{\Psi}'\Big[\widetilde{\Psi}\Big(\int_0^1 \big(B_k(r) - rB_k(1)\big)\big(B_k(r) - rB_k(1)\big)'dr\Big)\widetilde{\Psi}'\Big]^{-1}\widetilde{\Psi}B_k(1). \tag{A.2}$$

The limiting distribution in (A.2) is equivalent to U_m in distribution; see the proof of Theorem 1 of Kiefer et al. (2000). The proof is then complete. $\diamondsuit$

Existence of $\widehat{A}$ and A that satisfy PW1 and PW2:

Here we show PW1 and PW2 can be satisfied with the OLS estimates in (2.4). We consider $p = 2$ for simplicity in the following proof, but it is straightforward to extend the result to the general p. Let $\widehat{A}_1$ and $\widehat{A}_2$ be the OLS estimates in (2.4) with $p = 2$. Consider (2.4) with $\widehat{v}_t$ replaced by v_t as

$$v_t = \widetilde{A}_1 v_{t-1} + \widetilde{A}_2 v_{t-2} + \widetilde{\epsilon}_t, \quad t = 3, \cdots, T,$$

where $\widetilde{A}_1$, $\widetilde{A}_2$, and $\widetilde{\epsilon}_t$ are the corresponding OLS estimators and residuals. Denote $A = (A_1', A_2')'$, $\widehat{A} = (\widehat{A}_1', \widehat{A}_2')'$, and $\widetilde{A} = (\widetilde{A}_1', \widetilde{A}_2')'$, where A_1 and A_2 are defined in PW1. We define $B_1 = E\binom{v_2}{v_1}(\, v_2' \; v_1' \,)$, $B_2 = E\binom{v_2}{v_1} v_3'$, $\widehat{B}_1 = \frac{1}{T}\sum_{t=3}^T \binom{\widehat{v}_{t-1}}{\widehat{v}_{t-2}}(\, \widehat{v}_{t-1}' \; \widehat{v}_{t-2}' \,)$, $\widehat{B}_2 = \frac{1}{T}\sum_{t=3}^T \binom{\widehat{v}_{t-1}}{\widehat{v}_{t-2}} \widehat{v}_t'$, and $\widetilde{B}_1$ and $\widetilde{B}_2$ be the the same as $\widehat{B}_1$ and $\widehat{B}_2$ with $\widehat{v}_t$ replaced by v_t. The OLS estimators $\widehat{A}$ and $\widetilde{A}$ can be written as $\widehat{B}_1^{-1}\widehat{B}_2$ and $\widetilde{B}_1^{-1}\widetilde{B}_2$, respectively. Define $A = B_1^{-1}B_2$.

In what follows we present some primitive assumptions under which PW1 and PW2 hold. Let $Y_t = (u_t, X_t')$ and $X_{t,j}$ be the jth component of X_t..

A1. B_1 is invertible and $I_k - A_1 - A_2$ is nonsingular.

A2. $\{Y_t\}$ is stationary and α-mixing of size $r/(r-1)$, $r > 1$.

A3. $E(X_t'u_t) = \mathbf{0}$, $E|X_{t,j}u_t|^{r+\delta} < \infty$, and $E|X_{t,j}^2|^{r+\delta} < \infty$ for some $\delta > 0$, for $j = 1, \cdots, k$.

A4. $E(X_{t,j_1}X_{t,j_2}X_{t-l,j_3}X_{t-l,j_4}) < \infty$ for all $j_1, \cdots, j_4 = 1, \cdots, k$ and $l = 0, 1, 2$.

We assume A1 to guarantee that A is well-defined so PW2 holds. A2 and A3 are to ensure that $\{v_t\}$ is stationary and we can apply the law of large numbers (LLN) for stationary mixing random variables. See Exercise 3.51 in White (1984).

We claim that $\widetilde{A} - A = o_p(1)$ and $\widehat{A} - \widetilde{A} = o_p(1)$, which lead to $\widehat{A} - A = o_p(1)$ and hence PW1 holds. To prove the first claim, observe that $E(v_t v_{t-l}') = E(v_{l+1} v_1')$ for $t = l+1, \cdots, T$ and $l = 0, 1, 2$ due to the stationarity assumption. By the LLN for weakly dependent stationary sequences, we have $T^{-1} \sum_{t=3}^{T} \{v_t v_{t-l}' - E(v_{l+1} v_1')\} = o_p(1)$ for $l = 0, 1, 2$. It is easy to see that $\widetilde{A} - A = o_p(1)$ by Slutsky's theorem.

To prove $\widehat{A} - \widetilde{A} = o_p(1)$, we need to show that

$$T^{-1} \sum_{t=3}^{T} (\widehat{v}_t \widehat{v}_{t-l} - v_t v_{t-l}') = o_p(1), \qquad l = 0, 1, 2. \tag{A.3}$$

Once (A.3) is verified, we have $\widehat{B}_j - \widetilde{B}_j = o_p(1)$ for $j = 1, 2$. Note that $\widetilde{A} - A = o_p(1)$ and $\widetilde{B}_1 - B_1 = o_p(1)$ using the LLN. Then the desired result follows in view of the relation

$$\widehat{B}_2 - \widetilde{B}_2 = \widehat{B}_1 \widehat{A} - \widetilde{B}_1 \widetilde{A} = \widehat{B}_1(\widehat{A} - \widetilde{A}) + (\widehat{B}_1 - \widetilde{B}_1)\widetilde{A}.$$

To prove (A.3), note that for $l = 0, 1, 2$,

$$\sum_{t=3}^{T} (\widehat{v}_t \widehat{v}_{t-l}' - v_t v_{t-l}') = \sum_{t=3}^{T} X_t X_{t-l}' (\widehat{u}_t \widehat{u}_{t-l} - u_t u_{t-l}).$$

We can rewrite the latter part of the summands as

$$\widehat{u}_t \widehat{u}_{t-l} - u_t u_{t-l} = (\widehat{u}_t - u_t)(\widehat{u}_{t-l} - u_{t-l}) + (\widehat{u}_t - u_t) u_{t-l} + (\widehat{u}_{t-l} - u_{t-l}) u_t.$$

It can be seen that $T^{-1} \sum_{t=3}^{T} X_t X_{t-l}' (\widehat{u}_t - u_t)(\widehat{u}_{t-l} - u_{t-l})$, $T^{-1} \sum_{t=3}^{T} X_t X_{t-l}' (\widehat{u}_t - u_t) u_{t-l}$, and $T^{-1} \sum_{t=3}^{T} X_t X_{t-l}' (\widehat{u}_{t-l} - u_{t-l}) u_t$ are all $o_p(1)$, which are straightforward consequences of the LLN, $\sqrt{T}(\widehat{\beta} - \beta) = O_p(1)$, and the assumptions A3 and A4. The proof is thus complete.

Proof of Theorem 2.3.2. Let $\widehat{S}_t^{SN} = \sum_{j=1}^{t} \widehat{IF}_j = t(\widehat{\theta}_t - \widehat{\theta}_N)$. Under S1, we have

$$N^{-2} \sum_{t=1}^{N} \widehat{S}_t^{SN} \widehat{S}_t^{SN\prime} \longrightarrow_D \Delta \int_0^1 \big(B_k(r) - r B_k(1)\big)\big(B_k(r) - r B_k(1)\big)' dr \Delta'.$$

Using the similar argument as in (A.1) by replacing $\widehat{v}_t$ with $\widehat{IF}_t$ and T with N, we have

$$\begin{aligned}(N-p)^{-1/2}\widetilde{S}^{SN}_{\lfloor rN\rfloor} &= (N-p)^{-1/2}\textstyle\sum_{j=p+1}^{\lfloor rN\rfloor}(\widehat{IF}_j - \sum_{l=1}^{p}\widehat{A}_l\widehat{IF}_{j-l}) \\ &= (N-p)^{-1/2}\big(I_k - \textstyle\sum_{l=1}^{p}A_l\big)\widehat{S}^{SN}_{\lfloor rN\rfloor} + o_p(1).\end{aligned}$$

Under PW1 and PW2, it follows from the continuous mapping theorem that

$$RV^{SN}_{N,PW}R' \longrightarrow_D R\Delta\Big(\int_0^1 \big(B_k(r) - rB_k(1)\big)\big(B_k(r) - rB_k(1)\big)'dr\Big)\Delta' R'.$$

Define an $m\times k$ matrix $\widetilde{\Delta} = R\Delta$. Under the assumption S1, $N^{1/2}(\widehat{\theta}_N - \theta) \longrightarrow_D \Delta B_k(1)$, which implies $N^{1/2}R(\widehat{\theta}_N - \theta) \longrightarrow_D \widetilde{\Delta}B_k(1)$. Thus, by the continuous mapping theorem, we have

$$G^{SN}_{T,PW} \longrightarrow_D B_k(1)'\widetilde{\Delta}'\Big[\widetilde{\Delta}\Big(\int_0^1 \big(B_k(r) - rB_k(1)\big)\big(B_k(r) - rB_k(1)\big)'dr\Big)\widetilde{\Delta}'\Big]^{-1}\widetilde{\Delta}B_k(1) \overset{D}{=} U_m.$$

The proof is complete. $\diamondsuit$

Appendix B

Technical Details for Chapter 3

Denote by P^*, E^*, var^* the probability, expectation, and variance, respectively, conditional on data $\mathcal{X}_n$. Let $\stackrel{\mathcal{D}}{=}$ indicate the equality in distribution. For a vector $x = (x_1, \ldots, x_p)'$, let $||x|| = (\sum_{i=1}^p x_i^2)^{1/2}$ be the Euclidian norm. If A is a $p \times p$ symmetric nonnegative definite matrix with $A = P\Lambda P'$, $PP' = I_p$, where I_p is the $p \times p$ identity matrix, and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ is a diagonal matrix with $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$, we define the root of A as $A^{1/2} = P\Lambda^{1/2}P'$, where $\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_p^{1/2})$. We write $Y_n = o_p^*(1)$ if for any $\epsilon > 0$, $P^*\{|Y_n| > \epsilon\} \to 0$ in probability, and $Y_n = O_p^*(1)$ if there exists a constant $M > 0$ such that for all large n, $P^*\{|Y_n| > M\} < \epsilon$ with probability arbitrarily close to one, as defined in Chang and Park (2003). Define $Q_N(r) = \lfloor Nr \rfloor^{-1} \sum_{t=1}^{\lfloor Nr \rfloor + p - 1} F(t/n)F(t/n)'$ and $B_{N,F}(r) = N^{-1/2} \sum_{t=1}^{\lfloor Nr \rfloor + p - 1} F_{t,n}u_{t,n}$. Define $\overline{\omega} = \max\{\omega^*, \sup_{r\in[0,1]} \omega(r)\}$ and $\overline{\nu} = \sup_{r\in[0,1]} \nu(r)$. Notice that since $\omega(r)$ and $\nu(r)$ have only finitely many breaks on [0,1] interval, $\overline{\omega} < \infty$ by (A1) (i) and $\overline{\nu} < \infty$. In the following argument, all $\sup_r$ indicates supremum over $\{r \in [\epsilon, 1]\}$. The positive constant C is generic and may vary from place to place. We use $\mathcal{I}_j$ or $\mathcal{I}_{N,j}$ for $j = 1, 2, \ldots$ in different proofs to indicate different objects in a simple way.

The following lemmas contain some key results needed in the proofs of Theorems 3.3.1 and 3.3.2 and they may be of independent interest.

Lemma **B.0.1**. *Let* $A_p = (a_{p,ij})_{i,j\leq p}$ *and* $B_p = (b_{p,ij})_{i,j\leq p}$ *be* $p \times p$ *matrices with* $p = 1, 2, \ldots$ *and let* $C_p = \max_{i,j}\{a_{p,ij}^2, b_{p,ij}^2\}$. *Then*

$$\{\det(A_p) - \det(B_p)\}^2 \leq (2p^3C_p)^{p-1}||A_p - B_p||_F^2. \tag{B.1}$$

Proof of Lemma B.0.1. Let $A_{p,ij}$ be the $(p-1)\times(p-1)$ matrix of A_p with ith row and jth column deleted.

If $p = 1$, it is trivial. If (B.1) holds for p, then

$$
\begin{aligned}
|\det(A_{p+1}) - \det(B_{p+1})|^2 &= |\textstyle\sum_{j=1}^{p+1}(-1)^{j+1}\{a_{p+1,1j}\det(A_{p+1,1j}) - b_{p+1,1j}\det(B_{p+1,1j})\}|^2 \\
&\le \textstyle\sum_{j=1}^{p+1} a_{p+1,1j}^2\{\det(A_{p+1,1j}) - \det(B_{p+1,1j})\}^2 + \sum_{j=1}^{p+1}(a_{p+1,1j} - b_{p+1,1j})^2\{\det(B_{p+1,1j})\}^2 \\
&\le (2p^3C_{p+1})^{p-1}\{C_{p+1}\textstyle\sum_{j=1}^{p+1}||A_{p+1,1j} - B_{p+1,1j}||_F^2 + ||A_{p+1} - B_{p+1}||_F^2\sum_{j=1}^{p+1}||B_{p+1,1j}||_F^2\} \\
&\le (2p^3C_{p+1})^{p-1}\{C_{p+1}(p+1)||A_{p+1} - B_{p+1}||_F^2 + ||A_{p+1} - B_{p+1}||_F^2(p+1)p^2C_{p+1}\} \\
&\le \{2(p+1)^3C_{p+1}\}^p||A_{p+1} - B_{p+1}||_F^2.
\end{aligned}
$$

By induction, (B.1) holds for all p. ♢

Lemma **B.0.2**. *Assume (R1)-(R2).*

(i) $\sup_{r\in[\epsilon,1]}||Q_N(r) - Q_r||_F = O(n^{-1})$

(ii) $\sup_{r\in[\epsilon,1]}||Q_r^{-1}||_F < \infty$ *and* $\sup_{r\in[\epsilon,1]}||Q_N(r)^{-1}||_F < \infty$ *for all* N.

(iii) $\sup_{r\in[\epsilon,1]}||Q_N(r)^{-1} - Q_r^{-1}||_F = O(n^{-1})$

Proof of Lemma B.0.2. (i) Our goal is to show that, for all $i, j = 1, \ldots, p$,

$$\sup_r \left| r^{-1}\int_0^{\lfloor nr\rfloor/n} f_i(s)f_j(s)ds - (nr)^{-1}\sum_{t=1}^{\lfloor nr\rfloor} f_i(t/n)f_j(t/n)\right| = O(n^{-1}). \tag{B.2}$$

Once (B.2) is shown, we have $\sup_{r\in[\epsilon,1]}||Q_N(r) - Q_r||_F \le \left\{\sum_{i,j\le p}(Cn^{-1})^2\right\}^{1/2} = pCn^{-1} = O(n^{-1})$ so that the part (i) of this lemma is complete.

Now we prove (B.2). For each $i = 1, \ldots, p$ and $r \in [\epsilon, 1]$, define $\mathcal{S}_{n,i,r} = \{s : s = t/n$ for some $t = 1, \ldots, \lfloor nr\rfloor$, or s is a break point on $s \in [0, \lfloor nr\rfloor/n]\}$. Let $\{s_k\}_{k=1,\ldots,m_{i,r}}$ denote the elements of $\mathcal{S}_{n,i,r}$ in ascending order, where $m_{i,r}$ is the number of elements in $\mathcal{S}_{n,i,r}$. Let $s_0 = 0$. For each i, define a step function $f_{n,i}(s)$ on $s \in [0, 1]$ such that $f_{n,i}(s) = f_i(s_k)$ for $s \in [s_{k-1}, s_k)$, $k = 1, \ldots, m_{i,1}$, and $f_{n,i}(1) = f_i(1)$. Then this step function is very close to $f_i(s)$ in the sense that $\sup_{s\in[0,1]}|f_i(s) - f_{n,i}(s)| \le Cn^{-1}$, due to the piecewise Lipschitz continuity of $f_i(s)$. Then, for all $i, j = 1, \ldots, p$, $|f_i(s)f_j(s) - f_{n,i}(s)f_{n,j}(s)| \le |f_i(s)||f_j(s) - f_{n,j}(s)| + |f_i(s) - f_{n,i}(s)||f_{n,j}(s)|$ and the boundedness of $f_i(s)$ and $f_{n,j}(s)$ imply that

$$\sup_{r\in[\epsilon,1]}\left| r^{-1}\int_0^{\lfloor nr\rfloor/n} f_i(s)f_j(s)ds - r^{-1}\int_0^{\lfloor nr\rfloor/n} f_{n,i}(s)f_{n,j}(s)ds\right| = O(n^{-1}). \tag{B.3}$$

Also, notice that $\sup_r |\sum_{k=1}^{m_{i,r}} f_{n,i}(s_k) - \sum_{t=1}^{\lfloor nr \rfloor} f_i(t/n)| = O(1)$, since the number of breaks of $f_i(s)$ is finite. Similarly, for all $i, j = 1, \ldots, p$, $\sup_r |\sum_{k=1}^{m_{i,j,r}} f_{n,i}(s_k) f_{n,j}(s_k) - \sum_{t=1}^{\lfloor nr \rfloor} f_i(t/n) f_j(t/n)| = O(1)$, where $m_{i,j,r}$ is the number of elements in $\mathcal{S}_{n,i,r} \bigcup \mathcal{S}_{n,j,r}$. Thus we have

$$\sup_{r \in [\epsilon,1]} \left| (nr)^{-1} \sum_{k=1}^{m_{i,j,r}} f_{n,i}(s_k) f_{n,j}(s_k) - (nr)^{-1} \sum_{t=1}^{\lfloor nr \rfloor} f_i(t/n) f_j(t/n) \right| = O(n^{-1}). \tag{B.4}$$

Since $r^{-1} \int_0^{\lfloor nr \rfloor/n} f_{n,i}(s) f_{n,j}(s) ds = (nr)^{-1} \sum_{k=1}^{m_{i,j,r}} f_{n,i}(s_k) f_{n,j}(s_k)$, by the triangular inequality, (B.3) and (B.4) imply (B.2). This completes the proof.

(ii) Let $Q_{r,ij}$ be the $(p-1) \times (p-1)$ matrix is a submatrix of Q_r with ith row and jth column deleted. The (i, j) cofactor of Q_r is defined as $\text{cof}(Q_r, i, j) = (-1)^{i+j} \det(Q_{r,ij})$, which is a function of $(p-1)$ product of $\{r^{-1} \int_0^r f_i(s) ds\}_{1 \le i \le p}$. Thus we can write $\text{cof}(Q_r, i, j) = r^{-(p-1)} \int_0^r \cdots \int_0^r h_{i,j}\{F(s_1), \ldots, F(s_{p-1})\} ds_1 \cdots ds_{p-1}$ for a piecewise continuous function $h_{i,j}$ with at most finite breaks. Let $\overline{h}_{i,j} = \sup_{s_1, \ldots, s_{p-1}} h_{i,j}\{F(s_1), \ldots, F(s_{p-1})\}$ and $\overline{h} = \max_{i,j} \overline{h}_{i,j}$. The (i, j)th element of Q_r^{-1} can be written as

$$\{\det(Q_r)\}^{-1} r^{-(p-1)} \int_0^r \cdots \int_0^r h_{i,j}\{F(s_1), \ldots, F(s_{p-1})\} ds_1 \cdots ds_{p-1},$$

using the same $h_{i,j}$ above, and since $\overline{h}$ is bounded due to the piecewise Lipschitz continuity of f_i's, we have $||Q_r^{-1}||_F \le \{\det(Q_r)\}^{-1} \overline{h} p$ so that $\sup_r ||Q_r^{-1}||_F \le \overline{h} p \sup_{r \in [\epsilon,1]} \{\det(Q_r)\}^{-1}$, which is bounded due to (R1). Similarly, the (i, j)th element of $Q_N(r)^{-1}$ can be written as

$$[\det\{Q_N(r)\}]^{-1} \lfloor Nr \rfloor^{-(p-1)} \sum_{t_1, \ldots, t_{p-1}}^{\lfloor Nr \rfloor + p - 1} h_{i,j}\{F(t_1/n), \ldots, F(t_{p-1}/n)\}$$

so that $||Q_N(r)^{-1}||_F \le [\det\{Q_N(r)\}]^{-1} \overline{h} p$.

Now we need to show $\sup_r [\det\{Q_N(r)\}]^{-1} < \infty$. By Lemma B.0.1, we have $|\det\{Q_N(r)\} - \det(Q_r)| \le C ||Q_N(r) - Q_r||_F$, where C is bounded if all elements of $Q_N(r)$ and Q_r are bounded, which follow from (R1) and Lemma B.0.2 (i). In fact, the inequality holds uniformly over $r \in [\epsilon, 1]$, thus we have $\sup_r |\det\{Q_N(r)\} - \det(Q_r)| = o(1)$ by Lemma B.0.2 (i). It follows from (R1) and Lemma B.0.2 (i) that $\sup_r |[\det\{Q_N(r)\}]^{-1} - \{\det(Q_r)\}^{-1}| \le |\det\{Q_N(r)\} - \det(Q_r)| |\det\{Q_N(r)\}|^{-1} |\det(Q_r)|^{-1} = o(1)$. Thus $\sup_r [\det\{Q_N(r)\}]^{-1} \le \sup_r \{\det(Q_r)\}^{-1} + 1 < \infty$ for $N \ge N_0$ for some N_0.

(iii) Note that for any $p \times p$ matrices A and B, we have $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$. Due to the sub-multiplicative property of the Frobenius norm, we have $||Q_N(r)^{-1} - Q_r^{-1}||_F \le ||Q_N(r)^{-1}||_F ||Q_N(r) - Q||_F ||Q_r^{-1}||_F$. By Lemma B.0.2 (i) and (ii), $\sup_r ||Q_N(r)^{-1} - Q_r^{-1}||_F \le C n^{-1}$, which completes the proof.

$\diamondsuit$

REMARK **B.0.1**. Notice that the proof of Lemma B.0.2 (i) only works for $\lfloor nr \rfloor \geq 1$, which means this proof does not go through in general if we work on $r \in [0,1]$. For some special cases, e.g., $p = 1$ and $f_1(s) = c$ with constant c, then the uniformness can be relaxed to be over [0,1].

In the following proof of the invariance principle of the error process when the linear process (A1) is assumed, we use the BN decomposition (Beveridge and Nelson, 1981) technique introduced in Phillips and Solo (1992). The BN decomposition is

$$\mathbf{C}(L) = \mathbf{C}(1) - (1-L)\widetilde{\mathbf{C}}(L),$$

where $\widetilde{\mathbf{C}}(L) = \sum_{j=0}^{\infty} \widetilde{c}_j L^j$, $\widetilde{c}_j = \sum_{k=j+1}^{\infty} c_k$. Then the original error process can be decomposed into

$$u_{t,n} = \mathbf{C}(L)\varepsilon_{t,n} = \mathbf{C}(1)\varepsilon_{t,n} + \widetilde{\varepsilon}_{t-1,n} - \widetilde{\varepsilon}_{t,n}, \tag{B.5}$$

where $\widetilde{\varepsilon}_{t,n} = \widetilde{\mathbf{C}}(L)\varepsilon_{t,n} = \sum_{j=0}^{\infty} \widetilde{c}_j \varepsilon_{t-j,n}$.

LEMMA **B.0.3**. *Recall that $B_{N,F}(r) = N^{-1/2}\sum_{t=1}^{\lfloor Nr \rfloor + p-1} F_{t,n} u_{t,n}$. Assume (R1)-(R2). Then there exist iid standard normal random variables $V_1, \ldots, V_n$ and a centered Gaussian process $\overline{B}_{N,F}$ such that $\{\overline{B}_{N,F}(\lfloor N\epsilon \rfloor/N), \cdots, \overline{B}_{N,F}(N/N)\}$ have the same distribution as $\{B_{N,F}(\lfloor N\epsilon \rfloor/N), \cdots, B_{N,F}(N/N)\}$, and*

$$\sup_{r\in[\epsilon,1]} \left|\left|\overline{B}_{N,F}(r) - \widehat{B}_{N,F}(r)\right|\right| = o_p(1),$$

where $\widehat{B}_{N,F}(r) = \mathbf{C}(1)N^{-1/2}\sum_{t=1}^{\lfloor Nr \rfloor + p-1} F_{t,n}\omega_{t,n}V_t$ if (A1) and $\widehat{B}_{N,F}(r) = \Gamma N^{-1/2}\sum_{t=1}^{\lfloor Nr \rfloor + p-1} F_{t,n}\nu_{t,n}V_t$ if (A2).

Proof of Lemma B.0.3. We first prove for (A1). If $p = 1$ and $F(\cdot) \equiv 1$, from the BN decomposition (B.5), we have $\sum_{t=1}^{\lfloor nr \rfloor} u_{t,n} = C(1)\sum_{t=1}^{\lfloor nr \rfloor} \varepsilon_{t,n} + \widetilde{\varepsilon}_{0,n} - \widetilde{\varepsilon}_{\lfloor nr \rfloor,n}$ so that

$$\sup_{r\in[\epsilon,1]} \left| N^{-1/2} \sum_{t=1}^{\lfloor Nr \rfloor + p-1} u_{t,n} - C(1)N^{-1/2} \sum_{t=1}^{\lfloor Nr \rfloor + p-1} \omega_{t,n} e_t \right| \leq \frac{|\widetilde{\varepsilon}_{0,n}|}{N^{1/2}} + \frac{\sup_{r\in[\epsilon,1]} |\widetilde{\varepsilon}_{\lfloor Nr \rfloor + p-1,n}|}{N^{1/2}}. \tag{B.6}$$

Our goal is to show that the left-hand side of (B.6) is $o_p(1)$, which holds if

$$\widetilde{\varepsilon}^2_{0,n} = O_p(1) \tag{B.7}$$

and

$$\max_{1\leq k\leq n} n^{-1/2}|\widetilde{\varepsilon}_{k,n}| \xrightarrow{\mathcal{P}} 0. \tag{B.8}$$

Notice that $\sum_{j=0}^{\infty} |\widetilde{c}_j|^{\kappa} < \infty$ for any $\kappa \geq 1$ by (A1) (ii) and Lemma 2.1 of Phillips and Solo (1992). Since $E(\widetilde{\varepsilon}_{0,n}^2) = \sum_{j=0}^{\infty} \widetilde{c}_j^2 \omega_{-j,n}^2 \leq \overline{\omega}^2 \sum_{j=0}^{\infty} \widetilde{c}_j^2 < \infty$, (B.7) holds using the Markov inequality. For (B.8), by Minkowski's inequality, $\sup_{1\leq k\leq n} E|\widetilde{\varepsilon}_{k,n}|^4 \leq \sup_{1\leq k\leq n}(\sum_{j=0}^{\infty} ||\widetilde{c}_j \omega_{k-j,n} e_{k-j}||_4)^4 \leq C\overline{\omega}^4(\sum_{j=0}^{\infty} |\widetilde{c}_j|)^4 < \infty$. Then for any $M > 0$,

$$\begin{aligned} P(\max_{1\leq k\leq n} |\widetilde{\varepsilon}_{k,n}| > n^{1/2}M) &= P(\textstyle\bigcup_{k=1}^{n}\{|\widetilde{\varepsilon}_{k,n}| > n^{1/2}M\}) \leq \sum_{k=1}^{n} P(|\widetilde{\varepsilon}_{k,n}| > n^{1/2}M) \\ &\leq (n^{1/2}M)^{-4} \textstyle\sum_{k=1}^{n} E|\widetilde{\varepsilon}_{k,n}|^4 = O(n^{-1}) \end{aligned}$$

by Markov inequality, and (B.8) follows.

Also, since e_t are iid (0,1) with finite fourth moment, there exist iid N(0,1) random variables $\{V_1, \ldots, V_n\}$ on a richer probability space such that for any $M > 0$,

$$P\left\{\max_{1\leq k\leq n}\left|\sum_{t=1}^{k}\omega_{t,n}e_t - \sum_{t=1}^{k}\omega_{t,n}V_t\right| \geq n^{1/2}M\right\} \leq C(n^{1/2}M)^{-4}\sum_{t=1}^{n} E|e_t|^4 = O(n^{-1}) \tag{B.9}$$

using, for example, Lemma 2 of Csörgő et al. (2003). Then by (B.6), (B.7), (B.8), and (B.9), the result follows for (A1) with $p = 1$ and $F(\cdot) \equiv 1$.

For a general p and $F(\cdot)$, notice that $f_j(\cdot)$ is piecewise Lipschitz continuous for each $j = 1, \ldots, p$. If $\omega_{t,n}$ is replaced with $\omega_{t,n} f_j(t/n)$, the same proof goes through for all $j = 1, \ldots, p$, and we have

$$\sup_r ||B_{N,F}(r) - C(1)N^{-1/2}\textstyle\sum_{t=1}^{\lfloor Nr\rfloor+p-1} F(t/n)\omega_{t,n}V_t|| = o_p(1).$$

For (A2), if $p = 1$ and $F(\cdot) \equiv 1$, our model is a special case of Zhou (2013)'s setting and the result follows by Proposition 5 of Zhou (2013). For a general p and $F(\cdot)$, the same proof works for $\nu_{t,n} f_j(t/n)$, in place of $\nu_{t,n}$, for all $j = 1, \ldots, p$, which completes the proof. ♢

Proof of Theorem 3.3.1. We first prove for (A1). Define a step function $\omega_n(s) = \omega(t/n)$ for $s \in [t/n, (t+1)/n)$ for each $t = 0, \ldots, n-1$ if $\omega(s)$ has no break in the interval. If $\omega(s)$ has k break points $s_1, \ldots, s_k$ in an interval $(t/n, (t+1)/n)$, i.e., $s_0 := t/n < s_1 < \ldots < s_k < s_{k+1} := (t+1)/n$, then define $\omega_n(s) = \omega(s_i)$ for $s \in [s_i, s_{i+1}]$ and $i = 0, \ldots, k$. Let $\omega_n(1) = \omega(1)$. Similarly, for all $i = 1, \ldots, p$, define step functions $f_{n,i}(s)$ from $f_i(s)$, and let $F_n(s) = \{f_{n,1}(s), \ldots, f_{n,p}(s)\}'$. Let $\check{B}_{N,F,\mathbf{C},\omega}(r) = \mathbf{C}(1)\int_0^{(\lfloor Nr\rfloor+p-1)/n} F_n(s)\omega_n(s)dB(s)$, $\bar{B}_{N,F,\mathbf{C},\omega}(r) = \mathbf{C}(1)\int_0^r F_n(s)\omega_n(s)dB(s)$, and $B_{F,\mathbf{C},\omega}(r) = \mathbf{C}(1)\int_0^r F(s)\omega(s)dB(s)$.

Due to the triangular inequality of the sup norm, $\sup_r ||\check{B}_{N,F,\mathbf{C},\omega}(r) - B_{F,\mathbf{C},\omega}(r)|| \leq \sup_r ||\check{B}_{N,F,\mathbf{C},\omega}(r) -$

$\bar{B}_{N,F,\mathbf{C},\omega}(r)||+\sup_r ||\bar{B}_{N,F,\mathbf{C},\omega}(r)-B_{F,\mathbf{C},\omega}(r)|| =: \mathcal{I}_1+\mathcal{I}_2$. We have $\mathcal{I}_1 = \sup_r ||\check{B}_{N,F,\mathbf{C},\omega}(r)-\bar{B}_{N,F,\mathbf{C},\omega}(r)|| = o_p(1)$ because $\sup_r |r-(\lfloor Nr\rfloor+p-1)/n| \le p/n$ and $\sup_r || \int_{(\lfloor Nr\rfloor+p-1)/n}^{r} F_n(s)\omega_n(s)dB(s)|| \le C(p+1)\sup_{t=1,\cdots,n}|B(t/n)-B\{(t-1)/n\}| = o_p(1)$. Due to the way $\omega_n(s)$ is defined in the beginning of this proof, we have $\sup_r |\omega_n(r)-\omega(r)| = o(1)$ so that $\sup_r ||F_n(r)\omega_n(r)-F(r)\omega(r)|| = o(1)$ is implied by (R1). Therefore, $\mathcal{I}_2 = \sup_r ||\bar{B}_{N,F,\omega}(r)-B_{F,\omega}(r)|| = o_p(1)$ holds by Proposition 5.19 of Kurtz (2001). Then we have

$$\sup_{r\in[\epsilon,1]} ||\check{B}_{N,F,\mathbf{C},\omega}(r)-B_{F,\mathbf{C},\omega}(r)|| = o_p(1). \tag{B.10}$$

Note that $\sup_{s\in[0,1]} |f_i(s)\omega(s)| < C$ for some constant C which does not depend on i so that we have $\sup_r ||B_{F,\mathbf{C},\omega}(r)||^2 \le \mathbf{C}(1)^2 \sum_{i=1}^{p} \sup_r \left\{\int_0^r f_i(s)\omega(s)dB(s)\right\}^2 = O_p(1)$. By the triangular inequality of the Euclidean norm and the sub-multiplicative property of the Frobenius norm, we have $||Q_N(r)^{-1}\check{B}_{N,F,\mathbf{C},\omega}(r)- Q_r^{-1}B_{F,\mathbf{C},\omega}(r)|| \le ||Q_N(r)^{-1}-Q_r^{-1}||_F||B_{F,\mathbf{C},\omega}(r)|| + ||Q_N(r)^{-1}||_F||\check{B}_{N,F,\mathbf{C},\omega}(r)-B_{F,\mathbf{C},\omega}(r)||$. Thus by (B.10) and Lemma B.0.2 (ii) and (iii), we have

$$\sup_{r\in[\epsilon,1]} ||Q_N(r)^{-1}\check{B}_{N,F,\mathbf{C},\omega}(r)-Q_r^{-1}B_{F,\mathbf{C},\omega}(r)|| = o_p(1). \tag{B.11}$$

From Lemma B.0.3, there exist iid standard normal random variables $V_1,\ldots,V_n$ and a centered Gaussian process $\overline{B}_{N,F}$ such that $\{\overline{B}_{N,F}(\lfloor N\epsilon\rfloor/N),$
$\cdots,\overline{B}_{N,F}(N/N)\}$ have the same distribution as $\{B_{N,F}(\lfloor N\epsilon\rfloor/N),\cdots,B_{N,F}(N/N)\}$, and $\sup_{r\in[\epsilon,1]}||\overline{B}_{N,F}(r)$ $-\widehat{B}_{N,F}(r)|| = o_p(1)$, where $\widehat{B}_{N,F}(r) = \mathbf{C}(1)N^{-1/2}\sum_{t=1}^{\lfloor Nr\rfloor+p-1} F_{t,n}\omega_{t,n}V_t$. Thus we have from Lemma B.0.2 (ii)

$$\sup_{r\in[\epsilon,1]} \left|\left|Q_N(r)^{-1}\overline{B}_{N,F}(r)-Q_N(r)^{-1}\widehat{B}_{N,F}(r)\right|\right| = o_p(1). \tag{B.12}$$

Notice that

$$\{\widehat{B}_{N,F}(r)\}_{r=\epsilon}^1 \overset{\mathcal{D}}{=} \left\{\mathbf{C}(1)\sum_{t=1}^{\lfloor Nr\rfloor+p-1} F(t/n)\omega(t/n)\left[B(t/n)-B\{(t-1)/n\}\right]\right\}_{r=\epsilon}^{1} \overset{\mathcal{D}}{=} \left\{\check{B}_{N,F,\mathbf{C},\omega}(r)\right\}_{r=\epsilon}^{1}. \tag{B.13}$$

By (B.11) and (B.13), $Q_N(r)^{-1}\widehat{B}_{N,F}(r) \Rightarrow Q_r^{-1}B_{F,\mathbf{C},\omega}(r)$. Then (3.4) follows from (B.12) and the fact that $\{Q_N(r)^{-1}B_{N,F}(r)\}_{r=\epsilon}^1 \overset{\mathcal{D}}{=} \{Q_N(r)^{-1}\overline{B}_{N,F}(r)\}_{r=\epsilon}^1$. The proof is complete for (A1).

The same proof goes for (A2) by replacing $\omega(s)$ with $\nu(s)$ and $\mathbf{C}(1)$ with Γ. ♢

Now we prove the bootstrap consistency of the SN quantity. The following lemmas present some key steps.

LEMMA **B.0.4**. *If $\{u_{t,n}\}$ are generated from (A1),*

$$\sup_{r\in[\epsilon,1]}\left|n^{-1}\sum_{t=1}^{\lfloor nr\rfloor}u_{t,n}^2-\mathbf{D}(1)\int_0^r\omega^2(s)ds\right|=o_p(1),$$

where $\mathbf{D}(1)=\sum_{j=0}^{\infty}c_j^2$,

Proof of Lemma B.0.4. We define some notations first. Let $\mathbf{D}_j(L)=\sum_{k=0}^{\infty}c_kc_{k+j}L^k=\sum_{k=0}^{\infty}d_{jk}L^k$ for $j=0,1,\ldots$, where $d_{jk}=c_kc_{k+j}$. Define $\widetilde{\mathbf{D}}_j(L)$ from $\mathbf{D}_j(L)$ for each j in the same way as we define $\widetilde{\mathbf{C}}(L)$ from $\mathbf{C}(L)$, i.e., $\widetilde{\mathbf{D}}_j(L)=\sum_{k=0}^{\infty}\widetilde{d}_{jk}L^k$ and $\widetilde{d}_{jk}=\sum_{s=k+1}^{\infty}d_{js}=\sum_{s=k+1}^{\infty}c_sc_{s+j}$. Denote by $\mathbf{D}(L)=\mathbf{D}_0(L)$, which is consistent with the above definition for $\mathbf{D}(1)$. Let $\varepsilon_{t-1,n}^D=\sum_{j=1}^{\infty}\mathbf{D}_j(1)\varepsilon_{t-j,n}$, $\widetilde{u}_{at,n}=\widetilde{\mathbf{D}}_0(L)\varepsilon_{t,n}^2$, and $\widetilde{u}_{bt,n}=\sum_{j=1}^{\infty}\widetilde{\mathbf{D}}_j(L)\varepsilon_{t,n}\varepsilon_{t-j,n}$. Using the BN decomposition argument of (Phillips and Solo, 1992, pp. 978–9), we can write

$$u_{t,n}^2=\mathbf{D}(1)\varepsilon_{t,n}^2-(1-L)\widetilde{u}_{at,n}+2\varepsilon_{t,n}\varepsilon_{t-1,n}^D-2(1-L)\widetilde{u}_{bt,n}$$

and $n^{-1}\sum_{t=1}^{\lfloor nr\rfloor}u_{t,n}^2=\mathbf{D}(1)n^{-1}\sum_{t=1}^{\lfloor nr\rfloor}\varepsilon_{t,n}^2-n^{-1}(\widetilde{u}_{a\lfloor nr\rfloor}-\widetilde{u}_{a0})+2n^{-1}\sum_{t=1}^{\lfloor nr\rfloor}\varepsilon_{t,n}\varepsilon_{t-1,n}^D-2n^{-1}(\widetilde{u}_{b\lfloor nr\rfloor}-\widetilde{u}_{b0})=\mathcal{I}_1+\mathcal{I}_2+\mathcal{I}_3+\mathcal{I}_4$. Our goal is to show that

$$\sup_{r\in[\epsilon,1]}|\mathcal{I}_1-\mathbf{D}(1)\int_0^r\omega^2(s)ds|=o_p(1)\tag{B.14}$$

and

$$\sup_{r\in[\epsilon,1]}|\mathcal{I}_j|=o_p(1),\qquad j=2,3,4.\tag{B.15}$$

To prove (B.14), we need to show

$$\sup_{r\in[\epsilon,1]}|n^{-1}\sum_{t=1}^{\lfloor nr\rfloor}(\varepsilon_{t,n}^2-\omega_{t,n}^2)|=o_p(1).$$

Since $\{\varepsilon_{t,n}\}_{t=1}^n$ are independent, using Kolmogorov's inequality, $P(\max_{1\le k\le n}|\sum_{t=1}^k(\varepsilon_{t,n}^2-\omega_{t,n}^2)|>nM)\le(nM)^{-2}\sum_{k=1}^n\mathrm{var}(\varepsilon_{k,n}^2-\omega_{k,n}^2)=(nM)^{-2}\sum_{k=1}^n\omega_{k,n}^4(Ee_k^4-1)=O(n^{-1})$. Thus (B.14) holds.

Now we show (B.15). For $j=2$, observe that $E|\widetilde{\mathbf{D}}_0(L)\varepsilon_{t,n}^2|^2=E|\sum_{k=0}^{\infty}\widetilde{d}_{0k}\varepsilon_{t-k,n}^2|^2=\sum_{k=0}^{\infty}\widetilde{d}_{0k}^2E\varepsilon_{t-k,n}^4+\sum_{k\neq k'}\widetilde{d}_{0k}\widetilde{d}_{0k'}E\varepsilon_{t-k,n}^2E\varepsilon_{t-k',n}^2\le C\{\sum_{k=0}^{\infty}\widetilde{d}_{0k}^2+(\sum_{k=0}^{\infty}\widetilde{d}_{0k})^2\}<\infty$, where the last inequality is due to (A1) (ii) and Lemma 2.1 of Phillips and Solo (1992). Then for any $M>0$, $P(\sup_{r\in[\epsilon,1]}|n^{-1}\widetilde{u}_{a\lfloor nr\rfloor,n}|>M)\le P(\max_{1\le t\le n}|\widetilde{\mathbf{D}}_0(L)\varepsilon_{t,n}^2|>nM)\le\sum_{t=1}^nP(|\widetilde{\mathbf{D}}_0(L)\varepsilon_{t,n}^2|>nM)\le(nM)^{-2}\sum_{t=1}^n E|\widetilde{\mathbf{D}}_0(L)\varepsilon_{t,n}^2|^2=O(n^{-1})$, which proves (B.15) for $j=2$.

For $j=3$, notice that $\{\varepsilon_{t,n}\varepsilon_{t-1,n}^D\}$ is a martingale difference sequence with respect to $\{\mathcal{E}_t\}$, where $\mathcal{E}_t=\sigma(e_t,e_{t-1},\ldots)$ is the sigma field generated by $\{e_j\}_{j=-\infty}^t$, which implies that $E(\sum_{t=1}^n\varepsilon_{t,n}\varepsilon_{t-1,n}^D)^2=\sum_{t=1}^nE(\varepsilon_{t,n}\varepsilon_{t-1,n}^D)^2=O(n)$ because $\varepsilon_{t,n}$ and $\varepsilon_{t-1,n}^D$ are independent and have finite second moments, i.e., $E(\varepsilon_{t-1,n}^D)^2 = E(\sum_{j=1}^{\infty} D_j(1)\varepsilon_{t-j,n})^2 \le C\sum_{j=1}^{\infty} D_j(1)^2 < \infty$ due to (A1) (ii), (iii) and Lemma 3.6 of Phillips and Solo (1992). Then by Doob's martingale inequality, for any $M > 0$, $P\{\max_k |\sum_{t=1}^k \varepsilon_{t,n}\varepsilon_{t-1,n}^D| > nM\} \le (nM)^{-2}E(\sum_{t=1}^n \varepsilon_{t,n}\varepsilon_{t-1,n}^D)^2 = O(n^{-1})$.

For $j = 4$, $P(\sup_r |\widetilde{u}_{b\lfloor nr\rfloor,n}| > nM) \le (nM)^{-2}\sum_{t=1}^n E\widetilde{u}_{bt,n}^2 = O(n^{-1})$ if $\sup_{1\le t\le n} E\widetilde{u}_{bt,n}^2 < \infty$, which holds by the same argument used in the proof of Lemma 5.9 of Phillips and Solo (1992).

Thus (B.15) holds and the proof is complete. $\diamondsuit$

Lemma **B.0.5**. *Assume (B1). Fix $r \in [\epsilon, 1]$. If $\{u_{t,n}\}$ are generated from (A1),*

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t u_{t,n} = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t \mathbf{C}(L)\varepsilon_{t,n} \xrightarrow{\mathcal{D}} N\left(0, \mathbf{D}(1)\int_0^r \omega^2(s)ds\right)$$

in probability and if (A2),

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t u_{t,n} = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t \eta_t \nu_{t,n} \xrightarrow{\mathcal{D}} N\left(0, \int_0^r \nu^2(s)ds\right)$$

in probability.

Proof of Lemma B.0.5. We first prove for (A1), which is complete once we show the Lindeberg condition

$$n^{-1}\sum_{t=1}^{n} E^*\{W_t^2 u_{t,n}^2 \mathbf{1}(|W_t u_{t,n}| > n^{1/2}\delta)\} = o_p(1) \tag{B.16}$$

and the uniform consistency of the conditional variance

$$\sup_{r\in[\epsilon,1]}\left|\text{var}^*\left(n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t u_{t,n}\right) - \mathbf{D}(1)\int_0^r \omega^2(s)ds\right| = o_p(1). \tag{B.17}$$

Notice that $u_{t,n}^4 = O_p(1)$ since for any $t = 1, \ldots, n$, by Minkowski's inequality, $E(u_{t,n}^4) = E(\sum_{j=0}^{\infty} c_j\varepsilon_{t-j,n})^4 \le (\sum_{j=0}^{\infty} ||c_j\varepsilon_{t-j,n}||_4)^4 \le \overline{\omega}^4(Ee_0^4)(\sum_{j=0}^{\infty}|c_j|)^4 < \infty$, which is due to (A1) (ii) and (iii). Then $n^{-1}\sum_{t=1}^n E^*\{W_t^2 u_{t,n}^2 \mathbf{1}(|W_t u_{t,n}| > n^{1/2}\delta)\} \le (n\delta)^{-2}\sum_{t=1}^n E^*\{W_t^4 u_{t,n}^4 \mathbf{1}(|W_t u_{t,n}| > n^{1/2}\delta)\} \le (n\delta)^{-2}\sum_{t=1}^n E^*(W_t^4 u_{t,n}^4) \le (n\delta)^{-2}C\sum_{t=1}^n u_{t,n}^4 = O_p(n^{-1})$ so that the Lindeberg condition (B.16) holds. The uniform consistency of the conditional variance (B.17) follows from Lemma B.0.4 because $\text{var}^*(n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t u_{t,n}) = n^{-1}\sum_{t=1}^{\lfloor nr\rfloor} u_{t,n}^2$. Thus the proof is complete for (A1).

Now we prove for (A2). Notice that $E(u_{t,n}^4) = E(\nu_{t,n}^4\eta_t^4) \le \overline{\nu}^4 E(\eta_t^4) = \overline{\nu}^4 E(\eta_0^4) < \infty$ holds uniformly over all t, since $\overline{\nu} = \sup_{r\in[0,1]}\nu(s) < \infty$ and η_t is stationary with finite fourth moment. Thus by using the same argument as above, the Lindeberg condition (B.16) follows. Now we show the uniform consistency of the variance

$$\sup_{r\in[\epsilon,1]}\left|\text{var}^*\left(n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t \nu_{t,n}\eta_t\right) - \int_0^r \nu^2(s)ds\right| = o_p(1). \tag{B.18}$$

Since $\text{var}^*(n^{-1/2}\sum_{t=1}^{\lfloor nr \rfloor} W_t \nu_{t,n}\eta_t) = n^{-1}\sum_{t=1}^{\lfloor nr \rfloor} \nu_{t,n}^2\eta_t^2$, (B.18) follows if

$$\sup_{r\in[\epsilon,1]} \left| n^{-1/2}\sum_{t=1}^{\lfloor nr \rfloor} \nu_{t,n}^2(\eta_t^2-1)\right| = o_p(1) \tag{B.19}$$

is shown. Notice that $\max_t \nu_{t,n}^2 \leq \overline{\nu}^2$ so we can ignore $\nu_{t,n}$ part in (B.19). Define $S_k = \sum_{t=1}^k(\eta_t^2-1)$ and $S_k^* = \max_{1\leq t\leq k}|S_t|^2$. Let d be a smallest positive integer such that $n \leq 2^d$ and $q = 2+\delta/2$ wit $\delta > 0$ as in (A2) (ii). For some $m_1 \leq m_2$, by Minkowski's inequality, $||\sum_{t=m_1+1}^{m_2}(\eta_t^2-1)||_q \leq \sum_{t=m_1+1}^{m_2}||(\eta_t^2-1)||_q \leq C(m_2-m_1)$ since $||(\eta_0^2-1)||_q < \infty$ due to (A2) (ii). Thus we have, by Proposition 1(i) in Wu (2007), $E(\max_{1\leq k\leq n}|\sum_{t=1}^k(\eta_t^2-1)|)^q \leq ||S_{2^d}^*||_q^q \leq \{\sum_{z=0}^d(\sum_{m=1}^{2^{d-z}}||\sum_{t=2^z(m-1)+1}^{2^z m}(\eta_t^2-1)||_q^q)^{1/q}\}^q \leq C\{\sum_{z=0}^d(2^{d-z}2^z)^{1/q}\}^q = (d+1)^q 2^d = O\{(\log n)^q n\}$. Thus, for any $M>0$, $P(\sup_{r\in[\epsilon,1]}|\sum_{t=1}^{\lfloor nr\rfloor}(\eta_t^2-1)| > n^{1/2}M) \leq n^{-q/2}M^{-q}E(\max_{1\leq k\leq n}|\sum_{t=1}^k(\eta_t^2-1)|)^q = O\{(\log n)^q n^{1-q/2}\} = O\{(\log n)^{2+\delta/2}n^{-\delta/4}\} = o(1)$. Now (B.19) is proved, which completes the proof for (A2). ♢

Lemma **B.0.6**. *Assume (B1). If $\{u_{t,n}\}$ are generated from (A1) or (A2), then for $\epsilon \leq r_1 < r_2 \leq 1$ and $n \geq n_0$ for some positive integer n_0,*

$$E^*\left|n^{-1/2}\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} W_t u_{t,n}\right|^4 \leq \overline{C}(\mathcal{X}_n)\{(r_2-r_1)^2 + n^{-1}(r_2-r_1)\},$$

where $\overline{C}(\mathcal{X}_n) = O_p(1)$.

Proof of Lemma B.0.6. Notice that W_t's are iid (0,1),

$$\begin{aligned} n^{-2}E^*\left|\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} W_t u_{t,n}\right|^4 &\leq n^{-2}\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} E^*|W_t u_{t,n}|^4 + n^{-2}\left\{\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} E^*|W_t u_{t,n}|^2\right\}^2 \\ &\leq Cn^{-2}\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} u_{t,n}^4 + n^{-2}\left(\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} u_{t,n}^2\right)^2 := \mathcal{I}_1 + \mathcal{I}_2. \end{aligned}$$

Since $E(\mathcal{I}_1) = Cn^{-2}\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} Eu_{t,n}^4 \leq C\overline{\omega}^4 Ee_0^4 n^{-2}(\lfloor nr_2\rfloor - \lfloor nr_1\rfloor) = O(1)n^{-1}(r_2-r_1)$ and $E(\mathcal{I}_2^{1/2}) = n^{-1}\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} Eu_{t,n}^2 \leq C\overline{\omega}^4 n^{-1}(\lfloor nr_2\rfloor - \lfloor nr_1\rfloor) = O(1)(r_2-r_1)$, we have $\mathcal{I}_1+\mathcal{I}_2 = O_p(1)n^{-1}(r_2-r_1) + O_p(1)(r_2-r_1)^2$. Thus the proof is complete. ♢

Lemma **B.0.7**. *Assume (R1)-(R2) and (B1).*

(I) Assume (A1). Then

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t F_{t,n}u_{t,n} = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t F_{t,n}C(L)\varepsilon_t \Rightarrow \{\mathbf{D}(1)\}^{1/2}B_{F,\omega}(r) \qquad \textit{in probability}$$

(II) Assume (A2). Then

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t F_{t,n} u_{t,n} = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t F_{t,n}\nu_{t,n}\eta_t \Rightarrow B_{F,\nu}(r) \qquad \textit{in probability.}$$

Proof of Lemma C.0.15. First we show

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t u_{t,n} \Rightarrow \int_0^r \varpi(s)dB(s) \qquad \text{in probability,} \tag{B.20}$$

where $\varpi(s) = \{\mathbf{D}(1)\}^{1/2}\omega(s)$ for (A1) and $\varpi(s) = \nu(s)$ for (A2). The finite-dimensional convergence of (B.20),

$$\left(n^{-1/2}\sum_{t=1}^{\lfloor nr_1\rfloor} W_t u_{t,n}, \ldots, n^{-1/2}\sum_{t=1}^{\lfloor nr_k\rfloor} W_t u_{t,n}\right) \xrightarrow{\mathcal{D}} \left\{\int_0^{r_1}\varpi(s)dB(s), \ldots, \int_0^{r_k}\varpi(s)dB(s)\right\}$$

for any $k \in \mathbb{N}$ and $r_1, \ldots, r_k$, follows from a similar argument presented in Lemma B.0.5 and the Cramér-Wold device in $D[\epsilon, 1]$ space (see Theorem 29.16 of Davidson (1994)). The tightness follows from Lemma B.0.6 and the argument of Theorem 2.1 in Shao and Yu (1996). This completes the proof of (B.20).

Since $f_j(s)$ are piecewise Lipschitz continuous, the above argument works if $\omega(s)$ is replaced by $\omega(s)f_j(s)$ for (A1) and $\nu(s)$ is replaced by $\nu(s)f_j(s)$ for (A2), respectively, for all $j = 1, \ldots, p$. Then using the Cramér-Wold device, the proof is now complete. ♢

Proof of Theorem 3.3.2. Define $Q_{N,W}(r) = N^{-1/2}\sum_{t=1}^{\lfloor Nr\rfloor+p-1} F_{t,n}F'_{t,n}W_t$, $B_{F,\varpi}(r) = \int_0^r F(s)\varpi(s)dB(s)$, and $\varpi(s) = \{\mathbf{D}(1)\}^{1/2}\omega(s)$ if (A1) and $\varpi(s) = \nu(s)$ if (A2), as defined in the above proof for Lemma C.0.15. Observe that

$$N^{-1/2}\lfloor Nr\rfloor(\widehat{\beta}^*_{\lfloor Nr\rfloor} - \widehat{\beta}_N) = \mathcal{I}_{N,1}(r) + \mathcal{I}_{N,2}(r), \tag{B.21}$$

where $\mathcal{I}_{N,1}(r) = \{Q_N(r)\}^{-1}(N^{-1/2}\sum_{t=1}^{\lfloor Nr\rfloor+p-1} F_t u_{t,n}W_t)$ and $\mathcal{I}_{N,2}(r) = \{Q_N(r)\}^{-1}Q_{N,W}(r)(\beta - \widehat{\beta}_N)$. From Lemma C.0.15 and Lemma B.0.1 (i), we have

$$\mathcal{I}_{N,1}(r) \Rightarrow Q_r^{-1}B_{F,\varpi}(r) \qquad \text{in probability.} \tag{B.22}$$

Let $Q^{i,j}_{N,W}(r) = N^{-1/2}\sum_{t=1}^{\lfloor Nr\rfloor} W_{t,ij}$ be the (i,j)th component of $Q_{N,W}(r)$, where $W_{t,ij} = f_i(t/n)f_j(t/n)W_t$. Define $S_{k,ij} = \sum_{t=1}^{k} W_{t,ij}$ and $S^*_{k,ij} = \max_{t\le k}|S_{t,ij}|^2$. Let d be a smallest positive integer such that $n \le 2^d$. By Proposition 1(i) in Wu (2007) and the fact that $W_{t,ij}$ is an iid process, we have $E\max_{k\le n} S^2_{k,ij} \le ||S^*_{2^d}||_2^2 \le \{\sum_{z=0}^{d}(\sum_{m=1}^{2^{d-z}}||\sum_{t=2^z(m-1)+1}^{2^z m} W_{t,ij}||_2^2)^{1/2}\}^2 \le C(d+1)^2 2^d \le C(\log n)^2 n$, where the constant C does not depend on i or j. Thus we have $\sup_{r\in[\epsilon,1]}\{Q^{i,j}_{N,W}(r)\}^2 = N^{-1}\max_{k\le n} S^2_{k,ij} = O^*_p\{(\log n)^2\}$ uniformly on i and j, which leads to

$$\sup_{r\in[\epsilon,1]}||Q_{N,W}(r)||_F = O^*_p\{(\log n)^2\}. \tag{B.23}$$

Then by the sub-multiplicativity of the Frobenius norm, Lemma B.0.2 (ii), (B.23), and Theorem 3.3.1, we have

$$\sup_{r\in[\epsilon,1]} ||\mathcal{I}_{N,2}(r)|| \leq \sup_{r\in[\epsilon,1]} ||Q_N(r)^{-1}||_F \sup_{r\in[\epsilon,1]} ||Q_{N,W}(r)||_F ||\beta - \widehat{\beta}_N|| = O_p^*\{(\log n)^2 n^{-1/2}\} = o_p^*(1). \quad \text{(B.24)}$$

Then (3.5) and (3.6) are proved in view of (B.21), (B.22), and (B.24). The proof is therefore complete. ◇

Appendix C

Technical Details for Chapter 4

Denote by P^*, E^*, var^* the probability, expectation, and variance, respectively, conditional on data $\mathcal{X}_n$. For notational simplicity, we often suppress the dependence of $X_{t,n}$, $u_{t,n}$, and $W_{t,n}$ on n and write them as X_t, u_t, and W_t, respectively. For a sequence of random variables $\{Y_n\}$, we write $Y_n = o_p^*(1)$ if for any $\epsilon > 0$, $P^*\{|Y_n| > \epsilon\} \to 0$ in probability, and $Y_n = O_p^*(1)$ if there exists a constant $M > 0$ such that for all large n, $P^*\{|Y_n| > M\} < \epsilon$ with probability arbitrarily close to one as $n \to \infty$, as defined in Chang and Park (2003). We define $S_{t,n} = \sum_{i=1}^{t} u_{i,n}$, which is often denoted S_t. The positive constant C is generic and may vary from place to place. We often use $\mathcal{I}_j$ in different proofs to indicate different objects. For notational simplicity, we often write $G(s, \mathcal{F}_t) := G_{\zeta_s}(s, \mathcal{F}_t)$ and $c(s;h) := c_{\zeta_s}(s;h)$, omitting the subscript ζ_s, where $\zeta_s = j$ such that $s \in [b_j, b_{j+1})$ and $\zeta_s = \tau$ if $s = 1$. We define $\gamma_h(r) = \int_0^r c(s;h)ds$. Notice that by definition, $\gamma_0(1) = \sigma_u^2$, which are interchangeably used in the proofs.

Recall that $\mathcal{F}_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$ with ε_t iid (0,1), and $\{\varepsilon_t'\}$ is an iid copy of $\{\varepsilon_t\}$. Following Wu (2005), for $I \subset \mathbb{Z}$, define $\mathcal{F}_{t,I}$ be the same as $\mathcal{F}_t$ except that ε_js are replaced by ε_j's for $j \in I$. In particular, for $i \le t$, $\mathcal{F}_{t,\{i\}} = (\ldots, \varepsilon_{i-1}, \varepsilon_i', \varepsilon_{i+1}, \ldots, \varepsilon_t)$. Denote $\mathcal{F}_{t,i}^* = \mathcal{F}_{t,\{k\in\mathbb{Z}:k\le i\}}$.

To keep the proofs concise, we present the case with no deterministic trend functions, i.e., $\beta \equiv 0$. The statements in the main theorem hold by replacing $B_\sigma(r)$ with $B_{\sigma|Z}(r)$ and X_t with $\widehat{X}_t$.

The following four lemmas prove some basic properties of $\{u_t\}$ and $\{X_t\}$ that are useful in the subsequent proofs.

Lemma **C.0.8**. *Assume (A1)-(A4). Fix* $j \in \{0, 1, \ldots, \tau\}$.

(i) For any $t, t' \in [b_j n, b_{j+1} n)$, $\left|cov(u_t, u_{t'}) - c_j(t/n; |t - t'|)\right| \le C(|t - t'|/n)$.

(ii) For any $s \ne s' \in [b_j, b_{j+1}]$, $\left|c_j(s;h) - c_j(s';h)\right| \le C|s - s'|$ *uniformly over* $h \in \mathbb{N}$.

(iii) For any $\rho > 0$, $\sup_{s\in[b_j,b_{j+1}]} \sum_{h=0}^{\infty} \left|h^\rho c_j(s,h)\right| \le C \sum_{h=0}^{\infty} h^\rho \chi^h < \infty$.

(iv) $\sup_{b_j \le s \ne s' < b_{j+1}} \frac{|\sigma(s) - \sigma(s')|}{|s - s'|(-\log|s - s'| + 1)} \le C.$

In addition, if $j = \tau$, (i) and (iv) also hold for all $t, t' \in [b_\tau n, n]$ or for supremum over $\{b_\tau \le s \ne s' \le 1\}$.

Proof of Lemma C.0.8. (i) For all $t, t' \in [b_j n, b_{j+1} n)$, it can be written as $\mathrm{cov}(u_t, u_{t'}) = c_j(t/n; |t - t'|) - \mathrm{cov}\big\{G_j(t/n, \mathcal{F}_t), G_j(t/n, \mathcal{F}_{t'}) - G_j(t'/n, \mathcal{F}_{t'})\big\}$. From the Cauchy-Schwartz inequality and (A1), we have $\big|\mathrm{cov}\big\{G_j(t/n, \mathcal{F}_t), G_j(j/n, \mathcal{F}_{t'}) - G_j(t'/n, \mathcal{F}_{t'})\big\}\big| \le ||G_j(t/n, \mathcal{F}_t)||_2 \; ||G_j(t/n, \mathcal{F}_{t'}) - G_j(t'/n, \mathcal{F}_{t'})||_2 \le C(|t - t'|/n)$, which completes the proof. If $j = \tau$, the same argument holds for all $t, t' \in [b_\tau n, n]$.

(ii) It follows from the triangular inequality, Cauchy-Schwartz inequality, and (A1) that for any $s \ne s' \in [b_j, b_{j+1}]$, $|c_j(s; h) - c_j(s'; h)| \le ||G_j(s, \mathcal{F}_0)||_2 ||G_j(s, \mathcal{F}_h) - G_j(s', \mathcal{F}_h)||_2 + ||G_j(s', \mathcal{F}_h)||_2 ||G_j(s, \mathcal{F}_0) - G_j(s', \mathcal{F}_0)\}||_2 \le C|s - s'|$ holds uniformly over $h \in \mathbb{N}$.

(iii) It is a straightforward consequence of Lemma A.1 in Shao and Wu (2007), Theorem 1 in Wu (2005), and our assumption (A3).

(iv) It follows from (A3) that $|c_j(s; h) - c_j(s'; h)| \le 2C\chi^h$ for all $h \in \mathbb{N}$ and $s, s' \in [b_j, b_{j+1})$. Let m be the smallest positive integer such that $\chi^m \le |s - s'|$. Then using (ii) we have $|\sigma(s) - \sigma(s')| \le \sum_{h=-\infty}^{\infty} |c_j(s; h) - c_j(s'; h)| \le C(\sum_{|h| \le m-1} |s - s'| + \sum_{|h| \ge m} \chi^h) \le C\{m|s - s'| + \chi^m(1 - \chi)^{-1}\} \le C|s - s'|(-\log\chi)^{-1}(-\log|s - s'|) + C(1 - \chi)^{-1}|s - s'| \le C|s - s'|(-\log|s - s'| + 1)$. Notice that constant C's do not depend on s or s'. Thus the proof is complete. If $j = \tau$, the same argument holds for $s, s' \in [b_\tau, 1]$. ♢

Lemma **C.0.9**. *Under the conditions (A2)-(A3), for any $i = 1, \dots, n - h$, $h = 0, \dots, n - i$,*

$$|E(u_i u_{i+h})| \le C\chi^h,$$

where C is a constant that does not depend on h, i, or n.

Proof of Lemma C.0.9. By the definition, $\mathcal{F}_i$ and $\mathcal{F}^*_{i+h,i}$ are independent. Thus $E\{G(i/n, \mathcal{F}_i)G((i + h)/n, \mathcal{F}^*_{i+h,i})\} = 0$ so that we can write

$$\begin{aligned} E(u_i u_{i+h}) &= E\left[G(i/n, \mathcal{F}_i)\{G((i + h)/n, \mathcal{F}_{i+h}) - G((i + h)/n, \mathcal{F}_{i+h,\{i\}})\}\right] \\ &\quad + E\left[G(i/n, \mathcal{F}_i)\{G((i + h)/n, \mathcal{F}_{i+h,\{i\}}) - G((i + h)/n, \mathcal{F}^*_{i+h,i})\}\right] \end{aligned}$$

Then by the Cauchy-Schwartz inequality,

$$\begin{aligned}|E(u_i u_{i+h})| &\leq ||G(i/n, \mathcal{F}_i)||_2 ||G((i+h)/n, \mathcal{F}_{i+h}) - G((i+h)/n, \mathcal{F}_{i+h,\{i\}})||_2 \\ &\quad + ||G(i/n, \mathcal{F}_i)||_2 ||G((i+h)/n, \mathcal{F}_{i+h,\{i\}}) - G((i+h)/n, \mathcal{F}^*_{i+h,i})||_2\end{aligned}$$

By (A2), we have $||G(i/n, \mathcal{F}_i)||_2 < C < \infty$, and by (A3) $||G((i+h)/n, \mathcal{F}_{i+h}) - G((i+h)/n, \mathcal{F}_{i+h,\{i\}})||_2 < ||G((i+h)/n, \mathcal{F}_{i+h}) - G((i+h)/n, \mathcal{F}_{i+h,\{i\}})||_4 \leq C\chi^h$. Thus the first term is bounded by $C\chi^h$, where C does not depend on h, i, or n.

Now the proof is done once we show that

$$||G((i+h)/n, \mathcal{F}_{i+h,\{i\}}) - G((i+h)/n, \mathcal{F}^*_{i+h,i})||_4 \leq C\chi^h.$$

Define $\mathcal{F}^*_{i+h,\{i\},m} = \mathcal{F}_{i+h,A}$, where $A = \{k \in \mathbb{Z} : k \leq i-m-1\} \bigcup \{i\}$. In particular, if $m = 0$, $\mathcal{F}^*_{i+h,\{i\},0} = \mathcal{F}^*_{i+h,i}$. Then $||G((i+h)/n, \mathcal{F}_{i+h,\{i\}}) - G((i+h)/n, \mathcal{F}^*_{i+h,i})||_4 = ||\sum_{m=0}^{\infty} G((i+h)/n, \mathcal{F}^*_{i+h,\{i\},m}) - G((i+h)/n, \mathcal{F}^*_{i+h,\{i\},m+1})||_4 \leq \sum_{m=0}^{\infty} ||G((i+h)/n, \mathcal{F}^*_{i+h,\{i\},m}) - G((i+h)/n, \mathcal{F}^*_{i+h,\{i\},m+1})||_4 \leq C\sum_{m=0}^{\infty} \chi^{h+m+1}$ $= C\chi^{h+1}/(1-\chi) \leq C\chi^h$, where the last C does not depend on h, i, or n. Thus the proof is complete. ♢

Let $\text{cum}(Y_0, Y_1, Y_2, Y_3)$ denote the fourth order cumulant. When $EY_i = 0,\ i = 0, 1, 2, 3$, we often use the relation (see page 36 in Rosenblatt (1985), for example)

$$\text{cov}(Y_0Y_1, Y_2Y_3) = E(Y_0Y_2)E(Y_1Y_3) + E(Y_0Y_3)E(Y_1Y_2) + \text{cum}(Y_0, Y_1, Y_2, Y_3). \tag{C.1}$$

Lemma **C.0.10**. *Assume (A1)-(A4). Then*

$$\sup_{1 \leq t_1 \leq \ldots \leq t_4 \leq n} |\text{cum}(u_{t_1}, u_{t_2}, u_{t_3}, u_{t_4})| \leq C\chi^{(t_4 - t_1)/3},$$

where χ as in (A3).

Proof of Lemma C.0.10. Let $\mathcal{F}'_t = \mathcal{F}^*_{t,0}$ if $t > 0$, and $\mathcal{F}'_t = \mathcal{F}_t$ if $t \leq 0$. Define $\mathcal{F}'_{t,m} = \mathcal{F}_{t,\{k \in \mathbb{Z}: -m \leq k \leq 0\}}$ for $m \geq 0$ and $t > 0$. The argument is similar to the proof of Proposition 2 in Wu and Shao (2004). Let $1 \leq t_1 \leq \ldots \leq t_4 \leq n$, and $m_k = t_{k+1} - t_k$ for $k \in \{1, 2, 3\}$. Since for a fixed $s \in [0, 1]$, the process

$\{G(s, \mathcal{F}_t)\}_t$ is stationary, we can write

$$
\begin{aligned}
& \mathrm{cum}(u_{t_1}, u_{t_2}, u_{t_3}, u_{t_4}) \\
= \; & \mathrm{cum}\{G(t_1/n, \mathcal{F}_{t_1}), G(t_2/n, \mathcal{F}_{t_2}), G(t_3/n, \mathcal{F}_{t_3}), G(t_4/n, \mathcal{F}_{t_4})\} \\
= \; & \mathrm{cum}\{G(t_1/n, \mathcal{F}_{t_1-t_k}), G(t_2/n, \mathcal{F}_{t_2-t_k}), G(t_3/n, \mathcal{F}_{t_3-t_k}), G(t_4/n, \mathcal{F}_{t_4-t_k})\} \\
= \; & \mathrm{cum}\{G(t_1/n, \mathcal{F}_{t_1-t_k}), G(t_2/n, \mathcal{F}_{t_2-t_k}) - G(t_2/n, \mathcal{F}'_{t_2-t_k}), G(t_3/n, \mathcal{F}_{t_3-t_k}), G(t_4/n, \mathcal{F}_{t_4-t_k})\} \\
& + \mathrm{cum}\{G(t_1/n, \mathcal{F}_{t_1-t_k}), G(t_2/n, \mathcal{F}'_{t_2-t_k}), G(t_3/n, \mathcal{F}_{t_3-t_k}) - G(t_3/n, \mathcal{F}'_{t_3-t_k}), G(t_4/n, \mathcal{F}_{t_4-t_k})\} \\
& + \mathrm{cum}\{G(t_1/n, \mathcal{F}_{t_1-t_k}), G(t_2/n, \mathcal{F}'_{t_2-t_k}), G(t_3/n, \mathcal{F}'_{t_3-t_k}), G(t_4/n, \mathcal{F}_{t_4-t_k}) - G(t_4/n, \mathcal{F}'_{t_4-t_k})\} \\
& + \mathrm{cum}\{G(t_1/n, \mathcal{F}_{t_1-t_k}), G(t_2/n, \mathcal{F}'_{t_2-t_k}), G(t_3/n, \mathcal{F}'_{t_3-t_k}), G(t_4/n, \mathcal{F}'_{t_4-t_k})\} \\
:= \; & \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4,
\end{aligned}
$$

with the third equality due to the additive property of cumulants [the property (iii) on page 35 in Rosenblatt (1985)]. First we claim that $\mathcal{I}_4 = 0$. If $k = 1$, $\mathcal{F}_{t_1-t_1} = \mathcal{F}_0$ is independent from $\mathcal{F}'_{t_2-t_1}$, $\mathcal{F}'_{t_3-t_1}$, $\mathcal{F}'_{t_4-t_1}$, so $\mathcal{I}_4 = 0$ using the property (ii) on page 35 in Rosenblatt (1985). If $k = 2$, notice that $\mathcal{F}'_{t_2-t_2} = \mathcal{F}'_0 = \mathcal{F}_0$ by definition, and $\mathcal{F}_{t_1-t_2}$ and $\mathcal{F}_0$ are independent from $\mathcal{F}'_{t_3-t_2}$, $\mathcal{F}'_{t_4-t_2}$, which leads to $\mathcal{I}_4 = 0$. Similarly, if $k = 3$, $\mathcal{F}_{t_1-t_3}$, $\mathcal{F}'_{t_2-t_3} = \mathcal{F}_{t_2-t_3}$, and $\mathcal{F}_0$ are independent from $\mathcal{F}'_{t_4-t_3}$. Thus we have $\mathcal{I}_4 = 0$ for all $k = 1, 2, 3$. Also, notice that since $\mathcal{F}'_t = \mathcal{F}_t$ if $t \le 0$, we have $\mathcal{I}_1 = 0$ if $k = 2$ and $\mathcal{I}_1 = \mathcal{I}_2 = 0$ if $k = 3$. Thus the proof is done if we show that for each $k = 1, 2, 3$,

$$
\max_{k \le i \le 3} |\mathcal{I}_i| \le C\chi^{m_k}. \tag{C.2}
$$

Once (C.2) is shown, we have for each $k = 1, 2, 3$, $|\mathrm{cum}(u_{t_1}, u_{t_2}, u_{t_3}, u_{t_4})| \le C\chi^{m_k}$. Taking minimum over k for both sides yields $|\mathrm{cum}(u_{t_1}, u_{t_2}, u_{t_3}, u_{t_4})| \le C \min_{k=1,2,3} \chi^{m_k} = C\chi^{\max_{k=1,2,3} m_k} \le C\chi^{(t_4-t_1)/3}$, since $t_4 - t_1 = \sum_{j=2}^{4}(t_j - t_{j-1}) \le 3 \max_{j=2,3,4}(t_j - t_{j-1}) = 3 \max_{k=1,2,3} m_k$.

Now we show (C.2). For each $k = 1, 2, 3$, fix any $j = k+1, \ldots, 4$. Let $Y_0 = G(t_j/n, \mathcal{F}_{t_j-t_k}) - G(t_j/n, \mathcal{F}'_{t_j-t_k})$ and Y_1, Y_2, and Y_3 be the other variables in $\mathcal{I}_{j-1}$ so that we can write $\mathcal{I}_{j-1} = \mathrm{cum}(Y_0, Y_1, Y_2, Y_3)$. Since $||Y_0||_4 \le ||\{G(t_j/n, \mathcal{F}_{t_j-t_k}) - G(t_j/n, \mathcal{F}'_{t_j-t_k,0})\}||_4 + \sum_{m=0}^{\infty} ||\{G(t_j/n, \mathcal{F}'_{t_j-t_k,m}) - G(t_j/n, \mathcal{F}'_{t_j-t_k,m+1})\}||_4$ $\le C\{\chi^{t_j-t_k} + \sum_{m=0}^{\infty} \chi^{t_j-t_k+m+1}\} \le C\chi^{t_j-t_k}$ holds by the triangular inequality and (A3), we have

$$
||Y_0||_4 \le C\chi^{t_j-t_k}, \tag{C.3}
$$

where C is a constant that does not depend on t_j, j, or n. Observe that due to (C.1), $\mathcal{I}_{j-1} = E(Y_0Y_1Y_2Y_3) - E(Y_0Y_1)E(Y_2Y_3) - E(Y_0Y_2)E(Y_1Y_3) - E(Y_0Y_3)E(Y_1Y_2)$. By Hölder's inequality, (C.3), and (A2), we have $|E(Y_0Y_1Y_2Y_3)| \le ||Y_0||_4||Y_1Y_2Y_3||_{4/3} \le C\chi^{t_j-t_k}$ and $|E(Y_0Y_i)| \le ||Y_0||_2||Y_i||_2 \le C\chi^{t_j-t_k}$. Thus we have $|\mathcal{I}_{j-1}| \le C\chi^{t_j-t_k} \le C\chi^{m_k}$ and (C.2) is proved. ♢

Lemma **C.0.11**. *Assume (A1)-(A4). In the following statements, C is a positive constant that does not depend on n.*

(i) Under the null hypothesis $\rho = 1$, $\sup_{1\le t\le n}\{E(X_t^4)/t^2\} \le C$.

(ii) Under the alternative hypothesis $|\rho| < 1$, $\sup_{t=1,\dots,n} E(X_t^4) \le C$.

Proof of Lemma C.0.11. (i) Suppose $1 \le i_1 \le i_2 \le i_3 \le i_4 \le t$ for some $t = 1,\dots,n$. Due to (C.1), we can write $E(u_{i_1}u_{i_2}u_{i_3}u_{i_4}) = E(u_{i_1}u_{i_2})E(u_{i_3}u_{i_4}) + E(u_{i_1}u_{i_3})E(u_{i_2}u_{i_4}) + E(u_{i_1}u_{i_4})E(u_{i_2}u_{i_3}) + \mathrm{cum}(u_{i_1}, u_{i_2}, u_{i_3}, u_{i_4})$. Since Lemma C.0.9 implies that $E(u_{i_j}u_{i_k}) \le C\chi^{|i_j-i_k|}$ for any $j, k = 1,\dots,4$, and Lemma C.0.10 implies that $\mathrm{cum}(u_{i_1}, u_{i_2}, u_{i_3}, u_{i_4}) \le C\chi^{(i_4-i_1)/3}$, we have $E(u_{i_1}u_{i_2}u_{i_3}u_{i_4}) \le C\chi^{i_2-i_1}\chi^{i_4-i_3} + C\chi^{i_3-i_1}\chi^{i_4-i_2} + C\chi^{i_4-i_1}\chi^{i_3-i_2} + C\chi^{i_4-i_1}$. Since $\sum_{1\le i_1\le i_2\le i_3\le i_4\le t} C\chi^{(i_4-i_1)/3} = \sum_{h=0}^{t-1}(t-h)(h+1)^2\chi^{h/3} \le Ct$ from simple calculations, and $\{\sum_{i_1,i_2}\chi^{|i_1-i_2|}\}\{\sum_{i_3,i_4}\chi^{|i_3-i_4|}\} \le (Ct)^2$, where both Cs are constants that do not depend on t or n, we have $E(X_t^4) = 16\sum_{1\le i_1\le i_2\le i_3\le i_4\le t} E(u_{i_1}u_{i_2}u_{i_3}u_{i_4}) \le Ct^2$, which completes the proof.

(ii) Observe that under the alternative, $X_t = \sum_{i=1}^t \rho^{t-i}u_i$. Thus $E(X_t^4) = E(\sum_{i=1}^t \rho^{t-i}u_i)^4 = \sum_{i_1,i_2,i_3,i_4=1}^t \rho^{4t-i_1-i_2-i_3-i_4}E(u_{i_1}u_{i_2}u_{i_3}u_{i_4}) \le C\sum_{i_1,i_2,i_3,i_4=1}^t |\rho|^{4t-i_1-i_2-i_3-i_4} = C(\sum_{i=1}^t |\rho|^{t-i})^4 = C\{(1-|\rho|^t)/(1-|\rho|)\}^4 \le C$, where C is a constant that does not depend on t. Thus (ii) follows. ♢

The following lemmas contain some key results needed in the proof of Theorem 4.3.1 and Theorem 4.3.2 and they may be of independent interest.

Lemma **C.0.12**. *Assume (A1)-(A4).*

(i) $n^{-1/2}S_{\lfloor nr\rfloor} = n^{-1/2}\sum_{i=1}^{\lfloor nr\rfloor} u_i \Rightarrow B_\sigma(r) = \int_0^r \sigma(s)dB(s)$.

(ii) For a fixed $r \in (0,1]$ and a fixed integer $h \ge 0$, $|n^{-1}\sum_{i=1}^{\lfloor nr\rfloor\wedge(n-h)} u_iu_{i+h} - \gamma_h(r)| = o_p(1)$. Recall that $\gamma_h(r) = \int_0^r c_{\zeta_s}(s;h)ds$, where $\zeta_s = j$ such that $s \in [b_j, b_{j+1})$.

Proof of Lemma C.0.12. (i) Define a step function $\sigma_n(s) = \sigma(t/n)$ for $s \in [t/n, (t+1)/n)$ and $t = 0, 1, \ldots, n$ and $\sigma_n(1) = \sigma(1)$. Let $\check{B}_{n,\sigma}(r) = \int_0^{\lfloor nr \rfloor/n} \sigma_n(s)dB(s)$ and $\tilde{B}_{n,\sigma}(r) = \int_0^r \sigma_n(s)dB(s)$. Recall that $B_\sigma(r) = \int_0^r \sigma(s)dB(s)$.

By the triangular inequality, $\sup_{r\in[0,1]} |\check{B}_{n,\sigma}(r) - B_\sigma(r)| \le \sup_{r\in[0,1]} |\check{B}_{n,\sigma}(r) - \tilde{B}_{n,\sigma}(r)| + \sup_{r\in[0,1]} |\tilde{B}_{n,\sigma} - B_\sigma(r)| =: \mathcal{I}_1 + \mathcal{I}_2$. We have $\mathcal{I}_1 = o_p(1)$, because $\sup_{r\in[0,1]} |r - \lfloor nr \rfloor/n| \le 1/n$ and $\sup_{r\in[0,1]} | \int_{\lfloor nr \rfloor/n}^r \sigma_n(s)dB(s)| \le C \sup_{t=1,\ldots,n} |B(t/n) - B((t-1)/n)| = o_p(1)$. Notice that by Lemma C.0.8 (iv), $\sup_{r\in[0,1]} |\sigma_n(r) - \sigma(r)| = \sup_{0\le j\le\tau} \sup_{b_j \le s < b_{j+1}} |\sigma_n(s) - \sigma(s)| = \sup_{0\le j\le\tau} \sup_{b_j \le s < b_{j+1}} |\sigma(\lfloor ns \rfloor/n) - \sigma(s)| \le (\tau+1)C|\lfloor ns \rfloor/n - s|(-\log|\lfloor ns \rfloor/n - s| + 1) = O(n^{-1}\log n) = o(1)$. Thus $\mathcal{I}_2 = o_p(1)$ holds by (Kurtz, 2001, Proposition 5.19). It follows that

$$\sup_{r\in[0,1]} |\check{B}_{n,\sigma}(r) - B_\sigma(r)| = o_p(1). \tag{C.4}$$

From Proposition 5 in Zhou (2013), on a richer probability space, there exist iid standard normal random variables $V_1, \ldots, V_n$ such that

$$\sup_{r\in[0,1]} |n^{-1/2} S_{\lfloor nr \rfloor} - \hat{B}_{n,\sigma}(r)| = o_p(1), \tag{C.5}$$

where $\hat{B}_{n,\sigma}(r) = n^{-1/2} \sum_{i=1}^{\lfloor nr \rfloor} \sigma(i/n) V_i$. Since $\{\hat{B}_{n,\sigma}(r)\}_{r\in[0,1]} \overset{\mathcal{D}}{=} \{\sum_{t=1}^{\lfloor nr \rfloor} \sigma(t/n)[B(t/n) - B\{(t-1)/n\}]\}_{r\in[0,1]} \overset{\mathcal{D}}{=} \{\check{B}_{n,\sigma}(r)\}_{r\in[0,1]}$,

$$\hat{B}_{n,\sigma}(r) \Rightarrow B_\sigma(r) \tag{C.6}$$

by (C.4). Then (i) follows from (C.5) and (C.6).

(ii) We first define $Y_i = Y_{i,n} = u_i u_{i+h} - E(u_i u_{i+h})$ and claim that $|n^{-1} \sum_{i=1}^{\lfloor nr \rfloor \wedge (n-h)} Y_{i,n}| = o_p(1)$. Observe that by (C.1), we can write for $i \ge i'$, $E(Y_i Y_{i'}) = \mathrm{cov}(u_i u_{i+h}, u_{i'} u_{i'+h}) = E(u_i u_{i'}) E(u_{i+h} u_{i'+h}) + E(u_i u_{i'+h}) E(u_{i+h} u_{i'}) + \mathrm{cum}(u_i, u_{i+h}, u_{i'}, u_{i'+h}) \le C\chi^{2|i-i'|} + C\chi^{|i-i'-h| + |i+h-i'|} + C\chi^{|i+h-i'|/3} \le C\chi^{|i-i'|/3}$, where the first inequality is due to Lemmas C.0.9 and C.0.10. Then, by Chebyshev's inequality, for any $\delta > 0$, $P(|\sum_{i=1}^{\lfloor nr \rfloor \wedge (n-h)} Y_{i,n}| > n\delta) \le (n\delta)^{-2} E(\sum_{i=1}^{\lfloor nr \rfloor \wedge (n-h)} Y_{i,n})^2 \le (n\delta)^{-2} \sum_{i,i'=1}^{\lfloor nr \rfloor \wedge (n-h)} E(Y_{i,n} Y_{i',n}) \le C(n\delta)^{-2} \sum_{i,i'=1}^{\lfloor nr \rfloor \wedge (n-h)} \chi^{|i-i'|/3} \le (n\delta)^{-2} Cn = o(1)$. Thus we have $|n^{-1} \sum_{i=1}^{n-h} \{u_i u_{i+h} - E(u_i u_{i+h})\}| = o_p(1)$.

Now it remains to show that $|n^{-1} \sum_{i=1}^{\lfloor nr \rfloor \wedge (n-h)} E(u_i u_{i+h}) - \gamma_h(r)| = o(1)$. For $r \in (0, 1]$, let $\mathcal{B}_r = \{i : i/n < b_j < (i+h)/n \text{ for some } b_j, \text{ and } 1 \le i \le \lfloor nr \rfloor \wedge (n-h)\}$ and τ_r be the number of break points in

$(0,r)$. Since $n^{-1}\sum_{i=1}^{\lfloor nr\rfloor\wedge(n-h)}E(u_iu_{i+h}) = n^{-1}\sum_{i\notin\mathcal{B}_r}E(u_iu_{i+h}) + n^{-1}\sum_{i\in\mathcal{B}_r}E(u_iu_{i+h}) = \mathcal{I}_{1,r}+\mathcal{I}_{2,r}$, it suffices to show that

$$\sup_{r\in(0,1]}|\mathcal{I}_{1,r}-\gamma_h(r)| = o(1) \qquad \text{and} \qquad \sup_{r\in(0,1]}|\mathcal{I}_{2,r}| = o(1). \tag{C.7}$$

For $\mathcal{I}_{1,r}$ part in (C.7), $|\mathcal{I}_{1,r} - n^{-1}\sum_{i=1}^{\lfloor rn\rfloor\wedge(n-h)}c_{\zeta_{i/n}}(i/n;h)| \le n^{-1}\sum_{i\notin\mathcal{B}_r}|Eu_iu_{i+h} - c_{\zeta_{i/n}}(i/n;h)| + n^{-1}\sum_{i\in\mathcal{B}_r}|c_{\zeta_{i/n}}(i/n;h)| \le Ch/n$ holds for C that does not depend on r, by Lemma C.0.8 (i). For $\mathcal{I}_{2,r}$, $\sup_{r\in[0,1)}|\mathcal{I}_{2,r}| \le \sup_{r\in[0,1)}C\tau_r h/n \le C\tau h/n = O(h/n)$. Thus (C.7) holds, and the proof is complete. $\diamondsuit$

Lemma **C.0.13**. *Assume (A1)-(A4). Let $S_t = \sum_{i=1}^t u_{i,n}$. The following statements hold jointly.*

(i) For any $r\in(0,1]$, $n^{-2}\sum_{t=1}^{\lfloor nr\rfloor}S_{t-1}^2 \xrightarrow{\mathcal{D}} \int_0^r B_\sigma^2(s)ds$.

(ii) For any $r\in(0,1]$, $n^{-1}\sum_{t=1}^{\lfloor nr\rfloor}S_{t-1}u_t \xrightarrow{\mathcal{D}} 2^{-1}\left\{B_\sigma^2(r)-\gamma_0(r)\right\}$.

(iii) For any $r\in(0,1]$, $n^{-1}\sum_{t=1}^{\lfloor nr\rfloor\wedge(n-h)}S_{t-1}u_{t+h} \xrightarrow{\mathcal{D}} 2^{-1}\left\{B_\sigma^2(r)-\gamma_0(r)\right\} - \sum_{k=1}^h\gamma_k(r)$ for any fixed integer $h\ge 1$.

(iv) For any $r\in(0,1]$, $n^{-3/2}\sum_{t=1}^{\lfloor nr\rfloor}S_{t-1} \xrightarrow{\mathcal{D}} \int_0^r B_\sigma(s)ds$.

Proof of Lemma C.0.13. (i) $n^{-2}\sum_{t=1}^{\lfloor nr\rfloor}S_{t-1}^2 = \sum_{t=1}^{\lfloor nr\rfloor}\int_{(t-1)/n}^{t/n}(n^{-1/2}S_{\lfloor nr\rfloor})^2dr = \int_0^{\lfloor nr\rfloor/n}(n^{-1/2}S_{\lfloor ns\rfloor})^2ds \xrightarrow{\mathcal{D}} \int_0^r B_\sigma^2(s)ds$ by the continuous mapping theorem and Lemma C.0.12 (i).

(ii) Observe that $2S_{t-1}u_t = S_t^2 - S_{t-1}^2 - u_t^2$ for all $t=1,\dots,n$. Then by Lemma C.0.12 (i) and (ii), $n^{-1}\sum_{t=1}^{\lfloor nr\rfloor}S_{t-1}u_t = (2n)^{-1}(S_{\lfloor nr\rfloor}^2 - S_0^2 - \sum_{t=1}^{\lfloor nr\rfloor}u_t^2) \xrightarrow{\mathcal{D}} 2^{-1}\{B_\sigma^2(r)-\gamma_0(r)\}$.

(iii) Observe that for a fixed integer $h\ge 1$, $S_{t+h}^2 - S_{t+h-1}^2 = u_{t+h}(S_{t+h}+S_{t+h-1}) = u_{t+h}(2S_{t-1} + 2\sum_{i=0}^{h-1}u_{t+i} + u_{t+h})$ so that $2u_{t+h}S_{t-1} = S_{t+h}^2 - S_{t+h-1}^2 - 2u_{t+h}\sum_{k=0}^{h-1}u_{t+k} - u_{t+h}^2$. Thus we have for $r\in(0,1]$, $n^{-1}\sum_{t=1}^{\lfloor nr\rfloor\wedge(n-h)}u_{t+h}S_{t-1} = (2n)^{-1}(S_{\lfloor nr\rfloor\wedge(n-h)}^2 - S_h^2 - 2\sum_{k=0}^{h-1}\sum_{t=1}^{\lfloor nr\rfloor\wedge(n-h)}u_{t+h}u_{t+k} - \sum_{t=1}^{\lfloor nr\rfloor\wedge(n-h)}u_{t+h}^2) \xrightarrow{\mathcal{D}} 2^{-1}\{B_\sigma^2(r) - 2\sum_{k=1}^h\gamma_k(r) - \gamma_0(r)\}$ by Lemma C.0.12 (i) and (ii).

(iv) $n^{-3/2}\sum_{t=1}^{\lfloor nr\rfloor}S_{t-1} = \sum_{t=1}^{\lfloor nr\rfloor}\int_{(t-1)/n}^{t/n}(n^{-1/2}S_{\lfloor ns\rfloor})ds = \int_0^{\lfloor nr\rfloor/n}(n^{-1/2}S_{\lfloor ns\rfloor})ds \xrightarrow{\mathcal{D}} \int_0^r B_\sigma(s)ds$ by the continuous mapping theorem and Lemma C.0.12 (i). $\diamondsuit$

Lemma **C.0.14**. *Assume (A1)-(A4). Under the null hypothesis $\rho=1$,*

$$s_n^2 = (n-2)^{-1}\sum_{t=1}^n(X_t-\widehat{\rho}_nX_{t-1})^2 \xrightarrow{\mathcal{P}} \gamma_0(1) = \sigma_u^2.$$

Proof of Lemma C.0.14. $s_n^2 = (n-2)^{-1}\sum_{t=1}^n (X_t - \widehat{\rho}_n X_{t-1})^2 = (n-2)^{-1}\sum_{t=1}^n u_t^2 + (n-2)^{-1}(\widehat{\rho}_n - \rho)^2 \sum_{t=1}^n X_{t-1}^2 + 2(n-2)^{-1}(\rho - \widehat{\rho}_n)\sum_{t=1}^n X_{t-1}u_t := \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3$. Here $\mathcal{I}_1 \xrightarrow{\mathcal{P}} \gamma_0(1)$ by Lemma C.0.12 (ii). For $\mathcal{I}_2$, under the null hypothesis $\rho = 1$, notice that $S_t = X_t = \sum_{i=1}^t u_i$. By Lemma C.0.13 (i) and (ii), $\widehat{\rho}_n - \rho = O_p(n^{-1})$ and $\sum_{t=1}^n X_{t-1}^2 = O_p(n^2)$ so that $\mathcal{I}_2 = O_p(n^{-1})$. Under the null, we also have $\sum_{t=1}^n X_{t-1}u_t = O_p(n^{3/2})$ by the Cauchy-Schwartz inequality, which leads to $\mathcal{I}_3 = O_p(n^{-1/2})$. Thus the proof is complete. ♢

Proof of Theorem 4.2.1. Notice that under the null hypothesis $\rho = 1$, $X_t = S_t = \sum_{i=1}^t u_i$. Then (4.3) and (4.4) are direct consequences of Lemma C.0.13 (i), (ii), the continuous mapping theorem, and Lemma C.0.14. ♢

Proof of Theorem 4.2.2. First observe that $e^{c/n} = 1 + c/n + O(n^{-2})$ so that $\rho_n = e^{c/n} + O(n^{-2})$. Then X_t is asymptotically equivalent to $\sum_{j=1}^t e^{(t-j)c/n}u_j$, i,e, $X_t = \sum_{j=1}^t \rho_n^{t-j}u_j = X_t = \sum_{j=1}^t e^{(t-j)c/n}u_j + O(n^{-2})\sum_{j=1}^t u_j = \sum_{j=1}^t e^{(t-j)c/n}u_j + O_p(n^{-3/2})$. Following the argument in Phillips (1987b), page 539, and using Lemma C.0.12 (i), we obtain $n^{-1/2}\sum_{j=1}^{\lfloor nr \rfloor} e^{(t-j)c/n}u_j \Rightarrow J_{c,\sigma}(r)$, which implies

$$n^{-1/2}X_{\lfloor nr \rfloor} \Rightarrow J_{c,\sigma}(r).$$

Then we have, by the same argument as in Lemma C.0.13 (i) and using (4.5),

$$n^{-2}\sum_{t=1}^{\lfloor nr \rfloor} X_{t-1}^2 \xrightarrow{\mathcal{D}} \int_0^r J_{c,\sigma}^2(s)ds. \tag{C.8}$$

Now we are going to show

$$n^{-1}\sum_{t=1}^{n} X_{t-1}u_t \xrightarrow{\mathcal{D}} \int_0^1 J_{c,\sigma}(r)\sigma(r)dB(r) + 2^{-1}\left\{\int_0^1 \sigma^2(r)dr - \sigma_u^2\right\}. \tag{C.9}$$

By taking square of (4.2), we have $X_t^2 = (1+cn^{-1})^2 X_{t-1}^2 + u_t^2 + 2(1+cn^{-1})X_{t-1}u_t$ so that $\sum_{t=1}^n X_t^2 =$

$(1+2cn^{-1})\sum_{t=1}^{n}X_{t-1}^2+\sum_{t=1}^{n}u_t^2+2\sum_{t=1}^{n}X_{t-1}u_t+O_p(1)$. Thus

$$\begin{aligned} 2n^{-1}\textstyle\sum_{t=1}^{n}X_{t-1}u_t &= n^{-1}X_n^2-2cn^{-2}\textstyle\sum_{t=1}^{n}X_{t-1}^2-n^{-1}\sum_{t=1}^{n}u_t^2+O_p(n^{-1}) \\ &\xrightarrow{\mathcal{D}} J_{c,\sigma}^2(1)-2c\textstyle\int_0^1 J_{c,\sigma}^2(r)dr-\sigma_u^2 \\ &= 2\textstyle\int_0^1 J_{c,\sigma}(r)\sigma(r)dB(r)+\{\int_0^1\sigma^2(r)dr-\sigma_u^2\}, \end{aligned}$$

which implies (C.9). Here, the last equality is due to

$$J_{c,\sigma}^2(1)=\int_0^1\sigma^2(r)dr+2c\int_0^1 J_{c,\sigma}^2(r)dr+2\int_0^1 J_{c,\sigma}(r)\sigma(r)dB(r),$$

which follows from Itô's formula.[1] Now that (C.9) is shown, we have by the continuous mapping theorem and (C.8),

$$n(\widehat{\rho}_n-\rho)\xrightarrow{\mathcal{D}}\frac{\int_0^1 J_{c,\sigma}(r)\sigma(r)dB(r)+2^{-1}\{\int_0^1\sigma^2(r)dr-\sigma_u^2\}}{\int_0^1 J_{c,\sigma}^2(r)dr}.$$

The statement for $\mathbf{T}_n$ is complete by observing that $\mathbf{T}_n=n(\widehat{\rho}_n-1)=n(\widehat{\rho}_n-\rho)-n(1-\rho)=n(\widehat{\rho}_n-\rho)+c$.

For $\mathbf{t}_n$, notice that the proof for Lemma C.0.14 goes through for $\rho=1+c/n$, $c<0$. Thus by Slutsky's theorem, (4.7) follows. ♢

Now we prove the bootstrap consistency. The proof can be done using the large-block small-block argument as presented in the proof of Theorem 3.1 in Shao (2010a). Let $L_n=\lfloor(n/l_n)^{1/2}\rfloor$ be the length of a large-block and l_n be that of a small-block. Note that $L_n\to\infty$ and $l_n=o(L_n)$. Our goal is to assign points $t\in\{1,2,\ldots,\lfloor nr\rfloor\}$ to alternating large and small blocks. Let $K_n=K_{n,r}=\lfloor\lfloor nr\rfloor(L_n+l_n)^{-1}\rfloor$ be the number of the large (small) blocks. Define the kth large-block $\mathcal{L}_k=\{j\in\mathbb{N}:(k-1)(L_n+l_n)+1\le j\le k(l_n+L_n)-l_n\}$ for $1\le k\le K_n$, and the kth small-block $\mathcal{S}_k=\{j\in\mathbb{N}:k(L_n+l_n)-l_n+1\le j\le k(l_n+L_n)\}$ for $1\le k\le K_n-1$ and $\mathcal{S}_{K_n}=\{j\in\mathbb{N}:K_n(L_n+l_n)-l_n+1\le j\le\lfloor nr\rfloor\}$.

Let $U_k=\sum_{j\in\mathcal{L}_k}W_ju_j$ and $V_k=\sum_{j\in\mathcal{S}_k}W_ju_j$, $k=1,...,K_n$. Define $\mathcal{B}_L=\{k:\mathcal{L}_k$ contains a break point b_j for some $j=0,\ldots,\tau\}$ and $\mathcal{B}_S=\{k:\mathcal{S}_k$ contains a break point b_j for some $j=0,\ldots,\tau\}$. Notice that there are only finitely many (less than τ) elements in $\mathcal{B}_L$ and $\mathcal{B}_S$.

[1] Recall that $J_{c,\sigma}(r)$ is defined as $dJ_{c,\sigma}(r)=cJ_{c,\sigma}(r)dr+\sigma(s)dB(r)$. Using Itô's formula, we can derive $J_{c,\sigma}^2(r)=J_{c,\sigma}^2(0)+\int_0^r 2cJ_{c,\sigma}^2(s)ds+\int_0^r 2\sigma(s)J_{c,\sigma}(s)dB(s)+\int_0^r\sigma^2(s)ds$, which leads to the desired result.

LEMMA **C.0.15**. *Assume (A1)-(A4) and (B1)-(B2). Then*

$$\sup_{r\in[0,1]}\left|n^{-1}\sum_{k=1}^{K_n}\sum_{j,j'\in\mathcal{L}_k} cov(u_j,u_{j'})a\{(j-j')/l_n\}-\int_0^r\sigma^2(s)ds\right|=o(1). \tag{C.10}$$

Proof of Lemma C.0.15. Suppose $k\notin\mathcal{B}_L$. We first show that

$$\sup_{r\in[0,1]}\left|n^{-1}\sum_{k\notin\mathcal{B}_L}\sum_{j,j'\in\mathcal{L}_k}\text{cov}(u_j,u_{j'})a\{(j-j')/l_n\}-\int_0^r\sigma^2(s)ds\right|=o(1). \tag{C.11}$$

Recall that $\zeta_s=j$ such that $s\in[b_j,b_{j+1})$ and $\zeta_1=\tau$, and $c(s;h)=c_{\zeta_s}(s;h)$. Since $a(\cdot)=0$ outside of its support [-1,1], by Lemma C.0.8 (i) and (ii), we have $L_n^{-1}\sum_{j,j'\in\mathcal{L}_k}\text{cov}(u_j,u_{j'})a\{(j-j')/l_n\}=c(k/K_n;0)+O(L_n/n)+2\sum_{h=1}^{l_n}(1-h/L_n)a(h/l_n)\{c(k/K_n;h)+O(L_n/n)\}=\sigma^2(k/K_n)-2\sum_{h=1}^{\infty}d_hc(k/K_n;h)+O(l_nL_n/n)$, where $d_h=1-(1-h/L_n)a(h/l_n)$ if $0\le h\le l_n$ and 1 if $h>l_n$. Following from (B2) and Lemma C.0.8 (iii), we have $\sum_{h=1}^{\infty}d_hc(k/K_n;h)\le Cl_n^{-q}\{k_q+o(1)\}\sum_{h=1}^{\infty}h^qc(k/K_n;h)+C\overline{a}L_n^{-1}\sum_{h=1}^{\infty}hc(k/K_n;h)$ $\le C(l_n^{-q}+L_n^{-1})=o(1)$, where $\overline{a}=\sup_{s\in[-1,1]}a(s)$ and C is a constant that does not depend on k or r.

Thus we have

$$\sup_{k\notin\mathcal{B}_L}\left|L_n^{-1}\sum_{j,j'\in\mathcal{L}_k}\text{cov}(u_j,u_{j'})a\{(j-j')/l_n\}-\sigma^2(k/K_n)\right|\le C\{l_n^{-q}+L_n^{-1}\}=o(1) \tag{C.12}$$

so that $\sup_{r\in[0,1]}|n^{-1}\sum_{k\notin\mathcal{B}_L}\sum_{j,j'\in\mathcal{L}_k}\text{cov}(u_j,u_{j'})a\{(j-j')/l_n\}-n^{-1}\sum_{k\notin\mathcal{B}_L}\sigma^2(k/K_n)L_n|=o(1)$. Since $\sup_{r\in[0,1]}|n^{-1}\sum_{k\notin\mathcal{B}_L}\sigma^2(k/K_n)L_n-\sum_{k\notin\mathcal{B}_L}\int_{(k-1)/K_n}^{k/K_n}\sigma^2(s)ds|=o(1)$ by Lemma C.0.8 (iv) and $\sup_{r\in[0,1]}|\sum_{k\notin\mathcal{B}_L}\int_{(k-1)/K_n}^{k/K_n}\sigma^2(s)ds-\int_0^r\sigma^2(s)ds|=o(1)$, we have (C.11) proved. If $k\in\mathcal{B}_L$, we have, due to (A2),

$$\left|n^{-1}\sum_{k\in\mathcal{B}_L}\sum_{j,j'\in\mathcal{L}_k}\text{cov}(u_j,u_{j'})a\{(j-j')/l_n\}\right|=O(n^{-1}\tau L_n^2)=o(1). \tag{C.13}$$

Thus (C.10) follows from (C.11) and (C.13). ◊

LEMMA **C.0.16**. *Assume (A1)-(A4) and (B1)-(B2). Fix $r\in(0,1]$. Then we have*

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}W_tu_t\xrightarrow{\mathcal{D}}N\left(0,\int_0^r\sigma^2(s)ds\right)\quad\textit{in probability.} \tag{C.14}$$

Proof of Lemma C.0.16. The left-hand side of (C.14) can be decomposed into large- and small-block parts as $n^{-1/2}\sum_{t=1}^{\lfloor nr \rfloor} W_t u_t = n^{-1/2}\sum_{k=1}^{K_n} U_k + n^{-1/2}\sum_{k=1}^{K_n} V_k$. Note that $E^*(U_k) = 0$ for all $k = 1, ..., K_n$ and since W_t's are l_n-dependent, $U_1, ..., U_{K_n}$ are independent random variables conditional on $\mathcal{X}_n$. The same property holds for $V_1, ..., V_{K_n}$.

We first show that the large-block part converges to the limit in (C.14), i.e.,

$$n^{-1/2}\sum_{k=1}^{K_n} U_k \xrightarrow{\mathcal{D}} N\left(0, \int_0^r \sigma^2(s)ds\right) \quad \text{in probability.} \tag{C.15}$$

Using the same argument as in the equation (A.3) in Shao (2010a) and Hölder's inequality, we have

$$\sum_{k=1}^{K_n} E^*|U_k|^4 \le C l_n^2 L_n \sum_{k=1}^{K_n}\sum_{j\in\mathcal{L}_k} |u_j|^4. \tag{C.16}$$

Shao (2010a)'s argument works here because everything is conditional on $\mathcal{X}_n$, and the property of W_t remains the same. From (A2), we have $E|u_j|^4 \le C$ for $j = 1, \dots, n$ so that $\sum_{k=1}^{K_n}\sum_{j\in\mathcal{L}_k}|u_j|^4 \le \sum_{j=1}^n |u_j|^4 = O_p(n)$. Thus we have $\sum_{k=1}^{K_n} E^*|U_k|^4 = O_p(l_n^2 L_n n) = O_p\{(nl_n)^{3/2}\}$. Since for any $\epsilon > 0$, $E^*\{U_k^2 \mathbf{1}(|U_k| > n^{1/2}\epsilon)\} \le (n^{1/2}\epsilon)^{-2} E^*\{|U_k|^4 \mathbf{1}(|U_k| > n^{1/2}\epsilon)\} \le n^{-1}\epsilon^{-2}E^*|U_k|^4$ holds for all k, we have $n^{-1}\sum_{k=1}^{K_n} E^*\{U_k^2\mathbf{1}(|U_k| > n^{1/2}\epsilon)\} = O_p\{(l_n^3/n)^{1/2}\} = o_p(1)$. Then (C.15) follows from Lemma C.0.15.

Next we show that the contribution from small-blocks $n^{-1/2}\sum_{k=1}^{K_n} V_k$ is negligible, i.e.,

$$n^{-1/2}\sum_{k=1}^{K_n} V_k = o_p^*(1). \tag{C.17}$$

For $k \notin \mathcal{B}_S$, by Lemma C.0.8 (i) and (iii), we have $E\{E^*(V_k^2)\} = E\big[\sum_{j,j'\in\mathcal{S}_k} u_j u_{j'} a\{(j-j')/l_n\}\big] = \sum_{j,j'\in\mathcal{S}_k} \mathrm{cov}(u_j, u_{j'}) a\{(j-j')/l_n\} \le l_n \sum_{h=0}^{l_n-1}\{c(k/K_n; h) + C(l_n/n)\}a(h/l_n) \le Cl_n$. For $k = K_n$, using a similar argument, we have $E\{E^*(V_{K_n}^2)\} \le CL_n$. For $k \in \mathcal{B}_S$ and $k \ne K_n$, $E\{E^*(V_{K_n}^2)\} \le C\tau l_n^2$. Since $\tau < \infty$, we have $\sum_{k=1}^{K_n} E\{E^*(V_k^2)\} \le C(K_n l_n + l_n^2 + L_n) = o(n)$. Then (C.17) follows from the Markov inequality, independence of V_k's, and linearity of expectation. The proof is completed in view of (C.15) and (C.17). ♢

Before we prove Theorem 4.3.1, we first show the following two lemmas.

LEMMA **C.0.17**. *Assume (A1)-(A4) and (B1)-(B2). Then for* $0 < r_1 < r_2 \le 1$ *and* $n \ge n_0$ *for some* *positive integer n_0, conditional on the data $\mathcal{X}_n$,*

$$E^*\left|n^{-1/2}\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} W_t u_t\right|^4 \le \overline{C}(\mathcal{X}_n)\big\{(r_2-r_1)^2 + n^{-p_1}(r_2-r_1)\big\}, \qquad \text{(C.18)}$$

for some $p_1 > 0$, $\overline{C}(\mathcal{X}_n)$ that does not depend on r_1 or r_2, and $\overline{C}(\mathcal{X}_n) = O_p(1)$. Furthermore,

$$n^{-1/2}\sum_{t=1}^{\lfloor Nr\rfloor} W_t u_t \Rightarrow B_\sigma(r) \qquad \textit{in probability.} \qquad \text{(C.19)}$$

Proof of Lemma C.0.17. First we prove (C.18). We again use the large-block small-block argument. Recall that $U_k = \sum_{j\in\mathcal{L}_k} W_j u_j$ and $V_k = \sum_{j\in\mathcal{S}_k} W_j u_j$ for $k = 1, ..., K_n$, $L_n = \lfloor (n/l_n)^{1/2}\rfloor$, and $K_{n,r} = O(\lfloor\lfloor nr\rfloor (L_n + l_n)^{-1}\rfloor)$. Let $K_1 = K_{n,r_1}$ and $K_2 = K_{n,r_2}$ for convenience. Define $p_2 = (1-3\kappa)/2 > 0$ and $p_3 = \kappa q$, where κ and q are from (B1) and (B2), respectively. Denote by $p_1 = \min(p_2, p_3)$. By the Cr-inequality,

$$E^*\left|\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} W_t u_t\right|^4 = E^*\left|\sum_{k=K_1+1}^{K_2} U_k + \sum_{k=K_1+1}^{K_2} V_k\right|^4 \le 2^3\left(E^*\left|\sum_{k=K_1+1}^{K_2} U_k\right|^4 + E^*\left|\sum_{k=K_1+1}^{K_2} V_k\right|^4\right).$$

Since U_k and V_k are independent conditional on the data and have mean 0, the following inequality holds for U_k,

$$E^*\left|\sum_{k=K_1+1}^{K_2} U_k\right|^4 = \sum_{k=K_1+1}^{K_2} E^*(U_k^4) + \sum_{k\neq k'} E^*(U_k^2 U_{k'}^2) \le \sum_{k=K_1+1}^{K_2} E^*(U_k^4) + \left\{\sum_{k=K_1+1}^{K_2} E^*(U_k^2)\right\}^2$$

as well as for V_k.

For the large-block part, from (C.16) and (A2),

$$n^{-2}\sum_{k=K_1+1}^{K_2} E^* U_k^4 \le n^{-2} C l_n^2 L_n \sum_{k=K_1+1}^{K_2}\sum_{j\in\mathcal{L}_k} |u_j|^4 \le C_1(\mathcal{X}_n) n^{-p_2}(r_2 - r_1), \qquad \text{(C.20)}$$

where $C_1(\mathcal{X}_n) = O_p(1)$. Following from (C.10), (C.12), and (C.13), for any $0 \le r_1 < r_2 \le 1$, we have $E\left|n^{-1}\sum_{k=K_1+1}^{K_2} E^* U_k^2 - \int_{r_1}^{r_2}\sigma^2(s)ds\right| \le C\{l_n^{-q} + L_n^{-1}\} \le C(n^{-p_3} + n^{-p_2}) \le Cn^{-p_1}$. Note that the constant

C does not depend on r_1 or r_2. Thus we have

$$n^{-2}\left(\sum_{k=K_1+1}^{K_2} E^* U_k^2\right)^2 \leq C_2(\mathcal{X}_n)(r_2-r_1)^2 + C_3(\mathcal{X}_n)n^{-p_1}(r_2-r_1), \tag{C.21}$$

where $c = \{\sup_{s\in[0,1]}\sigma^2(s)\}^2 < \infty$ is a finite constant, $C_2(\mathcal{X}_n)$ and $C_3(\mathcal{X}_n) = O_p(1)$.

For the small block part, note that $K_2 - K_1 \leq Cn(r_2-r_1)/L_n = C(r_2-r_1)(nl_n)^{1/2}$ from the definition of K_1, K_2, and L_n, and $E^*V_k^4 = O_p(l_n^4)$ by (A2) and (B1). Thus we have

$$n^{-2}\sum_{k=K_1+1}^{K_2} E^* V_k^4 = O_p\{n^{-2}l_n^4(K_2-K_1)\} = C_4(\mathcal{X}_n)(l_n^3/n)n^{-p_2}(r_2-r_1), \tag{C.22}$$

where $C_4(\mathcal{X}_n) = O_p(1)$. Also, it has been shown that $n^{-1}\sum_{k=K_1+1}^{K_2} E^*V_k^2 = O_p\{(K_2-K_1)l_n/n\} = O_p(1)n^{-p_2}(r_2-r_1)$, which implies

$$\left(n^{-1}\sum_{k=K_1+1}^{K_2} E^* V_k^2\right)^2 = C_5(\mathcal{X}_n)n^{-2p_2}(r_2-r_1)^2, \tag{C.23}$$

where $C_5(\mathcal{X}_n) = O_p(1)$. It is worth noting that $C_j(\mathcal{X}_n)$'s $j = 1, \ldots, 5$ in (C.20), (C.21), (C.22), and (C.23) do not depend on r_1 or r_2. Therefore an upper bound for the left-hand side of (C.18) is

$$2^3\left[\left\{C_2(\mathcal{X}_n) + C_5(\mathcal{X}_n)n^{-2p_2}\right\}(r_2-r_1)^2 + \left\{C_1(\mathcal{X}_n) + C_3(\mathcal{X}_n) + C_4(\mathcal{X}_n)(l_n^3/n)\right\}n^{-p_1}(r_2-r_1)\right],$$

so that (C.18) holds for large enough n with $\overline{C}(\mathcal{X}_n) = 2^3\max\{C_2(\mathcal{X}_n), C_1(\mathcal{X}_n) + C_3(\mathcal{X}_n)\} + 1$.

Now we prove (C.19). The finite-dimensional convergence,

$$\left(n^{-1/2}\sum_{t=1}^{\lfloor nr_1\rfloor} W_t u_t, \ldots, n^{-1/2}\sum_{t=1}^{\lfloor nr_k\rfloor} W_t u_t\right) \xrightarrow{\mathcal{D}} \left\{\int_0^{r_1}\sigma(s)dB(s), \ldots, \int_0^{r_k}\sigma(s)dB(s)\right\}$$

in probability for any $k \in \mathbb{N}$, and $r_1, \ldots, r_k$ follows from a similar argument presented in Lemma C.0.16 and the Cramér-Wold device. The tightness follows from (C.18) and the argument of Theorem 2.1 in Shao and Yu (1996). Thus the proof is now complete. ♢

Lemma **C.0.18**. *Under the conditions (A1)-(A4) and (B1)-(B2),*

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} X_{t-1}W_t(\rho-\widehat{\rho}_n)\Rightarrow 0 \qquad \textit{in probability}$$

both under the null hypothesis $\rho = 1$ and the alternative hypothesis $|\rho| < 1$.

Proof of Lemma C.0.18. The proof is done once the following two statements are shown both under the null and alternative hypotheses.

$$\left| n^{-1/2}(\rho-\widehat{\rho}_n)\sum_{t=1}^{\lfloor nr\rfloor} X_{t-1}W_t \right| = o_p^*(1) \qquad \text{for any } r\in[0,1] \tag{C.24}$$

$$E^*\left| n^{-1/2}(\rho-\widehat{\rho}_n)\sum_{t=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} X_{t-1}W_t \right|^4 \le \overline{C}(\mathcal{X}_n)\{(r_2-r_1)^2+n^{-p_1}(r_2-r_1)\} \tag{C.25}$$

for some $p_1 > 0$, $\overline{C}(\mathcal{X}_n)$ does not depend on r_1 or r_2, and $\overline{C}(\mathcal{X}_n) = O_p(1)$. Note that $\rho-\widehat{\rho}_n = O_p(n^{-1})$ under the null by Theorem 2.1, and $\rho-\widehat{\rho}_n = O_p(n^{-1/2})$ under the alternative.

We first show (C.24) for $r = 1$. By Chebyshev's inequality, $P^*(|\sum_{t=1}^n X_{t-1}W_t| > \lambda) \le \lambda^{-2}E^*|\sum_{t=1}^n X_{t-1} W_t|^2 = C\lambda^{-2}\sum_{t=1}^n\sum_{h=0}^{l_n} X_{t-1}X_{t+h-1}a(h/l_n)$. Observe that $E|X_{t-1}X_{t+h-1}| \le ||X_{t-1}||_2||X_{t+h-1}||_2 \le C(t+h)$ by Lemma C.0.11 (i) under the null and $E|X_{t-1}X_{t+h-1}| \le C$ by Lemma C.0.11 (ii) under the alternative. Then under the null, $E\{P^*(|n^{-3/2}\sum_{t=1}^n X_{t-1}W_t| > \delta)\} \le Cn^{-3}\delta^{-2}\sum_{t=1}^n\sum_{h=0}^{l_n}(t+h) \le Cn^{-3}(n^2l_n) = O(n^{-1}l_n) = o(1)$ for any $\delta > 0$, and under the alternative, $E\{P^*(|n^{-1}\sum_{t=1}^n X_{t-1}W_t| > \delta)\} \le Cn^{-2}\delta^{-2}\sum_{t=1}^n\sum_{h=0}^{l_n} 1 \le Cn^{-2}(nl_n) = O(n^{-1}l_n) = o(1)$ for any $\delta > 0$. A similar proof works for general $r \in (0, 1]$. Thus (C.24) is proved.

Now we show (C.25). We first define indices for large and small blocks $\mathcal{S}_k$ and $\mathcal{L}_k$ as before. Decompose $\sum_{t=1}^{\lfloor nr\rfloor} X_{t-1}W_t = \sum_{k=1}^{K_{n,r}}\mathbf{U}_k + \sum_{k=1}^{K_{n,r}}\mathbf{V}_k$ into large and small blocks. Recall that $K_{n,r} = \lfloor\lfloor nr\rfloor(L_n+l_n)^{-1}\rfloor$ is the number of large and small blocks, $L_n = \lfloor(n/l)^{1/2}\rfloor$ is the length of the large block, and $l_n \asymp Cn^{-\kappa}$ with $\kappa \in (0, 1/3)$. Let $K_1 = K_{n,r_1}$ and $K_2 = K_{n,r_2}$.

Following the same argument as in the proof of (C.18), we need to examine the upper bounds of $\sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k^4)$, $\sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k^2)$, $\sum_{k=K_1+1}^{K_2} E^*(\mathbf{V}_k^4)$, and $\sum_{k=K_1+1}^{K_2} E^*(\mathbf{V}_k^2)$. In the subsequent argument, $C(\mathcal{X}_n)$, $C_1(\mathcal{X}_n)$, $C_2(\mathcal{X}_n)$, $C_3(\mathcal{X}_n)$, and $C_4(\mathcal{X}_n)$ are all $O_p(1)$ that do not depend on r_2 or r_1. In particular, $C(\mathcal{X}_n)$ may have different values in different places.

First, consider when the null hypothesis is true. Following the same argument as in (22) or Shao (2010a)'s (A.3), we have $\sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k^4) \leq Cl_n^2 L_n \sum_{k=K_1+1}^{K_2} \sum_{j\in\mathcal{L}_k} |X_{j-1}|^4 \leq C(\mathcal{X}_n) l_n^2 L_n \sum_{j=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} j^2$ $\leq C(\mathcal{X}_n) l_n^2 L_n (\lfloor nr_2\rfloor^3 - \lfloor nr_1\rfloor^3) \leq C(\mathcal{X}_n) l_n^2 L_n n^3 (r_2 - r_1)$, where the second inequality is due to Lemma C.0.11 (i). Since, $l_n^2 L_n n^{-3} = l^{3/2} n^{-5/2} = O(n^{-(3\kappa+5)/2})$, letting $p_1 = (3\kappa+5)/2$, we have

$$n^{-6} \sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k^4) \leq C_1(\mathcal{X}_n) n^{-p_1}(r_2 - r_1). \tag{C.26}$$

For the next object, notice that by Lemma C.0.11 (i), $E\{E^*(\mathbf{U}_k^2)\} = E\{E^*(\sum_{t\in\mathcal{L}_k} X_{t-1}W_t)^2\} \leq 2\sum_{t\in\mathcal{L}_k} \sum_{h=0}^{l} E(X_{t-1}X_{t-h+1})a(h/l) \leq C\sum_{t\in\mathcal{L}_k}\sum_{h=0}^{l} t$ so that $\sum_{k=K_1+1}^{K_2} E\{E^*(\mathbf{U}_k^2)\} \leq Cl_n(\lfloor nr_2\rfloor^2 - \lfloor nr_1\rfloor^2) \leq Cl_n n^2(r_2 - r_1)$ and

$$n^{-6}\left\{\sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k^2)\right\}^2 \leq l_n^2 n^{-2} C_2(\mathcal{X}_n)(r_2 - r_1)^2. \tag{C.27}$$

Now consider when the alternative hypothesis is true. The proof is similar to the null case. Note that $\sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k^4) \leq Cl_n^2 L_n \sum_{k=K_1+1}^{K_2} \sum_{j\in\mathcal{L}_k} |X_{j-1}|^4 \leq C(\mathcal{X}_n) l_n^2 L_n \sum_{j=\lfloor nr_1\rfloor+1}^{\lfloor nr_2\rfloor} 1 \leq C(\mathcal{X}_n) l_n^2 L_n (\lfloor nr_2\rfloor - \lfloor nr_1\rfloor) \leq C(\mathcal{X}_n) l_n^2 L_n n (r_2 - r_1)$ due to Lemma C.0.11 (ii), so that

$$n^{-4} \sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k^4) \leq C_3(\mathcal{X}_n) n^{-p_1}(r_2 - r_1). \tag{C.28}$$

Also, $E\{E^*(\mathbf{U}_k^2)\} = E\{E^*(\sum_{t\in\mathcal{L}_k} X_{t-1}W_t)^2\} \leq 2\sum_{t\in\mathcal{L}_k}\sum_{h=0}^{l} E(X_{t-1}X_{t-h+1})a(h/l) \leq C\sum_{t\in\mathcal{L}_k}\sum_{h=0}^{l_n} 1$ by Lemma C.0.11 (ii) so that $\sum_{k=K_1+1}^{K_2} E\{E^*(\mathbf{U}_k^2)\} \leq Cl_n(\lfloor nr_2\rfloor - \lfloor nr_1\rfloor) \leq Cl_n n(r_2 - r_1)$ and

$$n^{-4}\left(\sum_{k=K_1+1}^{K_2} E^*(\mathbf{U}_k)^2\right)^2 \leq l_n^2 n^{-2} C_4(\mathcal{X}_n)(r_2 - r_1)^2. \tag{C.29}$$

The same arguments work for small blocks replacing $\mathbf{U}_k$ in (C.26), (C.27), (C.28), and (C.29) with $\mathbf{V}_k$. Thus the proof is complete. $\diamondsuit$

Now we are ready to prove Theorem 4.3.1.

Proof of Theorem 4.3.1. Observe that $n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} u_t^* = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} \widehat{u}_t W_t = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}(X_t - \widehat{\rho}_n X_{t-1})$

$W_t = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}(\rho X_{t-1} + u_t - \widehat{\rho}_n X_{t-1})W_t = \left\{n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} X_{t-1}W_t\right\}(\rho - \widehat{\rho}_n) + n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} W_t u_t =: \mathcal{I}_{1,r} + \mathcal{I}_{2,r}$. Since we have $\mathcal{I}_{1,r} \Rightarrow 0$ in probability by Lemma C.0.18 and $\mathcal{I}_{2,r} \Rightarrow B_\sigma(r)$ in probability by Lemma C.0.17, the proof is complete. $\diamondsuit$

Proof of Theorem 4.3.2. We claim that under both the null and alternative,

$$n^{-1}\sum_{t=1}^{n}\{(u_t^*)^2 - E^*(u_t^*)^2\} = o_p^*(1), \quad \text{and} \tag{C.30}$$

$$n^{-1}\sum_{t=1}^{n}\{E^*(u_t^*)^2 - u_t^2\} = o_p(1). \tag{C.31}$$

Once (C.30) and (C.31) are shown, we have $n^{-1}\sum_{t=1}^{n}\{(u_t^*)^2 - u_t^2\} = o_p^*(1)$. Then using the similar argument as in the proof of Lemma C.0.13 (i) and (ii), Theorem 4.3.2 follows from the continuous mapping theorem, Theorem 4.3.1, and the fact that $n^{-1}\sum_{t=1}^{n} u_t^2 \xrightarrow{\mathcal{P}} \sigma_u^2$, which is due to Lemma C.0.12 (ii).

We first prove (C.31). It can be written as $n^{-1}\sum_{t=1}^{n}\{E^*(u_t^*)^2 - u_t^2\} = n^{-1}\sum_{t=1}^{n}(\widehat{u}_t^2 - u_t^2) = n^{-1}\sum_{t=1}^{n}[\{u_t + (\rho - \widehat{\rho}_n)X_{t-1}\}^2 - u_t^2] = (\rho - \widehat{\rho}_n)^2 n^{-1}\sum X_{t-1}^2 + 2(\rho - \widehat{\rho}_n)n^{-1}\sum X_{t-1}u_t =: \mathcal{I}_1 + \mathcal{I}_2$. Here, $\mathcal{I}_k = O_p(n^{-1})$ for all $k = 1, 2$ under both the null and the alternative, since $(\rho - \widehat{\rho}_n) = O_p(n^{-1})$, $n^{-1}\sum X_{t-1}^2 = O_p(n)$, and $n^{-1}\sum X_{t-1}u_t = O_p(1)$ under the null, and $(\rho - \widehat{\rho}_n) = O_p(n^{-1/2})$, $n^{-1}\sum X_{t-1}^2 = O_p(1)$, and $n^{-1}\sum X_{t-1}u_t = O_p(n^{-1/2})$ under the alternative. Thus (C.31) is complete.

Now we prove (C.30). Observe that $\sum_{t=1}^{n}\{(u_t^*)^2 - E^*(u_t^*)^2\} = \sum_{t=1}^{n}\widehat{u}_t^2(W_t^2 - 1)$. For any $\delta > 0$, $P^*\{|\sum_{t=1}^{n}\widehat{u}_t^2(W_t^2 - 1)| > n\delta\} \le (n\delta)^{-2}E^*\{\sum_{t=1}^{n}\widehat{u}_t^2(W_t^2 - 1)\}^2 \le (n\delta)^{-2}C\{\sum_{t=1}^{n}\sum_{h=0}^{l_n}\widehat{u}_t^2\widehat{u}_{t+h}^2\}$, and our goal is to show $\sum_{t=1}^{n}\sum_{h=0}^{l_n}\widehat{u}_t^2\widehat{u}_{t+h}^2 = o_p(n^2)$. Since $\widehat{u}_t = u_t + (\rho - \widehat{\rho}_n)X_{t-1}$, it can be written as

$$\begin{aligned}
\sum_{t=1}^{n}\sum_{h=0}^{l_n}\widehat{u}_t^2\widehat{u}_{t+h}^2 \quad = \quad & \sum_{t=1}^{n}\sum_{h=0}^{l_n} u_t^2 u_{t+h}^2 \\
& +2(\rho - \widehat{\rho}_n)\sum_{t=1}^{n}\sum_{h=0}^{l_n}\{u_t^2 u_{t+h}X_{t+h-1} + u_{t+h}^2 u_t X_{t-1}\} \\
& +(\rho - \widehat{\rho}_n)^2\sum_{t=1}^{n}\sum_{h=0}^{l_n}\{u_t^2 X_{t+h-1}^2 + u_{t+h}^2 X_{t-1}^2 + 4u_t u_{t+h}X_{t-1}X_{t+h-1}\} \\
& +2(\rho - \widehat{\rho}_n)^3\sum_{t=1}^{n}\sum_{h=0}^{l_n}\{u_{t+h}X_{t-1}^2 X_{t+h-1} + u_t X_{t+h-1}^2 X_{t-1}\} \\
& +(\rho - \widehat{\rho}_n)^4\sum_{t=1}^{n}\sum_{h=0}^{l_n} X_{t-1}^2 X_{t+h-1}^2 \\
=: \quad & \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4 + \mathcal{I}_5.
\end{aligned}$$

We claim that $\mathcal{I}_j = o_p(n^2)$ for all $j = 1, \ldots, 5$ under both the null and the alternative hypotheses.

First, under the alternative $|\rho| < 1$, note that due to the Cauchy-Schwartz inequality, (A2), and Lemma C.0.11 (ii), we have $\sup_{t_1,t_2} E|u_{t_1}^2 u_{t_2}^2| \le C$, $\sup_{t_1,t_2,t_3} E|u_{t_1}^2 u_{t_2} X_{t_3}| \le C$, $\sup_{t_1,t_2,t_3,t_4} E|u_{t_1} u_{t_2} X_{t_3} X_{t_4}| \le C$, $\sup_{t_1,t_2,t_3} E|u_{t_1} X_{t_2}^2 X_{t_3}| \le C$, and $\sup_{t_1,t_2} E|X_{t_1}^2 X_{t_2}^2| \le C$. Since $\rho - \widehat{\rho}_n = O_p(n^{-1/2})$, we have $\mathcal{I}_j = O_p(n^{(3-j)/2} l_n) = o_p(n^2)$ for all $j = 1, \ldots, 5$.

Now it is enough to show that $\mathcal{I}_j = o_p(n^2)$ for $j = 2, \ldots, 5$, under the null $\rho = 1$. Recall that $\rho - \widehat{\rho}_n = O_p(n^{-1})$ under the null.

We write $\mathcal{I}_2 =: \mathcal{I}_{2,1} + \mathcal{I}_{2,2}$. Observe that

$$\begin{aligned}\mathcal{I}_{2,1} &= \sum_{t=1}^n u_t^2 \left(\sum_{h=0}^{l_n} X_{t+h-1} u_{t+h}\right) \le \left\{\sum_{t=1}^n u_t^4\right\}^{1/2} \left\{\sum_{t=1}^n \left(\sum_{h=0}^{l_n} X_{t+h-1} u_{t+h}\right)^2\right\}^{1/2} \\ &= \{O_p(n)\}^{1/2} \left\{\sum_{t=1}^n \sum_{h=0}^{l_n} \sum_{h'=0}^{l_n} X_{t+h-1} u_{t+h} X_{t+h'-1} u_{t+h'}\right\}^{1/2}\end{aligned}$$

by Hölder's inequality. Since $E|X_{t+h-1}u_{t+h}X_{t+h'-1}u_{t+h'}| \le ||X_{t+h-1}u_{t+h}||_2 ||X_{t+h'-1}u_{t+h'}||_2 \le (||X_{t+h-1}^2||_2 ||u_{t+h}^2||_2 ||X_{t+h'-1}^2||_2 ||u_{t+h'}^2||_2)^{1/2} \le C\{(t+h-1)(t+h'-1)\}^{1/2}$ by Hölder's inequality, (A2), and Lemma C.0.11 (i), we have $E(\sum_{t=1}^n \sum_{h=0}^{l_n} \sum_{h'=0}^{l_n} X_{t+h-1}u_{t+h}X_{t+h'-1}u_{t+h'}|) \le C \sum_{t=1}^n \{(t+l_n)^{3/2} - t^{3/2}\}^2 = O(l_n^2 n^2 + l_n^6)$. Thus $\mathcal{I}_{2,1} = O_p\{n^{-1} n^{1/2} (l_n^2 n^2 + l_n^6)^{1/2}\} = O_p(n^{1/2} l_n + n^{-1/2} l_n^3) = o_p(n^2)$. Similarly, we can show that $\mathcal{I}_{2,2} = o_p(n^2)$, which leads to $\mathcal{I}_2 = o_p(n^2)$. For $\mathcal{I}_3$, $\mathcal{I}_4$, and $\mathcal{I}_5$, the arguments are done using Hölder's inequality, (A2), and Lemma C.0.11 (i) for all summands. For $\mathcal{I}_3 =: \mathcal{I}_{3,1} + \mathcal{I}_{3,2} + \mathcal{I}_{3,3}$, observe that for any $t_1, t_2, t_3, t_4 \in \{1, \ldots, n\}$,

$$E|u_{t_1} u_{t_2} X_{t_3} X_{t_4}| \le ||u_{t_1} u_{t_2}||_2 ||X_{t_3} X_{t_4}||_2 \le \{E(u_{t_1}^4) E(u_{t_2}^4) E(X_{t_3}^4) E(X_{t_4}^4)\}^{1/4} \le Cn.$$

Thus $\mathcal{I}_3 = O_p(n^{-2} n n l_n) = o_p(n)$. For $\mathcal{I}_4 =: \mathcal{I}_{4,1} + \mathcal{I}_{4,2}$, observe that for any $t_1, t_2, t_3 \in \{1, \ldots, n\}$

$$\begin{aligned}E|u_{t_1} X_{t_2}^2 X_{t_3}| &\le ||u_{t_1}||_4 ||X_{t_2}^2 X_{t_3}||_{4/3} \le C\{E(X_{t_2}^{8/3} X_{t_3}^{4/3})\}^{3/4} \le C(||X_{t_2}^{8/3}||_{3/2} ||X_{t_3}^{4/3}||_3)^{3/4} \\ &\le C\{(EX_{t_2}^4)^{2/3} (EX_{t_3}^4)^{1/3}\}^{3/4} = C(EX_{t_2}^4)^{1/2} (EX_{t_3}^4)^{1/4} \le C t_2 t_3 \le Cn^2.\end{aligned}$$

Thus $\mathcal{I}_4 = O_p(n^{-3} n^2 n l_n) = o_p(n)$. For $\mathcal{I}_5$, notice that $E(X_{t-1}^2 X_{t+h-1}^2) \le ||X_{t-1}^2||_2 ||X_{t+h-1}^2||_2 = \{E(X_{t-1}^4) E(X_{t+h-1}^4)\}^{1/2} \le C(t-1)(t+h-1) \le Cn^2$. Thus $\mathcal{I}_5 = O_p(n^{-4} n^2 n l_n) = o_p(n^2)$, which completes the proof. ♢

References

Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association 93*(444), 1488–1501.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), 953–966.

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics 7*(2), 151–174.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Bunzel, H., N. M. Kiefer, and T. Vogelsang (2001). Simple robust testing of hypotheses in nonlinear models. *Journal of the American Statistical Association 96*, 1088–1096.

Bunzel, H. and T. Vogelsang (2005). Powerful trend function tests that are robust to strong serial correlation, with an application to the prebisch-singer hypothesis. *Journal of Business and Economic Statistics 23*, 381–394.

Busetti, F. and R. Taylor (2003). Variance shifts, structural breaks, and stationarity tests. *Journal of Business and Economic Statistics 21*(4), 510–531.

Canjels, E. and M. W. Watson (1997). Estimating deterministic trends in the presence of serially correlated errors. *The Review of Economics and Statistics 79*(2), 184–200.

Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews 23*(3), 259–292.

Cavaliere, G. and A. M. R. Taylor (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics 140*(2), 919–947.

Cavaliere, G. and A. M. R. Taylor (2008a). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis 29*(2), 300–330.

Cavaliere, G. and A. M. R. Taylor (2009a). Bootstrap M unit root tests. *Econometric Reviews 28*(5), 393–421.

Cavaliere, G. and A. M. R. Taylor (2009b). Heteroskedastic time series with a unit root. *Econometric Theory 25*(5), 1228–1276.

Cavaliere, G. and A. R. Taylor (2008b). Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory 24*(1), 43–71.

Chang, Y. and J. Y. Park (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis 24*(4), 379–400.

Csörgő, M., B. Szyszkowicz, and Q. Wang (2003). Donsker's theorem for self-normalized partial sums processes. *The Annals of Probability 31*(3), 1228–1240.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics 25*(1), 1–37.

Davidson, J. E. H. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366), 427–431.

Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica 49*(4), 1057–1072.

Domínguez, M. A. and I. N. Lobato (2001). Size corrected power for bootstrap tests. Working Papers 102, Centro de Investigacion Economica, ITAM.

Draghicescu, D., S. Guillas, and W. B. Wu (2009). Quantile curve estimation and visualization for nonstationary time series. *Journal of Computational and Graphical Statistics 18*(1), 1–20.

Fitzenberger, B. (1998). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics 82*(2), 235–287.

Giurcanu, M. and V. Spokoiny (2004). Confidence estimation of the covariance function of stationary and locally stationary processes. *Statistics and Decisions 22*(4), 283–300.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2007). A simple, robust and powerful test of the trend hypothesis. *Journal of Econometrics 141*(2), 1302–1330.

Kiefer, N. M. and T. Vogelsang (2002). Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica 70*(5), 2093–2095.

Kiefer, N. M. and T. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 1130–1164.

Kiefer, N. M., T. Vogelsang, and H. Bunzel (2000). Simple robust testing of regression hypotheses. *Econometrica 68*(3), 695–714.

Kim, C.-J. and C. R. Nelson (1999). Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *The Review of Economics and Statistics 81*(4), 608–616.

Koop, G. and M. Steel (1994). A decision-theoretic analysis of the unit-root hypothesis using mixtures of elliptical models. *Journal of Business and Economic Statistics 12*(1), 95–107.

Kreiss, J.-P. (1988). Asymptotic statistical inference for a class of stochastic processes. Habilitationsschrift, Universität Hamburg.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics 17*(3), 1217–1241.

Kurtz, T. G. (2001). *Lectures on Stochastic Analysis.* University of Wisconsin-Madison: available at http://www.math.wisc.edu/ kurtz/735/main735.pdf.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data.* Springer.

Lee, W.-M. (2006). Robust tests of hypotheses in models with M-estimation. Working paper, National Chung Cheng University.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association 96*(455), 1066–1076.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* New York: Springer-Verlag.

Mallat, S., G. Papanicolaou, and Z. Zhang (1998). Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics 26*(1), 1–47.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics 21*(1), 255–285.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review 90*(5), 1464–1476.

Nelson, C. R. and C. I. Plosser (1982). Trends and random walks in macroeconmic time series : Some evidence and implications. *Journal of Monetary Economics 10*(2), 139–162.

Newey, W. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Palm, F. C., S. Smeekes, and J.-P. Urbain (2008). Bootstrap unit-root tests: Comparison and extensions. *Journal of Time Series Analysis 29*(2), 371–401.

Paparoditis, E. and D. N. Politis (2002). Local block bootstrap. *Comptes Rendus Mathematique 335*(11), 959–962.

Paparoditis, E. and D. N. Politis (2003). Residual-based block bootstrap for unit root testing. *Econometrica 71*(3), 813–855.

Paparoditis, E. and D. N. Politis (2005). Bootstrapping unit root tests for autoregressive time series. *Journal of the American Statistical Association 100*(470), 545–553.

Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. *Econometrica 57*(6), 1361–1401.

Perron, P. and S. Ng (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies 63*(3), 435–463.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica 55*(2), 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*(2).

Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics 20*(2), 971–1001.

Phillips, P. C. B. and Z. Xiao (1998). A primer on unit root testing. *Journal of Economic Surveys 12*(5), 423–469.

Phillips, P. C. B. and K.-L. Xu (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis 27*(2), 289–308.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. New York: Springer.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B 27*(2), 204–237.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Volume 1. New York: Academic Press.

Psaradakis, Z. (2001). Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors. *Journal of Time Series Analysis 22*(5), 577–594.

Rao, S. S. (2004). On multiple regression models with nonstationary correlated errors. *Biometrika 91*(3), 645–659.

Rho, Y. and X. Shao (2013). Improving the bandwidth-free inference methods by prewhitening. *Journal of Statistical Planning and Inference 143*(11), 1912–1922.

Romano, J. P. and M. Wolf (2006). Improved nonparametric confidence intervals in time series regressions. *Journal of Nonparametric Statistics 18*(2), 199–214.

Rosenblatt, M. (1985). *Stationary Squences and Random Fields*. Birkhũser.

Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika 71*(3), 599–607.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics 86*(3), 833–839.

Shao, Q.-M. and H. Yu (1996). Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability 24*(4), 2098–2127.

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 72*(3), 343–366.

Shao, X. and W. B. Wu (2007). Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics 35*(4), 1773–1801.

Sherman, M. (1997). Subseries methods in regression. *Journal of the American Statistical Association 92*(439), 1041–1048.

Smeekes, S. and R. Taylor (2012). Bootstrap union tests for unit roots in the presence of nonstationary volatility. *Econometric Theory 28*(2), 422–456.

Stock, J. and M. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R. Engle and H. White (Eds.), *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*, pp. 1–44. Oxford: Oxford University Press.

Swensen, A. R. (2003). Bootstrapping unit root tests for integrated processes. *Journal of Time Series Analysis 24*(1), 99–126.

Taniguchi, M. and Y. Kakizawa (2000). *Asymptotic Theory of Statistical Inference for Time Series.* New York: Springer.

Vogelsang, T. (1998). Trend function hypothesis testing in the presence of serial correlation. *Econometrica 66*(1), 123–148.

White, H. (1984). *Asymptotic Theory for Econometricians.* New York: Academic Press.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *The Annals of Statistics 14*(4), 1261–1350.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America 102*(40), 14150–14154.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability 35*(6), 2294–2320.

Wu, W. B. and X. Shao (2004). Limit theorems for iterated random functions. *Journal of Applied Probability 41*(2), 425–436.

Wu, W. B. and Z. Zhou (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sininca 21*(3), 1397–1413.

Xu, K.-L. (2008). Bootstrapping autoregression under non-stationary volatility. *Econometrics Journal 11*(1), 1–26.

Xu, K.-L. (2012). Robustifying multivariate trend tests to nonstationary volatility. *Journal of Econometrics 169*(2), 147–154.

Xu, K.-L. and P. C. Phillips (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics 142*(1), 265–280.

Zhang, T. and W. B. Wu (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika 98*(3), 599–614.

Zhang, X. and X. Shao (2013). Fixed-smoothing asymptotics for time series. *The Annals of Statistics 41*(3), 1329–1349.

Zhao, Z. (2011). A self-normalized confidence interval for the mean of a class of nonstationary processes. *Biometrika 98*(1), 81–90.

Zhao, Z. and X. Li (2013). Inference for modulated stationary processes. *Bernoulli 19*(1), 205–227.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*(502), 726–740.

Zhou, Z. and X. Shao (2013). Inference for linear models with dependent errors. *Journal of the Royal Statistical Society: Series B 75*(2), 323–343.

Zhou, Z. and W. B. Wu (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics 37*(5B), 2696–2729.