Inference in Mildly Explosive Autoregressions under Unconditional Heteroskedasticity

Xuewen Yu* **Mohitosh Kejriwal†**

Purdue University *Purdue University*

November 3, 2021

Abstract

Mildly explosive autoregressions have been extensively employed in recent theoretical and applied econometric work to model the phenomenon of asset market bubbles. An important issue in this context concerns the construction of confidence intervals for the autoregressive parameter that represents the degree of explosiveness. Existing studies rely on intervals that are justified only under conditional homoskedasticity/heteroskedasticity. This paper studies the problem of constructing asymptotically valid confidence intervals in a mildly explosive autoregression where the innovations are allowed to be unconditionally heteroskedastic. The assumed variance process is general and can accommodate both deterministic and stochastic volatility specifications commonly adopted in the literature. Within this framework, we show that the standard heteroskedasticity-autocorrelation consistent (HAC) estimate of the long-run variance converges in distribution to a nonstandard random variable that depends on nuisance parameters. Notwithstanding this result, the corresponding t-statistic is shown to still possess a standard normal limit distribution. To improve the quality of inference in small samples, we propose a dependent wild bootstrap-t procedure and establish its asymptotic validity under relatively weak conditions. Monte Carlo simulations demonstrate that our recommended approach performs favorably in finite samples relative to existing methods across a wide range of volatility specifications. Applications to international stock price indices and US house prices illustrate the relevance of the advocated method in practice.

Keywords: mildly explosive, heteroskedasticity, nonstationary volatility, HAC, long-run variance, bubbles

JEL Classification: C22

*Krannert School of Management, Purdue University, 403 West State Street, West Lafayette IN 47907 (yu656@purdue.edu).

†Krannert School of Management, Purdue University, 403 West State Street, West Lafayette IN 47907 (mkejriwa@purdue.edu).

1 Introduction

Over the past decade, the mildly explosive autoregressive (MEA) framework has emerged as a popular econometric device for modeling the phenomenon of asset market bubbles. The framework has been extensively utilized to develop a multitude of procedures for detecting and dating the origination and termination of bubbles as well as constructing asymptotically valid confidence intervals for the size of the bubble. These methods equip policymakers with a powerful set of econometric tools that can be employed to mitigate the potentially adverse consequences of a bubble and thereby maintain economic and financial stability. The techniques have been successfully applied to a variety of asset prices including stock prices, house prices, cryptocurrencies, commodity prices, etc. For detailed reviews of this literature, see inter alia, Phillips and Shi (2020) and Skrobotov (2021).

Introduced by Phillips and Magdalinos (2005, 2007a), the MEA framework posits that the autoregressive parameter which represents the degree of explosiveness (i.e., the size of the bubble) evolves as a function of the sample size (T) according to $\rho_T = 1 + c/k_T,\ c > 0,\ k_T \to \infty,\ k_T/T \to 0$, as $T \to \infty$. The motivation for adopting this specification emanated from the fact that modeling the parameter as fixed and independent of the sample size (i.e., $\rho_T = \rho > 1$) precludes the application of an invariance principle so that the limit distribution of its least squares estimate depends on the underlying error distribution that is typically unknown in practice (Anderson, 1959). Wang and Yu (2015) show that in the fixed parameter autoregression of order one with an intercept, the standard t-statistic on the autoregressive coefficient has a nonstandard limit distribution that depends on the initial value of the stochastic process as well as the true values of the model parameters. On the other hand, modeling the autoregressive coefficient as local-to-unity ($\rho_T = 1+c/T$) facilitates the application of functional central limit theory and leads to a limit distribution that is not reliant on the error distribution but depends on the local-to-unity parameter c which cannot be consistently estimated (Phillips, 1987). Phillips and Magdalinos (2007a) show that the MEA framework permits the application of an invariance principle that induces a Cauchy limit distribution for the least squares estimate without assuming Gaussian errors.

While the Cauchy limit was initially derived assuming independently and identically distributed (i.i.d.) innovations, subsequent work has demonstrated the same limit distribution continues to hold under weak or strong dependence in the errors (Magdalinos, 2012) and conditional heteroskedasticity (Arvanitis and Magdalinos, 2018). Fei (2018) and Liu and Peng (2019) show that the inclusion of a fixed non-zero drift in the first order MEA model

with i.i.d. errors leads to a standard normal limit distribution for the t-statistic. Assuming weakly dependent and conditionally homoskedastic innovations, Guo et al. (2019) show that the t-statistic based on a heteroskedasticity and autocorrelation robust (HAR) estimate of the standard error follows a Student's t distribution in large samples. They further establish the invariance of the limit distribution to a possible drift in the process, regardless of whether the drift dominates the explosive stochastic component. Chan et al. (2012) develop an empirical likelihood-based confidence interval in the first order autoregressive model with i.i.d. errors that is asymptotically valid for stationary, unit root, near-integrated and fixed parameter explosive processes thereby providing a unified approach to inference.

While the aforementioned inference methods are justified under conditional heteroskedasticity, they rule out the presence of unconditional heteroskedasticity. A plethora of empirical studies, however, document that several macroeconomic and financial time series exhibit time-varying unconditional volatility profiles. For example, Sensier and van Dijk (2004) find that approximately 80% of 214 macroeconomic time series over the period 1959-99 were subject to a break in unconditional volatility with the break date estimated at 1984 and associated with a reduction in volatility for a large number of series (the so-called "Great Moderation"). Harvey et al. (2016) reject the null hypothesis of stationary volatility in the prices of two types of crude oil, three precious metals (gold, silver and platinum) and two non-ferrous metals (aluminum and copper) using a battery of four tests developed in Cavaliere and Taylor (2007b). Based on the rejection patterns of the tests, they conclude that a single discrete break volatility model or a trending volatility model might be appropriate for these series. Using the same set of tests, Astill et al. (2018) also find statistically significant evidence against stationary volatility for three out of five major stock price indices: the FTSE All Share index (UK), the Nasdaq Composite index (USA) and the Nikkei 225 index (Japan). Kurozumi et al. (2020) use estimated variance profiles to document the presence of time-varying volatility in the twelve largest cryptocurrencies by capitalization.

Motivated by these considerations, this paper studies the problem of constructing asymptotically valid confidence intervals in a mildly explosive autoregression with weakly dependent innovations that are allowed to be unconditionally heteroskedastic. Our framework adopts a general specification for the volatility process that can accommodate both deterministic and stochastic volatility specifications commonly employed in the literature (see section 2). Within this framework, we show that the standard heteroskedasticity-autocorrelation consistent (HAC) estimate of the long-run variance converges in distribution to a nonstandard random variable that depends on nuisance parameters. Notwithstanding this result, the

corresponding t-statistic is shown to still possess a standard normal limit distribution. To improve the quality of inference in small samples, we propose a dependent wild bootstrap-t procedure that can simultaneously account for time-varying unconditional volatility and weak dependence in the errors. The large sample validity of the proposed approach is established under relatively weak conditions. The theoretical analysis does, however, rule out the possibility that the sign of the current shock affects future volatility, commonly referred to as leverage. This is due to the fact that conditional on the data, the bootstrap innovations are independent over time. Monte Carlo simulations demonstrate that our proposed approach performs favorably in finite samples relative to existing methods across a wide range of volatility specifications. In particular, the dependent wild bootstrap confidence interval is shown to be adept at maintaining coverage close to the nominal confidence level while controlling average length both for data generating processes with and without leverage effects. The relevance of the proposed method in practice is illustrated in two empirical applications concerning international stock price indices and US house prices, respectively.

The bootstrap approach has been employed in prior work concerning inference in fixed parameter explosive autoregressive processes. Basawa et al. (1989) establish the asymptotic validity of the standard i.i.d. bootstrap in the first order autoregressive process with i.i.d. errors. More recently, Cavaliere et al. (2020) develop bootstrap-based inference procedures in non-causal autoregressions with heavy-tailed innovations. They show that the asymptotic distribution of the least squares estimate in this framework is non-pivotal in that it depends on the tail behavior of the innovations. To address this issue, three alternative choices for the bootstrap are considered, namely, the wild bootstrap, the permutation bootstrap and a permutation wild bootstrap. Sufficient conditions for the validity of each of these choices in large samples are provided. In contrast, the goal of the present paper is to study the properties of asymptotic and wild bootstrap procedures for conducting inference within the MEA framework with weakly dependent errors and time-varying volatility.

Our paper is also closely related to a strand of the literature that studies stable and unit root autoregressions under unconditional heteroskedasticity. Working in a stable autoregressive framework with deterministic volatility, Phillips and Xu (2006) develop inference procedures based on a non-parametric kernel-based estimate of the variance function while Xu and Phillips (2008) employ the estimated variance function to propose adaptive least squares estimation of the autoregressive coefficients and demonstrate via simulations the efficiency gains achievable over ordinary least squares estimation. Goncalves and Kilian (2004) propose a wild bootstrap approach to inference in stable autoregressions under conditional

heteroskedasticity of unknown form while Xu (2008) extends their work by showing that the wild bootstrap remains valid under time-varying unconditional volatility. Xu (2008) also studies the effects of allowing for deterministic and stochastic volatility on the consistency, rate of convergence and limit distributions of the least squares estimates. The non-robustness of standard unit root tests to unconditional heteroskedasticity was demonstrated, inter alia, by Cavaliere (2005) and Cavaliere and Taylor (2007a), who show that the limit distributions of these tests are non-pivotal and depend on the time-varying variance profile. Cavaliere and Taylor (2007a) and Beare (2018) propose unit root tests that employ a non-parametric estimate of the variance profile where the former uses critical values simulated from the limit distribution while the latter is based on standard null asymptotic critical values. Cavaliere and Taylor (2008, 2009) develop wild bootstrap tests of the unit root hypothesis and establish their asymptotic validity while Boswijk and Zu (2018) propose adaptive wild bootstrap unit root tests based on non-parametric volatility estimation and show that they achieve the same asymptotic power envelope as in the known volatility case. Boswijk and Zu (2021) show that adaptation with respect to the volatility process also delivers more powerful tests of cointegration. It is useful to note that the use of the bootstrap in this strand of the literature is motivated by the non-pivotal nature of the limit distributions of standard test statistics that depend on the unknown volatility process. In contrast, the dependent wild bootstrap adopted in the current paper is motivated by its ability to provide a more reliable approximation to the finite sample distribution of the standard HAC-based t-statistic than that afforded by its standard normal limit distribution.

The rest of the paper is organized as follows. Section 2 lays out the modeling framework and associated assumptions. Section 3 discusses methods for constructing asymptotic confidence intervals for the autoregressive parameter. Section 4 proposes a dependent wild bootstrap-t procedure for inference and develops its large sample properties. Section 5 presents a set of Monte Carlo experiments comparing the finite sample adequacy of the different methods in terms of coverage and expected length for a variety of volatility specifications. Section 6 details two empirical applications to illustrate the practical relevance of the proposed approach. Section 7 concludes. Supplementary (not for publication) Appendices A and B contain the proofs of the theoretical results and additional Monte Carlo results are reported in Supplementary Appendix C. As a matter of notation, $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{w}$ denotes weak convergence, and $\xrightarrow{w}_p$ denotes weak convergence in probability under the bootstrap measure.

2 The Model and Assumptions

Consider a scalar random variable y_t generated by the following mildly autoregressive process with possibly non-zero drift:

$$y_t = \mu_T + \rho_T y_{t-1} + u_t \quad t = 1, ..., T \tag{1}$$

$$\rho_T = 1 + \frac{c}{k_T}, \quad c > 0 \tag{2}$$

$$u_t = C(L)e_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \quad e_t = \sigma_t \varepsilon_t \tag{3}$$

The data generating process (DGP) in (1)-(3) allows for weakly dependent errors modeled via the polynomial $C(.)$ and for conditional as well as unconditional heteroskedasticity modeled through the volatility function σ_t. A special case of this DGP with $\sigma_t = \sigma \ \forall \ t$, was considered by Phillips and Magdalinos (2007b) and Guo et al. (2019). Specifically, our analysis is based on the following assumptions:

Assumption 1: (a) $k_T \to \infty$ and $k_T/T \to 0$ as $T \to \infty$; (b) The drift term satisfies $\mu_T \sqrt{k_T} \to \nu \in [0, \infty]$ as $T \to \infty$.

Assumption 2: (a) The lag polynomial satisfies $C(z) \neq 0$ for all $|z| \leq 1$, $C(1) \in (0, \infty)$ and $\sum_{j=0}^{\infty} j|c_j| < \infty$; (b) $\varepsilon_t \sim i.i.d.(0, 1)$ with $\sup_t \mathrm{E}(\varepsilon_t^{4+\kappa_1}) \leq K_1 < \infty$ for some $\kappa_1 > 0$; (c) for some strictly positive deterministic sequence $\{a_T\}$, $\{\sigma_t\}$ satisfies $a_T^{-1}\sigma_{\lfloor Ts \rfloor} \xrightarrow{w} g(s), \ s \in [0, 1]$, with $0 < \underline{g} \leq g(s) \leq \bar{g} < \infty$ a.s. for all s, and $\sup_t \mathrm{E}(a_T^{-1}\sigma_t)^{4+\kappa_2} \leq K_2 < \infty$ for some $\kappa_2 > 0$; (d) σ_t is independent of ε_s for any t and s; (e) the initial value y_0 is independent of $\{u_t\}_{t=1}^T$ and $a_T^{-1} y_0 = o(k_T^{1/2})$.

Assumption 3: The sequence $\{a_T\}$ satisfies $a_T = \psi T^{\gamma}$, where ψ and γ are constants with $\psi > 0$ and $\gamma \in [0, \infty)$.

Assumption 1(a) characterizes the mildly explosive framework developed by Phillips and Magdalinos (2007a) whereby the explosive root approaches unity at a sufficiently slow rate relative to the sample size. Assumption 1(b) specifies the drift component μ_T following Guo et al. (2019) and allows the drift to be small ($\nu \in [0, \infty)$) or large ($v = \infty$). Given that the magnitude of the drift is typically unknown in practice, potential model misspecification can be avoided by including a constant in the estimated regression.

Assumption 2(a) imposes conditions on the lag polynomial that ensures that the errors u_t are weakly dependent and admits a Beveridge-Nelson decomposition (see Phillips and

Solo, 1992). Assumption 2(b) specifies the innovations ε_t to be i.i.d. with bounded fourth moments. While we adopt the i.i.d. assumption to simplify the theoretical analysis, we expect that the results in the paper will continue to hold under the weaker condition that $\{\varepsilon_t, \mathcal{F}_t\}$ is a martingale difference sequence with respect to $\mathcal{F}_t = \sigma\text{-}field\{\varepsilon_s, s \leq t\}$,satisfying (i) $\mathrm{E}(\varepsilon_t^2) = 1$ for all t; (ii) $T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} 1$; and (iii) $\sup_t \mathrm{E}(\varepsilon_t^{4+\eta}) \leq K_1 < \infty$ for some $\eta > 0$. Assumption 2(c) allows for a wide class of volatility processes including a variety of deterministic and non-deterministic specifications for $\{\sigma_t\}$ commonly employed in the literature. In the deterministic case, the assumption allows single and multiple volatility shifts, linearly and polynomially trending volatility (with an appropriate choice of a_T), and smooth transition breaks. In the non-deterministic case, the class of volatility processes allowed includes non-stationary autoregressive stochastic volatility (SV) models (Hansen, 1995), SV models with jumps (Georgiev, 2008; Perron and Qu, 2010), non-stationary nonlinear heteroskedastic models with stochastically trending volatility, and near-integrated GARCH models.(see Cavaliere and Taylor, 2009, for a detailed discussion of the class of volatility processes permissible under this assumption). Assumption 2(d) precludes the possibility that the sign of the current shock affects future volatilities, often referred to as leverage. This assumption is needed to ensure the asymptotic validity of the proposed dependent wild bootstrap approach since the wild bootstrap innovations cannot replicate any leverage effects that may be present in the original data. Nevertheless, we examine the sensitivity of the various methods considered in the paper to the failure of this assumption via simulations in Section 5. Assumption 2(e) guarantees the invariance of the limit theory to the initial condition.

The specification for a_T adopted in Assumption 3 nests all of the volatility models in the examples considered by Cavaliere and Taylor (2009). Specifically, when $\gamma = 0,\ \psi = 1$, i.e., $a_T = 1$, it incorporates the models in their examples 1-4 and when $\gamma > 0$, it incorporates the models in their examples 5 and 6. As can be seen in the subsequent analysis, a_T will have an non-negligible effect on the asymptotic theory, a feature also observed in Xu (2008) which studies the asymptotic properties of OLS estimators in stationary autoregressive models with nonstationary volatility.

3 Asymptotic Confidence Intervals

The objective of the paper is to analyze the properties of alternative methods for constructing confidence intervals for the autoregressive parameter ρ_T in (1) in the potential presence of unconditional heteroskedasticity of the form specified in Assumption 2. This section first discusses existing methods based on an asymptotic approximation to the sampling

distribution of the least squares estimate of ρ_T or the corresponding t-statistic under the assumption that the innovations are unconditionally homoskedastic, i.e., $\mathrm{E}(e_t^2) = \sigma^2$ for all t. Subsequently, we consider the standard t-statistic based on the usual heteroskedasticity and autocorrelation-consistent (HAC) estimate of the long-run variance of u_t (e.g., Andrews, 1991). We show that despite the nonstandard nature of the limit distribution of the HAC estimate, the t-statistic is still asymptotically standard normal. In what follows, we denote $z_t = (1, y_{t-1})'$, $Z_t = (z_1, ..., z_t)'$, $U_t = (u_1, ..., u_t)$, and $\iota_2 = (0, 1)'$.

3.1 Existing Inference Methods

Phillips and Magdalinos (2007a) considered a version of (1)-(3) with no drift under the assumption that the errors u_t are i.i.d. and square integrable. They establish that the following limit theory holds as $T \to \infty$:

$$\frac{k_T \rho_T^T}{2c}(\widetilde{\rho}_T - \rho_T) \xrightarrow{w} \mathcal{C} \text{ and } \frac{\rho_T^T}{\rho_T^2 - 1}(\widetilde{\rho}_T - \rho_T) \xrightarrow{w} \mathcal{C} \tag{4}$$

where $\widetilde{\rho}_T = (\sum_{t=1}^T y_{t-1}^2)^{-1}(\sum_{t=1}^T y_{t-1} y_t)$ denotes the least squares estimate and $\mathcal{C}$ denotes a standard Cauchy random variable. It follows that a $100(1-\delta)\%$ confidence interval for ρ_T can be constructed as

$$\left(\widetilde{\rho}_T \pm \frac{\widetilde{\rho}_T^2 - 1}{\widetilde{\rho}_T^T} \mathcal{C}_\delta\right) \tag{5}$$

where $\mathcal{C}_\delta$ is the two-tailed δ percentile critical value of the standard Cauchy distribution. For instance, a 95% confidence interval will use the critical value $C_{0.05} = 12.7$ compared to the corresponding Gaussian critical value of 1.96. We will refer to (5) as the PM interval.

Phillips and Magdalinos (2007b) showed that (4) remains valid even when the errors u_t are weakly dependent and satisfy Assumption 1(a) while imposing conditional homoskedasticity by assuming $e_t \sim i.i.d.(0, \sigma^2)$. Magdalinos (2012) extended the validity of (4) to include error processes that can be strongly dependent (i.e., exhibiting long memory) thereby demonstrating the robustness of the interval (5) to a general dependence structure in the innovation sequence. More recently, Arvanitis and Magdalinos (2018) established that the Cauchy limit theory is also invariant to a wide class of stationary conditionally heteroskedastic error processes with weak or strong dependence.[1]

[1] Lee (2017) considers a framework in which u_t is assumed to be strong mixing with exponentially decaying coefficients and finite fourth moments. In contrast, Arvanitis and Magdalinos (2018) does not require strong mixing and instead relies on a L_1-mixingale condition on e_t which does not place restrictions on the moments of u_t higher than order 2.

Based on the limit result (4), Guo et al. (2019) show that under the assumption that the errors u_t are i.i.d., the OLS t-statistic that does not correct for heteroskedasticity or autocorrelation has a standard normal limiting distribution. In the no drift case ($\mu_T = 0$), this t-statistic is given by

$$t_{PM} = \frac{\widetilde{\rho}_T - \rho_T}{\sqrt{\widetilde{s}_T^2 \left(\sum_{t=1}^T y_{t-1}^2\right)^{-1}}} \xrightarrow{w} N(0,1)$$

where $\widetilde{s}_T^2 = (T-1)^{-1}\sum_{t=1}^T (y_t - \widetilde{\rho}_T y_{t-1})^2$. When the estimated regression includes a constant, Guo et al. (2019) establish, under Assumptions 1 and 2(e), that

$$t_{MED} = \frac{\hat{\rho}_T - \rho_T}{\sqrt{\hat{s}_T^2 \iota_2' \left(\sum_{t=1}^T z_t z_t'\right)^{-1} \iota_2}} \xrightarrow{w} N(0,1)$$

where $(\hat{\mu}_T, \hat{\rho}_T)' = \left(\sum_{t=1}^T z_t z_t'\right)^{-1} \sum_{t=1}^T z_t y_t$, and $\hat{s}_T^2 = (T-2)^{-1}\sum_{t=1}^T (y_t - \hat{\mu}_T - \hat{\rho}_T y_{t-1})^2$.

In the case of weakly dependent errors u_t given by (3) where $e_t \sim i.i.d.(0, \sigma^2)$ with finite fourth moments and $C(.)$ satisfies Assumption 2(a), Guo et al. (2019) develop an inference procedure based on a orthonormal series long-run variance estimator that accounts for the dependence structure. Specifically, they propose the t-statistic

$$\widetilde{t}_{MED} = \frac{\hat{\rho}_T - \rho_T}{\sqrt{\hat{\lambda}_K^2 \iota_2' \left(\sum_{t=1}^T z_t z_t'\right)^{-1} \iota_2}} \tag{6}$$

where $\hat{\lambda}_K^2$ is the estimate of the long-run variance of u_t constructed from the estimated residuals $\hat{u}_t = y_t - \hat{\mu}_T - \hat{\rho}_T y_{t-1}$ as

$$\hat{\lambda}_K^2 = \frac{1}{K}\sum_{j=1}^K \left[\frac{1}{\sqrt{T}}\sum_{t=1}^T \phi_j\left(\frac{t}{T}\right)\hat{u}_t\right]^2 \tag{7}$$

In (7), K is an even constant and $\phi_j(x) = \sqrt{2}\sin(2\pi j x)$ and $\phi_{2j-1}(x) = \sqrt{2}\cos(2\pi j x)$ are the Fourier basis functions. Guo et al. (2019) show that under the fixed-K asymptotics where $T \to \infty$ for a given K, $\widetilde{t}_{MED} \xrightarrow{w} t_K$, where t_K is the Student's t distribution with K degrees of freedom.[2] The choice of K is data-dependent and based on the asymptotic mean squared error criterion implemented using the AR(1) plug-in procedure. This value of K is then rounded to the closest even number between 4 and T (see Phillips, 2005).

[2] Under joint asymptotics where $K \to \infty$ as $T \to \infty$ with $K/T \to 0$, $\widetilde{t}_{MED} \xrightarrow{w} N(0,1)$.

All of the aforementioned confidence intervals are predicated upon the assumption of unconditionally homoskedastic innovations. In Section 5, we will examine their finite sample performance for data generating processes that fail this assumption via simulations. These simulation results would allow us to assess the degree to which these methods are sensitive to the underlying homoskedasticity assumption.

3.2 HAC-Based Inference

We now consider an asymptotic approach to inference that, in contrast to the extant methods described in Section 3.1, remains valid even in the presence of unconditional heteroskdasticity of the form allowable under Assumption 2. Our approach is simply based on the standard t-statistic that employs a HAC estimate of the standard errors to account for heteroskedasticity and autocorrelation (e.g., Newey and West, 1987; Andrews, 1991). In order to define this statistic, we introduce the following notation. Let $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$, $\bar{y}_{-1} = T^{-1}\sum_{t=0}^{T-1} y_t$, $\bar{u} = T^{-1}\sum_{t=1}^{T} u_t$, $\dot{y}_{t-1} = y_{t-1} - \bar{y}_{-1}$, $\dot{u}_t = u_t - \bar{u}$ for $t = 1, \cdots, T$ and

$$\hat{\Omega} = \sum_{j=-(T-1)}^{T-1} w(j/b_T)\hat{\Gamma}(j), \quad \hat{\Gamma}(j) = T^{-1}\sum_{t=1}^{T-|j|} \dot{y}_{t-1}\hat{u}_t\dot{y}_{t-1+|j|}\hat{u}_{t+|j|} \tag{8}$$

with $\hat{u}_t$ the residuals defined as in (7), $w(\cdot)$ is a kernel function and b_T is the bandwidth. The conditions on $w(\cdot)$ and b_T will be specified later. Then, letting $Q_T = T^{-1}\sum_{t=1}^{T} \dot{y}_{t-1}^2$, and $\hat{\Lambda} = T^{-1}Q_T^{-2}\hat{\Omega}$, the HAC-based t-statistic can be expressed as

$$t_{hac} := \frac{\hat{\rho}_T - \rho_T}{\hat{\Lambda}^{\frac{1}{2}}} \tag{9}$$

We will now establish that under Assumptions 1 and 2 accompanied by suitable conditions on $w(\cdot)$ and b_T, the statistic t_{hac} has a standard normal limit distribution. To this end, we first decompose y_t in (1) into two parts, i.e., $y_t = d_t + \mu_T(\rho_T^t - 1)k_T/c$, where d_t follows

$$d_t = \rho_T d_{t-1} + u_t, \quad d_0 = y_0 \tag{10}$$

Now d_t is a mildly explosive process without drift while $\mu_T(\rho_T^t - 1)k_T/c$ is a deterministic nonlinear trend component when $\mu_T \neq 0$. The following result derives the limits of two random quantities which will be useful in the subsequent analysis, where $MN(0, V_x)$ and $MN(0, V_y)$ denote mixed Gaussian random variables with mixing variates V_x and V_y, respectively.

Theorem 1 *Under Assumptions 1-2,*

(a) $X_T := a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}u_t \xrightarrow{w} X \sim MN(0,V_x),\ V_x = C^2(1)\lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tc(1-s)}g^2(s)ds$

(b) $Y_T := a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\rho_T^{-t}u_t \xrightarrow{w} Y \sim MN(0,V_y),\ V_y = C^2(1)\lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tcs}g^2(s)ds$

(c) $0 < V_x, V_y < \infty$

Remark 1 *In the conditionally homoskedastic case where* $g(s) = \sigma, \forall s \in [0,1]$*, Theorem 1 degenerates to the results in Phillips and Magdalinos (2007a), as* $\lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tcs}g^2(s)ds = \lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tc(1-s)}g^2(s)ds = \frac{\sigma^2}{2c}$.

Next, we obtain the limit distribution of the least squares estimate $\hat{\rho}_T$ of ρ_T in (1). Note that

$$\hat{\rho}_T - \rho_T = \frac{\sum_{t=1}^{T}(y_{t-1}-\bar{y}_{-1})u_t}{\sum_{t=1}^{T}(y_{t-1}-\bar{y}_{-1})^2} = \frac{\sum_{t=1}^{T}y_{t-1}u_t - T^{-1}\sum_{t=1}^{T}y_{t-1}\sum_{t=1}^{T}u_t}{\sum_{t=1}^{T}y_{t-1}^2 - T^{-1}(\sum_{t=1}^{T}y_{t-1})^2} \tag{11}$$

The following theorem presents the limits of the sample statistics appearing in (11):

Theorem 2 *Under Assumptions 1-3, defining* $1/\infty = 0$*, we have the following joint convergence results:*

$$\begin{aligned}
&(a)\quad (a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T}y_{t-1}^2 \xrightarrow{w} \begin{cases} \frac{Y^2}{2c\nu^2} & \gamma > 0 \\ \frac{1}{2c}(\frac{Y}{\nu}+\frac{1}{\psi c})^2 & \gamma = 0 \end{cases} \\
&(b)\quad (a_T\mu_Tk_T^2\rho_T^T)^{-1}\sum_{t=1}^{T}y_{t-1} \xrightarrow{w} \begin{cases} \frac{Y}{c\nu} & \gamma > 0 \\ \frac{1}{c}(\frac{Y}{\nu}+\frac{1}{\psi c}) & \gamma = 0 \end{cases} \\
&(c)\quad (a_T^2\mu_Tk_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T}y_{t-1}u_t \xrightarrow{w} \begin{cases} \frac{XY}{\nu} & \gamma > 0 \\ X(\frac{Y}{\nu}+\frac{1}{\psi c}) & \gamma = 0 \end{cases} \\
&(d)\quad a_T^{-1}T^{-1/2}\sum_{t=1}^{T}u_t \xrightarrow{w} U \sim MN(0,\sigma_u^2), \quad \sigma_u^2 = C^2(1)\int_0^1 g(s)^2ds
\end{aligned} \tag{12}$$

In order to obtain the limit distribution of $\hat{\rho}_T$, we make the following additional assumption that restricts the rate at which volatility grows with the sample size:

Assumption 3′: $\gamma < 1/2$

The upper bound on the rate of growth of a_T specified in Assumption 3′ ensures the consistency of the intercept estimate $\hat{\mu}_T$. Interestingly, this upper bound condition coincides

with that in Xu (2008) to ensure the consistency of the intercept estimate in the stationary autoregressive framework. Now it is easy to derive, using the results of Theorem 2, that

$$\begin{aligned}
&(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\left(\sum_{t=1}^{T} y_{t-1}u_t - T^{-1}\sum_{t=1}^{T} y_{t-1}\sum_{t=1}^{T} u_t\right) \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1}u_t - T^{-1/2}k_T^{1/2}(a_T\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1} \times a_T^{-1}T^{-1/2}\sum_{t=1}^{T} u_t \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1}u_t - O_p(T^{-1/2}k_T^{1/2}) \xrightarrow{w} \frac{1}{2c}\left[\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0)\right]^2 \qquad (13)
\end{aligned}$$

$$\begin{aligned}
&(a_T^2\mu_T^2 k_T^3\rho_T^{2T})^{-1}\left(\sum_{t=1}^{T} y_{t-1}^2 - T^{-1}(\sum_{t=1}^{T} y_{t-1})^2\right) \\
= \;& (a_T^2\mu_T^2 k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T} y_{t-1}^2 - T^{-1}k_T[(a_T\mu_T k_T^2\rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1}]^2 \\
= \;& (a_T^2\mu_T^2 k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T} y_{t-1}^2 - O_p(T^{-1}k_T) \xrightarrow{w} X\left[\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0)\right] \qquad (14)
\end{aligned}$$

We thus have the asymptotic distribution of $\hat{\rho}_T$ as stated in the following corollary:

Corollary 1 *Under Assumptions 1-2 and 3′,*

$$\mu_T k_T^{3/2}\rho_T^T(\hat{\rho}_T - \rho_T) \xrightarrow{w} \frac{2cX}{\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0)} \qquad (15)$$

Remark 2 *In the special case $\nu = \infty$, the limit distribution in* (15) *reduces to $2c^2X$, which implies $\hat{\rho}_T$ is asymptotically mixed normal. A similar result assuming $\sigma_t = \sigma$ in* (3) *was obtained in Guo et al. (2019) where X reduces to a $N(0, \sigma^2/2c)$ random variable. In the more general case where the volatility structure follows Assumption 2, the mixing variate V_x takes a more complex form that depends on the unknown function $g(.)$.*

We now establish a result that applies when $\nu = 0$, i.e., the no drift case:

Theorem 3 *Under Assumptions 1-2 and 3′, with $\nu = 0$, we have the following joint con-*

vergence results:

$$
\begin{aligned}
&(a) \quad (a_T^2 k_T^2 \rho_T^{2T})^{-1} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{w} \frac{1}{2c} Y^2 \\
&(b) \quad (a_T k_T^{3/2} \rho_T^T)^{-1} \sum_{t=1}^{T} y_{t-1} \xrightarrow{w} \frac{1}{c} Y \\
&(c) \quad (a_T^2 k_T \rho_T^T)^{-1} \sum_{t=1}^{T} y_{t-1} u_t \xrightarrow{w} XY \\
&(d) \quad a_T^{-1} T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{w} U \sim MN(0, \sigma_u^2), \quad \sigma_u^2 = C^2(1) \int_0^1 g(s)^2 ds
\end{aligned}
\tag{16}
$$

The following corollary states the limit distribution of the OLS estimator when $\nu = 0$:

Corollary 2 *Under Assumptions 1-3, with $\nu = 0$, we have*

$$k_T \rho_T^T (\hat{\rho}_T - \rho_T) \xrightarrow{w} \frac{2cX}{Y}$$

Remark 3 *The limit distribution of $\hat{\rho}_T$ under $\nu = 0$ does not require Assumption 3′, i.e., the rate of growth of volatility need not be restricted to be slower than $O(T^{1/2})$. In the stationary autoregressive framework analyzed by Xu (2008), the limit distribution of the autoregressive estimates does not depend on the volatility scale a_T while the limit distribution of the estimate of the deterministic component depends on a_T. In the present context, the limit distribution of $\hat{\rho}_T$ depends on the magnitude of the drift component and when this component is large enough, the limit depends on the volatility scale as well.*

Remark 4 *As demonstrated in the above analysis, the limit distribution of $\hat{\rho}_T$ is non-pivotal regardless of the magnitude of the drift as it depends on the unknown volatility process $g(.)$. In particular, the standard inferential result* (4) *based on the Cauchy distribution as derived in Phillips and Magdalinos (2007a, 2007b) is no longer valid in the current context and thus the PM interval* (5) *does not have asymptotically correct coverage. The finite sample implications of this result are investigated via simulations in Section 5.*

The final step in establishing the limit distribution of the t-statistic (9) entails obtaining the limit of the HAC estimator $\hat{\Omega}$ defined in (8). We make the following assumption that governs the behavior of the weight function $w(.)$ and the bandwidth b_T:

Assumption 4: (i) The function $w(\cdot)$ is a continuous and even function with $|w(\cdot)| \leq 1$, $w(0) = 1$ and $\int_{-\infty}^{\infty} w^2(x) < \infty$; (ii) The bandwidth satisfies $b_T^{-1} + k_T^{-1/2} b_T \to 0$ as $T \to \infty$.

The following Lemma states a key result instrumental in deriving the asymptotic distribution of the long-run variance estimator $\hat{\Omega}$:

Lemma 1 *Define* $\Phi_T(j) = a_T^{-2} k_T^{-1} \sum_{t=1}^{T-|j|} \rho_T^{-2(T-t)+|j|-2} u_t u_{t+|j|}$. *Under Assumptions 1-2, 3′ and 4, as* $T \to \infty$, *the following result holds:* $\sum_{j=-(T-1)}^{T-1} w(j/b_T)\Phi_T(j) \xrightarrow{w} V_x$.

We can then state the following result regarding the limit behavior of $\hat{\Omega}$:

Theorem 4 *Under Assumptions 1-2, 3′ and 4, as* $T \to \infty$,

$$T(a_T^2 \mu_T k_T^{3/2} \rho_T^T)^{-2} \hat{\Omega} \xrightarrow{w} V_x \left[\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0) \right]^2 \tag{17}$$

Remark 5 *The limit of* $\hat{\Omega}$ *is non-standard and depends on nuisance parameters. In particular, the limit involves the volatility function* $g(.)$, *the localizing parameter* c, *and the drift magnitude* ν.

Remark 6 *The condition* $b_T / k_T^{1/2} \to 0$ *is stronger than the condition* $b_T / T^{1/2} \to 0$ *typically adopted to establish the consistency of the long-run variance estimator in the standard stationary framework (e.g., Jansson, 2002). This condition in turn restricts the allowable set of mildly explosive neighborhoods if a data dependent rule is used to select the bandwidth as in Andrews (1991). For instance, Andrews (1991) shows that using the Quadratic Spectral kernel yields an estimated bandwidth of order* $O_p(T^{1/5})$ *which, by Assumption 4, rules out neighborhoods in which* $k_T = O(T^{2/5})$.

Using the preceding results, it is straightforward to show that

$$\begin{aligned}
(\mu_T k_T^{3/2} \rho_T^T)^2 \hat{\Lambda} &= T^{-1} (\mu_T k_T^{3/2} \rho_T^T)^2 Q_T^{-2} \hat{\Omega} = \left[T(a_T^2 \mu_T^2 k_T^3 \rho_T^{2T})^{-1} Q_T \right]^{-2} \left[T(a_T^2 \mu_T k_T^{3/2} \rho_T^T)^{-2} \hat{\Omega} \right] \\
&\xrightarrow{w} \left(\frac{1}{2c} \left[\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0) \right]^2 \right)^{-2} \left(V_x \left[\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0) \right]^2 \right) \\
&= \frac{4c^2 V_x}{\left[\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0) \right]^2}
\end{aligned} \tag{18}$$

Combining the limit (18) with the limit distribution of $\hat{\rho}_T$ in (15), we finally have

$$\begin{aligned} t_{hac} &= \frac{\hat{\rho}_T - \rho_T}{\hat{\Lambda}^{\frac{1}{2}}} = \frac{(\mu_T k_T^{3/2} \rho_T^T)(\hat{\rho}_T - \rho_T)}{\left[(\mu_T k_T^{3/2} \rho_T^T)^2 \hat{\Lambda}\right]^{\frac{1}{2}}} \\ &\xrightarrow{w} \frac{2cX/(\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0))}{\left[4c^2 V_x/(\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0))^2\right]^{\frac{1}{2}}} = \frac{X}{V_x} \sim N(0,1) \end{aligned} \tag{19}$$

The standard normal limit of t_{hac} is formalized in the following theorem:

Theorem 5 *Under Assumptions 1-2, 3′ and 4, as $T \to \infty$, we have $t_{hac} \xrightarrow{w} N(0,1)$.*

Remark 7 *A pivotal limit of t_{hac} is attained since the limit of the standard error estimate, though nuisance parameter-dependent, is proportional to the same random variable that appears in the limit distribution of the least squares estimate $\hat{\rho}_T$. The cancellation of the non-pivotal terms in the numerator and denominator of the t-statistic effectuates a pivotal limiting distribution.*

4 Dependent Wild Bootstrap

The previous section established the large sample validity of the HAC-based t-statistic in the potential presence of nonstationary volatility as well as weak dependence in the noise component within the MEA framework. In small samples, however, the performance of HAC-based asymptotic confidence intervals may be less than satisfactory as illustrated via Monte Carlo simulations in Section 5. In response to this possibility, we propose an alternative, bootstrap-based approximation to the finite sample distribution of the t-statistic that can improve upon the asymptotic approximation provided by the standard normal distribution. In particular, as the ensuing Monte Carlo comparison demonstrates, the bootstrap-based interval is shown to achieve improved coverage while controlling average length, relative to existing asymptotic methods as well as the asymptotic interval (9) based on t_{hac}.

The bootstrap procedure we adopt is the so-called dependent wild bootstrap (DWB, henceforth), introduced by Shao (2010). The DWB is designed to simultaneously capture unconditional heteroskedasticity and potential temporal dependence in the errors, and thus is a natural extension of the wild bootstrap developed by Wu (1986) and Liu (1988) for serially uncorrelated errors. While originally proposed for stationary time series by Shao (2010), several recent studies have investigated its applicability in the nonstationary time

series setup. For instance, Smeekes and Urbain (2014) study several modified wild bootstrap methods, including the DWB, in a multivariate framework and prove its asymptotic validity in testing for unit roots. Rho and Shao (2019) propose the DWB in the unit root testing context with piecewise locally stationary errors and provide justification for its consistency. Our paper contributes to the DWB literature by further extending its validity to the MEA framework allowing for general and flexible forms of variance and dependence structures in the errors.[3]

The DWB is based on generating a series of random variables $\{\eta_t\}_{t=1}^T$ that are independent of the data in order to capture the heteroskedasticity in the errors. In the original wild bootstrap (Liu, 1988), the $\{\eta_t\}_{t=1}^T$ are independent while in the DWB, the $\{\eta_t\}_{t=1}^T$ are correlated to accommodate temporal dependence in the errors. Specifically, we make the following assumption on $\{\eta_t\}_{t=1}^T$ (Shao, 2010):

Assumption 5: The series $\{\eta_t\}_{t=1}^T$ is drawn independently of the data such that $\mathrm{E}(\eta_t) = 0$, $\mathrm{Var}(\eta_t) = 1$, $\mathrm{Cov}(\eta_s, \eta_t) = K(\frac{s-t}{l_T})$, where K: $\mathbb{R} \to [0,1]$ is a symmetric kernel function that satisfies $K(0) = 1$, $K(x) = 0$ for $x \geq 1$, $\lim_{x\to 0}[1 - K(x)]/|x|^q \neq 0$ for some $q \in (0, 2]$, and $\int_{-\infty}^{\infty} K(u)e^{-iux}du \geq 0$ for $x \in R$. The quantity l_T is a bandwidth parameter satisfying $l_T = O(T^g)$, $0 < g < 1/3$. Assume η_t is l_T-dependent and $\mathrm{E}(\eta_t^4) < \infty$.

In practice, the series $\{\eta_t\}_{t=1}^T$ can be obtained by drawing samples from a multivariate normal distribution with zero mean and covariance function $\mathrm{Cov}(\eta_s, \eta_t) = K(\frac{s-t}{l_T})$.[4] Several kernels popular in practice such as the Bartlett kernel (with $q = 1$) and the Parzen and Tukey-Hanning kernels (with $q = 2$) satisfy Assumption 5. Alternative choices for the bandwidth will be explored via simulations in Section 5. In addition to the restriction on the bandwidth l_T as specified in Assumption 5, our theoretical analysis is based on the following additional assumption which is akin to Assumption 4 in the preceding section:

Assumption 6: The bandwidth l_T satisfies $k_T^{-1/2} l_T \to 0$ as $T \to \infty$.

With the OLS residuals $\hat{u}_t = y_t - \hat{\mu}_T - \hat{\rho}_T y_{t-1}$ at hand, the DWB residuals are simply constructed as $u_t^* = \eta_t \hat{u}_t$. To analyze the properties of the bootstrap samples, we first derive

[3] Weak dependence in the errors can be alternatively captured using a block boostrap based approach (e.g., Carlstein, 1986; Kunsch, 1989) or a sieve bootstrap approach (Bühlmann, 1997). Therefore, apart from the DWB analyzed in this paper, heteroskedasticity-robust versions of certain block and sieve bootstrap methods may also be viable in the present context. A comparison of alternative bootstrap approaches within the MEA framework allowing for unconditionally heteroskedastic and weakly dependent errors is a potentially interesting topic for future research.

[4] As illustrated in Example 4.1 of Shao (2010), $\{\eta_t\}_{t=1}^T$ can be also be generated from a non-normal distribution.

the following invariance principle for $\{u_t^*\}$ which parallels that derived in Theorem 1 for the original errors $\{u_t\}$:

Theorem 6 *Under Assumptions 1-2, 3′, 4-6, as* $T \to \infty$,

$$X_T^* := a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \hat{\rho}_T^{-(T-t)-1} u_t^* \xrightarrow{w}_p X \sim N(0, V_x)$$

$$Y_T^* := a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \hat{\rho}_T^{-t} u_t^* \xrightarrow{w}_p Y \sim N(0, V_y) \tag{20}$$

This theorem reveals that the DWB is able to mimic the unknown heteroskedasticity and temporal dependence in the errors within the MEA framework, thereby generalizing the set of time series to which the procedure is applicable. However, it is not fully transferable from the stationary/unit root case since the additional Assumption 6 plays a crucial role in ensuring its validity (see Appendix B for details).

We now discuss how to apply the DWB in constructing a bootstrap-based confidence interval for the autoregressive parameter ρ_T. The following algorithm enumerates the steps involved in implementation of the DWB.

Residual-based DWB Algorithm

1. Generate T bootstrap innovations η_t, $t = 1, \ldots, T$ from a multivariate normal distribution with zero mean and covariance function $\text{Cov}(\eta_s, \eta_t) = K(\frac{s-t}{l_T})$, and construct the DWB residuals $u_t^* = \eta_t \hat{u}_t$, where $\hat{u}_t = y_t - \hat{\mu}_T - \hat{\rho}_T y_{t-1}$, $t = 1, \ldots, T$ are the OLS regression residuals.

2. Construct the bootstrap samples $\{y_t^*, t = 1, \ldots, T\}$, recursively as

$$y_t^* = \hat{\mu}_T + \hat{\rho}_T y_{t-1}^* + u_t^*, \ t = 1, \ldots, T \tag{21}$$

with $y_0^* = y_0$.

3. Calculate the t_{hac} statistic defined in (9) for the bootstrap data as $t^*_{\hat{\rho}_T, hac} = (\hat{\rho}_T^* - \hat{\rho}_T)/\hat{\Lambda}^{*\frac{1}{2}}$, where $\hat{\rho}_T^*$ is the bootstrap OLS estimate and $\hat{\Lambda}^*$ is the bootstrap analogue of $\hat{\Lambda}$ computed from the estimated bootstrap residuals $\hat{u}_t^* = y_t^* - \hat{\mu}_T^* - \hat{\rho}_T^* y_{t-1}^*$. Specifically, $\hat{\Lambda}^* = T^{-1} Q_T^{*-2} \hat{\Omega}^*$, where $Q_T^* = T^{-1} \sum_{t=1}^{T} \dot{y}_{t-1}^{*2}$, $\dot{y}_{t-1}^* = y_{t-1}^* - \bar{y}_{-1}^*$, $\bar{y}_{-1}^* = T^{-1} \sum_{t=1}^{T} y_{t-1}^*$, and $\hat{\Omega}^*$ is computed as in (8),

$$\hat{\Omega}^* = \sum_{j=-(T-1)}^{T-1} w(j/b_T^*) \hat{\Gamma}^*(j), \ \ \hat{\Gamma}^*(j) = T^{-1} \sum_{t=1}^{T-|j|} \dot{y}_{t-1}^* \hat{u}_t^* \dot{y}_{t-1+|j|}^* \hat{u}_{t+|j|}^* \tag{22}$$

4. Repeat steps (1)-(3) B times to approximate the the distribution of the original statistic t_{hac}. Obtain the $\delta/2$ and $(1-\delta/2)$ quantiles from the empirical distribution of $t^*_{\hat{\rho}_T,hac}$, denoted $t^*_{\delta/2}$ and $t^*_{1-\delta/2}$, respectively. Construct the equal-tailed $100(1-\delta)\%$ bootstrap confidence interval as

$$\left(\hat{\rho}_T - \hat{\Lambda}^{\frac{1}{2}} t^*_{1-\delta/2}, \hat{\rho}_T - \hat{\Lambda}^{\frac{1}{2}} t^*_{\delta/2}\right) \tag{23}$$

Remark 8 *The bandwidth parameter b^*_T in step 3 is determined in a data dependent way as b_T in the original statistic (9), albeit based on the bootstrap data $\{y^*_t\}$. It should be noted that using the same b_T as in the original statistic for all bootstrap replications is also a valid procedure. Nevertheless, the results were found to be qualitatively similar in simulations to those reported and are available upon request.*

Finally, the asymptotic validity of the residual-based DWB is formally established in the following theorem:

Theorem 7 *Under Assumptions 1-2, 3′, 4-6, as $T \to \infty$,*

$$t^*_{\hat{\rho}_T,hac} := \frac{\hat{\rho}^*_T - \hat{\rho}_T}{\hat{\Lambda}^{*\frac{1}{2}}} \xrightarrow{w}_p N(0,1) \tag{24}$$

Theorem 7 demonstrates that the residual-based DWB is consistent, i.e., the bootstrap t-statistic has the same first-order limiting distribution as the original test statistic t_{hac}. Thus, the bootstrap statistic achieves (asymptotically) correct size and the associated bootstrap confidence interval (23) achieves (asymptotically) correct coverage.

5 Monte Carlo Simulations

This section conducts a set of Monte Carlo experiments designed to assess the finite sample adequacy of the asymptotic approximations developed in the preceding section as well as provide a numerical comparison of the proposed approach with existing approaches. In particular, we evaluate the relative efficacy of the different procedures via the coverage rates and average effective length (i.e., length conditional on covering the true parameter value) of the resulting confidence intervals. The simulation design is similar to Guo et al. (2019).

The data generating process (DGP) is given by

$$\begin{aligned} y_t &= \mu_T + \rho_T y_{t-1} + u_t \quad t = 1, ..., T \\ \rho_T &= 1 + \frac{c}{T^\alpha}, \quad c = 0.5, \ \alpha \in \{0.5, 0.8\} \end{aligned}$$

Two specifications for the drift are considered: (i) $\mu_T = 0$; (ii) $\mu_T = T^{-\alpha/4}$. For the noise component u_t, we consider the case with no serial correlation ($u_t = e_t$) as well as cases with the following autoregressive (AR) and moving average (MA) structures:

$$\begin{aligned} u_t &= \phi u_{t-1} + \sqrt{1-\phi^2} e_t \\ u_t &= \sqrt{1-\theta^2} e_t + \theta e_{t-1} \end{aligned}$$

The time series for e_t is generated based on the following specifications that include homoskedastic, conditionally heteroskedastic and unconditionally heteroskedastic cases with $\varepsilon_t \overset{i.i.d.}{\sim} N(0,1)$ throughout:

- DGP-0 [constant volatility]: $e_t = \sigma_t \varepsilon_t,\ \sigma_t = 1$.
- DGP-1 [single volatility shift]: $e_t = \sigma_t \varepsilon_t,\ \sigma_t = \mathbf{1}(t \leq \tau_1 T) + \sigma \mathbf{1}(t > \tau_1 T), \tau_1 = 0.5,\ \sigma = 1/3$.
- DGP-2 [double volatility shift]: $e_t = \sigma_t \varepsilon_t,\ \sigma_t = \mathbf{1}(t \leq \tau_1 T) + \sigma \mathbf{1}(\tau_1 T < t \leq \tau_2 T) + \mathbf{1}(t > \tau_2 T),\ (\tau_1, \tau_2) = (0.3, 0.7),\ \sigma = 3$.
- DGP-3 [trending volatility]: $e_t = \sigma_t \varepsilon_t,\ \sigma_t = 1 + 5t/T$.
- DGP-4 [GARCH]: $e_t = \sqrt{h_t}\varepsilon_t,\ h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 e_{t-1}^2,\ (\beta_0, \beta_1, \beta_2) = (0.01, 0.9, 0.09)$.
- DGP-5,6 [Stochastic Volatility]: $e_t = v_t \exp(\frac{1}{2}(\omega_0 + \frac{\omega_1}{T^{1/2}} h_t)),\ h_t = (1 - c_1/T) h_{t-1} + \varepsilon_t,\ h_0 = 0,\ (v_t, \varepsilon_t) \overset{i.i.d.}{\sim} N(0, \Sigma_{v\epsilon}),\ \Sigma_{v\epsilon} = \begin{bmatrix} 1 & \bar{\omega} \\ \bar{\omega} & 1 \end{bmatrix}$. We set $\omega_0 = 0,\ \omega_1 = 5$, and $c_1 = 0$. DGP-5 and DGP-6 correspond to the cases with $\bar{\omega} = 0$ and $\bar{\omega} = -0.5$, respectively.

DGP-0 is the base case with constant volatility. DGP-1 and DGP-2 exhibit discrete jumps in volatility while DGP-3 is a case of trending volatility.[5] DGP-4 follows a GARCH specification adopted from Goncalves and Kilian (2004) which is in turn based on Engle and Ng (1993).[6] The stochastic volatility specification for DGP-5,6 is borrowed from Cavaliere

[5] The results for DGP-1 with $\sigma = 3$ (not reported) are qualitatively similar to those with $\sigma = 1/3$ while the results for DGP-2 with $\sigma = 1/3$ (not reported) are qualitatively similar to those with $\sigma = 3$. The full set of results is available upon request.

[6] Engle and Ng (1993, p.1760) consider two different configurations of parameter values: (i) "medium persistence"- $(\beta_0, \beta_1, \beta_2) = (0.05, 0.9, 0.05)$ (ii) "high persistence"- $(\beta_0, \beta_1, \beta_2) = (0.01, 0.9, 0.09)$. We only present results for (ii) since the results for (i) have a similar overall pattern. The latter set of results is available upon request.

and Taylor (2009). DGP-5 represents a case with no leverage while DGP-6 allows for leverage via a non-zero correlation between the shocks v_t and ε_t.[7] Note that DGP-6 is ruled out by Assumption 2 so that the results for this case serve as a check on the robustness of the various methods to the violation of this assumption.

Three alternative values for the sample size are considered: $T \in \{50, 100, 200\}$. The nominal level of the confidence intervals is set at 95%. The results for $\alpha = 0.5$ are reported in the main text while those for $\alpha = 0.8$ are presented in Appendix C. Except for the PM interval, the estimated regression always includes a constant regardless of whether the true drift is zero or not. All experiments are based on 10,000 Monte Carlo replications and 399 bootstrap replications.

Seven alternative methods for the construction of confidence intervals for ρ_T are considered. These include (i) the HAC-based interval based on the statistic (9), denoted "t_{hac}"; (ii) the Phillips and Magdalinos (2007) interval (5), denoted "PM";[8] (iii): the Guo et al. (2019) interval based on the t-statistic (6), denoted "GSW"; (iv)-(vi): the dependent wild bootstrap with bandwidth l, denoted DWB$_l$, $l \in \{3, 5, 10\}$; (vii) the dependent wild bootstrap with bandwidth chosen according to the deterministic rule $l = \lfloor 4.5(T/100)^{1/4} \rfloor$, denoted DWB$_r$. This rule yields bandwidths of 3,4,5 for $T = 50,\ T = 100,\ T = 200$, respectively. Rho and Shao (2019) propose an alternative rule in the context of unit root testing: $l = \lfloor 6(T/100)^{1/4} \rfloor$. In our simulations, we found this rule to generate bandwidths that are too large to deliver confidence intervals with adequate coverage. The Quadratic Spectral kernel is used to construct the HAC long-run variance estimate and, following Andrews (1991), a data dependent bandwidth rule based on an AR(1) approximating model for each element of the vector $z_t\hat{u}_t$ is used (see equation (6.4) of Andrews, 1991). To improve finite sample performance, we employ prewhitening as suggested by Andrews and Monahan (1992) based on a VAR(1) model for $z_t\hat{u}_t$.[9] The Bartlett kernel is adopted as the kernel function for implementing the dependent wild bootstrap procedure.

Table 1 presents the empirical coverage rates of the different methods for the case without

[7] Cavaliere and Taylor (2009) also considered other parameter values, namely, $c_1 \in \{10, 20\}$ and $\omega_1 = 10$. For brevity, we do not present these results given that the results reported are fairly representative of these cases.

[8] The results for the PM interval are conditioned on those realizations for which $\widetilde{\rho}_T > 1$ and increasing the number of Monte Carlo replications till 10,000 estimates satsifying this condition were obtained. Without conditioning, the method often yields poor (liberal) coverage rates, especially when ρ_T is close to (but greater than) unity (e.g., when $\alpha = 0.8$).

[9] We found the prewhitened HAC estimator to deliver considerably more accurate coverage rates relative to its non-prewhitened counterpart. The asymptotic results derived in Sections 3 and 4, however, remain valid for the pre-whitened estimator as well.

drift ($\mu_T = 0$). Panels A,B,C report the results with serially uncorrelated errors, AR errors with $\phi = 0.5$, and MA errors with $\theta = 0.5$, respectively. Consider first the coverage rates based on t_{hac}. With serially uncorrelated errors, the coverage rates of t_{hac} are liberal (i.e., less than the nominal level) regardless of whether the errors are heteroskedastic with the degree of undercoverage being especially severe when the sample size is small. For instance, when $T = 50$, the maximum coverage across all DGPs is only 85% while coverage is below 80% for five out of the seven DGPs considered, including the constant volatility case. The overall pattern of results with AR errors is similar to that in the serially uncorrelated case although the liberal nature of the confidence intervals is somewhat mitigated in the former case relative to the latter when $T = 50$. When errors are of the MA type, the performance of t_{hac} is considerably improved relative to the other two error structures with coverage exceeding 90% for six of the seven DGPs as long as $T \geq 100$. The t_{hac} intervals in the MA case are notably conservative in the presence of a single volatility shift (DGP-1).

Turning to the PM interval, we find that with constant volatility, its coverage can be quite conservative (>98%) when $T \leq 100$ regardless of the serial correlation structure but moves closer to the nominal level when $T = 200$. This is not surprising given that the interval is (asymptotically) justified in this case. When the errors are heteroskedastic, the performance of PM depends crucially on the specific form of heteroskedasticity. For DGP-1 and DGP-2 which are characterized by discrete volatility shifts, the interval continues to be conservative with discernible improvement in coverage observed only for DGP-2 as the sample size increases. In contrast, in the trending volatility case (DGP-3), coverage is at most 82% across the three different error structures when $T = 200$. In fact, coverage declines by at least nine percentage points as the sample size increases from $T = 100$ to $T = 200$, suggesting the inadequacy of the asymptotic approximation on which the PM interval is based. A deterioration in performance as the sample size increases is also observed for DGP 4-6, though to varying degrees.

For the GSW interval, the coverage rates are in excess of 90% in the constant volatility case for $T \geq 100$ and gradually approach the nominal level as the sample size increases, consistent with the asymptotic validity of the interval in this case. This is, however, no longer true with time-varying volatility, as exemplified by the results for DGP-1 to DGP-6. When volatility is subject to discrete shifts, the interval becomes more conservative as the sample size increases from $T = 100$ to $T = 200$. For instance, when $T = 200$, the coverage rates for DGP-1/DGP-2 are at least 98%. A similar lack of convergence towards the nominal level is also observed for DGP 3-6 although in these cases the coverage rates remain notably

liberal (<90%) regardless of the sample size and the serial correlation structure (except when $T = 200$ and errors are of the MA type).

Consider now the coverage rates of the intervals based on the dependent wild bootstrap. Several features of these results are noteworthy. First, coverage performance varies with the bandwidth employed with a smaller bandwidth generally leading to intervals with more accurate coverage. The proposed bandwidth rule exhibits coverage similar to that with the smallest bandwidth. Second, consistent with Theorem 7, the coverage rates typically improve as the sample size increases for each of the DGPs considered. Interestingly, this feature is also observed in the stochastic volatility case with leverage (DGP-6) despite the fact that this case is not allowed for in the theory. Third, the coverage rates of the bootstrap-based intervals are considerably less sensitive to the nature of serial correlation and the particular form of heteroskedasticity relative to the t_{hac} and PM intervals. Fourth, the performance of the bootstrap-based intervals is particularly impressive when the sample size is small ($T = 50$), where the t_{hac} and PM intervals often suffer from substantial under/over-coverage.

Table 2 reports the coverage rates when the DGP includes a drift. The performance of t_{hac} in this case generally improves relative to the no drift case by ameliorating the extent of undercoverage especially when the sample size is small. The PM interval, on the other hand, is now seen to be severely conservative in most cases with coverage being as high as 100% for four of the seven volatility specifications considered, including the constant volatility case. Further, the coverage rates do not necessarily approach the nominal level as the sample size increases for any of the volatility structures. This feature can be explained by the fact that when $\nu \neq 0$ in Assumption 2, the Cauchy limit distribution underlying the PM interval is no longer valid.[10] The coverage rates of the GSW interval remain inadequate when volatility is time-varying although some improvements may be noted for DGP-3 and DGP-4 in the serially uncorrelated case. In contrast, the bootstrap-based intervals are much more stable with coverage rates bearing a very similar pattern to those in the no drift case. In summary, the coverage results in Tables 1 and 2 make a favorable case for employing the dependent wild bootstrap based on a small bandwidth or the recommended bandwidth rule compared to the asymptotic approaches.

Table 3 presents the average effective lengths of the confidence intervals, normalized with respect to the t_{hac} interval. Thus, a ratio smaller (larger) than one indicates an interval

[10] When $\nu = \infty$, $\widetilde{\rho}_T$ converges to ρ_T at rate $\mu_T k_T^{3/2} \rho_T^T$, which is faster than the rate $k_T \rho_T^T$ in Phillips and Magdalinos (2007a) and $\widetilde{\rho}_T$ is asymptotically normal (see Fei, 2018; Liu and Peng, 2019). When $\nu \in (0, \infty)$, the limit distribution is mixed normal (see Guo et al., 2019).

with average effective length shorter (longer) than the t_{hac} interval. The results reveal the following notable patterns. First, the PM interval can be discernibly longer than the other intervals in cases where the errors have constant conditional variance or involve discrete shifts in volatility. In contrast, it typically delivers the shortest average length in the trending and non-deterministic volatility cases (DGPs 3-6) when $T = 200$. Thus, as with coverage, the length of the PM interval can be quite sensitive to the underlying volatility specification. Second, the length of the GSW interval depends to a considerable extent on both the serial correlation and volatility structures driving the true DGP. For instance, this interval is the shortest on average relative to the other intervals for DGP 3-6 when $T \leq 100$ but always longer than the interval based on t_{hac} for DGP-1 and DGP-2. In the constant volatility case, the GSW interval is longer than the HAC-based interval with serially uncorrelated errors but shorter than the same in the serially correlated scenarios. Third, for the bootstrap-based intervals, average length is generally shorter, the larger the bandwidth employed. The average length based on the bandwidth rule DWB_r typically lies between the average lengths for the smallest and largest bandwidths considered. Fourth, the length improvements offered by the bandwidth rule over the asymptotic methods are primarily concentrated in situations where the errors are serially correlated and volatility is subject to discrete shifts. Fifth, the performance of the rule-based bandwidth is substantially more stable across the different volatility specifications relative to the asymptotic procedures, a feature also previously observed for the coverage rates. Table 4 reports the corresponding length results in the drift case. These results paint a qualitatively similar overall picture as the results in the no drift case.

In summary, the Monte Carlo results indicate that while employing a relatively smaller bandwidth leads to more accurate coverage, it also leads to longer average effective lengths. The recommended bandwidth rule offers a reasonable approach to addressing the coverage-length trade-off by delivering relatively short intervals while retaining adequate coverage properties. Additional Monte Carlo results presented in Appendix C for the case $\alpha = 0.8$ further confirm the effectiveness of the proposed procedure for conducting inference within the mildly explosive autoregressive framework.

6 Empirical Applications

This section illustrate the proposed methodology on two sets of time series. Section 6.1 revisits the empirical application in GSW where the degree of explosiveness of ten major stock market indices in the pre-2008 financial exuberance period is studied. In particular, our analysis accounts for the potential nonstationary volatility pattern in the stock indices

and highlights the difference between our results and those in GSW which are based on assuming stationary volatility. Section 6.2 investigates the extent of explosive behavior in three monthly U.S. home price indices during the 2002-2006 housing bubble.

6.1 Stock Market Indices

GSW employ a two-step testing strategy to identify the degree of explosiveness in ten major stock indices over the period leading up to the 2008 financial crisis. The first step entails a pretest for detecting whether the time series is explosive and the second step uses their proposed method to construct a confidence interval if the pretest signals the presence of explosive behavior. They find limited evidence of explosive behavior with most series either only mildly explosive, or not explosive at all. To facilitate comparison with their results, our analysis employs the same data set as GSW.[11] The data are weekly and the sample size is $T = 100$ for all series. Specifically, the data are collected in a way that end at the (pre-selected) highest point in the pre-2008 financial crisis period and then span 100 periods before that highest point. All of the series peaked at some point during 2007-2008, thereby making the whole sample approximately span from 2005 to 2008. A plot of the indices is displayed in Figure 1.

We start with an assessment of the time series behavior of volatility in these series to justify the plausibility of allowing for unconditional heteroskedasticity. Following Cavaliere and Taylor (2007a), Figure 2 plots the estimated variance profile, defined by $\widehat{\mathrm{VP}}(s) = \big(\sum_{t=1}^{\lfloor sT \rfloor} \hat{e}_t^2 + (sT - \lfloor sT \rfloor)\hat{e}_{\lfloor sT \rfloor+1}^2\big)/\sum_{t=1}^{T} \hat{e}_t^2$, $0 \leq s \leq 1$, as well as the volatility estimates $\hat{\sigma}_t^2$, obtained by fitting a nonparametric regression to the squared residuals $\hat{e}_t^2$ as suggested by Xu and Phillips (2008)[12]. Specifically, to adjust for potential serial correlation in the errors, the residuals $\hat{e}_t$ are obtained by fitting an autoregressive model to the series with lags determined by BIC with a maximum of six lags. For the nonparametric estimates, a Gaussian kernel is used with the bandwidth chosen by cross validation, searching over bandwidths $h_i = c_i T^{-0.4}, i = 1, \ldots, 4$ with $\{c_1, \ldots, c_4\} = \{0.25, 0.4, 0.6, 0.75\}$. As observed from Figure 2, the estimated variance profile of several series, especially USA, Brazil, China and Hong Kong, deviates substantially from the 45° line which represents the constant volatility

[11] The data come from Wind Economic Database and consists of ten countries/districts, namely, USA, Brazil, China, Hong Kong, Australia, France, Germany, Italy, Egypt and Nigeria, which are representatives of the world stock markets in different continents: America, Asia-Pacific, Europe and Africa. See Guo et al. (2019) for further details.

[12] To make the estimated volatility curves comparable across the different time series, we plot the estimated volatility ratio over $t = 1, \ldots, T$: $\hat{\sigma}_t^2/\bar{\hat{\sigma}}^2$, where $\bar{\hat{\sigma}}^2 = \sum_{t=1}^{T} \hat{\sigma}_t^2/T$.

scenario. The corresponding nonparametric estimates of the volatility clearly depict the underlying nonstationary evolution of the sample volatility paths, indicating smooth trending changes for USA, Brazil, China and Hong Kong, and possibly single/multiple shifts for the remaining countries.

In addition to visualizing the sample volatility paths, we also conduct formal diagnostic tests for the stationarity of unconditional volatility proposed by Cavaliere and Taylor (2007b). They present four test statistics, $\mathcal{H}_{KS}$, $\mathcal{H}_{R}$, $\mathcal{H}_{CVM}$, $\mathcal{H}_{AD}$, and derive their asymptotic distributions under the stationarity null from which the relevant critical values of the tests are obtained. In implementing these tests, the squared residuals $\hat{e}_t^2$ are used in constructing the stationary volatility test statistics. A long-run variance estimator based on the Bartlett kernel with lag truncation parameter 4 is employed.[13] Table 5 presents the testing results along with the 10%, 5% and 1% critical values. It is clear that the first four series, namely, USA, Brazil, China and Hong Kong, show evidence of significant nonstationary volatility from a majority of the tests. As noted in the simulation evidence in Cavaliere and Taylor (2007b), when volatility exhibits trending behavior or a single abrupt break, $\mathcal{H}_{AD}$ and $\mathcal{H}_{CVM}$ usually have the highest finite sample power out of the four tests while the $\mathcal{H}_{R}$ test is the least powerful. In contrast, the $\mathcal{H}_{R}$ test is the most powerful in the presence of multiple discrete volatility breaks. Considering these facts together with the visual evidence presented in Figure 2, we believe a smooth trending variation of the volatility is more likely to prevail in these four series, as opposed to single/multiple discrete volatility break(s).

We now turn to the two-step testing strategy adopted by GSW. Their first step involves a pretest for explosiveness using the right-tailed augmented Dickey-Fuller (RADF) and the supremum augmented Dickey-Fuller (SADF) tests proposed by Phillips et al. (2011, PWY henceforth) and Phillips et al. (2015, PSY henceforth), both of which assume stationary volatility. In contrast, drawing upon recent developments in the literature, we employ two set of tests proposed in Harvey et al. (2018, 2020) which allow for nonstationary volatility. In the first set of tests, Harvey et al. (2018) modify the RADF statistic of PWY using a weighted least squares-based variant (supBZ) which is borrowed from Boswijk and Zu (2019). To further increase power, they propose a union of rejections test ($\mathcal{U}$) that combines the original PWY test statistic (supDF) and supBZ. In the second set of tests, Harvey et al. (2020) suggest a sign-based version (sPSY) of the PSY test for multiple bubbles and for the same reason also advocate a union of rejections test (uPSY) which consists of sPSY

[13] To conserve space, we omit the details pertaining to the construction of the test statistics and refer the interested reader to Cavaliere and Taylor (2007b).

and the original PSY test. Following these studies, we present the bootstrap p-values of these six tests in Table 6. It is evident that both the union tests reject the unit root null at the 10% significance level for all series except USA and Italy for the $\mathcal{U}$ test. As noted by Harvey et al. (2020), the first set of tests - $\mathcal{U}$, supBZ, and supDF tests, which build on the PWY testing approach, are usually less powerful than the second set of tests based on the PSY testing principle. Interestingly, we also observe a similar phenomenon here: for each case except Australia, uPSY test has a lower p-value than $\mathcal{U}$. Based on the overall pattern found in Table 6 and the superior power performance of uPSY test revealed in Harvey et al. (2020), all of the ten series are deemed to be explosive. However, to construct a meaningful confidence interval in the second step, we conclude a series to be explosive if and only if both of the following two conditions are satisfied - the pretest must reject the unit root null and the point estimate of the degree of explosiveness must exceed unity, i.e., $\hat{\rho}_T > 1$. Taking into account these conditions, we exclude Australia, France and Italy from the second step analysis since their estimates $\hat{\rho}_T \leq 1$ [14].

The second step entails constructing HAC-based confidence intervals for the parameter ρ_T that governs the degree of explosiveness. To this end, we follow the approach in GSW of testing over a certain grid of values, $H_0 : \rho \in \{1.001, 1.002, \ldots, 1.500\}$. In practice, this corresponds to constructing a $100(1-\delta)\%$ confidence interval $[\hat{\rho}_L, \hat{\rho}_U]$ such that

$$\hat{\rho}_L = \max\{1.001, \hat{\rho}_T - D_{1-\delta/2} \times \hat{\sigma}(\hat{\rho}_T)\}, \quad \hat{\rho}_U = \hat{\rho}_T - D_{\delta/2} \times \hat{\sigma}(\hat{\rho}_T) \tag{25}$$

where $D_{\delta/2}$, $D_{1-\delta/2}$ represents the $\delta/2$ and $1-\delta/2$ percentiles of the approximating distribution D, and $\hat{\sigma}(\hat{\rho}_T)$ is an estimate of the standard deviation of $\hat{\rho}_T$. In our case, D is either the standard normal distribution or the DWB distribution and $\hat{\sigma}(\hat{\rho}_T)$ is the HAC estimate as defined in (9). Table 7 presents the confidence intervals constructed by the HAC-based approach as well as the other methods compared in the simulations. Overall, our proposed DWB approach is supportive of the hypothesis in GSW that most series are mildly explosive ($\rho_T \in [1.004, 1.04]$). However, the DWB-based confidence intervals are in general a bit wider than those of GSW with a larger upper bound for the former, which is consistent with the preceding simulation evidence that GSW usually under-covers in most of the nonstationary volatility cases considered. Hence, the GSW interval tends to understate both the sampling

[14] It is worth noting that in GSW, the time series for Australia, France and Italy are also categorized as non-explosive but for a different reason - due to failing to reject their bubble detection tests that do not allow for nonstationary variance. In contrast, after adjusting for potential nonstationarity in the variance, the three series show significant evidence of explosive behavior, although the p-values for Australia (0.086) and Italy (0.112) are the largest in the uPSY and $\mathcal{U}$ tests, respectively.

uncertainty associated with the point estimate of ρ_T as well as the degree of explosiveness driving the time series. Moreover, for the four series which showed considerable evidence of time-varying volatility in the foregoing analysis, the difference between the GSW and DWB_r methods is more prominent than for countries such as Germany and Nigeria which do not show significant time variation in volatility. This pattern suggest that when volatility is not time-varying, our DWB-based t_{hac} method suffers little efficiency loss, further highlighting the advantages of using our proposed procedure relative to those that do not control for unconditional heteroskedasticity. Finally, the PM intervals are in general too wide and thus not particularly informative, again consistent with the simulation evidence that PM intervals tend to over-cover under nonstationary volatility.

In summary, we find that four out of the seven explosive stock market series analyzed in GSW show strong evidence of time varying volatility. For these explosive series, our DWB-based method produces wider confidence intervals than GSW while the two methods provide very similar intervals for series that do not exhibit time variation in volatility as indicated by the tests for stationary volatility. These patterns are consistent with the simulation results in Section 5 and confirm the effectiveness of our proposed method in constructing confidence intervals for the degree of explosiveness in time series analysis.

6.2 U.S. Housing Price Indices

The U.S. experienced a sharp increase in home prices during the years 1997 and 2006. While the increase was larger in some areas and smaller in others, real home prices went up by about 85% over this period for the country as a whole (Shiller, 2015). Figure 3 plots three housing price indices - the national home price index, the 20-city composite index, the 10-city composite index, all of which nearly doubled from 2002 to 2006.[15] As Shiller (2015) notes, there was a "rocket taking off" that eventually crashed in 2006 and caused the 2008 financial crisis, the most severe of its kind since the Great Depression of the 1930s. A variety of potential explanations has been advanced for this rapid escalation in home prices including lax lending standards, the Federal Reserve's low interest rate policy, promotion in the media,

[15] The three indices are the main indices from the well known Case-Shiller Index, which was developed in the 1980s by three economists - Allan Weiss, Karl Case and Robert Shiller. Specifically, the national home price index records the value of residential housing by tracking the purchase and resale price of single-family homes, which covers nine major census divisions in the U.S. The 10-city composite index covers Boston, Chicago, Denver, Las Vegas, Los Angeles, Miami, New York, San Diego, San Francisco, and Washington, D.C.. The 20-city composite index, further includes Atlanta, Charlotte, Cleveland, Dallas, Detroit, Minneapolis, Phoenix, Portland, Seattle, and Tampa. The data can be downloaded from the economic research data website of Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/release/tables?rid=199&eid=243552.

and excessively optimistic investor beliefs from a behavioral perspective (see Glaeser, 2013; Shiller, 2015; Mian and Sufi, 2015; Griffin, Kruger and Maturana, 2021; and the references therein). In our study, while not investigating the causes of the house price boom during 2002-2006, we conduct an econometric analysis aimed at measuring the extent of explosive behavior in U.S. housing prices to help practitioners and policymakers gauge the intensity of these dramatic price accelerations. We thus provide a simple yet robust and effective tool to identify the degree of house price exuberance in the presence of potential serial correlation and heteroskedasticity, the latter emphasized by Case and Shiller (2003).

Similar to our foregoing analysis of the stock indices, we collect $T = 50$ observations of the three monthly Case-Shiller indices displayed in Figure 3 by first pinning down the peak point and then gathering data backward until the T-th observation. The peak for the national index is at March 2006, while the peaks for the 10/20-city composite indices are at April 2006, which makes the starting month to be February or March 2002.[16] The middle and lower panels in Figure 3 present the variance profiles and estimated volatilities for the three series after fitting an autoregressive model using the same specifications as the stock indices in the previous section. These plots reveal considerable instability in the sample volatility paths for all three indices with a trending volatility specification appearing to provide a suitable characterization of the nature of the volatility process. Table 8 presents the results from the diagnostic tests for stationary volatility (Panel A) and the bubble detection tests (Panel B) as described in Section 6.1. The $\mathcal{H}_{AD}$ and $\mathcal{H}_{CVM}$ tests turn out to be significant at least at the 10% level for all series, thereby formally corroborating the nonstationarity of the volatility paths and its type observed in Figure 3. All of the bubble tests are significant at the 1% level except PSY (although the p-values for PSY are all below 2%) for each of the three series. Therefore, we move all three series to the second stage of constructing confidence intervals for the degree of explosiveness. Table 9 present the results. As with our analysis of the stock indices, GSW provides a tighter interval and PM provides a much wider interval than our DWB-based method, both of which are consistent with our earlier discussions on the impact of ignoring possible nonstationarities in the second moments of the time series.

In conclusion, our analysis shows that the three housing price indices, all of which appear to possess trending (increasing) volatility, exhibit mildly explosive behavior as opposed to a severe explosion during the 2002-2006 housing market boom. Existing methods that do not allow for nonstationary volatility tend to understate/overstate the sampling variability

[16] See Phillips and Yu (2011) and Fabozzi and Xiao (2018) for empirical evidence on the bubble's timeline.

around the point estimates. More recently, U.S. housing prices have been increasing at a record pace despite high unemployment during the COVID-19 pandemic, which is believed to be induced by the Federal Reserve's unlimited quantitative easing approach in response to the pandemic. In principle, our method can also be applied to analyze the intensity of this recent surge in home prices. Since the pandemic is still ongoing, we leave such an investigation for future research.

7 Conclusion

The recent upsurge of interest in the mildly explosive autoregressive framework has been spurred by its ability to provide a simple yet effective tool for modeling the presence of asset market bubbles. The development of this framework has been followed by a plethora of theoretical and empirical studies that have sought to generalize the original framework or apply it to study the time series evolution of several price indices that may potentially be subject to explosive behavior. This paper considers the problem of constructing asymptotically justified confidence intervals for the autoregressive parameter that represents the degree of explosiveness. Existing approaches typically employed in empirical practice are valid only under the assumption of conditional homoskedasticity/heteroskedasticity, notwithstanding extensive empirical evidence against the same for a wide range of important economic and financial time series. Our framework allows the noise component to be unconditionally heteroskedastic and sufficiently general to subsume a variety of volatility specifications common in the literature. We propose a dependent wild bootstrap-t procedure for inference that is shown to provide an improved approximation to the finite sample distribution of the t-statistic relative to asymptotic methods. Given that the t-statistic is asymptotically pivotal, it is possible that the bootstrap offers asymptotic refinements (Hall, 1992). A theoretical investigation of this possibility is outside the scope of the present paper but a potentially fruitful avenue for future research.

References

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

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

Arvanitis, S., Magdalinos, T. (2018), "Mildly explosive autoregression under stationary conditional heteroskedasticity," *Journal of Time Series Analysis* 39, 892-908.

Astill, S., Harvey, D. I., Leybourne, S. J., Sollis, R., Taylor, A.M.R. (2018), "Real-time monitoring for explosive financial bubbles," *Journal of Time Series Analysis* 39, 863-891.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Taylor, R. L. (1989), "Bootstrapping explosive autoregressive processes," *Annals of Statistics* 17, 1479-1486.

Beare, B. K. (2018), "Unit root testing with unstable volatility," *Journal of Time Series Analysis* 39, 816-835.

Boswijk, H. P., Zu, Y. (2018), "Adaptive wild bootstrap tests for a unit root with nonstationary volatility," *Econometrics Journal* 21, 87-113.

Boswijk, H. P., Zu, Y. (2021), "Adaptive testing for cointegration with nonstationary volatility," *Journal of Business and Economic Statistics*, forthcoming.

Bühlmann, P. (1997), "Sieve bootstrap for time series," *Bernoulli* 3, 123-148.

Carlstein, E. (1986), "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *Annals of Statistics* 14, 1171-1179.

Case, K. E., Shiller, R. J. (2003), "Is there a bubble in the housing market?," *Brookings Papers on Economic Activity* 2003, 299-362.

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

Cavaliere, G., Nielsen, H. B., Rahbek, A. (2020), "Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling," *Journal of Business and Economic Statistics* 38, 55-67.

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

Cavaliere, G., Taylor, A. M. R. (2007b), "Time-transformed unit root tests for models with non-stationary volatility," *Journal of Time Series Analysis* 29, 300-330.

Cavaliere, G., Taylor, A. M. R. (2008), "Bootstrap unit root tests for time series with nonstationary volatility," *Econometric Theory* 24, 43-71.

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

Chan, N. H., Li, D., Peng, L. (2012), "Toward a unified interval estimation of autoregressions," *Econometric Theory* 28(3), 705-717.

Engle, R. F., Ng, V. K. (1993), "Measuring and testing the impact of news on volatility," *Journal of Finance* 48, 1749-1778.

Fabozzi, F. J., Xiao, K. (2019), "The timeline estimation of bubbles: The case of real estate," *Real Estate Economics* 47, 564-594.

Fei, Y. (2018). Limit theory for mildly integrated process with intercept," *Economics Letters* 163, 98-101.

Georgiev, I. (2008), "Asymptotics for cointegrated processes with infrequent stochastic level shifts and outliers," *Econometric Theory* 24, 587-615.

Glaeser, E. L. (2013), "A nation of gamblers: Real estate speculation and American history," *American Economic Review* 103, 1-42.

Gonçalves, S., Kilian, L. (2004), "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form," *Journal of Econometrics* 123, 89-120.

Griffin, J. M., Kruger, S., Maturana, G. (2021), "What drove the 2003-2006 house price boom and subsequent collapse? disentangling competing explanations," *Journal of Financial Economics* 141, 1007-1035.

Guo, G., Sun, Y., Wang, S. (2019), "Testing for moderate explosiveness," *Econometrics Journal* 22, 73-95.

Hall, P. (1992), "The bootstrap and Edgeworth expansion," Springer Series in Statistics.

Hansen, B.E. (1995), "Regression with nonstationary volatility," *Econometrica* 63, 1113-1132.

Harvey, D. I., Leybourne, S. J., Sollis, R., Taylor, A. M. R. (2016), "Tests for explosive financial bubbles in the presence of non-stationary volatility," *Journal of Empirical Finance* 38, 548-574.

Harvey, D. I., Leybourne, S. J., Zu, Y. (2018), "Testing explosive bubbles with time-varying volatility," *Econometric Reviews* 38, 1131-1151.

Harvey, D. I., Leybourne, S. J., Zu, Y. (2020), "Sign-based unit root tests for explosive financial bubbles in the presence of deterministically time-varying volatility," *Econometric Theory* 36, 122-169.

Jansson, M. (2002), "Consistent covariance matrix estimation for linear processes," *Econometric Theory* 18, 1449-1459.

Kunsch, H. R. (1989), "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics* 17, 1217-1241.

Kurozumi, E., Skrobotov, A., Tsarev, A. (2020), "Time-Transformed Test for the Explosive Bubbles under Non-stationary Volatility," Working Paper.

Lee, J. H. (2018), "Limit theory for explosive autoregression under conditional heteroskedasticity," *Journal of Statistical Planning and Inference* 196, 30-55.

Liu, R. Y. (1988), "Bootstrap procedures under some non-iid models," *Annals of Statistics* 16, 1696-1708.

Liu, X., Peng, L. (2019), "Asymptotic theory and unified confidence region for an autoregressive model," *Journal of Time Series Analysis* 40, 43-65.

Magdalinos, T. (2012), "Mildly explosive autoregression under weak and strong dependence," *Journal of Econometrics* 169, 179-187.

Mayer, C. (2011), "Housing bubbles: A survey," *Annual Review of Economics* 3, 559-577.

Mian, A., Sufi, A. (2015), "House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again." University of Chicago Press.

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

Perron, P., Qu, Z. (2010), "Long-memory and level shifts in the volatility of stock market return indices," *Journal of Business and Economic Statistics* 28, 275-290.

Phillips, P. C. B. (2005), "HAC estimation by automated regression," *Econometric Theory* 21, 116-142.

Phillips, P. C. B. (1987), "Towards a unified asymptotic theory for autoregressio," *Biometrika* 74, 535-547.

Phillips, P. C. B., Magdalinos, T. (2005), "Limit theory for moderate deviations from a unit root under weak dependence," Cowles Foundation for Research in Economics, Yale University.

Phillips, P. C. B., Magdalinos, T. (2007a), "Limit theory for moderate deviations from a unit root," *Journal of Econometrics* 136, 115-130.

Phillips, P. C. B., Magdalinos, T. (2007b), "Limit theory for moderate deviations from a unit root under weak dependence," in G. D. A. Phillips and E. Tzavalis, eds., The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis. Cambridge University Press, 123-162.

Phillips, P. C. B., Shi, S. (2020), "Real time monitoring of asset markets: Bubbles and crises," *Handbook of Statistics* 42, 61-80.

Phillips, P. C. B., Shi, S., Yu, J. (2015), "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500," *International Economic Review* 56, 1043-1078.

Phillips, P. C. B., Solo, V. (1992), "Asymptotics for linear processes," *Annals of Statistics* 20, 971-1001.

Phillips, P. C. B., Wu, Y., and Yu, J. (2011), "Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?," *International Economic Review* 52, 201-226.

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

Phillips, P. C. B., Yu, J. (2011), "Dating the timeline of financial bubbles during the subprime crisis," *Quantitative Economics* 2, 455-491.

Rho, Y., Shao, X. (2019), "Bootstrap-assisted unit root testing with piecewise locally stationary errors," *Econometric Theory* 35, 142-166.

Sensier, M., van Dijk, D. (2004), "Testing for volatility changes in US macroeconomic time series," *Review of Economics and Statistics* 86, 833-839.

Shao, X. (2010), "The dependent wild bootstrap," *Journal of the American Statistical Association* 105, 218-235.

Shiller, R. J. (2015), "Irrational Exuberance." Princeton University Press.

Skrobotov, A. (2021), "Testing for explosive bubbles: a review," Working Paper.

Smeekes, S., Urbain, J. P. (2014), "A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing," Working Paper, Maastricht University.

Wang, X., Yu, J. (2015), "Limit theory for an explosive autoregressive process," *Economics Letters* 126, 176-180.

Wu, C. F. J. (1986), "Jackknife, bootstrap and other resampling methods in regression analysis," *Annals of Statistics* 14, 1261-1295.

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

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

Table 1: Empirical coverage rate of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.5$, $\mu_T = 0$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
t_{hac}	76.4	87.4	91.4	82.5	90.3	93.1	74.3	85.4	92.2	72.3	83.0	89.4	72.9	84.5	90.9	79.5	88.0	92.1	85.4	90.6	93.7
PM	99.2	98.8	95.8	99.8	100	98.5	99.6	99.8	95.1	93.2	92.9	81.7	95.7	96.0	88.8	94.9	94.9	84.9	91.6	91.8	86.8
GSW	84.2	91.8	94.4	90.7	96.5	99.1	88.2	94.6	98.6	76.0	79.5	82.6	77.1	85.3	89.8	83.5	86.4	87.7	87.1	88.5	88.6
DWB_3	86.9	91.3	92.2	91.1	92.6	93.5	88.8	90.9	92.5	84.5	88.7	91.6	84.9	89.1	92.5	91.7	92.3	93.4	92.7	93.1	94.3
DWB_5	85.1	89.3	90.8	89.7	91.9	92.8	88.4	90.4	92.0	82.9	87.2	90.1	83.4	87.6	91.5	90.5	91.5	92.9	91.8	92.2	93.7
DWB_{10}	81.1	86.2	88.0	86.8	89.9	91.3	86.8	89.5	90.7	80.4	84.2	87.1	80.9	84.8	88.9	88.4	90.0	91.7	89.9	90.7	92.4
DWB_r	88.3	90.9	91.9	91.9	92.7	93.6	89.7	91.0	92.8	85.8	88.6	91.8	86.2	89.1	92.8	92.1	92.3	93.4	93.0	92.7	94.1
Panel B: AR case, $\phi = 0.5$																					
t_{hac}	82.6	88.4	90.1	85.2	91.9	93.6	86.0	90.2	93.0	79.2	83.6	87.6	80.3	85.4	89.4	84.2	89.5	92.2	85.3	90.0	93.0
PM	98.3	98.3	95.3	99.8	99.8	98.5	99.6	99.6	95.4	91.1	90.9	82.0	94.2	94.6	88.8	92.3	92.8	85.5	89.2	89.7	86.7
GSW	87.4	91.2	92.9	90.0	95.6	98.4	96.8	97.9	98.9	86.6	82.9	81.8	84.0	85.7	88.9	86.2	87.5	86.9	84.9	86.0	87.4
DWB_3	86.1	88.2	89.5	89.7	91.2	92.2	89.9	89.9	91.5	84.5	86.4	89.1	84.9	86.4	90.1	90.6	90.7	92.2	90.4	90.9	92.7
DWB_5	84.0	86.9	88.7	89.1	91.0	92.1	88.9	89.6	91.3	81.3	84.7	87.6	82.6	84.9	89.3	89.1	90.0	92.0	89.2	89.9	92.1
DWB_{10}	79.5	83.6	86.5	86.0	89.0	91.1	87.2	88.6	90.2	77.1	81.4	85.1	79.1	81.8	87.1	86.3	88.0	90.9	86.6	88.1	91.0
DWB_r	87.1	88.1	89.6	89.7	91.4	92.1	90.6	90.1	91.5	86.1	86.5	88.8	86.5	86.5	90.3	91.1	90.5	92.1	90.9	90.9	92.7
Panel C: MA case, $\theta = 0.5$																					
t_{hac}	88.1	93.2	95.0	91.7	96.1	97.1	90.1	94.2	96.2	84.0	89.5	93.0	85.6	90.7	93.9	89.6	94.2	96.0	90.1	94.3	96.2
PM	98.8	98.6	95.4	99.9	99.9	98.5	99.5	99.5	95.0	91.7	91.9	80.8	94.8	95.2	88.5	93.4	93.8	85.0	90.5	89.9	86.8
GSW	89.8	92.8	94.3	93.7	97.1	99.2	95.4	97.2	99.1	85.7	81.7	81.7	86.5	87.2	90.5	88.1	88.2	88.1	86.8	87.5	87.4
DWB_3	90.4	92.9	93.5	93.1	94.8	95.5	92.4	94.1	94.9	88.4	91.0	92.6	88.7	91.3	93.8	93.6	94.0	95.3	93.5	94.5	95.7
DWB_5	87.4	90.5	91.3	91.6	93.2	94.1	91.3	92.6	93.6	85.5	88.5	90.3	86.0	88.5	92.1	91.8	92.7	93.8	92.3	92.9	94.2
DWB_{10}	82.9	86.4	87.9	88.2	90.6	92.1	88.8	90.8	91.4	81.5	84.5	86.7	82.4	84.9	89.1	89.1	90.6	92.1	89.6	91.0	92.5
DWB_r	92.4	92.8	93.4	94.2	94.8	95.3	93.9	94.0	94.9	90.6	91.0	92.8	90.8	91.3	93.9	94.6	94.1	95.3	94.6	94.3	95.8

Table 2: Empirical coverage rate of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.5$, $\mu_T = T^{-\alpha/4}$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
t_{hac}	88.9	91.3	92.0	91.9	94.0	94.1	93.2	93.9	93.8	81.9	88.6	91.3	87.2	90.0	92.0	87.4	91.0	93.4	84.2	90.6	93.6
PM	99.8	99.9	100	99.9	100	100	99.9	99.9	100	97.2	98.6	99.3	100	100	100	96.7	97.1	94.5	94.4	94.7	89.3
GSW	94.3	95.0	94.5	98.0	98.8	99.4	98.9	99.6	99.7	83.8	85.3	82.6	90.8	91.0	91.1	89.0	88.7	89.1	86.9	87.6	88.1
DWB_3	90.1	91.7	92.7	91.9	93.5	94.3	87.4	88.3	92.8	86.8	89.7	91.6	89.3	91.1	93.1	91.6	91.8	92.8	91.8	92.0	93.9
DWB_5	88.1	90.3	91.6	90.5	92.6	93.8	86.8	87.3	92.2	84.8	87.4	90.3	87.3	89.3	92.0	90.7	90.9	92.1	91.2	91.6	93.3
DWB_{10}	84.0	87.1	88.9	87.5	90.4	92.6	85.7	85.3	90.6	80.9	84.0	86.9	83.6	86.0	89.6	88.6	89.5	91.1	89.7	90.3	92.6
DWB_r	91.0	91.7	92.6	92.3	93.3	94.2	87.8	88.2	92.8	88.2	89.7	91.6	90.5	91.2	93.4	91.7	91.7	92.8	92.0	92.0	93.9
Panel B: AR case, $\phi = 0.5$																					
t_{hac}	85.7	89.5	91.1	89.3	92.9	93.9	91.6	93.4	93.3	80.1	86.3	88.3	84.1	87.9	89.9	85.6	90.4	92.0	84.8	89.9	92.6
PM	98.8	99.0	99.0	99.1	99.1	98.9	99.4	99.4	98.7	93.3	94.6	93.7	99.8	99.9	100	94.3	95.0	91.9	91.6	92.3	87.2
GSW	90.3	92.3	93.1	93.4	96.3	98.6	98.0	99.2	99.3	86.3	83.5	82.4	86.6	87.6	89.1	87.1	86.8	87.3	85.2	86.6	86.3
DWB_3	86.5	88.7	90.3	88.7	91.0	92.5	87.5	85.8	89.6	84.2	86.5	89.0	86.0	87.7	90.7	90.3	89.9	91.2	89.8	90.7	92.5
DWB_5	84.9	87.7	89.8	87.7	90.7	92.6	86.9	84.8	89.3	81.9	84.2	87.8	84.0	86.5	90.2	89.1	89.3	91.0	88.5	89.8	92.3
DWB_{10}	81.1	84.9	87.9	84.9	89.2	91.8	85.7	83.5	88.3	77.5	80.9	84.9	80.1	83.4	87.9	86.7	87.9	89.8	86.2	88.1	91.3
DWB_r	87.4	88.3	90.7	89.0	91.1	92.7	87.9	85.8	89.7	86.1	86.3	88.9	86.9	87.8	91.1	90.6	90.1	91.2	90.0	90.7	92.6
Panel C: MA case, $\theta = 0.5$																					
t_{hac}	92.7	94.3	95.1	95.2	96.8	97.2	95.7	97.1	97.1	86.9	91.6	93.6	90.0	93.2	94.6	91.6	94.6	96.1	89.5	94.4	96.1
PM	99.4	99.5	99.7	99.4	99.5	99.6	99.7	99.7	99.6	94.6	96.5	96.7	100	100	100	96.0	96.1	93.0	92.8	93.5	88.1
GSW	94.1	95.0	94.6	97.0	98.5	99.2	98.7	99.3	99.5	86.2	84.1	82.8	90.4	90.3	91.4	89.3	88.6	88.2	87.2	87.8	87.2
DWB_3	91.2	93.1	94.4	92.9	95.1	96.0	90.5	89.5	93.7	88.8	91.2	92.8	90.5	92.6	94.6	93.3	93.5	94.7	93.2	93.9	95.7
DWB_5	88.7	90.9	92.3	91.4	93.7	94.9	89.1	87.6	92.3	85.8	88.1	90.3	87.8	90.2	92.6	91.9	92.1	93.1	91.9	92.6	94.3
DWB_{10}	84.5	87.2	89.2	88.4	91.2	93.2	87.3	85.4	90.1	81.5	84.2	86.7	83.8	86.3	89.6	89.6	90.1	91.4	89.6	90.7	92.8
DWB_r	93.0	93.1	94.4	94.1	95.1	96.1	91.4	89.8	93.7	91.0	91.2	92.8	92.5	92.6	94.5	94.4	93.8	94.7	94.1	94.0	95.7

Table 3: Effective interval length ratio (benchmark: t_{hac}) of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.5$, $\mu_T = 0$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
PM	2.62	2.06	1.47	3.23	3.01	2.49	2.52	2.11	1.38	1.42	1.04	0.49	1.68	1.31	0.71	1.27	0.88	0.41	1.06	1.13	0.58
GSW	1.07	1.02	1.09	1.12	1.22	1.39	1.26	1.40	1.54	0.98	0.81	0.70	0.92	0.90	0.90	0.83	0.71	0.58	0.52	0.71	0.56
DWB_3	1.28	1.18	1.11	1.16	1.01	0.94	1.13	1.09	1.17	1.33	1.24	1.19	1.30	1.22	1.16	1.57	1.27	1.20	2.54	1.37	1.22
DWB_5	1.25	1.15	1.09	1.16	1.00	0.93	1.13	1.08	1.17	1.27	1.20	1.15	1.26	1.18	1.13	1.54	1.20	1.14	2.48	1.28	1.15
DWB_{10}	1.21	1.12	1.03	1.20	1.00	0.93	1.19	1.12	1.17	1.21	1.14	1.08	1.20	1.12	1.07	1.50	1.09	1.06	2.41	1.16	1.06
DWB_r	1.31	1.18	1.11	1.18	1.01	0.95	1.15	1.09	1.16	1.35	1.24	1.19	1.33	1.21	1.15	1.60	1.27	1.20	2.33	1.36	1.22
Panel B: AR case, $\phi = 0.5$																					
PM	1.90	1.62	1.22	2.67	2.32	1.92	2.02	1.53	1.03	0.47	0.73	0.45	1.15	0.89	0.56	0.85	0.70	0.29	1.33	0.98	0.48
GSW	0.99	0.92	0.97	1.09	1.05	1.07	1.43	1.40	1.35	0.44	0.80	0.78	0.80	0.74	0.77	0.62	0.68	0.50	0.70	0.64	0.54
DWB_3	1.24	1.02	1.01	0.97	0.83	0.77	0.92	0.85	0.90	1.23	1.09	1.04	1.23	1.05	1.03	1.37	1.15	1.05	1.66	1.42	1.11
DWB_5	1.25	1.00	1.00	0.96	0.81	0.78	0.91	0.82	0.88	1.21	1.05	1.03	1.20	1.02	1.02	1.12	1.10	1.02	1.54	1.49	1.08
DWB_{10}	1.33	0.97	0.96	0.96	0.79	0.75	0.93	0.83	0.89	1.15	0.98	0.98	1.10	0.95	0.97	0.94	1.02	0.97	1.39	1.35	1.02
DWB_r	1.20	1.01	1.01	0.99	0.83	0.77	0.93	0.86	0.90	1.25	1.09	1.04	1.22	1.05	1.02	1.34	1.14	1.03	1.68	1.48	1.11
Panel C: MA case, $\theta = 0.5$																					
PM	1.76	1.49	1.09	2.51	2.20	1.62	1.89	1.35	0.92	0.64	0.67	0.39	0.88	0.82	0.55	0.43	0.59	0.31	1.17	0.76	0.37
GSW	0.92	0.87	0.84	1.07	1.05	1.02	1.29	1.12	1.21	0.58	0.71	0.65	0.62	0.67	0.69	0.32	0.56	0.49	0.61	0.51	0.41
DWB_3	1.10	0.99	0.99	0.96	0.81	0.82	0.91	0.84	0.88	1.23	1.08	1.03	1.18	1.03	1.01	1.54	1.13	1.02	1.58	1.27	1.11
DWB_5	1.04	0.92	0.94	0.94	0.77	0.79	0.89	0.79	0.83	1.17	0.99	0.97	1.10	0.96	0.95	1.47	1.06	0.96	1.46	1.22	1.07
DWB_{10}	0.95	0.85	0.85	0.92	0.74	0.74	0.88	0.77	0.80	1.07	0.78	0.88	1.00	0.88	0.87	1.35	0.96	0.87	1.32	1.08	0.99
DWB_r	1.15	0.99	0.99	1.00	0.81	0.82	0.94	0.84	0.87	1.26	1.02	1.03	1.21	1.03	1.00	1.59	1.13	1.01	1.66	1.26	1.11

Table 4: Effective interval length ratio (benchmark: t_{hac}) of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.5$, $\mu_T = T^{-\alpha/4}$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
PM	11.5	16.0	20.1	18.3	27.2	38.5	6.55	10.7	14.9	2.16	2.42	3.16	11.0	15.2	19.1	1.88	1.59	1.23	1.33	1.31	0.71
GSW	1.11	1.06	1.03	1.31	1.33	1.49	1.58	1.58	1.62	0.96	0.81	0.78	0.96	0.93	0.93	0.80	0.74	0.67	0.72	0.82	0.72
DWB_3	1.14	1.08	1.06	1.04	1.02	1.01	0.99	0.95	1.01	1.39	1.19	1.11	1.22	1.12	1.08	1.68	1.33	1.14	2.20	1.65	1.20
DWB_5	1.07	1.03	1.02	0.99	0.99	0.99	0.98	0.92	0.98	1.38	1.14	1.07	1.13	1.06	1.03	1.62	1.26	1.11	2.06	1.52	1.14
DWB_{10}	0.95	0.95	0.96	0.92	0.95	0.95	1.00	0.87	0.94	1.37	1.05	0.98	1.00	0.96	0.96	1.51	1.18	1.04	1.94	1.45	1.10
DWB_r	1.18	1.08	1.06	1.07	1.02	1.01	1.01	0.95	1.01	1.38	1.19	1.11	1.27	1.12	1.08	1.71	1.32	1.15	2.14	1.56	1.20
Panel B: AR case, $\phi = 0.5$																					
PM	4.48	5.62	6.95	8.09	10.1	11.7	2.84	3.80	4.92	0.72	0.99	0.99	3.33	7.16	10.3	1.08	1.06	0.69	1.47	0.93	0.43
GSW	0.99	0.93	1.03	1.19	1.13	1.16	1.45	1.56	1.45	0.53	0.75	0.74	0.51	0.76	0.87	0.62	0.68	0.61	0.91	0.70	0.57
DWB_3	1.16	1.06	1.02	1.07	0.99	0.96	0.95	0.90	0.92	1.25	1.10	1.09	1.30	1.10	1.05	1.52	1.47	1.06	1.59	1.29	1.11
DWB_5	1.12	1.05	1.02	1.06	0.98	0.98	0.95	0.88	0.93	1.22	1.08	1.06	1.24	1.08	1.04	1.40	1.48	1.02	1.55	1.23	1.05
DWB_{10}	1.04	1.00	0.98	1.04	0.98	0.96	0.99	0.86	0.92	1.12	1.03	1.01	1.33	1.00	0.99	1.29	1.34	0.94	1.39	1.15	1.00
DWB_r	1.17	1.07	1.02	1.10	0.99	0.96	0.95	0.90	0.92	1.26	1.11	1.08	1.32	1.10	1.05	1.54	1.49	1.05	1.65	1.28	1.10
Panel C: MA case, $\theta = 0.5$																					
PM	5.33	7.71	10.3	9.33	13.0	20.0	3.27	5.05	7.13	1.12	1.18	1.39	4.99	8.66	11.6	0.48	0.39	0.73	0.93	0.77	0.42
GSW	0.89	0.84	0.87	1.13	1.04	1.40	1.35	1.38	1.40	0.75	0.67	0.67	0.63	0.75	0.78	0.25	0.22	0.55	0.58	0.54	0.50
DWB_3	1.14	1.05	1.00	1.03	0.98	0.96	0.92	0.88	0.91	1.29	1.12	1.04	1.21	1.08	1.02	1.67	1.18	1.12	1.66	1.31	1.07
DWB_5	1.06	0.97	0.95	0.96	0.92	0.92	0.90	0.83	0.84	1.24	1.05	0.98	1.11	1.02	0.96	1.55	1.11	1.02	1.62	1.18	0.99
DWB_{10}	0.96	0.88	0.87	0.89	0.84	0.88	0.90	0.78	0.80	1.12	0.93	0.89	0.96	0.90	0.87	1.40	1.01	0.97	1.52	1.09	0.90
DWB_r	1.16	1.05	1.00	1.07	0.99	0.96	0.96	0.88	0.89	1.31	1.11	1.05	1.26	1.08	1.02	1.68	1.17	1.08	1.69	1.26	1.06

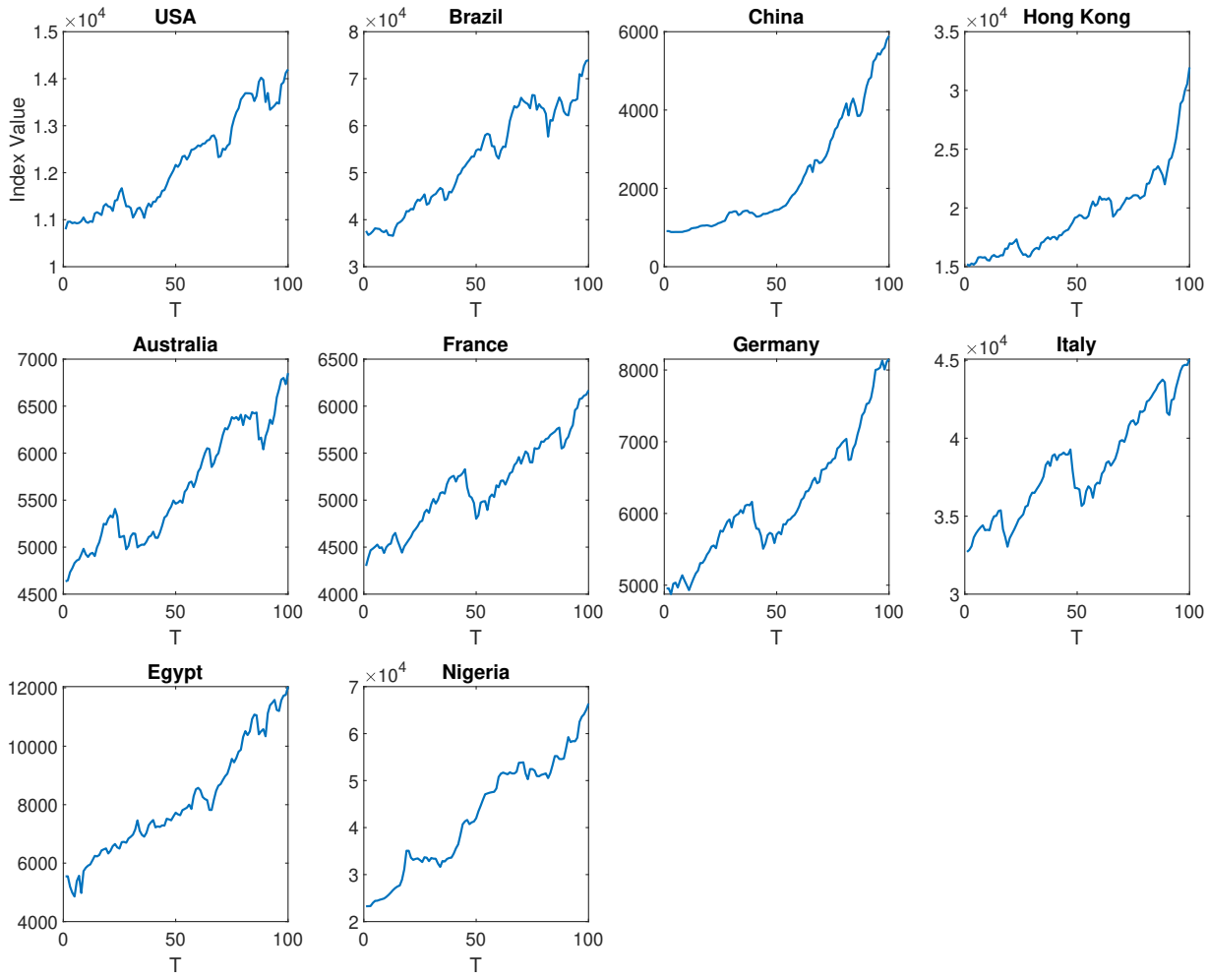

Figure 1: Plots of ten stock indices during 2005-2008.

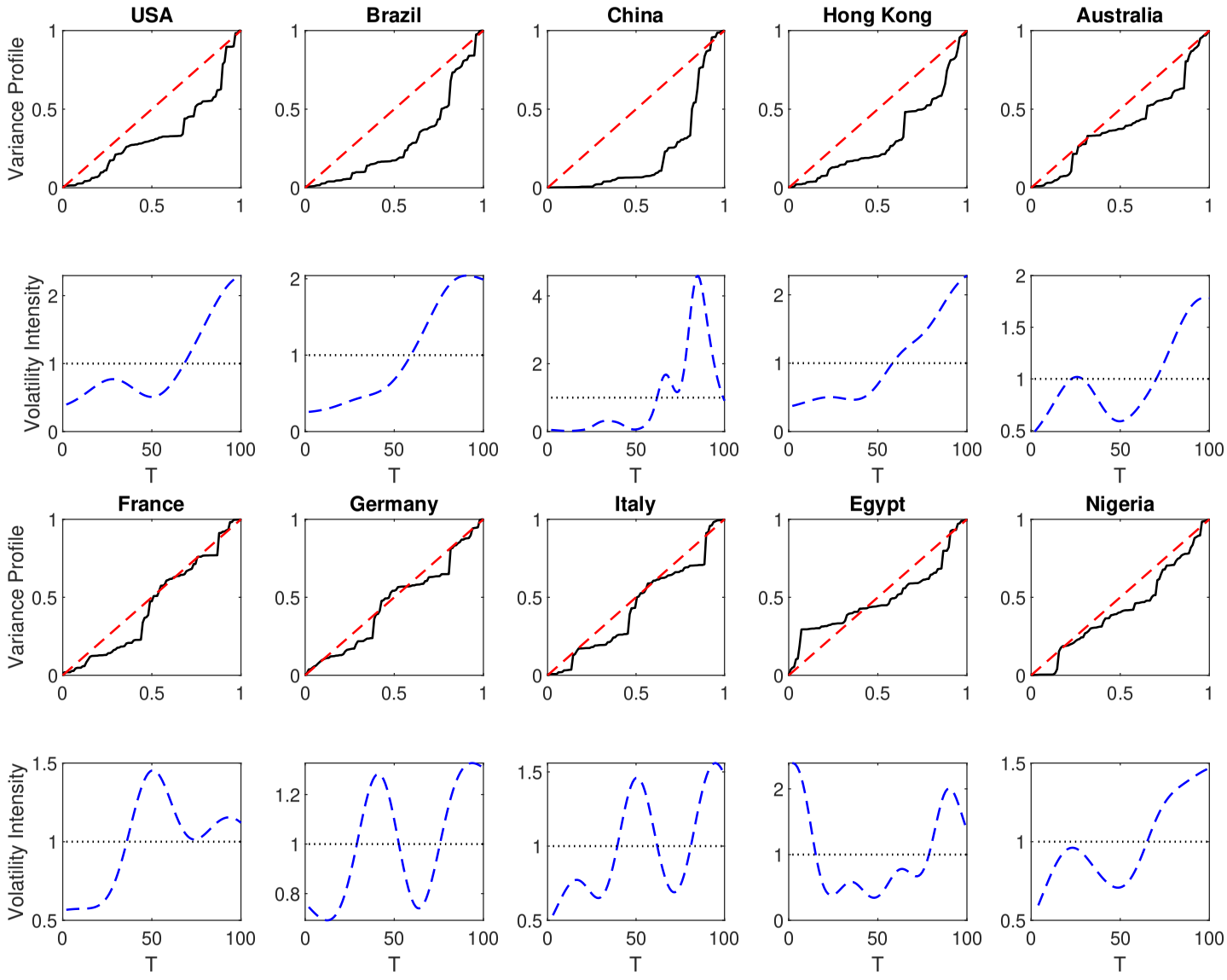

Figure 2: Variance profiles and nonparametric volatility estimates of ten stock indices.

Table 5: Nonstationary volatility tests (Cavaliere and Taylor, 2007b) and their critical values (C.V.) for ten stock indices.

	$\mathcal{H}_{KS}$	$\mathcal{H}_{R}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
USA	1.382	1.328*	0.541**	2.965**
Brazil	1.560	1.487**	0.870***	4.223***
China	1.717*	1.634***	0.993***	4.744***
Hong Kong	1.368	1.368**	0.728**	3.753**
Australia	1.024	0.993	0.280	1.582
France	1.264	1.055	0.174	0.890
Germany	0.968	0.747	0.098	0.523
Italy	0.879	0.716	0.108	0.660
Egypt	1.660*	0.914	0.190	1.387
Nigeria	1.070	0.941	0.205	1.190
C.V. (10%)	1.620	1.230	0.347	1.933
C.V. (5%)	1.750	1.360	0.461	2.492
C.V. (1%)	2.010	1.630	0.743	3.850

Note: *denotes 10%, **denotes 5%, and ***denotes 1% significance level for the above tests.

Table 6: Explosiveness estimates and p-values of bubble tests allowing for nonstationary volatility (Harvey et al., 2018, 2020) for ten stock indices.

	$\hat{\rho}_T$	$\mathcal{U}$	supBZ	supDF	uPSY	PSY	sPSY
USA	1.003	0.100	0.072	0.293	0.000	0.607	0.000
Brazil	1.001	0.070	0.034	0.210	0.006	0.555	0.006
China	1.025	0.022	0.002	0.022	0.000	0.016	0.000
Hong Kong	1.051	0.004	0.006	0.004	0.000	0.038	0.000
Australia	1.000	0.062	0.052	0.493	0.086	0.497	0.080
France	0.998	0.060	0.054	0.505	0.000	0.766	0.000
Germany	1.012	0.004	0.002	0.054	0.000	0.190	0.000
Italy	0.994	0.112	0.106	0.461	0.000	0.673	0.000
Egypt	1.007	0.012	0.008	0.044	0.010	0.337	0.010
Nigeria	1.002	0.024	0.030	0.014	0.000	0.020	0.000

Table 7: AR(1) estimates and 95% confidence intervals of various methods for ten stock indices.

	$\hat{\rho}_T$	t_{hac}	PM	GSW	DWB_3	DWB_5	DWB_{10}	DWB_r
USA	1.003	[1.001,1.031]	[1.001,1.053]	[1.001,1.028]	[1.001,1.039]	[1.001,1.038]	[1.001,1.035]	[1.001,1.037]
Brazil	1.001	[1.001,1.024]	[1.001,1.024]	[1.001,1.026]	[1.001,1.033]	[1.001,1.033]	[1.001,1.031]	[1.001,1.034]
China	1.025	[1.008,1.042]	[1.001,1.079]	[1.011,1.039]	[1.007,1.045]	[1.008,1.044]	[1.012,1.043]	[1.010,1.044]
Hong Kong	1.051	[1.021,1.081]	[1.042,1.060]	[1.026,1.076]	[1.022,1.088]	[1.022,1.092]	[1.020,1.093]	[1.020,1.090]
Australia	1.000	-	-	-	-	-	-	-
France	0.998	-	-	-	-	-	-	-
Germany	1.012	[1.001,1.031]	[1.001,1.104]	[1.001,1.033]	[1.001,1.035]	[1.001,1.036]	[1.001,1.038]	[1.001,1.035]
Italy	0.998	-	-	-	-	-	-	-
Egypt	1.007	[1.001,1.034]	[1.001,1.094]	[1.001,1.030]	[1.001,1.041]	[1.001,1.040]	[1.001,1.034]	[1.001,1.039]
Nigeria	1.004	[1.001,1.025]	[1.001,1.068]	[1.001,1.023]	[1.001,1.027]	[1.001,1.024]	[1.001,1.023]	[1.001,1.026]

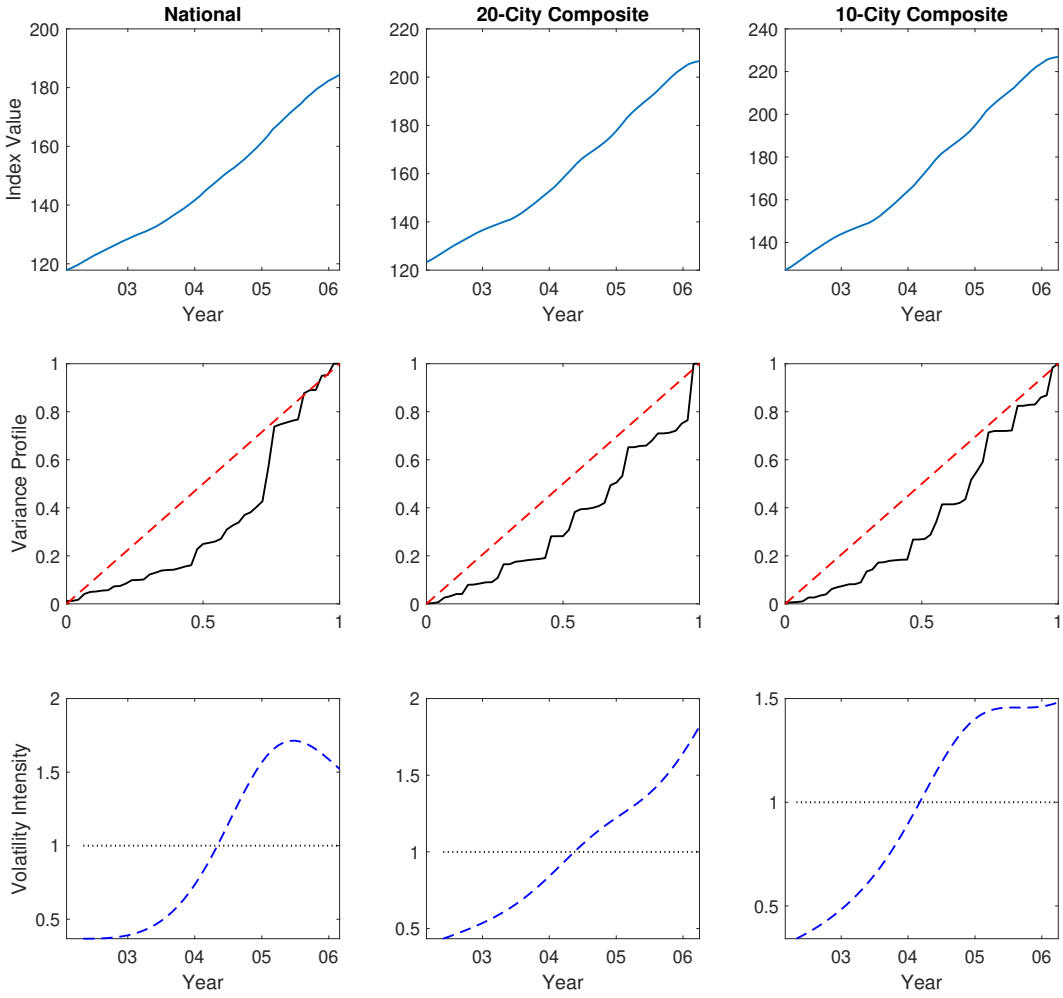

Figure 3: Plots of three U.S. housing price indices and their variance profile, nonparametric volatility estimates during 2002-2006.

Table 8: Nonstationary volatility tests (Cavaliere and Taylor, 2007b) and bubble tests (Harvey et al. 2018, 2020) results for U.S. housing price series.

Panel A: Tests for stationary volatility.			
	National	20-City Composite	10-City Composite
$\mathcal{H}_{KS}$	1.177	1.093	1.314
$\mathcal{H}_{R}$	1.101	1.010	1.294*
$\mathcal{H}_{CVM}$	0.478**	0.424*	0.567**
$\mathcal{H}_{AD}$	2.165*	2.620**	2.822**
Panel B: Explosiveness estimates and p-values from bubble tests.			
	National	20-City Composite	10-City Composite
$\hat{\rho}_T$	1.012	1.009	1.005
$\mathcal{U}$	0.002	0.000	0.000
supBZ	0.000	0.000	0.000
supDF	0.002	0.000	0.002
uPSY	0.000	0.002	0.002
PSY	0.018	0.010	0.016
sPSY	0.000	0.000	0.000

Note: *denotes 10%, **denotes 5%, and ***denotes 1% significance level for the above tests.

Table 9: AR(1) estimates and 95% confidence intervals of various methods for U.S. housing price series.

	National	20-City Composite	10-City Composite
$\hat{\rho}_T$	1.012	1.009	1.005
t_{hac}	[1.001,1.038]	[1.001,1.029]	[1.001,1.032]
PM	[1.001,1.184]	[1.001,1.151]	[1.001,1.108]
GSW	[1.001,1.027]	[1.001,1.025]	[1.001,1.024]
DWB_3	[1.001,1.042]	[1.001,1.026]	[1.001,1.027]
DWB_5	[1.001,1.040]	[1.001,1.023]	[1.001,1.026]
DWB_{10}	[1.001,1.044]	[1.001,1.023]	[1.001,1.026]
DWB_r	[1.001,1.043]	[1.001,1.026]	[1.001,1.032]

Supplementary Appendix A: Technical Lemmas

This Appendix contains a set of technical lemmas that will be subsequently used in the proofs of the main results in Appendix B. As a matter of notation, we will use $\mathcal{C} = \mathcal{C}[0,1]$ to denote the space of continuous functions on $[0,1]$ and $\mathcal{D}$ the space of right continuous with left limit processes on $[0,1]$, '$\xrightarrow{p}$' to denote convergence in probability, '$\xrightarrow{w}$' to denote weak convergence in the space $\mathcal{D}$ endowed with the Skorohod metric, $\lfloor . \rfloor$ to denote the integer part of its argument, and $\mathbf{1}(\cdot)$ to denote the indicator function. For a random quantity δ, we write $\delta = \delta_0 + o_p(\delta_0)$ as $\delta = \delta_0 + s.o.$, where $s.o.$ represents a term of smaller order in probability. Further, we define $\dot{Y} = \frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0)$.

Lemma A.1 *For any real numbers a_1, a_2, a_3 that satisfy $0 < a_1 < \infty$, $-\infty < a_2, a_3 < \infty$, $a_1 \geq a_2$, we have $\lim_{T\to\infty} k_T^{-1} \sum_{t=1}^{T-|j|} \rho_T^{-a_1(T-t)+a_2|j|+a_3} < \infty$ holds uniformly in $j = 1, \ldots, T$.*

Lemma A.2 *[Guo et al., 2019] Under Assumptions 1-2, the following limiting results hold jointly:*

(a) $(a_T k_T^{3/2} \rho_T^T)^{-1} \sum_{t=1}^{T} \sum_{j=t}^{T} \rho_T^{t-1-j} u_j = o_p(1)$;

(b) $(a_T k_T^{3/2} \rho_T^{2T})^{-1} \sum_{t=1}^{T} \sum_{j=t}^{T} \rho_T^{2(t-1)-j} u_j = o_p(1)$;

(c) $(a_T^2 k_T \rho_T^T)^{-1} \sum_{t=1}^{T} \sum_{j=t}^{T} \rho_T^{t-1-j} u_j u_t = o_p(1)$;

(d) $(a_T k_T \rho_T^T)^{-2} \sum_{t=1}^{T} d_{t-1}^2 = Y_T^2/2c + o_p(1)$;

(e) $(a_T k_T^{3/2} \rho_T^T)^{-1} \sum_{t=1}^{T} d_{t-1} = Y_T/c + o_p(1)$;

(f) $(a_T^2 k_T \rho_T^T)^{-1} \sum_{t=1}^{T} d_{t-1} u_t = X_T Y_T + o_p(1)$.

Lemma A.3 *Under Assumptions 1-2, the following limit results hold uniformly in j:*

(a) $(a_T^2 \mu_T k_T^{3/2} \rho_T^T)^{-2} \sum_{t=1}^{T-|j|} \Big[u_t u_{t+|j|} \big\{ \sum_{i_1=1}^{t-1} \rho_T^{t-1-i_1} u_{i_1} \big\} \big\{ \sum_{i_2=t+|j|}^{T} \rho_T^{t+|j|-1-i_2} u_{i_2} \big\} \Big] = o_p(k_T^{-1})$;

(b) $(a_T^2 \mu_T k_T^{3/2} \rho_T^T)^{-2} \sum_{t=1}^{T-|j|} \Big[u_t u_{t+|j|} \big\{ \sum_{i_1=t}^{T} \rho_T^{t-1-i_1} u_{i_1} \big\} \big\{ \sum_{i_2=1}^{t+|j|-1} \rho_T^{t+|j|-1-i_2} u_{i_2} \big\} \Big] = o_p(k_T^{-1})$;

(c) $(a_T^2 \mu_T k_T^{3/2} \rho_T^T)^{-2} \sum_{t=1}^{T-|j|} \Big[u_t u_{t+|j|} \big\{ \sum_{i_1=t}^{T} \rho_T^{t-1-i_1} u_{i_1} \big\} \big\{ \sum_{i_2=t+|j|}^{T} \rho_T^{t+|j|-1-i_2} u_{i_2} \big\} \Big] = o_p(k_T^{-1})$.

Lemma A.4 *Under Assumptions 1-2 and 3′, the following limit results hold uniformly in*

j:

(a) $\sum_{t=1}^{T-|j|} y_{t-1}^2 y_{t+|j|-1}^2 = O_p(a_T^4 \mu_T^4 k_T^5 \rho_T^{4T})$,

(b) $\sum_{t=1}^{T-|j|} y_{t-1}^2 y_{t+|j|-1} = O_p(a_T^3 \mu_T^3 k_T^4 \rho_T^{3T})$, $\sum_{t=1}^{T-|j|} y_{t-1} y_{t+|j|-1}^2 = O_p(a_T^3 \mu_T^3 k_T^4 \rho_T^{3T})$,

(c) $\sum_{t=1}^{T-|j|} y_{t-1} y_{t+|j|-1} = O_p(a_T^2 \mu_T^2 k_T^3 \rho_T^{2T})$, $\sum_{t=1}^{T-|j|} y_{t-1}^2 = O_p(a_T^2 \mu_T^2 k_T^3 \rho_T^{2T})$,

$\sum_{t=1}^{T-|j|} y_{t+|j|-1}^2 = O_p(a_T^2 \mu_T^2 k_T^3 \rho_T^{2T})$.

Lemma A.5 *Under Assumptions 1-2 and 3′, the following limit results hold uniformly in j:*

(a) $\sum_{t=1}^{T-|j|} y_{t-1} y_{t+|j|-1}^2 u_t = O_p(a_T^4 \mu_T^3 k_T^{7/2} \rho_T^{3T})$, $\sum_{t=1}^{T-|j|} y_{t-1}^2 y_{t+|j|-1} u_{t+|j|} = O_p(a_T^4 \mu_T^3 k_T^{7/2} \rho_T^{3T})$,

(b) $\sum_{t=1}^{T-|j|} y_{t-1} y_{t+|j|-1} u_t = O_p(a_T^3 \mu_T^2 k_T^{5/2} \rho_T^{2T})$, $\sum_{t=1}^{T-|j|} y_{t-1} y_{t+|j|-1} u_{t+|j|} = O_p(a_T^3 \mu_T^2 k_T^{5/2} \rho_T^{2T})$,

$\sum_{t=1}^{T-|j|} y_{t+|j|-1}^2 u_t = O_p(a_T^3 \mu_T^2 k_T^{5/2} \rho_T^{2T})$, $\sum_{t=1}^{T-|j|} y_{t-1}^2 u_{t+|j|} = O_p(a_T^3 \mu_T^2 k_T^{5/2} \rho_T^{2T})$,

(c) $\sum_{t=1}^{T-|j|} y_{t-1} u_t = O_p(a_T^2 \mu_T k_T^{3/2} \rho_T^T)$, $\sum_{t=1}^{T-|j|} y_{t-1} u_{t+|j|} = O_p(a_T^2 \mu_T k_T^{3/2} \rho_T^T)$,

$\sum_{t=1}^{T-|j|} y_{t+|j|-1} u_t = O_p(a_T^2 \mu_T k_T^{3/2} \rho_T^T)$, $\sum_{t=1}^{T-|j|} y_{t+|j|-1} u_{t+|j|} = O_p(a_T^2 \mu_T k_T^{3/2} \rho_T^T)$.

Lemma A.6 *Under Assumptions 1-2 and 3′, the following limit results hold uniformly in j:*

(a) $\hat{\rho}_T^{-2(T-t)+|j|-2} = \rho_T^{-2(T-t)+|j|-2} + o_p(1)$, $t = 1, ..., T - |j|$.

(b) $\sum_{t=1}^{T-|j|} \hat{\rho}_T^{-2(T-t)+|j|-2} y_{t-1} y_{t+|j|-1} = O_p(a_T^2 \mu_T^2 k_T^3 \rho_T^{2T})$,

(c) $\sum_{t=1}^{T-|j|} \hat{\rho}_T^{-2(T-t)+|j|-2} y_{t-1} = O_p(a_T \mu_T k_T^2 \rho_T^T)$, $\sum_{t=1}^{T-|j|} \hat{\rho}_T^{-2(T-t)+|j|-2} y_{t+|j|-1} = O_p(a_T \mu_T k_T^2 \rho_T^T)$,

(d) $\sum_{t=1}^{T-|j|} \hat{\rho}_T^{-2(T-t)+|j|-2} y_{t-1} u_{t+|j|} = O_p(a_T^2 \mu_T k_T^{3/2} \rho_T^T)$, $\sum_{t=1}^{T-|j|} \hat{\rho}_T^{-2(T-t)+|j|-2} y_{t+|j|-1} u_t = O_p(a_T^2 \mu_T k_T^{3/2} \rho_T^T)$,

(e) $\sum_{t=1}^{T-|j|} \hat{\rho}_T^{-2(T-t)+|j|-2} u_t = O_p(a_T k_T^{1/2})$, $\sum_{t=1}^{T-|j|} \hat{\rho}_T^{-2(T-t)+|j|-2} u_{t+|j|} = O_p(a_T k_T^{1/2})$,

Lemma A.7 *Under Assumptions 1-2 and 3′,* $\sum_{t=1}^{T} y_t \sigma_t^2 = O_p(a_T^3 \mu_T k_T^2 \rho_T^T)$.

Proof of Lemma A.1: We have

$$
\begin{aligned}
& k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-a_1(T-t)+a_2|j|+a_3} = \left|k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-a_1(T-t)+a_2|j|+a_3}\right| \\
= \; & k_T^{-1}\frac{\left|\rho_T^{-(a_1-a_2)|j|+a_3-a_1}-\rho_T^{-a_1(T-1)+a_2|j|+a_3}\right|}{\rho_T^{a_1}-1} = \frac{\left|\rho_T^{-(a_1-a_2)|j|+a_3-a_1}-\rho_T^{-a_1(T-|j|)-(a_1-a_2)|j|+a_1+a_3}\right|}{a_1c+O(k_T^{-1})} \\
\leq \; & \frac{\left|\rho_T^{-(a_1-a_2)|j|+a_3-a_1}\right|+\left|\rho_T^{-a_1(T-|j|)-(a_1-a_2)|j|+a_1+a_3}\right|}{a_1c+O(k_T^{-1})} \qquad (A.1)
\end{aligned}
$$

Taking the limit of both sides of $(A.1)$, we have

$$
\begin{aligned}
& \lim_{T\to\infty} k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-a_1(T-t)+a_2|j|+a_3} \leq \lim_{T\to\infty}\frac{\left|\rho_T^{-(a_1-a_2)|j|+a_3-a_1}\right|+\left|\rho_T^{-a_1(T-|j|)-(a_1-a_2)|j|+a_1+a_3}\right|}{a_1c+O(k_T^{-1})} \\
\leq \; & \lim_{T\to\infty}\frac{\left|\rho_T^{a_3-a_1}\right|+\left|\rho_T^{a_1+a_3}\right|}{a_1c+O(k_T^{-1})} = \frac{2}{a_1c} < \infty \;\; \blacktriangle \qquad (A.2)
\end{aligned}
$$

Proof of Lemma A.2: The proofs are essentially the same as Lemmas A.4-5 in Guo et al. (2019), with additional scaling factors a_T of certain order, either a_T^{-1} or a_T^{-2}, added obviously to ensure that the variance of e_t is well-behaved. We omit the proofs here but they are available upon request. ▲

Proof of Lemma A.3: (a). Note that

$$
\begin{aligned}
& \mathrm{E}\left|k_T(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\Big[u_tu_{t+|j|}\Big\{\sum_{i_1=1}^{t-1}\rho_T^{t-1-i_1}u_{i_1}\Big\}\Big\{\sum_{i_2=t+|j|}^{T}\rho_T^{t+|j|-1-i_2}u_{i_2}\Big\}\Big]\right| \\
\leq \; & k_T(\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\Big[\Big\{\sum_{i_1=1}^{t-1}\rho_T^{t-1-i_1}\Big\}\Big\{\sum_{i_2=t+|j|}^{T}\rho_T^{t+|j|-1-i_2}\Big\}\mathrm{E}\left|a_T^{-4}u_tu_{t+|j|}u_{i_1}u_{i_2}\right|\Big] \\
\leq \; & k_T(\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\Big[\Big\{\sum_{i_1=1}^{t-1}\rho_T^{t-1-i_1}\Big\}\Big\{\sum_{i_2=t+|j|}^{T}\rho_T^{t+|j|-1-i_2}\Big\}\Big] \\
& \times\{\mathrm{E}(a_T^{-1}u_t)^4\}^{1/4}\{\mathrm{E}(a_T^{-1}u_{t+|j|})^4\}^{1/4}\{\mathrm{E}(a_T^{-1}u_{i_1})^4\}^{1/4}\{\mathrm{E}(a_T^{-1}u_{i_2})^4\}^{1/4} \\
\leq \; & k_T(\mu_T k_T^{1/2})^{-2}\rho_T^{-2T}k_T^{-2}\left(\frac{(\rho_T^{T-|j|}-1)(1+\rho_T^{t-T}-\rho_T^{|j|-T})}{(\rho_T-1)^3}-\frac{T-|j|}{(\rho_T-1)^2}\right)\times(\sum_{j=0}^{\infty}|c_j|)^4K_2^{4/(4+\kappa)}K_1^{4/(4+\eta)} \\
= \; & \nu^{-2}\rho_T^{-2T}k_T^{-1}\left(O(k_T^3\rho_T^T)+O(k_T^2T)\right)\times O(1)\times K_2^{4/(4+\kappa)}K_1^{4/(4+\eta)} = O(k_T^2\rho_T^{-T}) = o(1). \qquad (A.3)
\end{aligned}
$$

holds uniformly in j, where the second inequality holds due to Cauchy-Schwarz inequality and the third inequality holds if $\sup_t \mathrm{E}[(a_T^{-1}u_t)^4]$ is bounded. This is indeed true as (see, Tanaka, 1996, pp. 501-502)

$$\begin{aligned} |a_T^{-1}u_t| &= |\sum_{j=0}^{\infty} c_j a_T^{-1}\sigma_{t+|j|}\varepsilon_{t+|j|}| \leq \sum_{j=0}^{\infty} |c_j|^{3/4} a_T^{-1}\sigma_{t+|j|}(|c_j||\varepsilon_{t+|j|}|^4)^{1/4} \\ &\leq \sup_t(a_T^{-1}\sigma_t) \times (\sum_{j=0}^{\infty}|c_j|)^{3/4}(\sum_{j=0}^{\infty}|c_j||\varepsilon_{t+|j|}|^4)^{1/4} \end{aligned} \tag{A.4}$$

so that

$$\begin{aligned} &\sup_t \mathrm{E}(a_T^{-1}u_t)^4 \leq \sup_t \mathrm{E}(a_T^{-1}\sigma_t)^4 \times (\sum_{j=0}^{\infty}|c_j|)^4 \times \sup_t \mathrm{E}(\varepsilon_t^4) \\ \leq\ & (\sum_{j=0}^{\infty}|c_j|)^4 K_2^{4/(4+\kappa)} K_1^{4/(4+\eta)} \leq (|c_0| + \sum_{j=0}^{\infty} j|c_j|)^4 K_2^{4/(4+\kappa)} K_1^{4/(4+\eta)} < \infty \end{aligned} \tag{A.5}$$

Then part (a) follows as we have shown its convergence in L^1, which implies convergence in probability. The proofs for (b) and (c) follow very similar steps as the proof for (a) and are

hence omitted, but they are available upon request. ▲

Proof of Lemma A.4: (a). Since $y_t = y_0\rho_T^t + \sum_{i=1}^t \rho_T^{t-i}u_i + \mu_T(\rho_T^t - 1)k_T/c$, we have

$$
\begin{aligned}
& (a_T^4\mu_T^4 k_T^5 \rho_T^{4T})^{-1} \sum_{t=1}^{T-|j|} y_{t-1}^2 y_{t+|j|-1}^2 \\
= \; & (a_T^4\mu_T^4 k_T^5 \rho_T^{4T})^{-1} \sum_{t=1}^{T-|j|} \Bigg[\Big(y_0\rho_T^{t-1} + \sum_{i=1}^{t-1} \rho_T^{t-1-i}u_i + \mu_T(\rho_T^{t-1} - 1)k_T/c \Big)^2 \\
& \times \Big(y_0\rho_T^{t+|j|-1} + \sum_{i=1}^{t+|j|-1} \rho_T^{t+|j|-1-i}u_i + \mu_T(\rho_T^{t+|j|-1} - 1)k_T/c \Big)^2 \Bigg] \qquad \text{(A.6)} \\
= \; & (a_T^4\mu_T^4 k_T^5 \rho_T^{4T})^{-1} \sum_{t=1}^{T-|j|} \Bigg[\Big(y_0\rho_T^{t-1} + \sum_{i=1}^{T} \rho_T^{t-1-i}u_i + \mu_T(\rho_T^{t-1} - 1)k_T/c \Big)^2 \\
& \times \Big(y_0\rho_T^{t+|j|-1} + \sum_{i=1}^{T} \rho_T^{t+|j|-1-i}u_i + \mu_T(\rho_T^{t+|j|-1} - 1)k_T/c \Big)^2 \Bigg] + s.o. \qquad \text{(A.7)} \\
= \; & (a_T^4\mu_T^4 k_T^5 \rho_T^{4T})^{-1} \sum_{t=1}^{T-|j|} \rho_T^{2(t-1)} \rho_T^{2(t+|j|-1)} \Big(\sum_{i=1}^{T} \rho_T^{-i}u_i + \mu_T k_T/c \Big)^4 + s.o. \\
= \; & (\mu_T^4 k_T^2)^{-1} \underbrace{\Big(k_T^{-1} \sum_{t=1}^{T-|j|} \rho_T^{-2[T-(t-1)]} \rho_T^{-2[T-(t+|j|-1)]} \Big)}_{=O(1), \text{ by Lemma A.1}} \Big(a_T^{-1}k_T^{-1/2} \sum_{i=1}^{T} \rho_T^{-i}u_i + a_T^{-1}c^{-1}\nu \Big)^4 + s.o. \\
\xrightarrow{w} \; & \nu^{-5} \times O(1) \times \dot{Y}^4 + o_p(1) = O_p(1) \qquad \text{(A.8)}
\end{aligned}
$$

where the approximation from $(A.6)$ to $(A.7)$ can be easily proved in a similar manner to $(A.3)$ in Lemma A.3. Thus we have proved (a) is stochastically bounded.
(b). Similar to (a), with some algebra, we have

$$
\begin{aligned}
(a_T^3\mu_T^3 k_T^4 \rho_T^{3T})^{-1} \sum_{t=1}^{T-|j|} y_{t-1}^2 y_{t+|j|-1} \; & \xrightarrow{w} \; \nu^{-4} \times O(1) \times \dot{Y}_T^3 + o_p(1) = O_p(1) \qquad \text{(A.9)} \\
(a_T^3\mu_T^3 k_T^4 \rho_T^{3T})^{-1} \sum_{t=1}^{T-|j|} y_{t-1} y_{t+|j|-1}^2 \; & \xrightarrow{w} \; \nu^{-4} \times O(1) \times \dot{Y}^3 + o_p(1) = O_p(1) \qquad \text{(A.10)}
\end{aligned}
$$

(c). Similar to (a) and (b), with some algebra, we have

$$(a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}y_{t+|j|-1} \xrightarrow{w} \nu^{-3}\times O(1)\times\dot{Y}^2+o_p(1)=O_p(1) \tag{A.11}$$

$$(a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}^2 \xrightarrow{w} \nu^{-3}\times O(1)\times\dot{Y}^2+o_p(1)=O_p(1) \tag{A.12}$$

$$(a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t+|j|-1}^2 \xrightarrow{w} \nu^{-3}\times O(1)\times\dot{Y}^2+o_p(1)=O_p(1) \ \blacktriangle \tag{A.13}$$

Proof of Lemma A.5: (a). The first result is calculated by

$$\begin{aligned}
&(a_T^4\mu_T^3k_T^{7/2}\rho_T^{3T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}y_{t+|j|-1}^2u_t\\
=\ &(a_T^4\mu_T^4k_T^5\rho_T^{4T})^{-1}\sum_{t=1}^{T-|j|}\Bigg[\Big(y_0\rho_T^{t-1}+\sum_{i=1}^{t}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t-1}-1)k_T/c\Big)\\
&\times\Big(y_0\rho_T^{t+|j|-1}+\sum_{i=1}^{t+|j|-1}\rho_T^{t+|j|-1-i}u_i+\mu_T(\rho_T^{t+|j|-1}-1)k_T/c\Big)^2u_t\Bigg]\\
=\ &(a_T^4\mu_T^3k_T^{7/2}\rho_T^{3T})^{-1}\sum_{t=1}^{T-|j|}\Bigg[\Big(y_0\rho_T^{t-1}+\sum_{i=1}^{T}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t-1}-1)k_T/c\Big)\\
&\times\Big(y_0\rho_T^{t+|j|-1}+\sum_{i=1}^{T}\rho_T^{t+|j|-1-i}u_i+\mu_T(\rho_T^{t+|j|-1}-1)k_T/c\Big)^2u_t\Bigg]+s.o.\\
=\ &(a_T^4\mu_T^3k_T^{7/2}\rho_T^{3T})^{-1}\sum_{t=1}^{T-|j|}\rho_T^{2(t+|j|-1)}\rho_T^{t-1}u_t\Big(\sum_{i=1}^{T}\rho_T^{-i}u_i+\mu_Tk_T/c\Big)^3+s.o.\\
=\ &(\mu_T^3k_T^{3/2})^{-1}\underbrace{\Big(a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T-|j|}\rho_T^{-2[T-(t+|j|-1)]}\rho_T^{-[T-(t-1)]}u_t\Big)}_{=O_p(1)}\Big(a_T^{-1}k_T^{-1/2}\sum_{i=1}^{T}\rho_T^{-i}u_i+a_T^{-1}c^{-1}\nu\Big)^3+s.o.\\
\xrightarrow{w}\ &\nu^{-4}\times O_p(1)\times\dot{Y}^3+o_p(1)=O_p(1)
\end{aligned} \tag{A.14}$$

where $\left(a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T-|j|}\rho_T^{-2[T-(t+|j|-1)]}\rho_T^{-[T-(t-1)]}u_t\right)=O_p(1)$ is due to

$$\begin{aligned}
&\mathrm{E}\left(a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T-|j|}\rho_T^{-2[T-(t+|j|-1)]}\rho_T^{-[T-(t-1)]}u_t\right)^2\\
\leq\ & C^2(1)\underbrace{\left(k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-4[T-(t+|j|-1)]}\rho_T^{-2[T-(t-1)]}\right)}_{=O(1),\text{ by Lemma A.1}}\times\{\sup_t\mathrm{E}(a_T^{-1}e_t)^4\}^{1/2}+s.o.\\
=\ & C^2(1)\times O(1)\times O(1)=O(1) \qquad (A.15)
\end{aligned}$$

which uses the result that $\sup_t\mathrm{E}[(a_T^{-1}e_t)^4]\leq\sup_t\mathrm{E}[\epsilon_t^4]\sup_t\mathrm{E}[(a_T^{-1}\sigma_t)^4]<\infty$. The second result is similar, i.e., $(a_T^4\mu_T^3k_T^{7/2}\rho_T^{3T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}^2y_{t+|j|-1}u_{t+|j|-1}=$
$(\mu_T^3k_T^{3/2})^{-1}\underbrace{\left(a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T-|j|}\rho_T^{-[T-(t+|j|-1)]}\rho_T^{-2[T-(t-1)]}u_{t+|j|-1}\right)}_{=O_p(1)}\left(a_T^{-1}k_T^{-1/2}\sum_{i=1}^{T}\rho_T^{-i}u_i+a_T^{-1}c^{-1}\nu\right)^3+$
$s.o.\xrightarrow{w}\nu^{-4}\times O_p(1)\times\dot{Y}^3+o_p(1)=O_p(1)$.
(b). Similar to (a), with some algebra, we have

$$(a_T^3\mu_T^2k_T^{5/2}\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}y_{t+|j|-1}u_t\ \xrightarrow{w}\ \nu^{-3}\times O_p(1)\times\dot{Y}^2+o_p(1)=O_p(1) \qquad (A.16)$$

$$(a_T^3\mu_T^2k_T^{5/2}\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}y_{t+|j|-1}u_{t+|j|-1}\ \xrightarrow{w}\ \nu^{-3}\times O_p(1)\times\dot{Y}^2+o_p(1)=O_p(1) \qquad (A.17)$$

$$(a_T^3\mu_T^2k_T^{5/2}\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t+|j|-1}^2u_t\ \xrightarrow{w}\ \nu^{-3}\times O_p(1)\times\dot{Y}^2+o_p(1)=O_p(1) \qquad (A.18)$$

$$(a_T^3\mu_T^2k_T^{5/2}\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}^2u_{t+|j|-1}\ \xrightarrow{w}\ \nu^{-3}\times O_p(1)\times\dot{Y}^2+o_p(1)=O_p(1) \qquad (A.19)$$

(c). For a given j, we have

$$
\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T-|j|} y_{t-1}u_t \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T-|j|}\Big(y_0\rho_T^{t-1}+\sum_{i=1}^{t}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t-1}-1)k_T/c\Big)u_t \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T-|j|}\Big(y_0\rho_T^{t-1}+\sum_{i=1}^{T}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t-1}-1)k_T/c\Big)u_t + s.o. \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T-|j|}\rho_T^{t-1}u_t\Big(\sum_{i=1}^{T}\rho_T^{-i}u_i+\mu_T k_T/c\Big)+s.o. \\
= \;& (\mu_T k_T^{1/2})^{-1}\underbrace{\Big(a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T-|j|}\rho_T^{-[T-(t-1)]}u_t\Big)}_{=O_p(1)}\Big(a_T^{-1}k_T^{-1/2}\sum_{i=1}^{T}\rho_T^{-i}u_i+a_T^{-1}c^{-1}\nu\Big)+s.o. \\
\xrightarrow{w} \;& \nu^{-2}\times O_p(1)\times\dot{Y}+o_p(1)=O_p(1)
\end{aligned}
\tag{A.20}
$$

where $a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T-|j|}\rho_T^{-[T-(t-1)]}u_t = O_p(1)$ can be easily proved in a similar manner to $(A.15)$. Using the same approach, it is easy to check that the rest of the three cases (b), (c) and (d) are of the stated orders. ▲

Proof of Lemma A.6: Using the fact that $\hat{\rho}_T = \rho_T + O_p((\mu_T k_T^{3/2}\rho_T^T)^{-1})$, a first order Taylor expansion yields

$$
\begin{aligned}
\hat{\rho}_T^{-2(T-t)+|j|-2}-\rho_T^{-2(T-t)+|j|-2} &= (\hat{\rho}_T-\rho_T)\times\underbrace{(-2(T-t)+|j|-2)}_{O(T)\text{ or smaller}}\underbrace{\rho_T^{-2(T-t)+|j|-3}}_{<\infty}+s.o. \\
&= O_p((\mu_T k_T^{3/2}\rho_T^T)^{-1})\times O(T)\times O(1)+s.o. \\
&= \nu^{-1}O_p(Tk_T^{-1}\rho_T^{-T})+s.o.=o_p(1)
\end{aligned}
\tag{A.21}
$$

For the remaining results (b)-(e), we only prove (b) as the rest of the cases can be proved in a similar manner to this case and cases in previous Lemmas A.4-A.5. We first show the

results hold if $\hat{\rho}_T$ is replaced with ρ_T:

$$
\begin{aligned}
& (a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}y_{t-1}y_{t+|j|-1} \\
= \;& (a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}\Bigg[\rho_T^{-2(T-t)+|j|-2}\Big(y_0\rho_T^{t-1}+\sum_{i=1}^{t}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t-1}-1)k_T/c\Big) \\
& \times\Big(y_0\rho_T^{t+|j|-1}+\sum_{i=1}^{t+|j|-1}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t+|j|-1}-1)k_T/c\Big)\Bigg] \\
= \;& (a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}\Bigg[\rho_T^{-2(T-t)+|j|-2}\Big(y_0\rho_T^{t-1}+\sum_{i=1}^{T}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t-1}-1)k_T/c\Big) \\
& \times\Big(y_0\rho_T^{t+|j|-1}+\sum_{i=1}^{T}\rho_T^{t-1-i}u_i+\mu_T(\rho_T^{t+|j|-1}-1)k_T/c\Big)\Bigg]+s.o. \\
= \;& (a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}\rho_T^{(t+|j|-1)}\rho_T^{t-1}\rho_T^{-2(T-t)+|j|-2}\Big(\sum_{i=1}^{T}\rho_T^{-i}u_i+\mu_Tk_T/c\Big)^2+s.o. \\
= \;& (\mu_T^2k_T)^{-1}\underbrace{\Big(k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-[T-(t+|j|-1)]}\rho_T^{-[T-(t-1)]}\rho_T^{-2(T-t)+|j|-2}\Big)}_{=O(1),\text{ by Lemma A.1}}\Big(a_T^{-1}k_T^{-1/2}\sum_{i=1}^{T}\rho_T^{-i}u_i+a_T^{-1}c^{-1}\nu\Big)^2+s.o. \\
\xrightarrow{w} \;& \nu^{-3}\times O(1)\times\dot{Y}^2+o_p(1)=O_p(1)
\end{aligned}
\tag{A.22}
$$

Next, using the results in (a), we have

$$
\begin{aligned}
& (a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}y_{t-1}y_{t+|j|-1} \\
= \;& (a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}y_{t-1}y_{t+|j|-1}+\underbrace{(a_T^2\mu_T^2k_T^3\rho_T^{2T})^{-1}\sum_{t=1}^{T-|j|}y_{t-1}y_{t+|j|-1}}_{=O_p(1),\text{ by Lemma A.4(c)}}\times o_p(1) \\
= \;& O_p(1)\quad\blacktriangle
\end{aligned}
\tag{A.23}
$$

Proof of Lemma A.7: Since $y_{t-1} = y_0\rho_T^{t-1} + \sum_{i=1}^{t-1}\rho_T^{t-1-i}u_i + \mu_T(\rho_T^{t-1}-1)k_T/c$, we have

$$
\begin{aligned}
& (a_T^3\mu_T k_T^2\rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1}\sigma_t^2 \\
= \;& (a_T^3\mu_T k_T^2\rho_T^T)^{-1}\sum_{t=1}^{T}\Big(y_0\rho_T^{t-1} + \sum_{i=1}^{t-1}\rho_T^{t-1-i}u_i + \mu_T(\rho_T^{t-1}-1)k_T/c\Big)\sigma_t^2 \\
= \;& (a_T^3\mu_T k_T^2\rho_T^T)^{-1}\sum_{t=1}^{T}\Big(\sum_{i=1}^{T}\rho_T^{t-1-i}u_i\Big)\sigma_t^2 + \sum_{t=1}^{T}\mu_T\rho_T^{t-1}k_T\sigma_t^2/c + s.o. \\
= \;& (\mu_T k_T^{1/2})^{-1}\Big(a_T^{-1}k_T^{-1/2}\sum_{i=1}^{T}\rho_T^{-i}u_i\Big)\Big(a_T^{-2}k_T^{-1}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}\sigma_t^2\Big) + (a_Tc)^{-1}a_T^{-2}k_T^{-1}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}\sigma_t^2 + s.o. \\
\xrightarrow{w} \;& \nu^{-1}X_T \times \lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-k_T^{-1}Tc(1-s)}g^2(s)ds + (a_Tc)^{-1} \times \lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-k_T^{-1}Tc(1-s)}g^2(s)ds \\
= \;& \Big[\nu^{-1}X_T + 1/\psi c\mathbf{1}(\gamma=0)\Big] \times \lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-k_T^{-1}Tc(1-s)}g^2(s)ds \;\blacktriangle
\end{aligned}
\tag{A.24}
$$

References

[1] Guo, G., Sun, Y., Wang, S. (2019), "Testing for moderate explosiveness," *Econometrics Journal* 22, 73-95.

[2] Tanaka, K. (1996), *Time Series Analysis*. Wiley, New York.

Supplementary Appendix B: Proofs of Main Results

This appendix contains proofs of the main results in the paper. Let P^* denote the bootstrap probability measure and E^* the expectation with respect to P^*. We will use $\mathcal{C} = \mathcal{C}[0,1]$ to denote the space of continuous functions on $[0,1]$ and $\mathcal{D}$ the space of right continuous with left limit processes on $[0,1]$, '$\xrightarrow{p}$' to denote convergence in probability, '$\xrightarrow{w}$' to denote weak convergence in the space $\mathcal{D}$ endowed with the Skorohod metric, '$\xrightarrow{w}_p$' to denote weak convergence in probability under the bootstrap measure (Giné and Zinn, 1990), $\lfloor . \rfloor$ to denote the integer part of its argument, and $\mathbf{1}(\cdot)$ to denote the indicator function. We will write $Z_T^* = o_{p^*}(1)$ if, for any $\epsilon_1 > 0, \epsilon_2 > 0$, $\lim_{T\to\infty} P[P^*(|Z_T^*| > \epsilon_1) > \epsilon_2] = 0$. Similarly, we write $Z_T^* = O_{p^*}(1)$ if, for all $\epsilon > 0$, there exists an $M_\epsilon < \infty$ such that $\lim_{T\to\infty} P[P^*(|Z_T^*| > M_\epsilon) > \epsilon] = 0$. Finally, for a random quantity δ, we write $\delta = \delta_0 + o_p(\delta_0)$ as $\delta = \delta_0 + s.o.$, where $s.o.$ represents a term of smaller order in probability.

Proof of Theorem 1: (a) Let $\widetilde{X}_T = a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \rho_T^{-(T-t)-1} e_t$. We first show that $X_T = C(1)\widetilde{X}_T + o_p(1)$. Under Assumption 2, the standard Beveridge-Nelson decomposition, see Phillips and Solo (1992), states that

$$u_t = C(1)e_t + \widetilde{e}_{t-1} - \widetilde{e}_t \tag{B.1}$$

where $\widetilde{e}_t = \widetilde{C}(L)e_t = \sum_{j=0}^{\infty} \widetilde{c}_j e_{t-|j|}$, $\widetilde{c}_j = \sum_{k=j+1}^{\infty} c_k$. Furthermore, the assumption $\sum_{j=0}^{\infty} j|c_j| < \infty$ ensures $\sum_{j=0}^{\infty} \widetilde{c}_j^2 < \infty$, which implies $\mathrm{E}(a_T^{-2}\widetilde{e}_t^2) \leq \sum_{j=0}^{\infty} \widetilde{c}_j^2 \mathrm{sup}_t \mathrm{E}(a_T^{-2}\sigma_t^2) \mathrm{sup}_t \mathrm{E}(\epsilon_t^2) \leq K_2^{2/(4+\kappa)} K_1^{2/(4+\eta)} \sum_{j=0}^{\infty} \widetilde{c}_j^2 < \infty$, $\forall t$. We first prove for X_T. Now we have

$$\begin{aligned} X_T &= a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \rho_T^{-(T-t)-1} u_t \\ &= a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \rho_T^{-(T-t)-1} C(1) e_t + a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \rho_T^{-(T-t)-1} (\widetilde{e}_{t-1} - \widetilde{e}_t) \\ &= C(1)\widetilde{X}_T + a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \rho_T^{-(T-t)-1} (\widetilde{e}_{t-1} - \widetilde{e}_t) \end{aligned} \tag{B.2}$$

where we need to show that the second term $a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \rho_T^{-(T-t)-1} (\widetilde{e}_{t-1} - \widetilde{e}_t)$ is $o_p(1)$.

Following standard calculations,

$$\begin{aligned} a_T^{-1}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}(\widetilde{e}_{t-1}-\widetilde{e}_t) &= a_T^{-1}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}\widetilde{e}_{t-1} - a_T^{-1}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}\widetilde{e}_t \\ &= a_T^{-1}\sum_{t=0}^{T-1}\rho_T^{-(T-t)}\widetilde{e}_t - a_T^{-1}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}\widetilde{e}_t \\ &= a_T^{-1}\rho_T^{-T}\widetilde{e}_0 - a_T^{-1}\rho_T^{-1}\widetilde{e}_T + a_T^{-1}\sum_{t=1}^{T-1}(\rho_T^{-(T-t)}-\rho_T^{-(T-t)-1})\widetilde{e}_t \\ &= a_T^{-1}\rho_T^{-T}\widetilde{e}_0 - a_T^{-1}\rho_T^{-1}\widetilde{e}_T + ca_T^{-1}k_T^{-1}\sum_{t=1}^{T-1}\rho_T^{-(T-t)-1}\widetilde{e}_t \qquad \text{(B.3)} \end{aligned}$$

Since $\mathrm{E}(a_T^{-2}\widetilde{e}_t^2)<\infty$, $\forall t$, we have $k_T^{-1/2}a_T^{-1}\rho_T^{-T}\widetilde{e}_0 = o_p(1)$ and $k_T^{-1/2}a_T^{-1}\rho_T^{-1}\widetilde{e}_T = o_p(1)$. Further,

$$\begin{aligned} \mathrm{E}\left[(ca_T^{-1}k_T^{-1}\sum_{t=1}^{T-1}\rho_T^{-(T-t)-1}\widetilde{e}_t)^2\right] &= c^2k_T^{-2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\left(\rho_T^{-(T-t)-1}\rho_T^{-(T-s)-1}\mathrm{E}(a_T^{-2}\widetilde{e}_t\widetilde{e}_s)\right) \\ &\le c^2k_T^{-2}O\left((\sum_{t=1}^{T-1}\rho_T^{-(T-t)-1})^2\right)\times O(1) \\ &= c^2k_T^{-2}O\left(k_T^2\right) = O(1) \qquad \text{(B.4)} \end{aligned}$$

where the second inequality comes from the fact that, for any $t,s=1,\ldots,T$, $\mathrm{E}(a_T^{-2}\widetilde{e}_t\widetilde{e}_s) = \sum_{j=0}^{\infty}\widetilde{c}_j\widetilde{c}_{|t-s|+j}\mathrm{E}(a_T^{-2}e_{t-|j|}^2) \le K_2^{2/(4+\kappa)}K_1^{2/(4+\eta)}(\sum_{j=0}^{\infty}\widetilde{c}_j)^2<\infty$. Thus we have $a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}(\widetilde{e}_{t-1}-\widetilde{e}_t) = k_T^{-1/2}O_p(1) = o_p(1)$. Next, we establish that as $T\to\infty$, $\widetilde{X}_T \xrightarrow{w} \widetilde{X} \sim MN(0,V_{\widetilde{x}})$, where $V_{\widetilde{x}} = \lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tc(1-s)}g^2(s)ds$. To see this, define $\widetilde{\sigma}_t = \rho_T^{-(T-t)-1}\sigma_t$, $\omega_t = \rho_T^{-(T-t)-1}e_t = \widetilde{\sigma}_t\epsilon_t$ and $\widetilde{a}_T = (k_T/T)^{1/2}a_T$. Then we have

$$\widetilde{a}_T^{-1}\widetilde{\sigma}_{\lfloor sT\rfloor} = (k_T/T)^{-1/2}\rho_T^{-(T-\lfloor sT\rfloor)-1}a_T^{-1}\sigma_t \xrightarrow{w} (k_T/T)^{-1/2}\rho_T^{-T(1-s)}g(s),\ s\in[0,1], \qquad \text{(B.5)}$$

Adapting Lemma 1 in Cavaliere and Taylor (2009), as $T\to\infty$, we have

$$\begin{aligned} \widetilde{X}_T &= T^{1/2}k_T^{-1/2}a_T^{-1}T^{-1/2}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}e_t = T^{-1/2}\sum_{t=1}^{T}\widetilde{a}_T^{-1}\omega_t = T^{-1/2}\sum_{t=1}^{T}\widetilde{a}_T^{-1}\widetilde{\sigma}_t\epsilon_t \\ &\xrightarrow{w} \lim_{T\to\infty}\int_0^1\widetilde{a}_T^{-1}\widetilde{\sigma}_{\lfloor sT\rfloor}dB(s) = \lim_{T\to\infty}\int_0^1(k_T/T)^{-1/2}\rho_T^{-T(1-s)}g(s)dB(s) := \widetilde{X} \qquad \text{(B.6)} \end{aligned}$$

where $\widetilde{X}$ is the sum of normal variables and thus is normal. The variance is

$$V_{\widetilde{x}} = \lim_{T\to\infty} \mathrm{E}\left[\left(\int_0^1 (k_T/T)^{-1/2}\rho_T^{-T(1-s)}g(s)dB(s)\right)^2\right] = \lim_{T\to\infty}\int_0^1 (k_T/T)^{-1}\rho_T^{-2T(1-s)}g^2(s)ds$$
$$= \lim_{T\to\infty}\int_0^1 (k_T/T)^{-1}\left((1+c/k_T)^{k_T/c}\right)^{-2k_T^{-1}Tc(1-s)} g^2(s)ds \xrightarrow{w} \lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tc(1-s)}g^2(s)ds \tag{B.7}$$

(b) Let $\widetilde{Y}_T = a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\rho_T^{-t}e_t$. Then, using arguments entirely analogous to those in (a), we can show that $Y_T = C(1)\widetilde{Y}_T + o_p(1)$ and $\widetilde{Y}_T \xrightarrow{w} \widetilde{Y} \sim MN(0, V_{\widetilde{y}})$, where $V_{\widetilde{y}} = \lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tcs}g^2(s)ds$.

(c) Note that by Assumption 2(c),

$$0 < \frac{\underline{g}^2}{2c} = \underline{g}^2\lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tc(1-s)}ds \le V_x \le \bar{g}^2\lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tc(1-s)}ds = \frac{\bar{g}^2}{2c} < \infty$$

A similar argument establishes $0 < V_y < \infty$. ▲

Proof of Theorem 2: The proofs of (a)-(c) are similar to the proofs of the results in Theorem 3.1 in Guo et al. (2019), with some differences, due to the additional scaling factor a_T. Specifically, we take (c) as an example to show how to deal with it. To prove (c), by Lemma A.1,

$$(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1}u_t = (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T}(d_{t-1} + \mu_T(\rho_T^{t-1}-1)k_T/c)u_t$$
$$= (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T} d_{t-1}u_t + (a_T^2 k_T^{1/2}\rho_T^T)^{-1}c^{-1}\sum_{t=1}^{T}\rho_T^{t-1}u_t - (a_T^2 k_T^{1/2}\rho_T^T)^{-1}c^{-1}\sum_{t=1}^{T}u_t$$
$$= X_TY_T/(\mu_T k_T^{1/2}) + a_T^{-1}c^{-1}(a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}u_t) + O_p(T^{1/2}(k_T^{1/2}\rho_T^T)^{-1}a_T^{-1}) \tag{B.8}$$
$$= X_TY_T/(\mu_T k_T^{1/2}) + X_T/(a_T c) + o_p(1) \xrightarrow{w} \begin{cases} X(\frac{Y}{\nu} + \frac{1}{\psi c}) & \gamma = 0 \\ \frac{XY}{\nu} & \gamma > 0 \end{cases} := X(\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0))$$

The proofs for (a) and (b) follow Guo et al. (2019) and the techniques used in proof (a) and are hence omitted. For (d), we have

$$a_T^{-1}T^{-1/2}\sum_{t=1}^{T} u_t = a_T^{-1}T^{-1/2}\sum_{t=1}^{T} C(1)e_t + a_T^{-1}T^{-1/2}\sum_{t=1}^{T}(\widetilde{e}_{t-1} - \widetilde{e}_t)$$
$$= C(1)\times a_T^{-1}T^{-1/2}\sum_{t=1}^{T} e_t + \underbrace{a_T^{-1}T^{-1/2}\widetilde{e}_0}_{=o_p(1)} - \underbrace{a_T^{-1}T^{-1/2}\widetilde{e}_T}_{=o_p(1)} \xrightarrow{w} C(1)\int_0^1 g(s)dB(s) := U$$

together with $Var(U) = C^2(1)\int_0^1 g(s)^2 ds$, where $B(\cdot)$ defines a standard Brownian motion on $[0,1]$ and $a_T^{-1}T^{-1/2}\sum_{t=1}^{T} e_t \xrightarrow{w} \int_0^1 g(s)dB(s)$ follows from Lemma 1 in Cavaliere and Taylor (2009). ▲

Proof of Theorem 3: Same as the proofs for Theorem 2, we only prove (c).

$$
\begin{aligned}
(a_T^2 k_T \rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1}u_t &= (a_T^2 k_T \rho_T^T)^{-1}\sum_{t=1}^{T}(d_{t-1} + \mu_T(\rho_T^{t-1}-1)k_T/c)u_t \\
= \; & (a_T^2 k_T \rho_T^T)^{-1}\sum_{t=1}^{T} d_{t-1}u_t + (\mu_T k_T^{1/2})(a_T^2 k_T^{1/2}\rho_T^T)^{-1}c^{-1}\sum_{t=1}^{T}\rho_T^{t-1}u_t - (\mu_T k_T^{1/2})(a_T^2 k_T^{1/2}\rho_T^T)^{-1}c^{-1}\sum_{t=1}^{T} u_t \\
= \; & X_T Y_T + (\mu_T k_T^{1/2})a_T^{-1}c^{-1}(a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\rho_T^{-(T-t)-1}u_t) + O_p(T^{1/2}(k_T^{1/2}\rho_T^T)^{-1}a_T^{-1}(\mu_T k_T^{1/2})) \\
= \; & X_T Y_T + \nu X_T/(a_T c) + o_p(1) = X_T Y_T + o_p(1) \xrightarrow{w} XY
\end{aligned} \tag{B.9}
$$

where the last equality holds due to the fact that $a_T = \psi T^{\gamma}$ is at least $O(1)$ under $\gamma \geq 0$. ▲

Proof of Lemma 1: Let $\widetilde{\Phi}_T(j) = a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty} c_k c_{k+|j|}\sigma_{t-k}^2$. We first establish $\sum_{j=-(T-1)}^{T-1} w(j/b_T)\Phi_T(j) = \sum_{j=-(T-1)}^{T-1} w(j/b_T)\widetilde{\Phi}_T(j) + o_p(1)$. By definition,

$$
\begin{aligned}
\Phi_T(j) - \widetilde{\Phi}_T(j) &= a_T^{-2}k_T^{-1}\sum_{k=0}^{\infty}\sum_{i=0}^{\infty} c_k c_i \sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}\sigma_{t+|j|-i}\epsilon_{t-k}\epsilon_{t+|j|-i} - \widetilde{\Phi}_T(j) \\
&= \underbrace{a_T^{-2}k_T^{-1}\sum_{k=0}^{\infty} c_k c_{k+|j|}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}^2(\epsilon_{t-k}^2-1)}_{H_{1T}(j)} \\
&\quad + \underbrace{a_T^{-2}k_T^{-1}\sum_{k=0}^{\infty} c_k c_{k+|j|}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}^2(1-\rho_T^{-|j|})}_{H_{2T}(j)} \\
&\quad + \underbrace{a_T^{-2}k_T^{-1}\sum_{k=0}^{\infty}\sum_{i=0}^{\infty} c_k c_i \sum_{t=1}^{T-|j|}\mathbf{1}(i\neq k+|j|)\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}\sigma_{t+|j|-i}\epsilon_{t-k}\epsilon_{t+|j|-i}}_{H_{3T}(j)}
\end{aligned} \tag{B.10}
$$

To that end, we establish some bounds for $H_{1T}(j)$, $H_{2T}(j)$ and $H_{3T}(j)$. Regarding $H_{1T}(j)$,

let $\zeta_{T,j} = a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}^2(\epsilon_{t-k}^2-1)$, we have

$$\begin{aligned}
\mathrm{E}[\zeta_T^2] &= \mathrm{E}\left(a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}^2(\epsilon_{t-k}^2-1)\right)^2 \\
&= k_T^{-2}\sum_{t=1}^{T-|j|}\rho_T^{-4(T-t)+2|j|-4}\left\{a_T^{-4}\mathrm{E}\sigma_{t-k}^4\right\}\left\{\mathrm{E}(\epsilon_{t-k}^2-1)^2\right\} \\
&\quad +2k_T^{-2}\sum_{s<t,s,t=1}^{T-|j|}\rho_T^{-4T+2(t+s)+2|j|-4}\left\{\mathrm{E}(a_T^{-4}\sigma_{t-k}^2\sigma_{s-k}^2)\right\}\underbrace{\left\{\mathrm{E}(\epsilon_{s-k}^2-1)\right\}}_{=0}\underbrace{\left\{\mathrm{E}(\epsilon_{t-k}^2-1)\right\}}_{=0} \\
&= k_T^{-2}\sum_{t=1}^{T-|j|}\rho_T^{-4(T-t)+2|j|-4}\left\{a_T^{-4}\mathrm{E}\sigma_{t-k}^4\right\}\left\{\mathrm{E}\epsilon_{t-k}^4-2\mathrm{E}\epsilon_{t-k}^2+1\right\} \\
&= k_T^{-2}\frac{\rho_T^{-2|j|}-\rho_T^{-4T+2|j|}}{\rho_T^4-1}\left\{a_T^{-4}\mathrm{E}\sigma_{t-k}^4\right\}\left\{\mathrm{E}\epsilon_{t-k}^4-1\right\} \\
&= C_1k_T^{-1}\rho_T^{-2|j|} \le C_1k_T^{-1}, \text{ uniformly in } j \qquad \text{(B.11)}
\end{aligned}$$

where C_1 is a finite positive constant which does not depend on j and T. Then (B.11) implies $\mathrm{E}|\zeta_{T,j}| \le (\mathrm{E}[\zeta_{T,j}^2])^{1/2} \le O(k_T^{-1/2})$, uniformly in j. Thus, $\mathrm{E}\Big|H_{1T}(j)\Big| = \mathrm{E}\Big|\sum_{k=0}^{\infty}c_kc_{k+|j|}\zeta_{T,j}\Big| \le \sum_{k=0}^{\infty}\Big|c_kc_{k+|j|}\Big|\mathrm{E}\Big|\zeta_{T,j}\Big| = k_T^{-1/2}C_1^{1/2}\sum_{k=0}^{\infty}\Big|c_kc_{k+|j|}\Big| = O(k_T^{-1/2})$. Regarding $H_{2T}(j)$,

$$\begin{aligned}
\mathrm{E}\Big|H_{2T}(j)\Big| &= \mathrm{E}\Big|a_T^{-2}k_T^{-1}\sum_{k=0}^{\infty}c_kc_{k+|j|}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}^2(1-\rho_T^{-|j|})\Big| \\
&\le \{\sup_t\mathrm{E}(a_T^{-1}\sigma_{t-k})^4\}^{1/2}\times\sum_{k=0}^{\infty}|c_kc_{k+|j|}|\times k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}(1-\rho_T^{-|j|}) \\
&= C_2\sum_{k=0}^{\infty}|c_kc_{k+|j|}|(\rho_T^{-|j|}-\rho_T^{-2|j|}) \qquad \text{(B.12)}
\end{aligned}$$

where C_2 is a finite positive constant which does not depend on j and T. Regarding $H_{3T}(j)$,

$$\begin{aligned}
&\mathrm{E}\Big|H_{3T}(j)\Big| = \mathrm{E}\Big|a_T^{-2}k_T^{-1}\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}c_kc_i\sum_{t=1}^{T-|j|}\mathbf{1}(i\neq k+|j|)\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}\sigma_{t+|j|-i}\epsilon_{t-k}\epsilon_{t+|j|-i}\Big| \\
\leq\ & \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}|c_kc_i|\left(\mathrm{E}\left[\mathbf{1}(i\neq k+|j|)\sum_{t=1}^{T-|j|}a_T^{-2}k_T^{-1}\rho_T^{-2(T-t)+|j|-2}\sigma_{t-k}\sigma_{t+|j|-i}\epsilon_{t-k}\epsilon_{t+|j|-i}\right]^2\right)^{1/2} \\
=\ & \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}|c_kc_i|\mathbf{1}(i\neq k+|j|)\Bigg(\mathrm{E}\left[\sum_{t=1}^{T-|j|}a_T^{-4}k_T^{-2}\rho_T^{-4(T-t)+2|j|-4}\sigma_{t-k}^2\sigma_{t+|j|-i}^2\epsilon_{t-k}^2\epsilon_{t+|j|-i}^2\right] \\
&+2\sum_{s<t,s,t=1}^{T-|j|}a_T^{-4}k_T^{-2}\rho_T^{-4T+2(t+s)+2|j|-4}\{\mathrm{E}(\sigma_{t-k}\sigma_{s-k}\sigma_{t+|j|-i}\sigma_{s+|j|-i})\}\underbrace{\{\mathrm{E}(\epsilon_{t-k}\epsilon_{s-k}\epsilon_{t+|j|-i}\epsilon_{s+|j|-i})\}}_{=0}\Bigg)^{1/2} \\
\leq\ & \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}|c_kc_i|\mathbf{1}(i\neq k+|j|)\Bigg(k_T^{-2}\sum_{t=1}^{T-|j|}\rho_T^{-4(T-t)+2|j|-4} \\
&\times\{\sup_t\mathrm{E}(a_T^{-1}\sigma_{t-k})^4\}^{1/2}\{\sup_t\mathrm{E}(a_T^{-1}\sigma_{t+|j|-i})^4\}^{1/2}\{\sup_t\mathrm{E}\epsilon_{t-k}^4\}^{1/2}\{\sup_t\mathrm{E}\epsilon_{t+|j|-i}^4\}^{1/2}\Bigg)^{1/2} \\
=\ & k_T^{-1/2}C_3\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}|c_kc_i|\mathbf{1}(i\neq k+|j|)\rho_T^{-|j|}\leq k_T^{-1/2}C_3(\sum_{k=0}^{\infty}|c_k|)^2, \text{ uniformly in } j \qquad \text{(B.13)}
\end{aligned}$$

where C_3 is a finite positive constant which does not depend on j and T. Now combining the above three bounds, we claim

$$\begin{aligned}
&\mathrm{E}\left|\sum_{j=-(T-1)}^{T-1}w(j/b_T)[\Phi_T(j)-\widetilde{\Phi}_T(j)]\right| \\
\leq\ & \mathrm{E}\left|\sum_{j=-(T-1)}^{T-1}w(j/b_T)H_{1T}(j)\right|+\mathrm{E}\left|\sum_{j=-(T-1)}^{T-1}w(j/b_T)H_{2T}(j)\right|+\mathrm{E}\left|\sum_{j=-(T-1)}^{T-1}w(j/b_T)H_{3T}(j)\right| \\
=\ & o(1)+o(1)+o(1)=o(1) \qquad \text{(B.14)}
\end{aligned}$$

Define $H_{iT}=\sum_{j=-(T-1)}^{T-1}w(j/b_T)H_{iT}(j)$, $i=1,2,3$. The claim (B.14) holds if $\mathrm{E}|H_{iT}|=o(1)$, $i=1,2,3$. We consider each of these terms in turn.

For $\mathrm{E}|H_{1T}|$,

$$\begin{aligned}
&\mathrm{E}\left|\sum_{j=-(T-1)}^{T-1}w(j/b_T)H_{1T}(j)\right|\leq k_T^{-1/2}\sup_x|w(x)|C_1^{1/2}\sum_{j=-(T-1)}^{T-1}\sum_{k=0}^{\infty}|c_kc_{k+|j|}| \\
\leq\ & k_T^{-1/2}\times 1\times C_1^{1/2}\sum_{k=0}^{\infty}|c_k|\sum_{j=-(T-1)}^{T-1}|c_{k+|j|}|\leq k_T^{-1/2}C_1^{1/2}(\sum_{k=0}^{\infty}|c_k|)^2=O(k_T^{-1/2}) \qquad \text{(B.15)}
\end{aligned}$$

For $\mathrm{E}|H_{2T}|$,

$$\begin{aligned}
&\mathrm{E}\left|\sum_{j=-(T-1)}^{T-1} w(j/b_T)H_{2T}(j)\right| \leq C_2 \sum_{j=-(T-1)}^{T-1} |w(j/b_T)| \sum_{k=0}^{\infty} |c_k c_{k+|j|}|(\rho_T^{-|j|} - \rho_T^{-2|j|}) \\
= \; & 2C_2 \left(\sum_{j=0}^{b_T} |w(j/b_T)| \sum_{k=0}^{\infty} |c_k c_{k+j}|(\rho_T^{-j} - \rho_T^{-2j}) + \sum_{j=b_T+1}^{T-1} |w(j/b_T)| \sum_{k=0}^{\infty} |c_k c_{k+j}| \underbrace{(\rho_T^{-j} - \rho_T^{-2j})}_{\leq 1} \right) \\
\leq \; & 2C_2 \left(\mathrm{sup}_x|w(x)|(\sum_{k=0}^{\infty} |c_k|)^2 \sum_{j=0}^{b_T} (\rho_T^{-j} - \rho_T^{-2j}) + \mathrm{sup}_x|w(x)| \sum_{j=b_T+1}^{T-1} \sum_{k=0}^{\infty} |c_k c_{k+j}| \right) \\
\leq \; & 2C_2 \left(\mathrm{sup}_x|w(x)|(\sum_{k=0}^{\infty} |c_k|)^2 \times O(k_T^{-1} b_T^2) + \mathrm{sup}_x|w(x)| \sum_{k=0}^{\infty} |c_k| \sum_{j=b_T+1}^{T-1} |c_{k+j}| \right) \\
\leq \; & O(k_T^{-1} b_T^2) + \mathrm{sup}_x|w(x)| \sum_{k=0}^{\infty} |c_k| \sum_{j=b_T+1}^{T-1} |c_j| = o(1) + o(1) = o(1) \qquad \text{(B.16)}
\end{aligned}$$

which uses the fact that, by binomial expansion

$$\begin{aligned}
&\sum_{j=0}^{b_T} (\rho_T^{-j} - \rho_T^{-2j}) = \sum_{j=0}^{b_T} \rho_T^{-j} - \sum_{j=0}^{b_T} \rho_T^{-2j} = \frac{1 - \rho_T^{-b_T-1}}{1 - \rho_T^{-1}} - \frac{1 - \rho_T^{-2b_T-2}}{1 - \rho_T^{-2}} \\
= \; & \frac{\rho_T + \rho_T^{-2b_T} - \rho_T^{-b_T+1} - \rho_T^{-b_T}}{\rho_T^2 - 1} = \frac{[\rho_T^{2b_T+1} + 1] - [\rho_T^{b_T}(1 + \rho_T)]}{\rho_T^{2b_T}(\rho_T^2 - 1)} \\
= \; & \frac{[2 + c(2b_T + 1)k_T^{-1} + O(b_T^2 k_T^{-2})] - [(1 + cb_T k_T^{-1} + O(b_T^2 k_T^{-2}))(2 + ck_T^{-1})]}{O(ck_T^{-1})} \\
= \; & \frac{O(b_T^2 k_T^{-2})}{O(ck_T^{-1})} = O(k_T^{-1} b_T^2) \qquad \text{(B.17)}
\end{aligned}$$

and $\sum_{j=b_T+1}^{T-1} |c_j| = o(b_T^{-1})$, (see Chang and Park, 2002, pp. 434).

For $\mathrm{E}|H_{3T}|$,

$$\begin{aligned}
&\mathrm{E}\left|\sum_{j=-(T-1)}^{T-1} w(j/b_T)H_{3T}(j)\right| \leq \sum_{j=-(T-1)}^{T-1} |w(j/b_T)| k_T^{-1/2} C_3 (\sum_{k=0}^{\infty} |c_k|)^2 \\
= \; & \left(k_T^{-1/2} \sum_{j=-(T-1)}^{T-1} |w(j/b_T)| \right) \times C_3 (\sum_{k=0}^{\infty} |c_k|)^2 = o(1) \qquad \text{(B.18)}
\end{aligned}$$

since $k_T^{-1/2} \sum_{j=-(T-1)}^{T-1} |w(j/b_T)| = \left(k_T^{-1/2} b_T\right) \times \left(b_T^{-1} \sum_{j=-(T-1)}^{T-1} |w(j/b_T)|\right) = O(k_T^{-1/2} b_T) = o(1)$ using the fact that $b_T^{-1} \sum_{j=-(T-1)}^{T-1} |w(j/b_T)| \to \int_{-\infty}^{+\infty} |w(x)| dx < +\infty$. Therefore, to

prove the lemma, it suffices to show that $\sum_{j=-(T-1)}^{T-1} w(j/b_T)\widetilde{\Phi}_T(j) \xrightarrow{w} V_x$. We can write

$$\begin{aligned}
&\sum_{j=-(T-1)}^{T-1} w(j/b_T)\widetilde{\Phi}_T(j) = \sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}\sigma_{t-k}^2 \\
= &\sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}\sigma_t^2 \qquad \text{(B.19)} \\
&+\sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}(\sigma_{t-k}^2-\sigma_t^2) := G_{1T}+G_{2T}
\end{aligned}$$

The proof is then completed if we prove $G_{1T} \xrightarrow{w} V_x$, and $G_{2T} = o_p(1)$. Regarding G_{1T}, we have

$$\begin{aligned}
&\sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}\sigma_t^2 \\
= &\sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=1}^{T}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}\sigma_t^2 \\
&-\sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=T-|j|+1}^{T}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}\sigma_t^2 \qquad \text{(B.20)}
\end{aligned}$$

By Lemma 6 of Jansson (2002), i.e., $\sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{k=0}^{\infty}c_k c_{k+|j|} \to C^2(1)$, we have the first term in (B.20)

$$\begin{aligned}
&\sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=1}^{T}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}\sigma_t^2 \\
= &\left(\sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{k=0}^{\infty}c_k c_{k+|j|}\right)\left(a_T^{-2}k_T^{-1}\sum_{t=1}^{T}\rho_T^{-2(T-t)-2}\sigma_t^2\right) \\
\xrightarrow{w} &\ C^2(1)\lim_{T\to\infty}\int_0^1 k_T^{-1}Te^{-2k_T^{-1}Tc(1-s)}g^2(s)ds = V_x \qquad \text{(B.21)}
\end{aligned}$$

The second term in (B.20) converges in probability to zero since

$$\mathrm{E}\left|\sum_{j=-(T-1)}^{T-1} w(j/b_T)a_T^{-2}k_T^{-1}\sum_{t=T-|j|+1}^{T}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}c_k c_{k+|j|}\sigma_t^2\right| \tag{B.22}$$

$$\leq k_T^{-1}\sup_x|w(x)|\{\sup_t \mathrm{E}(a_T^{-1}\sigma_t)^4\}^{1/2}\sum_{j=-(T-1)}^{T-1}\sum_{t=T-|j|+1}^{T}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}|c_k c_{k+|j|}|$$

$$\leq k_T^{-1}K_2^{2/(4+\kappa)}\sum_{j=-(T-1)}^{T-1}|j|\sum_{k=0}^{\infty}|c_k c_{k+|j|}| \leq k_T^{-1}K_2^{2/(4+\kappa)}\sum_{k=0}^{\infty}|c_k|\sum_{j=-(T-1)}^{T-1}|j||c_{k+|j|}| = O(k_T^{-1})$$

where the second inequality in (B.22) uses the fact that $\rho_T^{-2(T-t)-2} \leq 1$ for $t = T-|j|+1,\ldots,T$, and the last equality in (B.22) holds since by Assumption 2(a), $\sum_{j=0}^{\infty} j|c_j| < \infty$, we have for any $k \geq 0$, $\sum_{j=0}^{\infty} j|c_{k+j}| \leq \sum_{j=0}^{\infty}(k+j)|c_{k+j}| \leq \sum_{j=0}^{\infty} j|c_j| < \infty$.

Regarding G_{2T}, let $m_T = K_3 T^{\omega}$, $0 < \omega < 1$ and K_3 is a constant which does not depend on T. It follows

$$
\begin{aligned}
& \mathrm{E}\left(\sum_{j=-(T-1)}^{T-1}|w(j/b_T)|a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\sum_{k=0}^{\infty}|c_kc_{k+|j|}|\left|\sigma_{t-k}^2-\sigma_t^2\right|\right) \\
= \quad & \mathrm{E}\left(\sum_{j=-(T-1)}^{T-1}|w(j/b_T)|\sum_{k=0}^{\infty}|c_kc_{k+|j|}|a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\left|\sigma_{t-k}^2-\sigma_t^2\right|\right) \\
= \quad & \mathrm{E}\left(\sum_{j=-(T-1)}^{T-1}|w(j/b_T)|\sum_{k=0}^{m_T}|c_kc_{k+|j|}|a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\left|\sigma_{t-k}^2-\sigma_t^2\right|\right) \\
& +\mathrm{E}\left(\sum_{j=-(T-1)}^{T-1}|w(j/b_T)|\sum_{k=m_T+1}^{\infty}|c_kc_{k+|j|}|a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\left|\sigma_{t-k}^2-\sigma_t^2\right|\right) \\
\leq \quad & \sum_{j=-(T-1)}^{T-1}|w(j/b_T)|\sum_{k=0}^{m_T}|c_kc_{k+|j|}|\mathrm{E}\left(\sup_{j<m_T}a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\left|\sigma_{t-k}^2-\sigma_t^2\right|\right) \\
& +\left(\sum_{j=-(T-1)}^{T-1}|w(j/b_T)|\sum_{k=m_T+1}^{\infty}|c_kc_{k+|j|}|k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\right)\times 2\{\sup_t\mathrm{E}(a_T^{-1}\sigma_t)^4\}^{1/2} \\
\leq \quad & \sup_x|w(x)|(\sum_{k=0}^{\infty}|c_k|)^2\mathrm{E}\left(\sup_{j<m_T,k\leq m_T}a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\left|\sigma_{t-k}^2-\sigma_t^2\right|\right) \\
& +2\{\sup_t\mathrm{E}(a_T^{-1}\sigma_t)^4\}^{1/2}\sup_x|w(x)|\sum_{k=m_T+1}^{\infty}|c_k|\sum_{j=-(T-1)}^{T-1}|c_{k+|j|}|\times C_4 \qquad \text{(B.23)} \\
= \quad & o(1)+o(1)=o(1)
\end{aligned}
$$

since for the first term in (B.23), following Lemma A.1 in Cavaliere and Taylor (2009), $\sup_{j<m_T,k\leq m_T}a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\left|\sigma_{t-k}^2-\sigma_t^2\right| \leq k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2}\times\sup_{t\leq T-|j|,j<m_T,k\leq m_T}a_T^{-2}|\sigma_{t-k}^2-\sigma_t^2\Big| = k_T^{-1}\times O(k_T)\times o_p(1)=o_p(1)$.[1] For the second term in (B.23), it follows $k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)-2} = C_4\times\rho_T^{-2|j|}\leq C_4$, where C_4 is a finite constant independent of T and j, and $\sum_{k=m_T+1}^{\infty}|c_k|\sum_{j=-(T-1)}^{T-1}|c_{k+|j|}|\leq$

[1] This directly follows from the fact that if $\sigma^2(\cdot)$ has continuous sample paths almost surely, then $\sup_{t\leq T-|j|,j<m_T,k\leq m_T}a_T^{-2}|\sigma_{t-k}^2-\sigma_t^2\Big| = o_p(1)$; if $\sigma^2(\cdot)$ does not have continuous sample paths almost surely, a similar proof of Lemma A.1 in Cavaliere and Taylor (2009) can be given to establish this result.

$(\sum_{k=m_T+1}^{\infty}|c_k|)(\sum_{k=0}^{\infty}|c_k|) \to 0$ as $T \to \infty$. ▲

Proof of Theorem 4: The proof of this theorem builds extensively on Lemmas A.3-A.4. By the definition of $\hat{\Omega}$, we have

$$T(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\hat{\Omega} = \sum_{j=-(T-1)}^{T-1} w(j/b_T)(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}T\hat{\Gamma}(j) \tag{B.24}$$

We first work out the details of $(\mu_T k_T^{3/2}\rho_T^T)^{-2}T\hat{\Gamma}(j)$, using the fact that $\hat{u}_t = \dot{u}_t - \dot{y}_{t-1}(\hat{\rho}_T - \rho_T)$. We have

$$\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}T\hat{\Gamma}(j) \\
= \; & (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\dot{y}_{t-1}\hat{u}_t\dot{y}_{t+|j|-1}\hat{u}_{t+|j|} \\
= \; & (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\Big[\dot{y}_{t-1}\dot{u}_t\dot{y}_{t+|j|-1}\dot{u}_{t+|j|} - (\hat{\rho}_T-\rho_T)\dot{y}_{t-1}^2\dot{y}_{t+|j|-1}\dot{u}_{t+|j|} \\
& -(\hat{\rho}_T-\rho_T)\dot{y}_{t-1}\dot{y}_{t+|j|-1}^2\dot{u}_t + (\hat{\rho}_T-\rho_T)^2\dot{y}_{t-1}^2\dot{y}_{t+|j|-1}^2\Big] \\
:= \; & A_j - B_j - C_j + D_j
\end{aligned} \tag{B.25}$$

Specifically, by Lemma A.3 and $\bar{y}_{-1} = O_p(T^{-1}a_T\mu_T k_T^2\rho_T^T)$ which is implied by Theorem 2(b), we have
B_j:

$$\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}(\hat{\rho}_T-\rho_T)\dot{y}_{t-1}^2\dot{y}_{t+|j|-1}\dot{u}_{t+|j|} \qquad (B.26) \\
= \; & (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}(\hat{\rho}_T-\rho_T)(y_{t-1}-\bar{y}_{-1})^2(y_{t+|j|-1}-\bar{y}_{-1})u_{t+|j|} \\
& -(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}(\hat{\rho}_T-\rho_T)(y_{t-1}-\bar{y}_{-1})^2(y_{t+|j|-1}-\bar{y}_{-1})\bar{u} := B_{1j} - B_{2j}
\end{aligned}$$

B_{1j}:

$$
\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}(\hat\rho_T-\rho_T)\dot y_{t-1}^2\dot y_{t+|j|-1}u_{t+|j|} \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}(\hat\rho_T-\rho_T)\sum_{t=1}^{T-|j|}(y_{t-1}-\bar y_{-1})^2(y_{t+|j|-1}-\bar y_{-1})u_{t+|j|} \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}(\hat\rho_T-\rho_T)\sum_{t=1}^{T-|j|}\Big[y_{t-1}^2y_{t+|j|-1}u_{t+|j|}-y_{t-1}^2u_{t+|j|}\bar y_{-1}-2y_{t-1}y_{t+|j|-1}u_{t+|j|}\bar y_{-1} \\
& +2y_{t-1}u_{t+|j|}\bar y_{-1}^2+y_{t+|j|-1}u_{t+|j|}\bar y_{-1}^2-u_{t+|j|}\bar y_{-1}^3\Big] \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\times O_p((\mu_T k_T^{3/2}\rho_T^T)^{-1})\times\Big[O_p(a_T^4\mu_T^3k_T^{7/2}\rho_T^{3T})+O_p(a_T^3\mu_T^2k_T^{5/2}\rho_T^{2T})O_p(T^{-1}a_T\mu_Tk_T^2\rho_T^T) \\
& +O_p(a_T^3\mu_T^2k_T^{5/2}\rho_T^{2T})O_p(T^{-1}a_T\mu_Tk_T^2\rho_T^T)+O_p(a_T^2\mu_Tk_T^{3/2}\rho_T^T)O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^2) \\
& +O_p(a_T^2\mu_Tk_T^{3/2}\rho_T^T)O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^2)+O_p(a_TT^{1/2})O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^3)\Big] \\
= \;& O_p(k_T^{-1})+O_p(T^{-1})+O_p(T^{-2}k_T)+O_p(T^{-5/2}k_T^{3/2})=O_p(k_T^{-1})
\end{aligned}
\tag{B.27}
$$

B_{2j}:

$$
\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}(\hat\rho_T-\rho_T)\dot y_{t-1}^2\dot y_{t+|j|-1}\bar u \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}(\hat\rho_T-\rho_T)\bar u\sum_{t=1}^{T-|j|}(y_{t-1}-\bar y_{-1})^2(y_{t+|j|-1}-\bar y_{-1}) \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}(\hat\rho_T-\rho_T)\bar u\sum_{t=1}^{T-|j|}\Big[y_{t-1}^2y_{t+|j|-1}-y_{t-1}^2\bar y_{-1}-2y_{t-1}y_{t+|j|-1}\bar y_{-1} \\
& +2y_{t-1}\bar y_{-1}^2+y_{t+|j|-1}\bar y_{-1}^2-\bar y_{-1}^3\Big] \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\times O_p((\mu_T k_T^{3/2}\rho_T^T)^{-1})\times O_p(a_TT^{-1/2})\times\Big[O_p(a_T^3\mu_T^3k_T^4\rho_T^{3T}) \\
& +O_p(a_T^2\mu_T^2k_T^3\rho_T^{2T})O_p(T^{-1}a_T\mu_Tk_T^2\rho_T^T)+O_p(a_T^2\mu_T^2k_T^3\rho_T^{2T})O_p(T^{-1}a_T\mu_Tk_T^2\rho_T^T) \\
& +O_p(a_T\mu_Tk_T^2\rho_T^T)O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^2)+O_p(a_T\mu_Tk_T^2\rho_T^T)O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^2) \\
& +O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^3)\Big] \\
= \;& O_p(T^{-1/2}k_T^{-1/2})+O_p(T^{-3/2}k_T^{1/2})+O_p(T^{-5/2}k_T^{3/2})=o_p(k_T^{-1})
\end{aligned}
\tag{B.28}
$$

which gives $B_j=O_p(k_T^{-1})$. $C_j=O_p(k_T^{-1})$, follows the same proof as B_j.

D_j:

$$
\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}(\hat{\rho}_T-\rho_T)^2\dot{y}_{t-1}^2\dot{y}_{t+|j|-1}^2 \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}(\hat{\rho}_T-\rho_T)^2\Big[y_{t-1}^2y_{t+|j|-1}^2-2y_{t-1}^2y_{t+|j|-1}\bar{y}_{-1}+y_{t-1}^2\bar{y}_{-1}^2-2y_{t-1}y_{t+|j|-1}^2\bar{y}_{-1} \\
& +4y_{t-1}y_{t+|j|-1}\bar{y}_{-1}^2-2y_{t-1}\bar{y}_{-1}^3+y_{t+|j|-1}^2\bar{y}_{-1}^2-2y_{t+|j|-1}\bar{y}_{-1}^3+\bar{y}_{-1}^4\Big] \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\times O_p((\mu_T k_T^{3/2}\rho_T^T)^{-2})\times\Big[O_p(a_T^4\mu_T^4k_T^5\rho_T^{4T})+O_p(a_T^3\mu_T^3k_T^4\rho_T^{3T})O_p(T^{-1}a_T\mu_Tk_T^2\rho_T^T) \\
& +O_p(a_T^2\mu_T^2k_T^3\rho_T^{2T})O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^2)+O_p(a_T^3\mu_T^3k_T^4\rho_T^{3T})O_p(T^{-1}a_T\mu_Tk_T^2\rho_T^T) \\
& +O_p(a_T^2\mu_T^2k_T^3\rho_T^{2T})O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^2)+O_p(a_T\mu_Tk_T^2\rho_T^T)O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^3) \\
& +O_p(a_T^2\mu_T^2k_T^3\rho_T^{2T})O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^2)+O_p(a_T\mu_Tk_T^2\rho_T^T)O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^3) \\
& +O_p((T^{-1}a_T\mu_Tk_T^2\rho_T^T)^4)\Big] \\
= \;& O_p(k_T^{-1})+O_p(T^{-1})+O_p(T^{-2}k_T)+O_p(T^{-3}k_T^2)+O_p(T^{-4}k_T^2)=O_p(k_T^{-1})
\end{aligned}
\tag{B.29}
$$

Thus B_j, C_j and D_j are all $O_p(k_T^{-1})$. We next calculate A_j.

A_j:

$$
\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\dot{y}_{t-1}\dot{u}_t\dot{y}_{t+|j|-1}\dot{u}_{t+|j|} \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\dot{y}_{t-1}(u_t-\bar{u})\dot{y}_{t+|j|-1}(u_{t+|j|}-\bar{u}) \\
= \;& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\dot{y}_{t-1}u_t\dot{y}_{t+|j|-1}u_{t+|j|}+(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\bar{u}^2\sum_{t=1}^{T-|j|}\dot{y}_{t-1}\dot{y}_{t+|j|-1} \\
& -(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\bar{u}\sum_{t=1}^{T-|j|}\Big[\dot{y}_{t-1}\dot{y}_{t+|j|-1}u_t+\dot{y}_{t-1}\dot{y}_{t+|j|-1}u_{t+|j|}\Big] \\
:= \;& A_{1j}+A_{2j}-A_{3j}
\end{aligned}
\tag{B.30}
$$

In what follows, we derive the orders for A_{2j}, A_{3j}, and claim they are asymptotically negli-

gible. Similar to the foregoing analysis, we have

$$A_{2j} = (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\bar{u}^2\sum_{t=1}^{T-|j|}\dot{y}_{t-1}\dot{y}_{t+|j|-1} \tag{B.31}$$
$$= (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\times O_p(a_T^2T^{-1})\times O_p(a_T^2\mu_T^2k_T^3\rho_T^{2T}) + s.o. = O_p(T^{-1}) = o_p(k_T^{-1}),$$
$$A_{3j} = (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\bar{u}\sum_{t=1}^{T-|j|}\left[\dot{y}_{t-1}\dot{y}_{t+|j|-1}u_t + \dot{y}_{t-1}\dot{y}_{t+|j|-1}u_{t+|j|}\right] \tag{B.32}$$
$$= (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\times O_p(a_TT^{-1/2})\times O_p(a_T^3\mu_T^2k_T^{5/2}\rho_T^{2T}) + s.o. = O_p(T^{-1/2}k_T^{-1/2}) = o_p(k_T^{-1})$$

Then we analyze A_{1j}:

$$(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\dot{y}_{t-1}u_t\dot{y}_{t+|j|-1}u_{t+|j|}$$
$$= (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}y_{t-1}u_ty_{t+|j|-1}u_{t+|j|} + (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}u_tu_{t+|j|}\bar{y}_{-1}^2$$
$$-(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}y_{t-1}u_tu_{t+|j|}\bar{y}_{-1} - (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}u_ty_{t+|j|-1}u_{t+|j|}\bar{y}_{-1}$$
$$:= A_{1j,1} + A_{1j,2} - A_{1j,3} - A_{1j,4} \tag{B.33}$$

where $A_{1j,2}$:

$$(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}u_tu_{t+|j|}\bar{y}_{-1}^2 = T^{-1}k_T(T^{-1}a_T\mu_Tk_T^2\rho_T^T)^{-2}\bar{y}_{-1}^2\times a_T^{-2}T^{-1}\sum_{t=1}^{T-|j|}u_tu_{t+|j|}$$
$$= O(T^{-1}k_T)O_p(1)\times a_T^{-2}T^{-1}\sum_{t=1}^{T-|j|}u_tu_{t+|j|} = O_p(T^{-1}k_T)\times a_T^{-2}T^{-1}\sum_{t=1}^{T-|j|}u_tu_{t+|j|} \tag{B.34}$$

$A_{1j,3}$:

$$(a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}y_{t-1}u_tu_{t+|j|}\bar{y}_{-1}$$
$$= T^{-1/2}k_T^{1/2}\underbrace{(T^{-1}a_T\mu_Tk_T^2\rho_T^T)^{-1}\bar{y}_{-1}}_{=O_p(1)}\times(T^{1/2}a_T^3\mu_Tk_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T-|j|}y_{t-1}u_tu_{t+|j|}\bar{y}_{-1}$$
$$= O_p(T^{-1/2}k_T^{1/2})\times(T^{1/2}a_T^3\mu_Tk_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T-|j|}y_{t-1}u_tu_{t+|j|} \tag{B.35}$$

$A_{1j,4} = O_p(T^{-1/2}k_T^{1/2}) \times (T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T-|j|} u_t y_{t+|j|-1}u_{t+|j|}$ follows the same way as $A_{1j,3}$.

To calculate $A_{1j,1}$, let $\dot{Y}_T := (\mu_T k_T^{1/2})^{-1}Y_T + 1/(a_T c)$. It then follows $\dot{Y}_T \xrightarrow{w} Y/\nu + 1/(\psi c)\mathbf{1}(\gamma = 0)$. By Lemma A.2 and the assumption $a_T^{-1}y_0 = o_p(k_T^{1/2})$, with some algebra, we have
$A_{1j,1}$:

$$
\begin{aligned}
& (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|} y_{t-1}u_t y_{t+|j|-1}u_{t+|j|} \\
= \; & (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\Big[\big\{y_0\rho_T^{t-1} + \sum_{i_1=1}^{t-1}\rho_T^{t-1-i_1}u_{i_1} + \mu_T(\rho_T^{t-1}-1)k_T/c\big\}u_t \times \\
& \big\{y_0\rho_T^{t+|j|-1} + \sum_{i_2=1}^{t+|j|-1}\rho_T^{t+|j|-1-i_2}u_{i_2} + \mu_T(\rho_T^{t+|j|-1}-1)k_T/c\big\}u_{t+|j|}\Big] \\
= \; & (a_T^2\mu_T k_T^{3/2}\rho_T^T)^{-2}\sum_{t=1}^{T-|j|}\Big[\big\{\sum_{i_1=1}^{T}\rho_T^{t-1-i_1}u_{i_1} + \mu_T\rho_T^{t-1}k_T/c\big\}u_t \times \\
& \big\{\sum_{i_2=1}^{T}\rho_T^{t+|j|-1-i_2}u_{i_2} + \mu_T\rho_T^{t+|j|-1}k_T/c\big\}u_{t+|j|}\Big] + o_p(k_T^{-1}) \\
= \; & \left(a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}u_t u_{t+|j|}\right)\left((\mu_T k_T^{1/2})^{-1}a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\rho_T^{-t}u_t + a_T^{-1}c^{-1}\right)^2 + o_p(k_T^{-1}) \\
= \; & \Phi_T(j)\dot{Y}_T^2 + o_p(k_T^{-1}) \qquad \text{(B.36)}
\end{aligned}
$$

Now, we combine A_j-D_j and use Lemma 1 to obtain

$$\begin{aligned}
&\sum_{j=-(T-1)}^{T-1} w(j/b_T)(A_j - B_j - C_j + D_j) \qquad \text{(B.37)}\\
=\ &\sum_{j=-(T-1)}^{T-1} w(j/b_T)(A_{1j,1} + A_{1j,2} - A_{1j,3} - A_{1j,4} \underbrace{+A_{2j} - A_{3j}}_{o_p(k_T^{-1})} \underbrace{-B_j - C_j + D_j}_{O_p(k_T^{-1})})\\
=\ &\sum_{j=-(T-1)}^{T-1} w(j/b_T)(A_{1j,1} + A_{1j,2} - A_{1j,3} - A_{1j,4}) + o_p(1)\\
=\ &\sum_{j=-(T-1)}^{T-1} w(j/b_T)\big\{ \underbrace{\Phi_T(j)\dot{Y}_T^2 + o_p(k_T^{-1})}_{A_{1j,1}} \big\} + \underbrace{\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,2}}_{\xrightarrow{p}0} - \underbrace{\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,3}}_{\xrightarrow{p}0}\\
&- \underbrace{\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,4}}_{\xrightarrow{p}0} + o_p(1) = \dot{Y}_T^2 \sum_{j=-(T-1)}^{T-1} w(j/b_T)\Phi_T(j) + o_p(1) \xrightarrow{w} V_x(\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0))^2
\end{aligned}$$

Note in the above $\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,2} \xrightarrow{p} 0$ holds due to

$$\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,2} = O_p(T^{-1}k_T) \times \underbrace{a_T^{-2}T^{-1} \sum_{j=-(T-1)}^{T-1} w(j/b_T) \sum_{t=1}^{T-|j|} u_t u_{t+|j|}}_{\xrightarrow{w}\sigma_u^2} \xrightarrow{p} 0$$

where the result $a_T^{-2}T^{-1}\sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{t=1}^{T-|j|} u_t u_{t+|j|} \xrightarrow{w} \sigma_u^2$ can be proved analogously to Lemma 1. Next, $\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,3} \xrightarrow{p} 0$ holds due to

$$\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,3} = O_p(T^{-1/2}k_T^{1/2}) \times \underbrace{(T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1} \sum_{j=-(T-1)}^{T-1} w(j/b_T) \sum_{t=1}^{T-|j|} y_{t-1}u_t u_{t+|j|}}_{=O_p(T^{-1/2}k_T^{1/2})} \xrightarrow{p} 0$$

since

$$
\begin{aligned}
& (T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{t=1}^{T-|j|} y_{t-1}u_t u_{t+|j|} \\
= \; & (T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{t=1}^{T-|j|} y_{t-1}\sum_{k=0}^{\infty}\sum_{i=0}^{\infty} c_k c_i \sigma_{t-k}\sigma_{t+|j|-i}\epsilon_{t-k}\epsilon_{t+|j|-i} \\
= \; & (T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{t=1}^{T} y_{t-1}\sum_{k=0}^{\infty} c_k c_{k+|j|}\sigma_{t-k}^2 + o_p(1) \qquad \text{(B.38)} \\
= \; & (T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{t=1}^{T} y_{t-1}\sum_{k=0}^{\infty} c_k c_{k+|j|}\sigma_t^2 + o_p(1) \qquad \text{(B.39)} \\
& \xrightarrow{w} \sum_{j=-(T-1)}^{T-1} w(j/b_T)\sum_{k=0}^{\infty} c_k c_{k+|j|} \times (T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1}\sum_{t=1}^{T} y_{t-1}\sigma_t^2 \\
& \xrightarrow{w} C^2(1) \times (T^{1/2}a_T^3\mu_T k_T^{3/2}\rho_T^T)^{-1} O_p(a_T^3\mu_T k_T^2\rho_T^T) = O_p(T^{-1/2}k_T^{1/2}) \qquad \text{(B.40)}
\end{aligned}
$$

where $(B.38)$ and $(B.39)$ can be established in a similar manner to the proof of Lemma 1, $(B.40)$ holds because of Lemma A.6. Finally, $\sum_{j=-(T-1)}^{T-1} w(j/b_T)A_{1j,4} \xrightarrow{p} 0$ follows from the same arguments used for $A_{1j,3}$.The result of Theorem 4 follows. ▲

Proof of Theorem 5: This is proved in the main text, see (18)-(19). ▲

Proof of Theorem 6: We will prove the result for X_T^*. First we note that conditionally on $\hat{u}_t$, $\hat{u}_t^*$, $t = 1, \ldots, T$ is normally distributed over time with zero mean, thus the partial sum process is normally distributed with zero mean. In what follows, we focus on deriving

its variance. Using the fact that $\mathrm{Cov}(\eta_s, \eta_t) = \mathrm{E}(\eta_s \eta_t) = K(\frac{s-t}{l_T})$, it follows

$$\begin{aligned}
\mathrm{E}^*\left[X_T^{*2}\right] &= \mathrm{E}^*\left[a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\hat{\rho}_T^{-(T-t)-1}u_t^*\right]^2 = \mathrm{E}^*\left[a_T^{-1}k_T^{-1/2}\sum_{t=1}^{T}\hat{\rho}_T^{-(T-t)-1}\eta_t\hat{u}_t\right]^2 \\
&= a_T^{-2}k_T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\rho}_T^{-[2T-(t+s)]-2}\hat{u}_t\hat{u}_s\mathrm{E}(\eta_s\eta_t) \\
&= a_T^{-2}k_T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}K(\frac{s-t}{l_T})\hat{\rho}_T^{-[2T-(t+s)]-2}\hat{u}_t\hat{u}_s \\
&= \sum_{j=-(T-1)}^{T-1}K(\frac{j}{l_T})\Big[a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}\hat{u}_t\hat{u}_{t+|j|}\Big] \\
&= \sum_{j=-(T-1)}^{T-1}K(\frac{j}{l_T})\Big[a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}u_tu_{t+|j|}\Big] + o_p(1) \qquad \text{(B.41)} \\
&= \sum_{j=-(T-1)}^{T-1}K(\frac{j}{l_T})\Big[a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\rho_T^{-2(T-t)+|j|-2}u_tu_{t+|j|}\Big] \\
&\quad + o_p(1)\times\underbrace{\sum_{j=-(T-1)}^{T-1}K(\frac{j}{l_T})\Big[a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}u_tu_{t+|j|}\Big]}_{\xrightarrow{w}\sigma_u^2} + o_p(1) \qquad \text{(B.42)} \\
&\xrightarrow{w} V_x \qquad \text{(B.43)}
\end{aligned}$$

where the $o_p(1)$ in $(B.41)$ is not surprising as, under $\hat{u}_t = \dot{u}_t - \dot{y}_{t-1}(\hat{\rho}_T - \rho_T)$,

$$\begin{aligned}
&a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}(\hat{u}_t\hat{u}_{t+|j|} - u_tu_{t+|j|}) \\
= \; &a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}\Big[(\dot{u}_t\dot{u}_{t+|j|} - u_tu_{t+|j|}) - (\hat{\rho}_T-\rho_T)\dot{y}_{t-1}\dot{u}_{t+|j|} \\
&-(\hat{\rho}_T-\rho_T)\dot{y}_{t+|j|-1}\dot{u}_t + (\hat{\rho}_T-\rho_T)^2\dot{y}_{t-1}\dot{y}_{t+|j|-1}\Big] \\
:= \; &\widetilde{A}_j - \widetilde{B}_j - \widetilde{C}_j + \widetilde{D}_j \qquad \text{(B.44)}
\end{aligned}$$

where, using the results from Lemma A.6, we can easily derive

$\widetilde{B}_j$:

$$\begin{aligned}
& a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}(\hat{\rho}_T-\rho_T)\dot{y}_{t-1}\dot{u}_{t+|j|} \\
= \;& a_T^{-2}k_T^{-1}(\hat{\rho}_T-\rho_T)\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}\Big[y_{t-1}u_{t+|j|}+\bar{y}_{-1}\bar{u}-y_{t-1}\bar{u}-\bar{y}_{-1}u_{t+|j|}\Big] \\
= \;& a_T^{-2}k_T^{-1}\times O_p((\mu_T k_T^{3/2}\rho_T^T)^{-1})\times O_p(a_T^2\mu_T k_T^{3/2}\rho_T^T)+s.o.=O_p(k_T^{-1}) \qquad \text{(B.45)}
\end{aligned}$$

$\widetilde{C}_j = O_p(k_T^{-1})$, which follows the same proof as $\widetilde{B}_j$.

$\widetilde{D}_j$:

$$\begin{aligned}
& a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}(\hat{\rho}_T-\rho_T)^2\dot{y}_{t-1}\dot{y}_{t+|j|-1} \\
= \;& a_T^{-2}k_T^{-1}(\hat{\rho}_T-\rho_T)^2\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}\Big[y_{t-1}y_{t+|j|-1}-y_{t-1}\bar{y}_{-1}-y_{t+|j|-1}\bar{y}_{-1}+\bar{y}_{-1}^2\Big] \\
= \;& a_T^{-2}k_T^{-1}\times O_p((\mu_T k_T^{3/2}\rho_T^T)^{-2})\times O_p(a_T^2\mu_T^2 k_T^3\rho_T^{2T})+s.o.=O_p(k_T^{-1}) \qquad \text{(B.46)}
\end{aligned}$$

$\widetilde{A}_j$:

$$\begin{aligned}
& a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}(\dot{u}_t\dot{u}_{t+|j|}-u_tu_{t+|j|}) \\
= \;& a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}\bar{u}^2-a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}u_{t+|j|}\bar{u}-a_T^{-2}k_T^{-1}\sum_{t=1}^{T-|j|}\hat{\rho}_T^{-2(T-t)+|j|-2}u_t\bar{u} \\
= \;& (a_T^{-2}k_T^{-1})O_p(a_T^2T^{-1})O(k_T)+O_p(a_TT^{-1/2})O_p(a_T^{-1}k_T^{-1/2})+O_p(a_TT^{-1/2})O_p(a_T^{-1}k_T^{-1/2}) \\
= \;& O_p(T^{-1})+O_p(T^{-1/2}k_T^{-1/2})=o_p(k_T^{-1}) \qquad \text{(B.47)}
\end{aligned}$$

Therefore, the sum of the above four terms in $(B.44)$ is at most $O_p(k_T^{-1})$. Under Assumption 6, $k_T^{-1/2}l_T \to 0$ as $T \to \infty$, the equality in $(B.41)$ holds. Moreover, $(B.42)$ follows Lemma A.6(a), the final limit result $(B.43)$ follows from Lemma 1 since l_T and $K(\cdot)$ also satisfy the conditions needed in the proof of Lemma 1 (Assumption 5). ▲

Proof of Theorem 7: To prove this theorem, it suffices to show

(i) $\mu_T k_T^{3/2}\hat{\rho}_T^T(\hat{\rho}_T^*-\hat{\rho}_T) \xrightarrow{w}_p 2cX/\Big(\frac{Y}{\nu}+\frac{1}{\psi c}\mathbf{1}(\gamma=0)\Big)$;

(ii) $(\mu_T k_T^{3/2}\hat{\rho}_T^T)^2\hat{\Lambda}^* \xrightarrow{w}_p 4c^2V_x/\Big(\frac{Y}{\nu}+\frac{1}{\psi c}\mathbf{1}(\gamma=0)\Big)^2$;

First, consider (i). In light of the OLS formula for $(\hat{\rho}_T^* - \hat{\rho}_T)$, we need to work out the asymptotics for the bootstrap data - we first prove a similar result as in Theorem 2 for the bootstrap data. Denoting $1/\infty = 0$, we have the following joint convergence results:

$$
\begin{aligned}
(a') \quad & (a_T^2 \mu_T^2 k_T^3 \hat{\rho}_T^{2T})^{-1} \sum_{t=1}^{T} y_{t-1}^{*2} \xrightarrow{w}_p \frac{1}{2c}\Big[\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0)\Big]^2 \\
(b') \quad & (a_T \mu_T k_T^2 \hat{\rho}_T^{T})^{-1} \sum_{t=1}^{T} y_{t-1}^{*} \xrightarrow{w}_p \frac{1}{c}\Big[\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0)\Big] \\
(c') \quad & (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^{T})^{-1} \sum_{t=1}^{T} y_{t-1}^{*} u_t^{*} \xrightarrow{w}_p X\Big[\frac{Y}{\nu} + \frac{1}{\psi c}\mathbf{1}(\gamma = 0)\Big] \\
(d') \quad & a_T^{-1} T^{-1/2} \sum_{t=1}^{T} u_t^{*} \xrightarrow{w}_p U \sim MN(0, \sigma_u^2), \quad \sigma_u^2 = C^2(1) \int_0^1 g(s)^2 ds
\end{aligned}
\tag{B.48}
$$

where $y_t^* = \hat{\mu}_T + \hat{\rho}_T y_{t-1}^* + u_t^*$, $t = 1, \ldots, T$, $y_0^* = y_0$. Note we can define

$$
d_t^* = \hat{\rho}_T d_{t-1}^* + u_t^*, \quad d_0^* = y_0^* \tag{B.49}
$$

so that $y_t^* = d_t^* + \hat{\mu}_T(\hat{\rho}_T^t - 1)k_T/\hat{c}$. The results in $(B.48)$ are the bootstrap counterparts to those stated in Theorem 2, which build upon the following results that are entirely analogous to Lemma A.1,

(a'') $(a_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} \sum_{j=t}^{T} \hat{\rho}_T^{t-1-j} u_j^* = o_{p^*}(1)$; (b'') $(a_T k_T^{3/2} \hat{\rho}_T^{2T})^{-1} \sum_{t=1}^{T} \sum_{j=t}^{T} \hat{\rho}_T^{2(t-1)-j} u_j^* = o_{p^*}(1)$;

(c'') $(a_T^2 k_T \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} \sum_{j=t}^{T} \hat{\rho}_T^{t-1-j} u_j^* u_t^* = o_{p^*}(1)$; (d'') $(a_T k_T \hat{\rho}_T^T)^{-2} \sum_{t=1}^{T} d_{t-1}^{*2} = Y_T^{*2}/2c + o_{p^*}(1)$;

(e'') $(a_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} d_{t-1}^* = Y_T^*/c + o_{p^*}(1)$; (f'') $(a_T^2 k_T \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} d_{t-1}^* u_t^* = X_T^* Y_T^* + o_{p^*}(1)$.

The proofs of (a'')-(f'') are entirely the same as the proofs of (a)-(f) in Lemma A.1, with an additional result $\hat{c} - c = k_T(\hat{\rho}_T - \rho_T) = (\mu_T k_T^{1/2} \rho_T^T)^{-1} \mu_T k_T^{3/2} \rho_T^T (\hat{\rho}_T - \rho_T) = \rho_T^{-T}\left(\frac{2cX}{Y + \nu/\psi c} + o_p(1)\right) = O_p(\rho_T^{-T}) = o_p(1)$, so we omit them for brevity. Given (a'')-(f''), to see how the proofs for the (a')-(d') are unchanged or follow with minor modifications to that of Theorem 2, we take

(c') as an example. Using the fact that $\frac{\hat{\mu}_T k_T^{1/2}}{\mu_T k_T^{1/2}} = 1 + \nu^{-1} O_p(a_T T^{-1/2} k_T^{1/2})$,

$$
\begin{aligned}
& (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} y_{t-1}^* u_t^* = (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} (d_{t-1}^* + \hat{\mu}_T (\hat{\rho}_T^{t-1} - 1) k_T / \hat{c}) u_t^* \\
= \; & (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} d_{t-1}^* u_t^* + \frac{\hat{\mu}_T k_T^{1/2}}{\mu_T k_T^{1/2}} (a_T^2 k_T^{1/2} \hat{\rho}_T^T)^{-1} \hat{c}^{-1} \sum_{t=1}^{T} \hat{\rho}_T^{t-1} u_t^* - \frac{\hat{\mu}_T k_T^{1/2}}{\mu_T k_T^{1/2}} (a_T^2 k_T^{1/2} \hat{\rho}_T^T)^{-1} \hat{c}^{-1} \sum_{t=1}^{T} u_t^* \\
= \; & X_T^* Y_T^* / (\mu_T k_T^{1/2}) + (1 + \nu^{-1} O_p(a_T T^{-1/2} k_T^{1/2})) a_T^{-1} \hat{c}^{-1} a_T^{-1} k_T^{-1/2} \sum_{t=1}^{T} \hat{\rho}_T^{-(T-t)-1} u_t^* \\
& - (1 + \nu^{-1} O_p(a_T T^{-1/2} k_T^{1/2})) \times O_{p*}(T^{1/2} (k_T^{1/2} \hat{\rho}_T^T)^{-1} a_T^{-1}) \\
= \; & X_T^* Y_T^* / (\mu_T k_T^{1/2}) + X_T^* / (a_T c) + \nu^{-1} O_{p*}(T^{-1/2} k_T^{1/2}) + o_{p*}(1) \xrightarrow{w}_p \frac{1}{2c} [\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0)]^2 \qquad \text{(B.50)}
\end{aligned}
$$

With the above results (a')-(d') at hand, we thus have,

$$
\begin{aligned}
& (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} (y_{t-1}^* - \bar{y}_{-1}^*) u_t^* = (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \left(\sum_{t=1}^{T} y_{t-1}^* u_t^* - T^{-1} \sum_{t=1}^{T} y_{t-1}^* \sum_{t=1}^{T} u_t^* \right) \\
= \; & (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} y_{t-1}^* u_t^* - \underbrace{T^{-1/2} k_T^{1/2} (a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} y_{t-1}^* \times a_T^{-1} T^{-1/2} \sum_{t=1}^{T} u_t^*}_{O_{p*}(T^{-1/2} k_T^{1/2})} \\
& \xrightarrow{w}_p X \Big[\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0) \Big] \qquad \text{(B.51)}
\end{aligned}
$$

$$
\begin{aligned}
& a_T^{-2} (\mu_T k_T^{3/2} \hat{\rho}_T^T)^{-2} \sum_{t=1}^{T} (y_{t-1}^* - \bar{y}_{-1}^*)^2 = a_T^{-2} (\mu_T k_T^{3/2} \hat{\rho}_T^T)^{-2} \left(\sum_{t=1}^{T} y_{t-1}^{*2} - T^{-1} (\sum_{t=1}^{T} y_{t-1}^*)^2 \right) \qquad \text{(B.52)} \\
= \; & a_T^{-2} (\mu_T k_T^{3/2} \hat{\rho}_T^T)^{-2} \sum_{t=1}^{T} y_{t-1}^{*2} - \underbrace{T^{-1} k_T [(a_T \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} y_{t-1}^*]^2}_{O_{p*}(T^{-1} k_T)} \xrightarrow{w}_p \frac{1}{2c} \Big[\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0) \Big]^2
\end{aligned}
$$

which gives

$$
\mu_T k_T^{3/2} \hat{\rho}_T^T (\hat{\rho}_T^* - \hat{\rho}_T) = \frac{(a_T^2 \mu_T k_T^{3/2} \hat{\rho}_T^T)^{-1} \sum_{t=1}^{T} (y_{t-1}^* - \bar{y}_{-1}^*) u_t^*}{a_T^{-2} (\mu_T k_T^{3/2} \hat{\rho}_T^T)^{-2} \sum_{t=1}^{T} (y_{t-1}^* - \bar{y}_{-1}^*)^2} \xrightarrow{w}_p 2cX / \Big(\frac{Y}{\nu} + \frac{1}{\psi c} \mathbf{1}(\gamma = 0) \Big)
$$

The proof of (ii) is similar in spirit to the proof of (i) in terms of showing that the limit result based on the original data is transferable to the bootstrap data. We outline the main

steps of the argument. In particular, we need to show a similar result as Theorem 4 holds for the bootstrap data, i.e.,

$$T(a_T^2\mu_T k_T^{3/2}\hat{\rho}_T^T)^{-2}\hat{\Omega}^* \xrightarrow{w}_p V_x\left[\frac{Y}{\nu}+\frac{1}{\psi c}\mathbf{1}(\gamma=0)\right]^2 \tag{B.53}$$

The proof shares entirely the same logic as the proof of Theorem 4, and it builds on several preliminary results which are analogous to Lemmas A.2-A.4 and A.6 with corresponding quantities replaced by their bootstrap analogues. Next, in view of $(B.52)$,

$$T(a_T^2\mu_T^2 k_T^3\hat{\rho}_T^{2T})^{-1}Q_T^* = a_T^{-2}(\mu_T k_T^{3/2}\hat{\rho}_T^T)^{-2}\sum_{t=1}^{T}\dot{y}_{t-1}^{*2} \xrightarrow{w}_p \frac{1}{2c}\left[\frac{Y}{\nu}+\frac{1}{\psi c}\mathbf{1}(\gamma=0)\right]^2 \tag{B.54}$$

Finally, combining $(B.53)$ and $(B.54)$,

$$\begin{aligned}(\mu_T k_T^{3/2}\hat{\rho}_T^T)^2\hat{\Lambda}^* &= T^{-1}(\mu_T k_T^{3/2}\hat{\rho}_T^T)^2 Q_T^{*-2}\hat{\Omega}^* = \left[T(a_T^2\mu_T^2 k_T^3\hat{\rho}_T^{2T})^{-1}Q_T^*\right]^{-2}\left[T(a_T^2\mu_T k_T^{3/2}\hat{\rho}_T^T)^{-2}\hat{\Omega}^*\right] \\ &\xrightarrow{w}_p 4c^2 V_x/\Big(\frac{Y}{\nu}+\frac{1}{\psi c}\mathbf{1}(\gamma=0)\Big)^2 \end{aligned} \tag{B.55}$$

which gives (ii). The theorem then follows. ▲

References

[1] Cavaliere, G., Taylor, A. M. R. (2009), "Heteroskedastic time series with a unit root," *Econometric Theory* 25, 1228-1276.

[2] Chang, Y., Park, J. Y. (2002), "On the asymptotics of ADF tests for unit roots," *Econometric Reviews* 21, 431-447.

[3] Giné, E., Zinn, J. (1990), "Bootstrapping general empirical measures," *Annals of Probability* 18, 851-869.

[4] Guo, G., Sun, Y., Wang, S. (2019), "Testing for moderate explosiveness," *Econometrics Journal* 22, 73-95.

[5] Jansson, M. (2002), "Consistent covariance matrix estimation for linear processes," *Econometric Theory* 18, 1449-1459.

[6] Phillips, P. C. B., Solo, V. (1992), "Asymptotics for linear processes," *Annals of Statistics* 20, 971-1001.

Supplementary Appendix C: Additional Monte Carlo Results

Notes to Tables

1. Table C.1 reports empirical coverage rate of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = 0$.

2. Table C.2 reports empirical coverage rate of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = T^{-\alpha/4}$.

3. Table C.3 reports effective interval length ratio (benchmark: t_{hac}) of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = 0$.

4. Table C.4 reports effective interval length ratio (benchmark: t_{hac}) of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = T^{-\alpha/4}$.

Table C.1: Empirical coverage rate of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = 0$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
t_{hac}	57.6	61.3	63.0	50.9	55.6	61.8	66.3	67.4	65.9	72.0	74.5	77.2	67.9	70.6	70.0	64.2	69.4	73.3	67.4	72.1	77.1
PM	98.6	98.5	98.7	99.0	99.1	99.2	98.9	99.1	98.9	92.9	92.7	93.1	95.9	96.2	97.0	93.6	93.9	94.1	90.3	89.6	90.3
GSW	67.9	66.1	65.9	54.8	56.8	59.2	78.5	77.2	76.5	82.8	77.9	74.5	76.6	74.1	71.1	64.8	62.5	62.2	67.5	67.2	69.0
DWB_3	83.4	83.3	84.0	80.5	81.3	83.5	89.5	87.2	86.7	89.5	88.3	89.3	89.3	87.6	87.1	88.3	87.6	88.3	89.5	89.9	90.8
DWB_5	81.1	81.8	82.7	77.3	78.7	81.6	88.4	85.9	86.6	88.0	87.2	88.7	87.9	87.0	86.5	85.8	85.7	87.2	87.7	88.6	89.9
DWB_{10}	78.8	79.3	80.6	73.2	74.8	78.5	86.4	84.3	85.4	86.1	85.1	86.9	86.2	85.6	85.9	82.2	82.4	84.5	85.3	86.8	88.7
DWB_r	85.0	83.5	84.0	82.5	81.2	83.5	90.6	87.2	87.1	90.7	88.6	89.6	90.3	87.3	87.2	89.9	87.5	88.2	90.6	89.8	90.8
Panel B: AR case, $\phi = 0.5$																					
t_{hac}	82.9	87.7	90.0	71.7	80.1	86.5	91.8	94.6	95.2	91.2	94.9	96.8	90.8	94.2	95.2	80.2	87.5	91.6	77.7	86.3	90.7
PM	97.9	97.9	97.8	99.4	99.3	99.5	99.4	99.3	99.3	87.8	87.8	88.3	92.6	93.8	94.5	90.6	90.9	90.9	87.5	87.5	87.7
GSW	89.3	90.7	91.2	76.3	80.5	84.3	97.5	98.2	98.8	95.8	95.9	96.1	96.0	96.8	96.3	77.9	80.2	83.5	76.0	78.7	82.1
DWB_3	88.8	89.2	88.8	85.0	87.3	88.3	94.7	95.0	94.1	92.5	93.4	93.3	93.5	93.4	91.9	90.1	91.7	92.1	89.0	91.4	92.9
DWB_5	86.2	86.2	85.5	82.3	84.6	86.4	93.6	92.8	91.0	90.4	90.1	89.1	91.6	90.7	88.0	87.8	89.7	90.3	87.0	89.5	91.9
DWB_{10}	82.0	82.0	81.5	77.3	80.1	82.6	91.5	89.8	88.1	86.9	86.0	83.7	88.1	87.3	84.5	84.1	85.7	86.9	83.2	86.4	89.6
DWB_r	90.2	89.1	88.9	86.4	87.0	88.3	95.6	94.9	94.2	93.8	93.4	93.3	94.6	93.5	92.1	91.6	92.0	92.2	90.4	91.2	93.0
Panel C: MA case, $\theta = 0.5$																					
t_{hac}	83.9	90.2	92.1	74.6	85.4	90.2	90.2	94.6	95.4	91.7	95.9	97.6	91.6	95.5	96.1	82.1	89.9	93.8	81.3	88.9	93.7
PM	98.4	98.2	98.4	99.2	99.3	99.3	99.2	99.1	99.3	90.3	89.4	89.4	93.7	94.5	95.6	91.4	92.1	92.2	88.0	88.4	88.8
GSW	89.7	89.0	87.8	78.8	80.8	83.0	95.0	95.0	95.2	95.0	94.2	94.0	95.6	95.7	94.4	78.4	79.5	80.9	77.2	78.3	81.0
DWB_3	90.3	90.9	90.5	87.1	90.2	90.8	94.7	94.6	93.9	93.7	94.2	94.0	94.1	94.1	92.6	91.8	93.1	93.6	91.5	93.7	94.9
DWB_5	87.5	87.7	86.4	83.8	86.2	87.4	93.1	92.9	91.3	91.5	91.5	91.1	92.2	92.0	89.6	89.1	90.7	91.3	89.3	92.0	93.3
DWB_{10}	83.6	83.8	82.3	79.3	81.2	82.9	90.9	90.0	88.6	88.5	88.4	87.3	89.6	89.1	86.7	85.7	86.8	88.0	86.3	88.9	91.1
DWB_r	92.5	90.9	90.1	89.7	89.9	90.8	95.7	94.7	94.1	95.0	93.9	94.1	95.2	94.1	92.6	93.5	93.2	93.7	93.3	93.7	95.0

Table C.2: Empirical coverage rate of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = T^{-\alpha/4}$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
t_{hac}	89.2	92.5	93.9	91.6	93.8	94.0	81.6	86.4	90.1	74.8	78.7	83.8	87.2	91.7	93.4	83.6	88.4	90.2	78.1	83.3	86.5
PM	99.8	99.9	100	100	100	100	99.7	99.8	100	95.3	96.2	97.2	99.9	100	100	96.5	97.0	97.2	93.1	94.3	94.3
GSW	93.8	94.5	94.9	96.2	96.6	96.7	92.2	95.5	98.3	84.0	82.4	82.4	92.6	93.3	94.0	88.3	87.8	87.4	82.5	83.1	83.3
DWB_3	90.1	91.7	92.7	91.9	93.5	94.3	87.4	88.3	92.8	86.8	89.7	91.6	89.3	91.1	93.1	91.6	91.8	92.8	91.8	92.0	93.9
DWB_5	88.1	90.3	91.6	90.5	92.6	93.8	86.8	87.3	92.2	84.8	87.4	90.3	87.3	89.3	92.0	90.7	90.9	92.1	91.2	91.6	93.3
DWB_{10}	84.0	87.1	88.9	87.5	90.4	92.6	85.7	85.3	90.6	80.9	84.0	86.9	83.6	86.0	89.6	88.6	89.5	91.1	89.7	90.3	92.6
DWB_r	91.0	91.7	92.6	92.3	93.3	94.2	87.8	88.2	92.8	88.2	89.7	91.6	90.5	91.2	93.4	91.7	91.7	92.8	92.0	92.0	93.9
Panel B: AR case, $\phi = 0.5$																					
t_{hac}	86.7	91.9	93.8	88.7	92.7	94.7	91.8	94.0	95.1	90.1	93.4	94.8	86.7	90.7	92.9	87.4	91.8	94.7	84.2	90.4	93.4
PM	98.2	98.6	99.1	99.5	99.6	99.7	99.4	99.5	99.6	89.8	90.6	91.6	98.2	99.1	99.6	93.2	93.9	94.5	89.4	90.6	91.0
GSW	90.5	91.9	93.2	90.9	94.1	95.3	97.9	98.6	99.2	94.6	94.7	94.6	89.9	91.7	92.5	87.2	88.8	89.5	83.6	86.5	87.8
DWB_3	86.5	88.7	90.3	88.7	91.0	92.5	87.5	85.8	89.6	84.2	86.5	89.0	86.0	87.7	90.7	90.3	89.9	91.2	89.8	90.7	92.5
DWB_5	84.9	87.7	89.8	87.7	90.7	92.6	86.9	84.8	89.3	81.9	84.2	87.8	84.0	86.5	90.2	89.1	89.3	91.0	88.5	89.8	92.3
DWB_{10}	81.1	84.9	87.9	84.9	89.2	91.8	85.7	83.5	88.3	77.5	80.9	84.9	80.1	83.4	87.9	86.7	87.9	89.8	86.2	88.1	91.3
DWB_r	87.4	88.3	90.7	89.0	91.1	92.7	87.9	85.8	89.7	86.1	86.3	88.9	86.9	87.8	91.1	90.6	90.1	91.2	90.0	90.7	92.6
Panel C: MA case, $\theta = 0.5$																					
t_{hac}	92.9	95.9	96.9	94.4	96.7	97.4	93.3	95.7	96.7	91.6	94.8	96.1	92.0	95.3	96.8	91.2	95.4	96.8	87.7	93.5	95.8
PM	99.0	99.4	99.6	99.8	99.7	99.9	99.6	99.8	99.8	92.0	92.9	94.4	99.4	99.8	99.9	94.5	95.1	95.6	91.1	92.0	92.2
GSW	93.6	94.5	94.7	95.4	96.1	97.0	97.0	98.0	99.0	94.1	93.0	92.4	92.6	93.5	93.8	89.3	89.9	91.2	86.1	87.0	88.2
DWB_3	91.2	93.1	94.4	92.9	95.1	96.0	90.5	89.5	93.7	88.8	91.2	92.8	90.5	92.6	94.6	93.3	93.5	94.7	93.2	93.9	95.7
DWB_5	88.7	90.9	92.3	91.4	93.7	94.9	89.1	87.6	92.3	85.8	88.1	90.3	87.8	90.2	92.6	91.9	92.1	93.1	91.9	92.6	94.3
DWB_{10}	84.5	87.2	89.2	88.4	91.2	93.2	87.3	85.4	90.1	81.5	84.2	86.7	83.8	86.3	89.6	89.6	90.1	91.4	89.6	90.7	92.8
DWB_r	93.0	93.1	94.4	94.1	95.1	96.1	91.4	89.8	93.7	91.0	91.2	92.8	92.5	92.6	94.5	94.4	93.8	94.7	94.1	94.0	95.7

Table C.3: Effective interval length ratio (benchmark: t_{hac}) of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = 0$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
PM	2.01	2.08	2.22	1.97	2.06	2.29	2.20	2.20	2.28	1.32	1.29	1.34	1.56	1.62	1.73	1.06	1.18	1.15	1.18	1.28	1.25
GSW	1.10	1.03	1.01	0.92	0.85	0.86	1.16	1.04	0.99	1.11	0.93	0.87	1.01	0.94	0.90	0.79	0.79	0.68	0.71	0.69	0.61
DWB_3	1.20	1.10	1.09	1.20	1.06	1.01	1.07	0.94	0.91	1.18	1.05	1.03	1.14	1.04	1.02	1.38	1.21	0.99	1.63	1.22	1.05
DWB_5	1.17	1.08	1.07	1.18	1.03	0.99	1.06	0.91	0.90	1.16	1.03	1.01	1.12	1.02	1.00	1.32	1.17	1.01	1.47	1.14	1.00
DWB_{10}	1.13	1.04	1.04	1.15	1.00	0.97	1.08	0.90	0.87	1.13	0.99	0.98	1.09	1.00	0.98	1.28	1.11	0.99	1.41	1.09	0.94
DWB_r	1.22	1.11	1.09	1.23	1.06	1.02	1.08	0.94	0.92	1.20	1.05	1.03	1.16	1.04	1.02	1.41	1.20	1.05	1.64	1.19	1.02
Panel B: AR case, $\phi = 0.5$																					
PM	1.79	1.74	1.75	1.96	1.84	1.84	2.12	1.97	1.89	1.10	1.15	1.17	1.21	1.39	1.47	1.02	1.11	1.12	1.23	0.96	1.22
GSW	1.14	1.04	0.99	1.01	0.93	0.88	1.39	1.19	1.09	1.09	1.04	0.95	0.89	0.94	0.94	0.76	0.79	0.75	0.73	0.54	0.67
DWB_3	1.05	0.91	0.84	1.03	0.87	0.78	0.86	0.75	0.68	1.15	0.97	0.90	1.07	0.91	0.83	1.42	1.14	0.94	1.70	1.26	0.96
DWB_5	1.02	0.88	0.81	1.01	0.84	0.75	0.85	0.71	0.64	1.13	0.94	0.88	1.05	0.88	0.80	1.42	1.14	0.90	1.69	1.27	0.93
DWB_{10}	0.97	0.84	0.77	0.97	0.81	0.71	0.86	0.69	0.61	1.07	0.89	0.84	1.00	0.85	0.76	1.41	1.10	0.87	1.48	1.22	0.89
DWB_r	1.07	0.91	0.84	1.05	0.87	0.78	0.88	0.75	0.68	1.16	0.97	0.90	1.09	0.91	0.84	1.44	1.15	0.93	1.54	1.26	0.96
Panel C: MA case, $\theta = 0.5$																					
PM	1.59	1.55	1.54	1.72	1.65	1.63	1.89	1.72	1.68	0.99	1.03	1.02	1.25	1.21	1.26	1.09	0.95	0.91	0.98	1.04	1.05
GSW	1.06	0.96	0.89	0.97	0.86	0.80	1.23	1.01	0.94	1.00	0.93	0.82	0.96	0.85	0.82	0.85	0.69	0.63	0.62	0.60	0.58
DWB_3	1.04	0.90	0.82	1.05	0.88	0.78	0.87	0.76	0.68	1.12	0.94	0.87	1.06	0.89	0.81	1.57	1.70	0.96	3.83	1.33	0.99
DWB_5	0.99	0.85	0.76	1.00	0.83	0.73	0.85	0.71	0.63	1.09	0.89	0.81	1.01	0.84	0.76	1.52	1.60	0.90	3.73	1.22	0.93
DWB_{10}	0.91	0.78	0.71	0.95	0.78	0.68	0.83	0.67	0.59	1.04	0.82	0.76	0.94	0.78	0.71	1.51	1.43	0.85	3.31	1.14	0.87
DWB_r	1.08	0.90	0.82	1.08	0.88	0.78	0.90	0.76	0.68	1.17	0.95	0.87	1.09	0.89	0.81	1.61	1.63	0.95	3.83	1.27	0.92

Table C.4: Effective interval length ratio (benchmark: t_{hac}) of various inferential methods for $\rho = 1 + c/T^{\alpha}$, 95% nominal rate, $T = 50, 100, 200$, $c = 0.5$, $\alpha = 0.8$, $\mu_T = T^{-\alpha/4}$.

DGP	0			1			2			3			4			5			6		
T	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200	50	100	200
Panel A: no serial correlation																					
PM	5.65	7.27	9.48	8.10	10.5	13.9	3.37	4.33	6.06	1.67	1.83	2.15	5.44	7.26	9.55	2.01	2.24	2.51	1.81	1.76	1.76
GSW	1.11	1.05	1.02	1.17	1.12	1.11	1.30	1.32	1.48	1.09	0.94	0.87	1.03	1.02	1.01	0.96	0.89	0.82	0.98	0.79	0.67
DWB_3	1.14	1.08	1.06	1.04	1.02	1.01	0.99	0.95	1.01	1.39	1.19	1.11	1.22	1.12	1.08	1.68	1.33	1.14	1.60	1.28	1.20
DWB_5	1.07	1.03	1.02	0.99	0.99	0.99	0.98	0.92	0.98	1.38	1.14	1.07	1.13	1.06	1.03	1.62	1.26	1.11	1.51	1.21	1.14
DWB_{10}	0.95	0.95	0.96	0.92	0.95	0.95	1.00	0.87	0.94	1.37	1.05	0.98	1.00	0.96	0.96	1.51	1.18	1.04	1.44	1.15	1.10
DWB_r	1.18	1.08	1.06	1.07	1.02	1.01	1.01	0.95	1.01	1.38	1.19	1.11	1.27	1.12	1.08	1.71	1.32	1.15	1.66	1.26	1.20
Panel B: AR case, $\phi = 0.5$																					
PM	3.05	3.68	4.84	4.34	5.46	7.30	2.46	2.53	2.87	1.05	1.21	1.39	2.49	3.67	5.05	1.66	1.45	1.06	1.39	1.47	1.48
GSW	1.08	0.99	0.98	1.12	1.06	1.05	1.43	1.26	1.26	0.95	0.97	0.92	0.87	0.93	0.95	0.96	0.77	0.49	0.85	0.82	0.77
DWB_3	1.16	1.06	1.02	1.07	0.99	0.96	0.95	0.90	0.92	1.25	1.10	1.09	1.30	1.10	1.05	1.52	1.47	1.06	1.59	1.29	1.11
DWB_5	1.12	1.05	1.02	1.06	0.98	0.98	0.95	0.88	0.93	1.22	1.08	1.06	1.24	1.08	1.04	1.40	1.48	1.02	1.55	1.23	1.05
DWB_{10}	1.04	1.00	0.98	1.04	0.98	0.96	0.99	0.86	0.92	1.12	1.03	1.01	1.33	1.00	0.99	1.29	1.34	0.94	1.39	1.15	1.00
DWB_r	1.17	1.07	1.02	1.10	0.99	0.96	0.95	0.90	0.92	1.26	1.11	1.08	1.32	1.10	1.05	1.54	1.49	1.05	1.65	1.28	1.10
Panel C: MA case, $\theta = 0.5$																					
PM	3.30	4.16	5.59	4.83	6.18	8.27	2.37	2.58	3.11	1.06	1.10	1.36	2.87	4.13	5.75	1.22	1.55	1.73	1.06	1.28	1.27
GSW	0.99	0.89	0.88	1.03	0.98	0.95	1.29	1.14	1.13	0.94	0.81	0.81	0.81	0.84	0.86	0.65	0.74	0.75	0.63	0.70	0.62
DWB_3	1.14	1.05	1.00	1.03	0.98	0.96	0.92	0.88	0.91	1.29	1.12	1.04	1.21	1.08	1.02	1.67	1.18	1.12	1.95	1.29	1.07
DWB_5	1.06	0.97	0.95	0.96	0.92	0.92	0.90	0.83	0.84	1.24	1.05	0.98	1.11	1.02	0.96	1.55	1.11	1.02	1.80	1.19	0.99
DWB_{10}	0.96	0.88	0.87	0.89	0.84	0.88	0.90	0.78	0.80	1.12	0.93	0.89	0.96	0.90	0.87	1.40	1.01	0.97	1.73	1.10	0.90
DWB_r	1.16	1.05	1.00	1.07	0.99	0.96	0.96	0.88	0.89	1.31	1.11	1.05	1.26	1.08	1.02	1.68	1.17	1.08	1.99	1.21	0.99