4-2020

Three essays on nonstationary time series econometrics

Yiu Lim LUI
Singapore Management University, yllui@smu.edu.sg

SINGAPORE MANAGEMENT UNIVERSITY

PHD DISSERTATION

Three Essays on Nonstationary Time Series Econometrics

Yiu Lim Lui

supervised by

Professor JUN YU

June 22, 2020

Three Essays on Nonstationary Time Series Econometrics

by

Yiu Lim Lui

Submitted to School of Economics in partial fulfillment of

the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Jun Yu (Supervisor/Chair)
Lee Kong Chian Professor of Economics and Finance
Singapore Management University

Peter C.B. Phillips
Sterling Professor of Economics and Statistics
Yale University
Distinguished Term Professor of Economics
Singapore Management University

Thomas J. Sargent
Professor of Economics
New York University
Distinguished Term Professor of Economics
Singapore Management University

Weilin Xiao
Associate Professor
School of Management
Zhejiang University

Singapore Management University

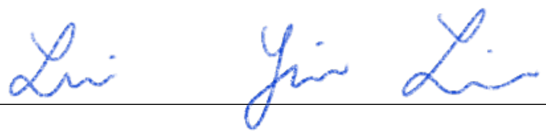
2020

Declaration

I hereby declare that this PhD dissertation is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information which have been used in this dissertation.

This dissertation has also not been submitted for any degree in any university previously.



Yiu Lim Lui

22 June 2020

Abstract

This dissertation comprises three papers that separately study different nonstationary time series models.

The first paper, titled as ”The Grid Bootstrap for Continuous Time Models”, is a joint work with Professor Jun Yu and Professor Weilin Xiao. It considers the grid bootstrap for constructing confidence intervals for the persistence parameter in a class of continuous-time models driven by a Lévy process. Its asymptotic validity is discussed under the assumption that the sampling interval (h) shrinks to zero, the time span (N) goes to infinity or both. Its improvement over the in-fill asymptotic theory is achieved by expanding the coefficient-based statistic around its in-fill asymptotic distribution which is non-pivotal and depends on the initial condition. Monte Carlo studies show that the grid bootstrap method performs better than the in-fill asymptotic theory and much better than the long-span asymptotic theory. Empirical applications to U.S. interest rate data and volatility data suggest significant differences between the bootstrap confidence intervals and the confidence intervals obtained from the in-fill and long-span asymptotic distributions.

The second paper, ”Mildly Explosive Autoregression with Anti-persistent Errors” is another joint work with Professor Yu and Professor Xiao. It studies a mildly explosive autoregression model with Anti-persistent Errors. An asymptotic distribution is derived for the least squares (LS) estimate of a first-order autoregression with a mildly explosive root and anti-persistent errors. While the sample moments depend on the Hurst parameter asymptotically, the Cauchy limiting distribution theory remains valid for the LS estimates in the model without intercept and a model with an asymptotically negligible intercept. Monte Carlo studies are designed to check the precision of the Cauchy distribution in finite samples. An empirical study based on the monthly NASDAQ index highlights the usefulness of the model and the new limiting distribution.

The third paper ”Testing for Rational Bubbles under Strongly Dependent Errors” considers testing procedures for rational bubbles under strongly dependent errors. A heteroskedasticity and autocorrelation robust (HAR) test statistic is proposed to detect the presence of rational bubbles in financial assets when errors are strongly dependent. The asymptotic theory of the test statistic is developed. Unlike conventional test statistics that lead to a too large type I error under strongly dependent errors, the new test does not suffer from the same size problem. In addition, it can consistently timestamp the origination and termination dates of a rational bubble. Monte Carlo studies are conducted to check the finite sample performance of the proposed test and estimators. An empirical application to the S&P 500 index highlights the usefulness of the proposed test statistic and estimators.

Contents

Acknowledgement

First and foremost, I want to thank God for His blessings and graces.

I would also like to thank my supervisor Professor Jun Yu for his guidance and patience over the last 5 years. Besides academic training, there were times I felt stressed during my PhD study and Professor Yu always provides encouragement and care. Moreover, I would like to express my gratitude to my thesis committee members: Professor Peter C.B. Phillips and Professor Thomas J. Sargent. Professor Phillips taught me so many techniques that allow me to understand and appreciate frontier econometrics and think of new research ideas. Professor Sargent provided me a deeper understanding of macroeconomics and shared life experience with me. Their support is essential for my completion of the program.

Additionally, I am indebted to my external committee member, Professor Weilin Xiao, who is very friendly and introduced Young integral to me, and all participants in SMU's weekly econometrics workshop. A lot of my research ideas indeed originated from the presentations and discussions in this workshop.

I am extremely grateful to my parents, Chun Kuen Lui and Yee Man Chow, my brother Yiu Him Lui for their love, understanding, and sacrifice. Special thanks go to my fiancé Cuiling Zhang for her support throughout these years.

Finally, I want to express my appreciation to the PhD program director Professor Anthony Tay for his leadership and the administrative staff Qiu Ling Thor, Amelia Tan, and Pei Qi Cheong for their hard work and contributions to the program.

1 The Grid Bootstrap for Continuous Time Models

1.1 Introduction

A popular model to describe the evolution of an economic time series $y(t)$ is given by the following Ornstein-Uhlenbeck (OU) diffusion process:

$$dy(t) = \kappa(\mu - y(t))dt + \sigma dW(t), y(0) = y_0, \tag{1}$$

where κ, μ, and σ are all constants, y_0 is the initial condition, and $W(t)$ is a standard Brownian motion. In this model, κ captures the persistence of $y(t)$ and is the parameter of interest in the present paper. Consider the case when a discrete sample of observations for $y(t)$ is available as y_t with $t = h, 2h, ..., Th\,(:= N)$, where h is the sample interval and T is the sample size. Clearly, N is the time span over which the discrete-sampled data is available.

Typically κ is estimated by the least squares (LS) method. Denote the LS estimator by $\hat{\kappa}$. To make the statistical inference about κ, one needs to obtain the exact finite sample distribution of $\hat{\kappa}$. Unfortunately, the exact finite sample distribution of $\hat{\kappa}$ is not analytically available. It has to be obtained by simulations (as was done in Yu (2014) and Zhou and Yu (2015)) or by numerical integrations when $\kappa \geqslant 0$ (as was done in Bao et al. (2017)). It generally depends on the initial condition (whether it is fixed or random) and the random behavior of the stochastic term in the model (whether it is a Brownian motion or a Lévy process). Not surprisingly, econometricians often rely on asymptotic theory to approximate the exact finite sample distribution.

Three sampling schemes can be used to obtain a limiting distribution, namely "in-fill", or "long-span" or "double", corresponding to the assumption of $h \to 0$, or $N \to \infty$, or $h \to 0$ together with $N \to \infty$, respectively. In practice, of course, no matter how small, h is always strictly positive; and no matter how large, N is always finite. Hence, all three asymptotic distributions are merely approximations to the finite sample distribution. Clearly, the double scheme cannot provide a more accurate approximation than the other two schemes due to an added restriction.

Different schemes lead to different limiting distributions for $\hat{\kappa}$. The long-span and double schemes lead to a Gaussian distribution for $\hat{\kappa}$ when $\kappa > 0$ but to a Dickey-Fuller-Phillips type distribution when $\kappa = 0$. The in-fill scheme leads to a non-standard limiting distribution for $\hat{\kappa}$ and there is no discontinuity in the limiting distribution when κ passes through zero.

The limiting distributions obtained from the three schemes have their advantages and drawbacks when they are used to make statistical inferences. The long-span and double limiting distributions are well-known, facilitating the statistical inferences. However, the limiting distribution typically has poor finite sample performance when $\kappa > 0$. On the other hand, the in-fill limiting distribution is continuous at $\kappa = 0$ and outperforms the long-span and double counterparts in finite samples. Unfortunately, the in-fill limiting distribution depends on nuisance parameters and is non-standard.

Our paper introduces a method for making statistical inferences about κ. Our method is based on the grid bootstrap and does not use any limiting distribution. It has three advantages over the asymptotic theory. First, its validity can be justified by any asymptotic scheme. This feature frees empirical researchers from using a limit distribution which depends critically on the asymptotic scheme. Second, the method is uniformly valid for $\kappa \geqslant 0$. Third, the method provides excellent finite-sample performance.

This approach is to use the grid bootstrap method to construct confidence intervals (CIs) for κ. The grid bootstrap was initially introduced by Hansen (1999) to construct CIs for the autoregressive (AR) coefficient in the AR(1) model. Hansen (1999) showed that the method is asymptotically valid in the stationary and local-to-unit-root case. Mikusheva (2007) showed that under the long-span scheme the grid bootstrap leads to CIs that have correct coverage uniformly over the parameter space, including the mildly stationary and the local-to-unit-root case. It is the results obtained in Hansen (1999) and Mikusheva (2007) that motivates us to make use of the grid bootstrap to construct CIs for κ since our model is closely related to a local-to-unit-root model, as will be shown later.

This paper justifies the grid bootstrap procedure under the three asymptotic schemes. In particular, it is shown that CIs for κ obtained by the grid bootstrap have correct coverage uniformly over the parameter space, including any value of $\kappa > 0$ and $\kappa = 0$. Moreover, we show that the grid bootstrap provides finite sample improvement over the in-fill asymptotic distribution. This finding is interesting as the in-fill asymptotic distribution already outperforms the other two limiting distributions in finite samples. We show this by applying stochastic expansion which uses the in-fill asymptotic distribution as the leading term.

Our setup and approach have a few attractive features. First, we can justify the bootstrap method under the in-fill scheme, the long-span scheme, or the double scheme. This is particularly important in empirical work since the CI remains the same regardless of which asymptotic scheme is adopted. Second, consistent estimation of κ and μ is not required for constructing a valid CI of κ under the in-fill scheme. Third, we show that the bootstrap CIs perform better than CIs based on the in-fill asymptotic distribution and much better than those based on the long-span asymptotic distribution. Fourth, since the grid bootstrap has correct coverage uniformly for all values of $\kappa \geqslant 0$, our method can be used to test for the unit root as well as for a stationary root. This is in sharp contrast to the approaches based on the long-span asymptotic scheme where the test statistics and their asymptotic distributions under the unit root null hypothesis (such as the Dickey-Fuller test and the Phillips-Perron test) are very different from those under the stationary null hypothesis (such as the KPSS test of Kwiatkowski et al (1992) and the test proposed in Chang et al. (2019)). Finally, the grid bootstrap method, with a simple modification, is applicable in the presence of heteroskedasticity.

We organize the paper as follows. Section 1.2 reviews some relevant results in the literature on the continuous-time model given by (1) and relates some of them to those in the discrete-time AR(1) model. The concept of a bootstrap CI is also reviewed. In Section 1.3, a more general continuous-time model is introduced. The LS estimator of κ and its in-fill asymptotic distribution are also discussed. Section 1.4 develops the grid bootstrap method to construct CIs for κ and provides the asymptotic justification to the procedure. Probabilistic expansions, which use the in-fill asymptotic distribution as the leading term, and the grid bootstrap under heteroskedasticity are also reported. Simulation studies, which aim to check the finite sample performance of the bootstrap method, are carried out in Section 1.5. Section 1.6 reports CIs for κ based on US interest rate data and Chicago Board Options Exchange's volatility index data. Section 1.7 concludes. Proofs of the main results in the paper are given in the Appendix.

We use the following notations throughout the paper, "$\Rightarrow$"means weak convergence in distribution, "$\overset{p}{\Rightarrow}$" means weak convergence in probability, "$\to$" means convergence in real sequence, "$\sim$" means asymptotic equivalence, "$\overset{d}{=}$" means distributional equivalence, "$\to_p$", "$\to_d$", and "$\to_{a.s.}$" mean convergence in probability, distribution, and almost surely, respectively.

1.2 A Literature Review

In this section, we review some relevant results in the literature on the continuous-time model given by (1). We also relate some of the results to those in the discrete-time literature. Then we review the concept of CI based on alternative distributions, including the bootstrap distributions.

Assume $Y := \{y_{th}\}_{t=1}^{T}$ to be data generated from the continuous-time model given by (1). The exact discrete model corresponding to (1) is given by

$$y_{th} = e^{-\kappa h} y_{(t-1)h} + \mu \left(1 - e^{-\kappa h}\right) + \sqrt{(1 - e^{-2\kappa h})/(2\kappa)}\epsilon_t,$$

where $\epsilon_t \sim i.i.d.N(0, \sigma^2)$, $t = 1, ..., T$. Clearly, T can be made to go to infinity by either increasing N (the long-span scheme) or decreasing h (the in-fill scheme) or both (the double scheme). Dividing both sides by $\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}$ gives rise to

$$x_{th} = e^{-\kappa h} x_{(t-1)h} + \frac{\mu \left(1 - e^{-\kappa h}\right)}{\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}} + \epsilon_t, x_0 = \frac{y_0}{\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}}, \tag{2}$$

where $x_{th} = y_{th}/\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}$.

Model (2) has the same structure as the popular discrete-time AR(1) model with $\rho_h(\kappa) = e^{-\kappa h}$ being the AR coefficient. Let the LS estimator of $\rho_h(\kappa)$ be $\widehat{\rho}_h$ and the LS estimator of κ be $\widehat{\kappa} = -\ln\left(\widehat{\rho}_h\right)/h$. If $\kappa = 0$, then $\rho_h(\kappa) = 1$, implying the presence of a unit root. If $h \to 0$ but N is finite, then $e^{-\kappa h} \sim 1 + (-\kappa h) = 1 + (-\kappa N/T)$. So the in-fill asymptotic scheme implies that Model (2) has a root which is local-to-unity with the local parameter being $c := -\kappa N$ and the initial condition $x_0 \sim O(1/\sqrt{h})$ if $y_0 \neq 0$ and $x_0 = 0$ if $y_0 = 0$. In the local-to-unity literature, the initial condition is typically assumed to be $O_p(1)$ and the corresponding long-span asymptotic distribution involves functionals of the OU process but is independent of the initial condition.[1] When $y_0 \neq 0$ in Model (2), it is expected that the in-fill asymptotic distribution of $\widehat{\rho}_h$ performs better than the usual long-span asymptotic distribution developed in the local-to-unity literature.

Phillips (1987b) developed the in-fill asymptotic distribution for $\widehat{\rho}_h$ when $y_0 = 0$ and μ is known (i.e. $\mu = 0$). In the same paper, Phillips showed that this in-fill asymptotic distribution is the same as the long-span asymptotic distribution in the local-to-unity model with the initial condition of $O_p(1)$. Perron (1991) extended the results in Phillips (1987b) by allowing for a general initial condition for y_0. Yu (2014) and Zhou and Yu (2015) developed the in-fill asymptotic distribution for $\widehat{\kappa}$ when μ is known and unknown, respectively. Unless $y_0 = 0$ the in-fill asymptotic distribution explicitly depends on the initial condition, and hence is different from the long-span asymptotic distribution in the local-to-unity model with the initial condition of $O_p(1)$.

It is straightforward to derive the long-span asymptotic distribution for $\widehat{\kappa}$ by applying the Delta method to the long-span asymptotic distribution for $\widehat{\rho}_h$. For example, when $\kappa > 0$, $\sqrt{T}(\hat{\kappa}-\kappa) \Rightarrow N\left(0, \left(\exp(2\kappa h) - 1\right)/h\right)$; see Tang and Chen (2009). When $\kappa = 0$, $N(\hat{\kappa} - \kappa) \Rightarrow -\int_0^1 \overline{W}(r)dW(r)/\int_0^1 \overline{W}(r)^2 dr$ with $\overline{W}(r) = W(r) - \int_0^1 W(s)ds$. The discontinuity in the long-span limiting distribution of κ echoes that of ρ in the discrete-time AR(1) model.

When κ is positive but reasonably close to zero, Yu (2014) and Zhou and Yu (2015) obtained the exact finite sample distribution of $\hat{\kappa}$ by simulations. Bao et al. (2017) approximated the finite sample distribution of $\hat{\kappa}$ via numerical integrations. All these studies find that the in-fill distribution is much closer to the finite sample distribution than the long-span and the double asymptotic distributions, even when 10 years or 50 years of monthly data are used. The superiority of the in-fill distribution over the long-span distribution is not surprising as the in-fill distribution depends explicitly on the initial condition and is asymmetric. While these two features can be found in the finite sample distribution, they are lost in the long-span asymptotic distribution.

[1]From Mikusheva (2015), it can be easily shown that as $T \to \infty$, in the local-to-unity model with intercept, $T(\widehat{\rho} - \rho) \Rightarrow \int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2 dr$ where $\overline{J}_c(r) = J_c(r) - \int_0^1 J_c(s)ds$ is the de-meaned OU process with $J_c(r) = \int_0^r \exp(-c(r-s))dW(s)$.

For the discrete-time AR(1) model, the in-fill scheme is not available. When the AR coefficient is in the stationary region (that is, it is less than one in absolute value), the long-span asymptotic distribution of the LS estimator of the AR coefficient is Gaussian. However, the finite sample distribution may be far away from Gaussianity, especially when the AR coefficient is close to one and the sample size is small or moderately large. This motivates Phillips (1977) and Tanaka (1983) to develop the Edgeworth expansions to approximate the finite sample distribution of the LS estimator of the AR coefficient. While the leading term in the Edgeworth expansions is a normal distribution, a departure from normality manifests in higher-order terms. Alternatively, the finite sample distribution can be approximated by bootstrap. Bose (1988) showed the linkage between the Edgeworth expansions and the bootstrap method.

When the AR(1) model has a unit root, the long-span asymptotic distribution is non-standard. Basawa et al. (1991) and Park (2003) introduced bootstrap procedures which improve upon the long-span asymptotic theory. In an important study, Park (2003) justified the bootstrap procedure by obtaining expansions for the Dickey-Fuller unit root test where the leading term is the Dickey-Fuller-Phillips distribution and showed that the bootstrap offers a second-order asymptotic refinement for the Dickey-Fuller test. Under the local-to-unity AR(1) model, Hansen (1999) introduced the grid bootstrap approach. Mikusheva (2015) obtains expansions of the t-statistic about the local-to-unity asymptotic distribution and shows that the grid bootstrap procedure of Hansen (1999) achieves a second-order refinement of the local-to-unity asymptotic approximation. The results of Mikusheva (2015) are important because, when the AR(1) coefficient is less than but close to one, the local-to-unity asymptotic distribution tends to give better approximations to the finite sample distribution than the normal distribution even when the sample size is moderately large. However, since the initial condition is assumed to be $O_p(1)$ in the model of Mikusheva (2015), the local-to-unity asymptotic distribution is independent of the initial condition.

We now review the concept of CI based on alternative distributions. Assume ρ is the parameter of interest in a statistical model. Without loss of generality, assume ρ is a scalar. Let T denote the sample size of available data Y and $t_T(Y, \rho)$ be a test statistic with sampling distribution $F_T(x|\rho) = \Pr(t_T(Y,\rho) < x|\rho)$. For $q \in (0,1)$, let $c_T(q|\rho)$ be the quantile function of $t_T(Y,\rho)$, that is, $F_T(c_T(q|\rho)|\rho) = q$. Define a q-level CI for ρ by

$$CI_q := \{\rho \in R : c_T(x_1|\rho) \leqslant t_T(Y,\rho) \leqslant c_T(x_2|\rho)\}, \tag{3}$$

where $x_1 = (1-q)/2$ and $x_2 = 1-(1-q)/2$. If ρ_0 is the true parameter value of ρ, by definition, $\Pr(\rho_0 \in CI_q) = q$, and hence, the coverage probability is exactly q, the intended level.

Suppose, as $T \to \infty$, $F_T(x|\rho)$ converges to an asymptotic distribution (call it $F(x|\rho)$) which is often pivotal. In this case both F and the corresponding quantile function (call it $c(q|\rho)$) are independent of T and the asymptotic CI will have a correct probability coverage. For example, if the asymptotic distribution is standard normal, then a 95% asymptotic CI is $CI^A_{95\%} = \{\rho \in R : -1.96 \leqslant t_T(Y,\rho) \leqslant 1.96\}$.

If the asymptotic distribution of $F_T(x|\rho)$ is not pivotal, say, depending on a set of unknown parameters θ (call the limit distribution $F(x,\theta|\rho)$ and the corresponding quantile function $c(q,\theta|\rho)$), replacing $c_T(x_i|\rho)$ with $c(x_i,\theta|\rho)$ in Equation (3) does not work because θ is not known. If θ can be consistently estimated, say by $\widehat{\theta}$, then we can replace $c_T(x_i|\rho)$ with $c\left(x_i,\widehat{\theta}|\rho\right)$ in Equation (3) to obtain an asymptotic CI, CI^A_q. It is easy to show that $\lim_{T\to\infty} \Pr\left(\rho_0 \in CI^A_q\right) = q$. If $c_T(x_i|\rho)$ is approximated by the quantile function corresponding to a bootstrap distribution, denoted by $c^*_T(x_i|\rho)$, then the CI is called a bootstrap confidence interval (BCI), CI^B_q. For example, a BCI given by the standard bootstrap procedure is given by

$$CI^B_q := \{\rho \in R : c^*_T(x_1|\widehat{\rho}) \leqslant t_T(Y,\rho) \leqslant c^*_T(x_2|\widehat{\rho})\},$$

where $\widehat{\rho}$ denotes an estimate of ρ.

There are some advantages to using BCIs. First, BCIs are obtained by re-sampling the data. Although asymptotic justification of bootstrap methods requires the knowledge of asymptotic theory, generating a BCI does not require an asymptotic scheme; see Section 1.4. Second, bootstrap methods are known to provide a finite sample refinement to asymptotic theory in the sense that the bootstrap distribution provides better approximations to the finite sample distribution than asymptotic distributions; see Hall (2013) and

Chang and Hall (2015). Not surprisingly, bootstrap methods often lead to CIs that have a more accurate coverage probability than the traditional asymptotic theory.

1.3 The Model and In-fill Theory

The present paper extends Model (1) by allowing for non-normality in errors. Such an extension makes the analytical approach of Bao et al. (2017) not applicable. We then develop the in-fill asymptotic distribution and the long-span asymptotic distribution for the coefficient-based statistic based on the LS estimator of κ. We then propose the grid bootstrap to obtain BCIs for κ and discuss its validity under different asymptotic schemes. Motivated by the better performance of the in-fill asymptotic distribution relative to the long-span asymptotic distribution, an asymptotic expansion for the coefficient-based statistic, with the in-fill asymptotic distribution as the leading term, is developed. The expansion shows that the bootstrap method offers a refinement of the in-fill asymptotic distribution and theoretically explains the superiority of the bootstrap method over the in-fill distribution.

1.3.1 The model

Following Wang and Yu (2016), we consider the following continuous-time model:

$$dy(t) = \kappa(\mu - y(t))dt + \sigma dL(t), y(0) = y_0 = O_p(1), \tag{4}$$

where σ is a strictly positive constant, κ is a non-negative constant, $L(t)$ is a Lévy process defined on a probability space $(\Sigma, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0}, P)$, with $L(0) = 0$ $a.s.$, $\mathcal{F}_t = \sigma\left\{\{y(s)\}_{s=0}^{t}\right\}$. The generalization from $W(t)$ to $L(t)$ is important in empirical applications for many financial variables; see Madan and Setena (1990) for equity prices, Bai and Ng (2005) for interest rates, and Aït-Sahalia and Jacod (2014) for an excellent textbook explanation of why $L(t)$ is important.

In this paper, we are interested in obtaining CIs for the persistence parameter κ from discrete-sampled observations $\{y_{th}\}_{t=1}^{T}$, μ, σ and parameters in $L(t)$ are treated as nuisance parameters. The exact discrete-time version of (4) is

$$y_{th} = e^{-\kappa h}y_{(t-1)h} + \mu(1 - \exp(-\kappa h)) + \sigma\int_{(t-1)h}^{th}\exp(-\kappa(th - s))dL(s), \tag{5}$$

where $t = 1, 2, \ldots, T$.

Note that, the characterization of the Lévy process makes $\left\{\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s)\right\}_{t=1}^{N/h}$ an i.i.d. sequence with the distribution depending on the specification of the Lévy measure. Let the characteristic function of $L(t)$ be of the form of $E(\exp\{isL(t)\}) = \exp\{-t\psi(s)\}$, where i is the imaginary unit and the function $\psi : R \to C$ is the Lévy exponent of $L(t)$.

Assuming that $L(t)$ is square-integrable, the error term has the following moments:

$$E\left(\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s)\right) = \sigma i\psi'(0)\frac{1-\exp(-\kappa h)}{\kappa},$$

$$Var\left(\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s)\right) = \sigma^2\psi''(0)\frac{1-\exp(-2\kappa h)}{2\kappa}.$$

To simplify notations, let

$$\begin{aligned}
&\rho_h(\kappa) := \exp(-\kappa h),\ \lambda_h := \sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}},\ \sigma_\psi^2 := \sigma^2\psi''(0),\\
&g_h := \left[\mu + \frac{\sigma i\psi'(0)}{\kappa}\right](1-\exp(-\kappa h)),\\
&u_{th} := (\sigma_\psi\lambda_h)^{-1}\left(\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s) - \sigma i\psi'(0)\frac{1-\exp(-\kappa h)}{\kappa}\right) \quad (6)
\end{aligned}$$

Note that $\{u_{th}\}_{t=1}^{T}$ is a sequence of i.i.d. variables with mean zero and variance 1. When there is no confusion, we simply omit h in y_{th} and u_{th}. Using notations in (6), we can rewrite (5) as:

$$y_t = \rho_h(\kappa)y_{t-1} + g_h + \epsilon_t, \epsilon_t = \sigma_\psi\lambda_h u_t, y(0) = y_0 = O_p(1). \quad (7)$$

1.3.2 Estimation

In Model (5), we use the LS method to estimate $\rho_h(\kappa)$ and then obtain the estimator of κ

$$\hat{\kappa}_h = -\ln(\widehat{\rho}_h)/h,$$

where $\widehat{\rho}_h$ is the LS estimator for $\rho_h(\kappa)$.

The coefficient-based statistic and the t statistic for $\rho_h(\kappa)$ are, respectively

$$z\left(Y, \rho, T\right) = T\left(\widehat{\rho}_h - \rho_h(\kappa)\right) \text{ and } t\left(Y, \rho, T\right) = \frac{\widehat{\rho}_h - \rho_h(\kappa)}{\hat{\sigma}_{\widehat{\rho}_h}},$$

where $\hat{\sigma}_{\widehat{\rho}_h} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{g}_h - \widehat{\rho}_h y_{t-1})^2 \times \left(\sum_{t=1}^{T} y_{t-1}^2 - \frac{1}{T}\left(\sum_{t=1}^{T} y_{t-1}\right)^2\right)^{-1}}$ is the standard error of $\widehat{\rho}_h$. The normalization in $z(Y, \rho, T)$ is T not $\sqrt{T}$; see Phillips (1987b).

Following Perron (1991) and Zhou and Yu (2015), we define the coefficient-based statistic for κ as

$$z(Y, \kappa, h) = N(\hat{\kappa}_h - \kappa). \tag{8}$$

Remark 1 *Although in this paper we use the coefficient-based statistic for κ to construct CIs, we can also define the t statistic as $t_T(Y, \kappa) = h(\hat{\kappa}_h - \kappa)/\hat{\sigma}_{\widehat{\rho}_h}$, and construct CIs accordingly. However, this may not be a standard t statistic as the standard error of $\hat{\kappa}_h$ is not defined clearly in the context.*

1.3.3 In-fill asymptotic theory

The in-fill asymptotic theory has gained much prominence in recent years. Studies which have developed the in-fill asymptotics for different econometric models include Li and Xiu (2016), Jiang et al. (2018, 2020). In this section we extend the in-fill asymptotic result of Zhou and Yu (2015) to Model (4).

Lemma 1 *For Model (4), define $z(Y, \kappa, h)$ by (8). Then, as $h \to 0$,*

$$z(Y, \kappa, h) \Rightarrow z^{y_0}(\kappa, \theta) := -\frac{\Upsilon_3 - \Upsilon_2 \int_0^1 dW(r)}{\Upsilon_1 - \Upsilon_2^2}, \tag{9}$$

where $\theta = (\mu, \sigma, \psi'(0), \psi''(0))$, Υ_1, Υ_2 and Υ_3 are defined in the Appendix.

This limiting distribution in (9) allows us to invert the coefficient-based statistic and construct CIs for κ. It can be seen that when an error term involves a Lévy process, the Lévy exponent enters the limiting distribution through σ_ψ and $\psi'(0)$. The approach is infeasible as there are some unknown parameters in the limiting distribution in (9), including κ, μ, σ, $\psi'(0)$, $\psi''(0)$.

Remark 2 *If Model (4) is driven by a standard Brownian motion (i.e. $L(t) = W(t)$), then $\psi'(0) = 0$, $\psi''(0) = 1$, and the in-fill distribution of $\hat{\kappa}$ given in (9) is the same as that obtained from Zhou and Yu (2015). In addition, if μ is known and equal to 0, the in-fill* *distribution of $\hat{\kappa}$ is identical to that in Perron (1991). By further assuming $y_0 = 0$, the in-fill distribution of $\hat{\kappa}$ is the same as that in Phillips (1987b).*

Remark 3 *If Model (4) is driven by a standard Brownian motion, unless $y_0 = 0$, the in-fill distribution of $\hat{\kappa}$ depends on the initial condition via γ_0. If $y_0 = 0$ and $\mu = 0$, then γ_0 and b are both equal to 0 in Lemma 1. If $y_0 = \mu$, subtracting y_0 from both sides of equation (5), we obtain $\tilde{y}_{th} = e^{-\kappa h}\tilde{y}_{(t-1)h} + \epsilon_t$, with $\tilde{y}_{th} = y_{th} - y_0$. In this case, Lemma 1 implies that*

$$z^{y_0}(\kappa, \theta) = -\frac{\int_0^1 J_c(r)dW(r) - \int_0^1 J_c(r)dr \int_0^1 dW(r)}{\int_0^1 J_c^2(r)dr - \left(\int_0^1 J_c(r)dr\right)^2} = -\frac{\int_0^1 \overline{J}_c(r)dW(r)}{\int_0^1 \overline{J}_c(r)^2 dr},$$

where $\overline{J}_c(r) = J_c(r) - \int_0^1 J_c(s)ds$ is the de-meaned OU process with $J_c(r) = \int_0^r \exp(-c(r-s))dW(s)$. Similarly, if we further impose $\kappa = 0$, we obtain $z^{y_0}(\kappa, \theta) = -\int_0^1 \overline{W}(r)dW(r)/\int_0^1 \overline{W}(r)^2 dr$ where $\overline{W}(r)$ is the de-mean Brownian motion.

The in-fill distribution of $\hat{\kappa}$ (i.e. $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2 dr$) is closely related to the long-span asymptotic distribution of the coefficient-based statistic for $\widehat{\rho}$ in the local-to-unity model with the initial condition of $O_p(1)$; see Remark 3.1 in Mikusheva (2015). The reason why the initial condition explicitly enters the asymptotic distribution is that Model (2) corresponds to a local-to-unity model with the initial condition diverges to infinity as $h \to 0$. Clearly, the in-fill distribution of $\hat{\kappa}$ given in (9) is expected to outperform $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2 dr$ when the initial condition is not zero.

To see the impact of the initial condition, we perform a small Monte Carlo experiment. The following parameter settings are considered: $\kappa = 0.5$, $\mu \in \{0, 0.1\}$, $y_0 \in \{0, 1, 2, 3\}$. The number of replications is always set at 10,000. Let z^0 denote $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2 dr$.

Table 1 reports the percentiles of z^0 and the in-fill distribution $z^{y_0}(\kappa, \theta)$. Making a statistical inference from the discrete-time local-to-unity model with intercept is similar to making a statistical inference in the continuous-time model (4) by restricting $\mu = 0$, $y_0 = 0$ or $y_0 = \mu$. From Table 1, it can be clearly seen that the distribution depends on the initial condition, and it is expected that the in-fill distribution $z^{y_0}(\kappa, \theta)$ outperforms z^0 in finite samples, as the finite sample distribution depends on the initial condition.

Table 1: Percentile of z^0 and $z^{y_0}(\kappa,\theta)$ when $\kappa = 0.5$

	Percentiles	1%	5%	10%	50%	90%	95%	99%
	z^0	-2.007	-0.746	0.035	4.219	11.669	14.673	21.084
$\mu = 0, y_0 = 1$	$z^{y_0}(\kappa,\theta)$	-2.209	-0.930	-0.155	3.766	11.036	13.921	20.148
$\mu = 0, y_0 = 2$	$z^{y_0}(\kappa,\theta)$	-2.415	-1.264	-0.565	2.815	9.222	11.608	17.867
$\mu = 0, y_0 = 3$	$z^{y_0}(\kappa,\theta)$	-2.486	-1.466	-0.842	1.910	6.826	8.963	14.346
$\mu = 0.1, y_0 = 0$	$z^{y_0}(\kappa,\theta)$	-1.984	-0.745	0.026	4.193	11.663	14.627	21.170
$\mu = 0.1, y_0 = 1$	$z^{y_0}(\kappa,\theta)$	-2.303	-0.995	-0.182	3.887	11.344	14.278	20.587
$\mu = 0.1, y_0 = 2$	$z^{y_0}(\kappa,\theta)$	-2.542	-1.333	-0.601	2.937	9.717	12.267	18.770
$\mu = 0.1, y_0 = 3$	$z^{y_0}(\kappa,\theta)$	-2.585	-1.537	-0.898	2.015	7.242	9.517	15.333

Remark 4 *If we define the t statistic for κ as $t\left(Y,\kappa,h\right) = h(\hat{\kappa}_h - \kappa)/\hat{\sigma}_{\widehat{\rho}_h}$, then as $h \to 0$,*

$$t\left(Y,\kappa,h\right) \Rightarrow t^{y_0}(\kappa,\theta) := -\frac{\Upsilon_3 - \Upsilon_2 \int_0^1 dW(r)}{\sqrt{\Upsilon_1 - \Upsilon_2^2}}.$$

Remark 5 *By assuming $N \to \infty$ with h fixed , it can be shown that the long-span asymptotic distribution of $t\left(Y,\kappa,h\right)$ is $N(0,1)$ when $\kappa > 0$. It becomes $-\int_0^1 \overline{W}(r)dW(r)/\sqrt{\int_0^1 \overline{W}(r)^2 dr}$ with $\overline{W}(r) = W(r) - \int_0^1 W(s)ds$ being the de-meaned Brownian motion when $\kappa = 0$.*

Remark 6 *As shown in Phillips (1987b), when $c \to -\infty$, $\int_0^1 J_c(r)dW(r)/\sqrt{\int_0^1 J_c(r)^2 dr} \Rightarrow N(0,1)$. It implies that, when N is fixed but $\kappa \to \infty$, $t\left(Y,\kappa,h\right)$ converges to $N(0,1)$, since all the terms that involve $\exp(c)$ and $1/c$ vanish, and so does the initial condition.*

1.4 Confidence Interval for κ

1.4.1 Grid bootstrap confidence interval

The grid bootstrap was first proposed by Hansen (1999) under the local-to-unity AR(1) model. The grid bootstrap is considered for three reasons. First, Basawa et al. (1991) showed the conventional residual-based bootstrap methods fail to give correct first-order asymptotic coverage when the AR parameter is local-to-unity. An implication is that the conventional residual-based bootstrap methods are not valid in our model unless we assume $\kappa > 0$, h is fixed and $N \to \infty$. On the other hand, Hansen (1999) showed that the grid bootstrap CIs have asymptotically correct coverage under the local-to-unity case. Under the in-fill scheme, Model (7) is a local-to-unity AR(1) model, giving us strong motivations to use the grid bootstrap for Model (4). Second, the grid bootstrap method is used due to its uniform validity in the parameter space, as pointed out by Mikusheva (2007). Third, as we will show, the grid bootstrap method has the property of uniform validity across the three asymptotic schemes and obtains the same valid CI across the three asymptotic schemes.

Here we show how to use the grid bootstrap procedure to generate bootstrap samples. Consider generating the following AR(1) pseudo time series $\{y_t^*\}_{t=0}^T$ with error u_t^* conditional on κ:

$$y_t^* = \rho_h(\kappa)y_{t-1}^* + \tilde{g}_h + \hat{\sigma}_c\lambda_h u_t^*, y^*(0) = y_0 = O_p(1), \tag{10}$$

where $\rho_h(\kappa) = \exp(-\kappa h)$. Let $\hat{\sigma}_c := \sqrt{\frac{1}{Th}\sum_{t=1}^T(y_t - \hat{g}_h - \widehat{\rho}_h y_{t-1})^2}$, $\lambda_h := \sqrt{\frac{1-\exp(-2\kappa h)}{2\kappa}}$, and $\tilde{g}_h$ is obtained from regressing $y_t - \rho_h(\kappa)y_{t-1}$ on a constant. This way of obtaining $\tilde{g}_h$ is crucial since g_h explicitly depends on κ in our model, unlike the usual discrete-time AR(1) model with intercept. It is important to point out that when we generate the pseudo time series data, we explicitly retain the initial condition by letting $y^*(0) = y_0$. This is different from the bootstrap procedure in the usual discrete-time model where some initially simulated data are burned-in to avoid the dependence of the initialization. Since the initial condition shows in the in-fill asymptotic distribution, we design the bootstrap procedure so that it explicitly depends on the initial condition.

The error u_t^* is generated in the following way. We first define x_t as y_t/λ_h (conditional on a value of κ). Then we regress x_t on x_{t-1} and a constant by LS. Let $\{e_{x,t}\}_{t=1}^T$ be the LS residuals. We first scale residuals $\{e_{x,t}\}_{t=1}^T$ by multiplying $1/\hat{\sigma}_c$, then we re-center the scaled residuals. Finally, we independently draw u_t^* from the empirical distribution function of these re-centered and scaled residuals with replacement. Clearly, model (10) is a bootstrap version of Model (7) conditional on κ and with the same initial condition y_0.

We can then apply LS to the bootstrap samples to obtain $\hat{\rho}^*$, $\hat{\kappa}^*(:= -\ln(\hat{\rho}^*)/h)$ and the bootstrap coefficient-based statistic $z\left(Y^*, \kappa, h\right) = N\left(\hat{\kappa}_h^* - \kappa\right)$ where $Y^* = \{y_{th}^*\}_{t=1}^T$ is a bootstrap sample. We define the BCI as in (3). Since κ is our parameter of interest, we express the BCI for κ as $CI_q^* = \{\kappa \in R : c_T^*(x_1|\kappa) \leqslant z\left(Y, \kappa, h\right) \leqslant c_T^*(x_2|\kappa)\}$, and $c_T^*(q|\kappa)$ is the quantile function of $z\left(Y^*, \kappa, h\right)$, $x_1 = (1-q)/2$ and $x_2 = 1-(1-q)/2$.

The following lemma shows that $\hat{\sigma}_c^2$ is a consistent estimator of σ_ψ^2 under the in-fill scheme.

Lemma 2 *Under Model (7), as $h \to 0$,*

$$\sup_{\sigma>0}\sup_{\kappa\in R}\Pr\left(\left|\frac{\hat{\sigma}_c^2}{\sigma_\psi^2}-1\right|>\epsilon\right)\to 0,\ \textit{for any}\ \epsilon>0.$$

1.4.2 Asymptotic validity of grid bootstrap confidence interval

The following theorem shows that the grid bootstrap can produce BCIs which are asymptotically valid under the in-fill asymptotic scheme.

Theorem 1 *Let κ_0 be the true value of κ, and $\Pr^*$ be the bootstrap distribution with the error term drawn from our resampling method. Assume that*

1. $\kappa_0 \in K = [0, +\infty)$.
2. *The increment of the Lévy process $L(t+h)-L(t)$ has a finite variance and bounded r^{th} absolute moment with $r \in (2, 4]$.*
3. *$\mu, \sigma, i\psi'(0)$ and $\psi''(0)$ are all bounded by $C < \infty$.*
4. *Let $Y = \{y_{th}\}_{t=1}^{T}$,*

$$\begin{aligned}&(S^*(T,\kappa), R^*(T,\kappa))\\ =&\left(\frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^{T} y_{t-1}^*\epsilon_t^* - \frac{1}{\hat{\sigma}T}\sum_{t=1}^{T} y_{t-1}^*\frac{1}{\hat{\sigma}T}\sum_{t=1}^{T}\epsilon_t^*\ ,\ \frac{1}{\hat{\sigma}^2T^2}\sum_{t=1}^{T} y_{t-1}^{*2} - \left(\frac{1}{\hat{\sigma}T^{\frac{3}{2}}}\sum_{t=1}^{T} y_{t-1}^*\right)^2\right),\end{aligned}$$

with $\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{g}_h - \widehat{\rho}_h(\kappa)y_{t-1})^2}$.[2] *Let $E\left(R^*(T,\kappa)\right) = J$. Assume that the pair of statistics $(S^*(T,\kappa), R^*(T,\kappa))$ has a uniformly continuous distribution over the parameter space K, such that for any $\epsilon > 0$, there exists a constant $M > 0$ such that for all $\delta_1 < \varepsilon, \delta_2 < \varepsilon, |b - J| > 2\varepsilon$ and all $\kappa \in K$, we have*

$$\begin{aligned}\Pr{}^*\{(S^*(T,\kappa), R^*(T,\kappa)) &\in [a-\delta_1, a+\delta_1]\times[b-\delta_2, b+\delta_2]|Y\} \leqslant M\delta_1\delta_2,\\ \Pr{}^*\{S^*(T,\kappa) &\in [a-\delta_1, a+\delta_1]|Y\} \leqslant M\delta_1.\end{aligned}$$

[2]Note that $S^*(T,\kappa)/R^*(T,\kappa) = z\,(Y^*, \rho, T)$.

Under these assumptions, we have, as $h \to 0$,

- $\sup_{\kappa \in K} \sup_{x} |\Pr\{z(Y, \kappa, h) < x|\kappa\} - \Pr^*\{z(Y^*, \kappa, h) < x|\kappa, Y\}| \to 0.$
- $\inf_{\kappa \in K} \Pr\{\kappa_0 \in CI_q^*\} \to x_2 - x_1 = q.$

The first assumption requires the parameter space of κ in the nonnegative half-line. While the in-fill asymptotic theory does not require κ to be nonnegative, for most economic and financial time series, the focus has been on cases where $\kappa \geqslant 0$. Therefore, we restrict our attention to the nonnegative region of κ. Assumptions 2 and 3 effectively regulate the error term in the exact discrete-time model, enabling us to apply the invariance principle. Assumption 4, which restricts the component of our test statistic to be (jointly) uniformly continuous, is also used in Mikusheva (2007).

The first result shows that the distribution of the bootstrap statistic is close to the finite sample distribution uniformly over the parameter space K, when the sampling interval is smaller. In the limit of $h \to 0$, the bootstrap statistic behaves like a random variable whose distribution is the in-fill asymptotic distribution. The second result shows that the coverage probability of CI_q^* converges to q when $h \to 0$.

Remark 7 *If we replace $z(Y, \kappa, h)$ and $z(Y^*, \kappa, h)$ in Theorem 1 by $t(Y, \kappa, h)$ and $t(Y^*, \kappa, h)$, Theorem 1 remains valid. This implies that we can use the t statistic to obtain BCIs which are also justifiable under the in-fill scheme.*

Remark 8 *In Model (4), only the consistency of σ_ψ is required to ensure the asymptotic validity of BCI. No consistent estimation for $(\kappa, \mu, \sigma, \psi'(0), \psi''(0))$ is needed for the purpose of constructing an asymptotically valid BCI for κ as $h \to 0$. This is because, in our bootstrap method, the exact discretization (7) of model (4) is what we try to mimic, and we only need to ensure the consistent estimation for the discretized parameters such as g_h/λ_h and σ_ψ.*[3]

Remark 9 *While the asymptotic justification of the grid bootstrap has been made under the in-fill scheme, it can be also made under the long-span scheme and the double scheme.*

[3]We establish the consistency of g_h/λ_h conditional on κ.

Under the long-span scheme, if $\kappa > 0$, $\rho_h(\kappa) = \rho(\kappa) < 1$, leading to a strictly stationary AR model. In this case the validity of the grid bootstrap was proven in Theorem 1 of Hansen (1999). Under the double scheme, if $\kappa > 0$ and $N = O(\log T)$, $\rho_h = \exp(-\kappa h)$ and $T(1-\rho_h) = T(\kappa h + o(h^2) \sim \kappa N = O(\log T)$. In this case the AR coefficient falls into the mildly stationary region defined in Mikusheva (2007) where the validity of the grid bootstrap was shown. When $\kappa = 0$, $\rho_h(\kappa) = 1$ and model (4) can be expressed by the following AR(1) model:

$$y_{th} = y_{(t-1)h} + \sigma_\psi \lambda_h u_{th}, \text{ with } y_0 = O_p(1).$$

Under the long-span scheme, the model is a usual unit root AR(1) model without intercept. Under the double scheme, we can write the model as

$$x_{th} = x_{(t-1)h} + u_{th},\ x_{th} = x_0 + \sum_{i=1}^{t} u_{ih}, \tag{11}$$

with $x_{th} = y_{th}/(\sigma_\psi \lambda_h)$ and $x_0 = O_p(h^{-1/2})$. Since $T^{-1/2}x_{th} = T^{-1/2}\sum_{i=1}^{t} u_{ih} + T^{-1/2}x_0$ with $T^{-1/2}x_0 = O_p(N^{-1/2}) = o_p(1)$, we have the usual limiting Brownian motion. Model (11) is a unit root model with an asymptotically negligible initial condition under the double scheme. In both cases, the model can be regarded as an AR(1) model with a local-to-unit-root. Hansen (1999) and Mikusheva (2007) showed the validity of the grid bootstrap in the local-to-unit-root region. In sum, for $\kappa \geqslant 0$, the grid bootstrap can be justified under the long-span scheme and the double scheme.

Remark 10 *Under the in-fill scheme, we define the coefficient-based statistic as $N(\hat{\kappa}_h - \kappa)$. Under the long-span scheme and the double scheme, the normalized statistic should be defined as $\sqrt{T}(\hat{\kappa} - \kappa)$ and $\sqrt{N}(\hat{\kappa} - \kappa)$, respectively. However, using the grid bootstrap method, The BCI is obtained by inverting the coefficient-based statistic. Therefore, the construction of BCI is independent of the selected normalization.*

1.4.3 Expansions and refinements

An important advantage of bootstrap methods over asymptotic distributions is that bootstrap methods often provide refinements in finite samples. This feature also holds true in our model. To prove refinements, we follow Park (2003) and Mikusheva (2015) by developing the second-order probabilistic expansions of the coefficient-based test statistic. The expansions were obtained in Park (2003) for both the t statistic and the coefficient-based statistic around their respective Dickey-Fuller-Phillips distributions which are pivotal. The expansions were obtained in Mikusheva (2015) for the t statistic around $\frac{\int_0^1 J_c(r)dW}{\sqrt{\int_0^1 J_c(r)^2 dr}}$ which is non-pivotal but independent of the initial condition. Our leading term is the in-fill asymptotic distribution, which is not only non-pivotal but also dependent on the initial condition. Although we only report the results for the coefficient-based test statistic, it can be shown that similar expansions can be developed for the t statistic for κ.

Theorem 2 *Assume that in Model (4), the assumptions in Theorem 1 hold, and additionally, the increment of the Lévy process $L(t+h) - L(t)$ has a bounded r^{th} moment for some $r \geqslant 8$. We have the following probabilistic expansions for $z(Y, \kappa, h)$*

$$z(Y, \kappa, h) = z^{y_0}(\kappa, \theta) + T^{-1/4}A + T^{-1/2}B + o_p(T^{-1/2}), \tag{12}$$

where the leading term $z^{y_0}(\kappa, \theta)$ is the in-fill asymptotic distribution given in (9), and the full expressions of the higher order terms A and B which are all $O_p(1)$, are provided in the appendix.

Furthermore, for grid bootstrap method, we have the following results for distributional expansions

$$\sup_x |\Pr{}^*(z(Y^*, \kappa, h) < x|\kappa, Y) - \Pr(z(Y, \kappa, h) < x|\kappa)| = o(T^{-1/2}), \tag{13}$$

where $Y^* = \{y_t^*\}_{t=0}^T$ is our bootstrap sample.

Remark 11 *When $\psi'(0) = 0$, $\psi''(0) = 1$, $y_0 = \mu$, $\kappa = 0$, $z^{y_0}(\kappa, \theta) = -\int_0^1 \overline{W}(r)dW(r)/\int_0^1 \overline{W}(r)^2 dr$. Equation (12) extends the result on G_n in Park (2003) from the unit root model without intercept to the unit root model with intercept. When $\psi'(0) = 0$, $\psi''(0) = 1$, $y_0 = \mu$, $z^{y_0}(\kappa, \theta) = -\int_0^1 \overline{J}(r)dW(r)/\int_0^1 \overline{J}(r)^2 dr$. Equation (12) extends the result on $t(y, n, \rho_n)$ in Mikusheva (2015) from the local-to-unity model with negligible initial condition to the local-to-unity model with divergent initial condition.*

Remark 12 *According to (12), we have*

$$\Pr(z(Y,\kappa,h) < x|\kappa) = \Pr(z^{y_0}(\kappa,\theta) < x|\kappa) + O(T^{-1/2}), \tag{14}$$

uniformly in x. This suggests that our second-order asymptotic expansions of $z(Y,\kappa,h)$, that is, $z^{y_0}(\kappa,\theta) + T^{-1/4}A + T^{-1/2}B(:= \xi)$, provide refinements of the in-fill asymptotic distribution up to order $o(T^{-1/2})$ since

$$\Pr(z(Y,\kappa,h) < x|\kappa) = \Pr(\xi < x|\kappa) + o(T^{-1/2}).$$

Remark 13 *Comparing equation (13) with equation (14), the grid bootstrap provides a second-order improvement over the in-fill asymptotic distribution.*

Remark 14 *Under an AR(1) model with the AR parameter $\rho = 1 + cm/T$, Phillips et al. (2010) obtained the local-to-unit-root distribution when $T \to \infty$ with a fixed m. They also showed that the local-to-unit-root distribution makes a first-order refinement of the double asymptotic distribution when $m \to \infty$ sequentially. This sequential asymptotic scheme ($T \to \infty$ followed by $m \to \infty$) is the scheme where $N \to \infty$ followed by $h \to 0$ in Model (7). Therefore, it is expected that the grid bootstrap provides an improvement over the double asymptotic distribution.*

1.4.4 Extensions to heteroskedastic models

It is possible to extend the grid bootstrap methods to more general model specifications. Here we discuss a model with time-varying volatility given by

$$dy(t) = \kappa(\mu - y(t))dt + \sigma(t)dL(t), \tag{15}$$

where $\sigma(t) = \omega(t/T)$ and ω is a measurable function on the interval $(0,1]$ such that both the infimum and the supremum of ω over $(0,1]$ is a bound strictly above 0 and below infinity and ω satisfies the Lipschitz condition except at a finite number of points of discontinuity. To keep our exposition simple, we assume $\psi'(0) = 0$. The exact discrete-time model is given by

$$y_t = \rho_h(\kappa)y_{t-1} + g_h + \sigma_t\lambda_h u_t, \tag{16}$$

As noted in Xu and Phillips (2008), a general deterministic function for ω and, hence, unconditional heteroskedasticity are allowed in the model. However, a general stochastic volatility process is not allowed.

The in-fill asymptotic distribution for $N(\hat{\kappa}_h - \kappa)$ is developed and documented in Lemma 16 of Appendix. It turns out that one can apply the wild bootstrap principle with the grid bootstrap method to generate a bootstrap sample. Analogous to (10), conditional on κ, we now generate pseudo data in the following manner:

$$y_t^* = \rho_h(\kappa) y_{t-1}^* + \tilde{g}_h + \lambda_h \frac{e_t}{\sqrt{h}} z_t^*, y^*(0) = y_0 = O_p(1), \tag{17}$$

where e_t and z_t^* are the LS residual and an i.i.d. N(0,1) random variable, respectively. The remaining steps are identical to the case where σ is a constant. The following proposition documents the asymptotic validity of the grid bootstrap method under the in-fill scheme.

Proposition 1 *Under model (15), suppose we generate pseudo data as in (17), and construct a grid BCI. As* $h \to 0$, *we have*

$$\lim_{h\to 0} \Pr\{\kappa \in CI_q^*\} \to x_2 - x_1 = q.$$

In proving Proposition 1, as we do not have the i.i.d. assumption for the error term due to the time-varying volatility, we cannot use the stochastic embedding with strong approximation as we did for Theorem 1. As a result, we do not have a uniform convergence result as in Theorem 1.

1.5 Simulation Studies

1.5.1 Implementation

Before we design experiments to check the performance of the grid bootstrap, we give the following 7 steps to construct a grid bootstrap CI for κ:

1. Given the data $\{y_{th}\}_{t=0}^T$, we run the following regression by LS:

$$y_{th} = \hat{\rho}_h y_{(t-1)h} + \hat{g}_h + e_{th},$$

where e_{th} is the LS residual. And we use $\{e_{th}\}_{t=1}^T$ to construct the consistent estimator for σ_ψ^2 by $\frac{1}{Th}\sum_{t=1}^T e_{th}^2$ (denoted as $\hat{\sigma}_c^2$).

2. Construct a grid of ρ_h, $A_G = \{\rho_{h1}, \rho_{h2}, ...\rho_{hG}\}$, centered at $\hat{\rho}_h$, with the first and last grid point being calculated from $\hat{\rho}_h \pm 5 \times se(\hat{\rho}_h)$.

3. Given a point in the grid ($\rho_{hG} \in A_G$), perform the second regression:

$$y_{th} - \rho_{hG} y_{(t-1)h} = \tilde{g}_h + \nu_t,$$

   where ν_t is the residual of the second regression. Note that $\tilde{g}_h$ is a function of ρ_{hG}.

4. Let $\kappa_G = -\frac{\ln(\rho_{hG})}{h}$, $\lambda_{hG} = \sqrt{\frac{1-\exp(-2\kappa_G h)}{2\kappa_G}}$, and u^*_{th} be generated according to section 1.4. We generate the bootstrap data $\{y^{*b}_{th}\}^T_{t=1}$ based on $\{u^*_{th}\}^T_{t=1}$ and the same initial condition as the observed data, i.e.,

$$y^*_{th} = \rho_{hG} y^*_{(t-1),h} + \tilde{g}_h + \hat{\sigma}_c \lambda_{hG} u^*_{th}, y^*_0 = y_0.$$

5. Generate B sets of bootstrap data, such that we have $\left\{\{y^{*b}_{th}\}^T_{t=1}\right\}^B_{b=1}$. For every set of bootstrap data, obtain the LS estimator of κ (denoted by $\hat{\kappa}^*_h$) and calculate the bootstrap coefficient-based statistic $z\left(Y^*, \kappa_G, h\right) = N(\hat{\kappa}^*_h - \kappa_G)$. Calculate the x^{th} quantile of the bootstrap statistic $z\left(Y^*, \kappa_G, h\right)$ to obtain $c^*_T(x|\kappa_G)$.

6. Following Hansen (1999), we estimate the quantile function $c^*_T(x|\kappa)$ by applying the kernel regression:

$$c^*_T(x|\kappa) = \frac{\sum_{g=1}^G K\left(\frac{\kappa - \kappa_G}{\delta}\right) c^*_T(x|\kappa_G)}{\sum_{g=1}^G K\left(\frac{\kappa - \kappa_G}{\delta}\right)},$$

   where $K(\cdot)$ is a kernel function and δ is a bandwidth. In our application and simulation, we use the Epanechnikov kernel ($K(x) = \frac{3}{4}(1 - x^2)1(|x| \leqslant 1)$) and choose the bandwidth by LS cross-validation.

7. The CI for κ is obtained by inverting the coefficient-based statistic:

$$CI^B_q = \left\{\kappa \in R : c^*_T(x_1|\kappa) \leqslant z\left(Y, \kappa, h\right) \leqslant c^*_T(x_2|\kappa)\right\}.$$

1.5.2 Comparing CIs in finite samples

To evaluate the performance of the proposed bootstrap methods in the continuous-time model, we construct CIs with the 95% coverage probability using the long-span asymptotic distribution, the in-fill asymptotic distribution, and the grid bootstrap method. To do so, we consider three-parameter settings to generate data (called DGP1 to DGP3) and simulate discrete-time observations with sampling interval h from Model (4) where the Lévy process is the variance gamma process with $v = 0.5$ in DGP1 and DGP2 and $v = 1$ in DGP3. In particular we set $\psi'(0) = 0$ and $\psi''(0) = 1$ in DGP1, $i\psi'(0) = 0.05$ and $\psi''(0) = 1$ in DGP2 and $i\psi'(0) = 0.2$ and $\psi''(0) = 3$ in DGP3. For DGPs 1-3, the in-fill asymptotic distribution is not feasible as we do not have estimates of $\psi'(0)$ and $\psi''(0)$. The following parameter settings are considered, $\kappa \in \{0.01, 0.1, 1\}$, $h = 1/12$, $N = 5$, $\mu = 0.1$, $\sigma = 1$, $y_0 = 0.1$. The number of replications is always set at 10,000.

We use the following methods to construct the 95% CI for κ:

1. In-fill asymptotic distribution. Since the in-fill distribution depends on κ, μ, σ and the 2 derivatives of ψ, we simply set the values of these parameters to their true values. This approach is infeasible in practice. It is only considered as a benchmark for the purpose of evaluating the performance of other methods.

2. Grid bootstrap method. To calculate BCIs we set the number of bootstrap iterations $B = 399$ with grid size $G = 50$.

3. Long-span asymptotic distribution, that is, $N\left(0, \left(\exp(2\kappa h) - 1\right)/h\right)$.

The Monte Carlo average is used to calculate the empirical coverage of the true value (κ_0), i.e., $\frac{1}{10000}\sum_{m=1}^{10000} 1\left(\kappa_L^{(m)} \leqslant \kappa_0 \leqslant \kappa_U^{(m)}\right)$, where $\kappa_L^{(m)}$ and $\kappa_U^{(m)}$ are the bounds of a CI in the m^{th} replication, $1(\cdot)$ is the indicator function which indicates whether the true value κ_0 is contained in the interval. The closer the empirical coverage to 95%, the better the performance of the method. Table 2 reports the empirical coverage and the absolute difference between the nominal coverage and the empirical coverage for alternative methods when $h = 1/12$. Numbers in the bold-face indicate that the corresponding methods have the best performance (in terms of the absolute difference) in each of the parameter settings.

Several interesting conclusions can be found from Table 2. First, it can be seen that CIs obtained from the long-span asymptotic distribution have very bad performance across all

Table 2: 95% Confidence Intervals ($h = 1/12$)

		$\kappa_0 = 0.01$		$\kappa_0 = 0.1$		$\kappa_0 = 1$	
DGP1	Long-span	0.019	(0.931)	0.062	(0.889)	0.272	(0.679)
	In-fill	0.941	(0.009)	0.940	(0.060)	0.919	(0.031)
	Grid bootstrap	**0.952**	(0.002)	**0.953**	(0.003)	**0.948**	(0.002)
DGP2	Long-span	0.020	(0.93)	0.063	(0.887)	0.270	(0.68)
	In-fill	0.943	(0.007)	0.940	(0.060)	0.921	(0.029)
	Grid bootstrap	**0.953**	(0.003)	**0.951**	(0.001)	**0.949**	(0.001)
DGP3	Long-span	0.021	(0.929)	0.070	(0.880)	0.281	(0.669)
	In-fill	0.940	(0.010)	0.936	(0.014)	0.922	(0.028)
	Grid bootstrap	**0.954**	(0.006)	**0.952**	(0.002)	**0.946**	(0.004)

DGPs. Although the difference between the nominal and the actual coverage diminishes when κ_0 increases, the problem of under-coverage is very serious. The simulation results simply suggest that, in these empirically realistic settings, the long-span asymptotic theory should not be used to construct a CI for κ. This conclusion echoes that in Zhou and Yu (2015) and Bao et al. (2017). Second, for the in-fill asymptotic theory, the empirical coverage is much closer to the nominal one. Again, this conclusion echoes that in Zhou and Yu (2015) and Bao et al. (2017). However, This method is infeasible. Finally, the grid bootstrap method always performs the best and the coverage is always very close to 95%. Regardless of κ_0, it tends to outperform the in-fill asymptotic distribution in all DGPs, consistent with the prediction of Theorem 2.

1.6 Empirical Studies

In this section, we apply the proposed grid bootstrap method to construct BCIs for κ in Model (1) and in Model (4) when these two models are fitted using the monthly Federal fund effective rate and the logarithmic volatility index of Chicago Board Options Exchange's (VIX). In addition to BCI, we also obtain two CIs of κ, one based on the long-span asymptotic distribution and the other based on the in-fill asymptotic distribution when the model is assumed to be (1).[4] We assume that the initial condition y_0 is the same

[4]In this case, we obtain the CI by replacing the unknown κ, μ, and σ with their estimates in the in-fill asymptotic distribution.

as the first observation.

1.6.1 Federal fund effective rate

The Federal fund effective rate data are available from H-15 Federal Reserve Statistical Release and cover the period from July 1954 to December 2017. In total, there are 762 observations with $T = 762$, $h = 1/12$ and $N = 63.5$. Similar datasets over different sample periods were used in Aït-Sahalia (1999) and Zhou and Yu (2015).

The LS estimates of ρ_h, g_h μ, and κ in Model (1) are: $\hat{\rho}_h = 0.99$, $\hat{g}_h = 0.0005$, $\hat{\mu} = 0.0493$, and $\hat{\kappa}_h = 0.1201$. The constructed CIs for κ are reported in Table 3. It can be seen that the CI constructed from the long-span distribution is very different from the other two CIs. It excludes zero, suggesting that we have to reject the unit root null hypothesis under the long-span scheme. However, the other two CIs all contain zero, suggesting that we cannot reject the unit root null hypothesis. While both the BCI and the CI implied by the in-fill asymptotic distribution contain zero, BCI is much narrower than the CI implied by the in-fill asymptotic distribution.

Table 3: Coverage of 90% and 95% confidence intervals for the interest rate data

	90% C.I.	95% C.I.
Long-span	(0.0908, 0.1495)	(0.0852, 0.1551)
In-fill	(-0.1505, 0.2191)	(-0.2050, 0.2448)
Grid bootstrap	(-0.0319, 0.1785)	(-0.0435, 0.2005)

1.6.2 VIX

The CBOE VIX data is available from yahoo.finance.com and contains daily observations from 4^{th} January 2010 to 31^{st} December 2019. In total, there are 2516 observations with $T = 2516$, $h = 1/252$ and $N = 9.98$.

The LS estimates of ρ_h, g_h, μ, and κ in Model (1) are: $\hat{\rho}_h = 0.965$, $\hat{g}_h = 0.098$, $\hat{\mu} = 2.775$, and $\hat{\kappa}_h = 9.0736$. The constructed CIs for κ are reported in Table 4. It can be seen that the CI constructed from the in-fill distribution is very different from the other two CIs. It includes zero, suggesting that we cannot reject the unit root null hypothesis under the in-fill scheme. However, the other two CIs all exclude zero, suggesting that we have found evidence of stationarity in log-VIX. While the BCI and the CI implied by the long-span asymptotic distribution exclude zero, the BCI is much wider than the CI implied by the long-span asymptotic distribution.

Table 4: Coverage of 90% and 95% confidence intervals for the VIX data

	90% C.I.	95% C.I.
Long-span	(8.9314, 9.2159)	(8.9041, 9.2431)
In-fill	(-13.2442, 29.6948)	(-17.5557, 34.3721)
Grid bootstrap	(6.5000, 11.0678)	(6.1042, 11.4316)

1.7 Conclusion

In this paper, we discuss the advantages and drawbacks of using three asymptotic distributions obtained from the long-span, double and in-fill schemes for constructing CIs of persistence parameter κ under a Lévy-driven OU model. The long-span and double schemes provide a poor finite sample performance. Moreover, the long-span and double schemes lead to an asymptotic distribution which is not continuous in κ as κ passes zero. On the other hand, although the in-fill scheme leads to an asymptotic distribution which is closer to the finite sample distribution than their long-span and double counterparts and is continuous in κ, it is infeasible.

We propose to use the grid bootstrap method for three reasons. First, unlike asymptotic methods, which depend on a particular scheme, the grid bootstrap can be justified by any of the three asymptotic schemes. Second, it is asymptotically valid when κ is close to or equal to zero. Finally, it provides a finite sample improvement over the in-fill distribution. To show this finite sample improvement, we follow Park (2003) and Mikusheva (2015) by developing probabilistic expansions to the coefficient-based statistic around the in-fill distribution. Via the second-order expansion we show that the grid bootstrap method provides refinement of the in-fill asymptotic distribution up to order $o(T^{-1/2})$. The in-fill asymptotic justification of the grid bootstrap only requires the consistency of σ_ψ which is ensured under the in-fill scheme. No consistent estimation of other parameters in the model is needed under the scheme.

Monte Carlo studies reveal several important results. First, the CIs implied by the long-span asymptotic distribution lead to serious under-coverage in all cases considered. Second, the gird bootstrap method performs better than the in-fill asymptotic theory and much better than the long-span theory.

The empirical application to the U.S. interest rate data shows that the unit root hypothesis cannot be rejected by the bootstrap CIs and the CI obtained from the in-fill asymptotic distribution, but has to be rejected by the CI obtained from the long-span asymptotic distribution. The empirical application to CBOE's VIX data shows that the unit root hypothesis is rejected by the bootstrap CIs and the CI obtained from the long-span asymptotic distribution, but cannot be rejected by the CI obtained from the in-fill asymptotic distribution.

2 Mildly Explosive Autoregression with Anti-persistent Errors

2.1 Introduction

The autoregressive (AR) model with an explosive root was first studied in White (1958) and Anderson (1959) where the following process was considered:

$$y_t = \rho y_{t-1} + u_t,\ \rho > 1,\ t = 1, 2, ..., n. \tag{18}$$

Under the assumptions of independent and identically distributed (iid) Gaussian errors (i.e. $u_t \overset{iid}{\sim} N(0, \sigma^2)$) and the zero initial condition (i.e. $y_0 = 0$), White (1958) and Anderson (1959) showed that the least squares (LS) estimate of ρ (denoted by $\hat{\rho}$) has the following Cauchy limiting distribution:

$$\frac{\rho^n}{\rho^2 - 1}(\hat{\rho} - \rho) \overset{as}{\to} C, \text{ as } n \to \infty, \tag{19}$$

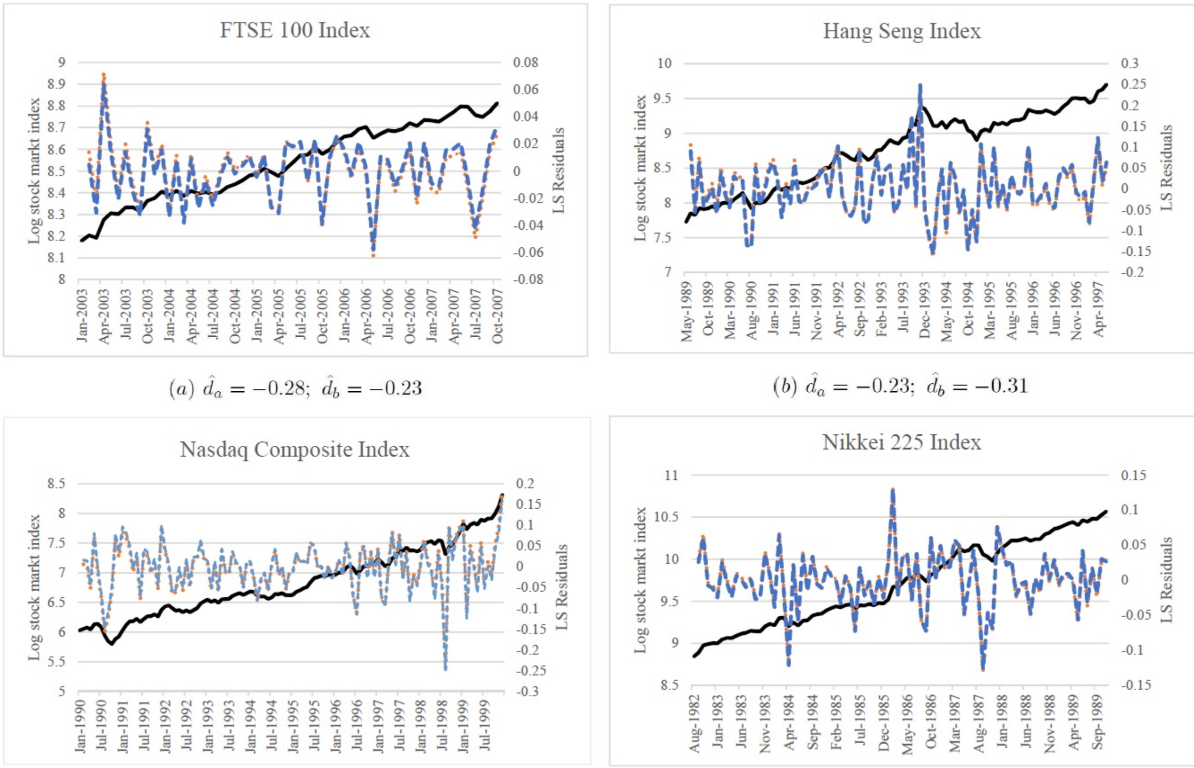
where $\overset{as}{\to}$ denotes the convergence almost surely and C is a standard Cauchy variate.

It is noteworthy that the above limit theory is not obtained from an invariance principle because the distributional assumption $u_t \overset{iid}{\sim} N(0, \sigma^2)$ cannot be relaxed.[5] To relax the assumption of Gaussian errors, and, in the meantime, to allow for a non-zero initial condition, Phillips and Magdalinos (2007a) (PM hereafter) and Phillips, Magdalinos and Giraitis (2010) (PMG hereafter) considered two variations which are analogous to Model (18). PM designed a mildly explosive AR model by letting $\rho = \rho_n = 1 + c/n^\alpha$, $c > 0$, $\alpha \in (0, 1)$, while PMG introduced a mildly explosive model by letting $\rho = \rho_{m,n} = 1 + cm/n$, $c > 0$. Under some suitable assumptions but without the requirements of Gaussian errors and the zero initial condition, applying different forms of the central limit theorem and the martingale convergence theorem, PM and PMG obtained the asymptotic theory:

$$\frac{\rho_n^n}{\rho_n^2 - 1}(\hat{\rho} - \rho_n) \Rightarrow C, \text{ as } n \to \infty; \tag{20}$$

$$\frac{\rho_{m,n}^n}{\rho_{m,n}^2 - 1}(\hat{\rho} - \rho_{m,n}) \Rightarrow C, \text{ as } n \to \infty \text{ followed by } m \to \infty. \tag{21}$$

[5]This is because what is used to derive Equation (19) is the martingale convergence theorem which gives the almost sure convergence.



$(a)\ \hat{d}_a = -0.28;\ \hat{d}_b = -0.23$

$(b)\ \hat{d}_a = -0.23;\ \hat{d}_b = -0.31$

$(c)\ \hat{d}_a = -0.32;\ \hat{d}_b = -0.36$

$(d)\ \hat{d}_a = -0.31;\ \hat{d}_b = -0.22$

Figure 1: Time series plot of four logarithmic stock market indices (left axis) and their residuals obtained from the fitted AR(1) model with and without intercept by LS (right axis)

The pivotalness of the Cauchy distribution suggests that it is easy to test a hypothesis about the AR coefficient. Not surprisingly, it has been used in the literature to test the presence of rational bubbles in asset prices; see Phillips, Wu and Yu (2011). Moreover, considerable efforts have been made in the literature to explore the explosive-type AR models with dependent errors. The errors could be weakly dependent as in Phillips and Magdalinos (2007b), or strongly dependent as in Magdalinos (2012), or could involve conditional heteroskedasticity as in Arvanitis and Magdalinos (2018). These generalizations are important as the explosive-type model with dependent errors can potentially better describe the movement of real data than the pure explosive AR(1) model. A number of related studies in the literature allow for m-dependent errors (Pedersen and Schütte, 2017), errors with deterministic time-varying volatilities (Harvey, Leybourne and Zu, 2019a, 2019b).

To the best of our knowledge, no limit theory has been developed to cover any explosive-type AR model with anti-persistent errors. The goal of this paper is to fill the gaps in the context of the explosive-type AR model of PMG. Why are the gaps important? To see the empirical relevance of an explosive model with anti-persistent errors, Figure 3 presents time series plots of four logarithmic stock market indices (left axis) and the residuals obtained from the fitted AR(1) model with and without intercept (right axis). In particular, we consider four monthly indices over different sampling periods, namely FTSE 100 Index from January 2003 to October 2007, Hang Seng Index from May 1989 to June 1997, NASDAQ Composite Index from January 1990 to December 1999, and Nikkei 225 Index from August 1982 to November 1989. The sampling periods are selected as these markets experienced exuberance over the respective periods, as it can be seen from the solid black lines in Figure 3. After fitting the AR(1) model with and without intercept to each time series by LS, we obtain two residual series with and without intercept and plot them in the blue and red dotted lines in Figure 3. These plots show that there is strong anti-persistence in the residuals.[6] When we apply the local Whittle (LW) method of Robinson (1994) to estimate the memory parameter d in the residuals, we find that the estimated d is always in the range $(-0.5, 0)$ in all cases. The estimated d is reported in Figure 3 with $\hat{d}_a$ and $\hat{d}_b$ corresponding to the model without and with intercept, respectively. These exercises strongly suggest that the explosive-type AR model with anti-persistent errors is not only of theoretical interest but also of empirical realism, making important the development of limit theory for an explosive-type AR model with anti-persistent errors.

The paper is organized as follows. Section 2.2 briefly reviews several forms of serially dependent error processes and mildly explosive AR models. Section 2.3 studies the mildly explosive AR model of PMG but with anti-persistent errors and develops the limiting distribution for the LS estimate of the AR coefficient under a sequential limit. Simulation studies are carried out in Section 2.4 to check the precision of the limiting distribution in finite samples. Section 2.5 provides an empirical study of a rational bubble in the NASDAQ index. Proofs of the main results in the paper are given in the Appendix.

[6] A detailed discussion on anti-persistence is provided in the next section where we also relate anti-persistence to the memory parameter d.

We use the following notations throughout the paper: $\xrightarrow{p}$, $\xrightarrow{as}$, $\Rightarrow$, $\overset{a}{\sim}$, $\overset{d}{=}$ and $\overset{iid}{\sim}$ denote convergence in probability, convergence almost surely, weak convergence, asymptotic equivalence, equivalence in distribution, and iid, respectively.

2.2 Literature Review

2.2.1 A review of serially correlated errors

Although our paper focuses on anti-persistent errors, to facilitate discussion and comparison, we first review the concepts of weakly dependent errors and strongly dependent errors. Suppose that the error process admits a Wold-decomposition such that

$$u_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, c_0 = 1, \ \epsilon_t \overset{iid}{\sim} (0, \sigma^2), \tag{22}$$

where $\{c_j\}_{j=0}^{\infty}$ are real coefficients. Denote $\psi(k)$ the k^{th} order autocovariance function of u_t, that is, $\psi(k) := E(u_t u_{t-k})$.

Weakly dependent errors require $\sum_{j=0}^{\infty} |c_j| < \infty$ and $\sum_{j=0}^{\infty} c_j \neq 0$. These conditions imply that $\sum_{k=-\infty}^{\infty} |\psi(k)| \in (0, \infty)$ and $\sum_{k=-\infty}^{\infty} \psi(k) \neq 0$. For strongly dependent errors, it is assumed that c_j in (22) has a slow decay rate, such as $c_j \sim j^{-1+d}$ with $d \in (0, 0.5)$ when j is large. This leads to a violation of the summability condition of the linear coefficients and the autocovariance function as $\sum_{j=0}^{\infty} c_j = \infty$ and $\sum_{k=-\infty}^{\infty} |\psi(k)| = \infty$.

Anti-persistent errors are remarkably different from weakly dependent errors and strongly dependent errors. First, they are different from strongly dependent errors as c_j has a fast decay rate for anti-persistent errors, such as $c_j \sim j^{-1+d}$ with $d \in (-0.5, 0)$ when j is large. Second, they are different from weakly dependent errors in the sense that $\sum_{j=0}^{\infty} c_j = 0$ and $\sum_{k=-\infty}^{\infty} \psi(k) = 0$. Moreover, for any $k \neq 0$, $\psi(k)$ has a negative sign (see Proposition 3.2.1 (3) in Giraitis, Koul, and Surgailis (2012)), giving rise to the name of anti-persistence. These properties make the interpretation of corresponding stochastic integrals different from that when the errors are weakly dependent or strongly dependent. From the theoretical viewpoint, therefore, it is important to develop the limit theory for anti-persistent errors.

We now formally introduce the definition of anti-persistence.

Assumption 1 (AP) *Under (22) and let γ be a constant. Assume $c_j \overset{a}{\sim} \gamma j^{-1+d}$ for $j \to \infty$ with $d \in (-0.5, 0)$, $\sum_{j=0}^{\infty} c_j = 0$ and $\sum_{k=-\infty}^{\infty} \psi(k) = 0$.*

Assumption AP is general enough to include stationary ARFIMA(p, d, q) processes where $u_t = (1-L)^{-d}\phi(L)^{-1}\theta(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$, $\phi(L) = 1 - \sum_{j=0}^{p} \phi_j L^j$, $\theta(L) = 1 + \sum_{j=0}^{p} \theta_j L^j$ and L is the lag operator. We can show that c_j can be asymptotically approximated by $\frac{\theta(1)}{\phi(1)\Gamma(d)} j^{-1+d}$, where $\Gamma(\cdot)$ is a gamma function. When $d \in (-0.5, 0)$, the stationary ARFIMA process has the zero-sum for the linear coefficients, that is, $\sum_{j=0}^{\infty} c_j = 0$. It is well-known that u_t corresponds to a fractional Brownian motion (fBM) with the Hurst parameter $H = 1/2 + d$; see Giraitis, Koul and Surgailis (2012). When $H = 0.5$, an fBM becomes the standard Brownian motion. When $H \in (0, 0.5)$ which corresponds to the case of interest in the present paper, an fBM has a rough sample path and is anti-persistent. When $H \in (0.5, 1)$, an fBM has a smooth sample path in the sense that it is $1/2 - \epsilon$-Hölder continuous for any $\epsilon > 0$. The empirical relevance of anti-persistent processes in financial time series was recently documented in Gatheral, Jaisson, and Rosenbaum (2018) and Wang, Xiao, and Yu (2019). The empirical relevance of anti-persistent errors in an explosive model was shown earlier in Figure 3. Assuming a continuous record of observations is available, Xiao and Yu (2019a, 2019b) recently developed the limit theory for the persistence parameter in the fractional Vasicek model which corresponds to the AR coefficient in the discrete-time representation.

2.2.2 A mildly explosive model

PMG considered the following mildly explosive model:

$$y_t = \left(1 + \frac{cm}{n}\right) y_{t-1} + u_t, c > 0, u_t \overset{iid}{\sim} (0, \sigma^2), y_0 = O_p(1). \tag{23}$$

As suggested in PMG, one way of thinking of the model specification is that the total number of observations (n) is partitioned into m blocks with K samples so that $n = m \times K$. Thus, the chronological time for y_t becomes $t = \lfloor Kj \rfloor + k$, for $k \in \{1, ..., K\}$ and $j \in \{0, 1, ...m - 1\}$. This model is closely related to the model proposed in Park (2003) where it was assumed that $c = -1 < 0$.

It is easy to see that as $n \to \infty$ with fixed m, Model (23) is a local-to-unity model with the noncentrality parameter cm and hence, the standard local-to-unity asymptotic theory is applicable. That is,

$$n(\hat{\rho} - \rho_{n,m}) \Rightarrow \int_0^1 J_{cm}(s)dW(s) / \int_0^1 J_{cm}^2(s)ds,$$

where $J_{cm}(s) = \int_0^s e^{cm(s-r)}dW(r)$ and $W(\cdot)$ denotes a standard Brownian motion.

However, since $c > 0$, if one assumes $n \to \infty$ followed by $m \to \infty$, Model (23) is akin to a mildly explosive AR model of PM whose root is in a larger neighborhood of unity than a local-to-unit-root. The second asymptotic ($m \to \infty$) creates a departure from the local-to-unit-root region; see Park (2003) and PMG for detailed discussions. With this sequential asymptotic scheme, we have

$$\begin{aligned} \frac{1}{2c}\frac{n}{m}e^{cm}\left(\hat{\rho} - \rho_{n,m}\right) &\Rightarrow \frac{e^{-cm}\int_0^1 J_{cm}(s)dW(s)}{2ce^{-2cm}\int_0^1 J_{cm}^2(s)ds}, \text{ as } n \to \infty \text{ with fixed } m \\ &= \frac{e^{-cm}\int_0^m \widetilde{J}_c(s)d\widetilde{W}(s)}{2ce^{-2cm}\int_0^m \widetilde{J}_c^2(s)ds} \\ &\Rightarrow C, \text{ as } m \to \infty, \end{aligned} \tag{24}$$

where $\widetilde{W}(t) = \sqrt{m}W(t/m)$ and $\widetilde{J}_c(t) = \int_0^t e^{c(t-s)}d\widetilde{W}(s)$. To see the link between this sequential asymptotic result in (24) and the asymptotic results in (19) and (20), note that $e^{cm} = \exp\left(\frac{cm}{n}\right)^n \overset{a}{\sim} \rho_{n,m}^n$ and $\rho_{n,m}^2 - 1 \overset{a}{\sim} 2c\frac{m}{n}$.[7]

2.3 Mildly Explosive Model with Anti-persistent Errors

We now extend the model of PMG to the following model:

$$y_t = \mu_n + \rho y_{t-1} + u_t, t = 1, ..., n, \tag{25}$$

where $y_0 = o_p(n^{1/2+d})$, $\mu_n = \mu/n^{\vartheta}$, $\rho = \rho_{n,m} = \left(1 + \frac{cm}{n}\right)$, $\vartheta > 1/2 - d$, and u_t satisfies Assumption AP.

[7] Although the limiting distribution in PM is the same as that in PMG, the techniques used to develop the limiting distribution are different in these two studies. PM uses a Lindeberg-Feller CLT while PMG uses the local-to-unit-root theory together with the martingale convergence theorem. Our proof follows that of PMG, but there are technical difficulties that we need to deal with in our proof.

Model (25) is different from Model (23) in two aspects. First and foremost, instead of assuming an iid error process, we allow for anti-persistent errors in Model (25). Second, when $\mu \neq 0$, a non-zero intercept μ_n, which is asymptotically negligible, enters the model. Following Phillips, Shi and Yu (2014), we assume $\mu_n = \mu/n^{\vartheta}$ so that, in finite samples, the model can generate a linear trend. Also, the specification of $\mu_n = \mu/n^{\vartheta}$ with $\vartheta > 1/2 - d$ regulates μ_n so that the localized drift cannot dominate the random component introduced by u_t. However, if $\mu = 0$, then $\mu_n = 0$ and the intercept vanishes.

In this section, we aim to develop the limiting distribution for the centered LS estimate with and without intercept. To be more precise, we define the LS estimate without intercept by $\hat{\rho}_a$ and the LS estimate with intercept by $\hat{\rho}_b$. Thus, we can express the centered LS estimates as

$$\hat{\rho}_a - \rho = \frac{\sum_{t=1}^n y_{t-1}u_t}{\sum_{t=1}^n y_{t-1}^2}, \tag{26}$$

and

$$\hat{\rho}_b - \rho = \frac{\sum_{t=1}^n y_{t-1}u_t - \frac{1}{n}\sum_{t=1}^n y_{t-1}\sum_{t=1}^n u_t}{\sum_{t=1}^n y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n y_{t-1}\right)^2}. \tag{27}$$

Before we develop the asymptotic theory, we first review the functional central limit theorem due to Giraitis, Koul, Surgailis (2012) which extends Donsker's theorem.

Lemma 3 (Corollary 4.4.1 in Giraitis, Koul and Surgailis (2012)) *Let u_t be as in (22). Assume $c_j \overset{a}{\sim} \gamma j^{-1+d}$ as $j \to \infty$ with γ being a constant and $d \in (-0.5, 0)$, $E|\epsilon_t|^p < \infty$ with $p > (0.5+d)^{-1}$ and $\sum_{j=0}^{\infty} c_j = 0$. Then, as $n \to \infty$,*

$$n^{-H}\sum_{t=1}^{\lfloor nr \rfloor} u_t \Rightarrow \varsigma B^H(r), \tag{28}$$

in $\mathcal{D}[0,1]$ with the uniform metric, where $\varsigma = \sqrt{\sigma^2\gamma^2\frac{\mathcal{B}(d,1-2d)}{d(1+2d)}}$ with $\mathcal{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, $H = \frac{1}{2} + d$, $B^H(r)$ is an fBM with the Hurst parameter H.

An fBM with the Hurst parameter $H \in (0,1)$ is a Gaussian process with zero mean and the following covariance,

$$E(B^H(r)B^H(s)) = \frac{1}{2}\left(|r|^{2H} + |s|^{2H} - |r-s|^{2H}\right).$$

Clearly, if $H = 1/2$, $B^H(t)$ becomes the standard Brownian motion $W(t)$. Unlike $W(t)$, $B^H(t)$ is not a semi-martingale if $H \neq 1/2$. Therefore, we cannot interpret the stochastic integral with respect to fBM as an Itô integral. In this paper, we interpret the stochastic integral with respect to fBM as a Young integral when we study the asymptotic theory for the error process under Assumption AP, where the mathematical techniques are related to those used in El Machkouri, Es-Sebaiy and Ouknine (2016) and Xiao and Yu (2019a, 2019b). This interpretation is in contrast to PMG where $\widetilde{J}_c(t) = \int_0^t e^{c(t-s)} d\widetilde{W}(s)$ is viewed as an Itô integral. Moreover, we need a different asymptotic theory to obtain a sequential limit. The following lemma obtains the asymptotic behavior of the sample moments.

Lemma 4 *In Model (25) with $\{u_t\}$ satisfying Assumption AP, we assume $E|\epsilon_t|^p < \infty$ with $p > (0.5+d)^{-1}$. As $n \to \infty$ with m fixed, we have the local-to-unit-root asymptotic results:*

1. $\frac{1}{n^{1/2+d}} y_{\lfloor nr \rfloor} \Rightarrow \varsigma J_{cm}^H(r)$;
2. $\frac{1}{n^{3/2+d}} \sum_{t=1}^n y_t \Rightarrow \varsigma \int_0^1 J_{cm}^H(r) dr$;
3. $\frac{1}{n^{2+2d}} \sum_{t=1}^n y_t^2 \Rightarrow \varsigma^2 \int_0^1 (J_{cm}^H(r))^2 dr$;
4. $\frac{1}{n^{1+2d}} \sum_{t=1}^n y_{t-1} u_t + \frac{1}{n^{1+2d}} \frac{1}{2} \sum_{t=1}^n u_t^2 \Rightarrow \varsigma^2 \left[cmZ(1) \int_0^1 e^{cms} dB^H(s) + R(1) \right]$,

   where

$$\begin{aligned}
\varsigma &= \sqrt{\sigma^2 \gamma^2 \frac{\mathcal{B}(d, 1-2d)}{d(1+2d)}}, \mathcal{B}(d, 1-2d) = \frac{\Gamma(d)\Gamma(1-2d)}{\Gamma(1-d)}, \\
J_{cm}^H(r) &= \int_0^r e^{cm(r-s)} dB^H(s),\ Z(1) = \int_0^1 e^{-cms} B^H(s) ds, \\
R(1) &= \frac{1}{2} \left[B^H(1) \right]^2 - cm \int_0^1 (B^H(s))^2 ds + (cm)^2 \int_0^1 \int_0^s e^{cm(r-s)} B^H(r) B^H(s) dr ds.
\end{aligned}$$

Since $B^H(s)$ is not a semi-martingale, in the present paper, we treat $J_{cm}^H(r)$ as a Young integral. For details about the Young integral, see (A.1) in El Machkouri, Es-Sebaiy and Ouknine (2016).

Remark 15 *The results in Lemma 4 are closely related to Lemma 1 in Phillips (1987), which can be used to show that for Model (25) with weakly dependent errors, when $n \to \infty$ with m fixed,*

$$\frac{1}{n^{1/2}} y_{\lfloor nr \rfloor} \Rightarrow \sigma J_{cm}(r),$$

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} y_t \Rightarrow \sigma\int_0^1 J_{cm}(r)dr,$$

$$\frac{1}{n^2}\sum_{t=1}^{n} y_t^2 \Rightarrow \sigma^2\int_0^1 (J_{cm}(r))^2dr,$$

$$\frac{1}{n}\sum_{t=1}^{n} y_{t-1}u_t \Rightarrow \frac{1}{2}\left[\sigma^2 J_{cm}(1)^2 - 2cm\sigma^2\int_0^1 (J_{cm}(r))^2dr - E(u_t^2)\right],$$

where $J_{cm}(r) = \int_0^r e^{(r-s)cm}dW(s)$.

Remark 16 *For Model (25) with strongly dependent errors, the first three claims in Lemma 4 remain valid, while for the last claim, we have*

$$\frac{1}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1}u_t \Rightarrow \varsigma^2\left[cmZ(1)\int_0^1 e^{cms}dB^H(s) + R(1)\right],$$

because the term $\frac{1}{n^{1+2d}}\frac{1}{2}\sum_{t=1}^{n} u_t^2$ *appearing in Lemma 4.4 asymptotically vanishes as* $n \to \infty$*. This difference makes the development of the limiting distribution in the mildly explosive model with anti-persistent errors more difficult. In particular, when* $n \to \infty$ *with* m *fixed, the centered LS involves an additional term where* $\frac{1}{n^{1+2d}}\frac{1}{2}\sum_{t=1}^{n} u_t^2$ *appears in the numerator. Additional rate condition is needed to make sure this additional term vanishes asymptotically, as shown in the following theorem.*

Theorem 3 *Let* $c > 0$ *in Model (25), under the same set of assumptions as in Lemma 4, if* $n \to \infty$ *followed by* $m \to \infty$ *with* $m = \delta \ln n$ *and* $\delta > -\frac{2d}{c}$*, we have*

$$\frac{1}{2c}\frac{n}{m}e^{cm}\left(\hat{\rho}_j - \rho\right) \Rightarrow C\,,\ \frac{\rho^n}{\rho^2-1}\left(\hat{\rho}_j - \rho\right) \Rightarrow C\,,\ j \in \{a, b\}\,. \tag{29}$$

Theorem 3 suggests that the centered LS estimates $\hat{\rho}_a$ and $\hat{\rho}_b$ in Model (25) have the Cauchy limiting distribution upon the correct normalization. Since the Cauchy distribution is pivotal and ρ can be consistently estimated by either $\hat{\rho}_a$ or $\hat{\rho}_b$, the limit theory provides a convenient way for hypothesis testing for ρ.

Remark 17 *The rate condition* $m = \delta \ln n$ *with* $\delta > -\frac{2d}{c}$ *suggests that* m *cannot go to infinity too slowly relative to* n*. This condition ensures that* $\frac{1}{n^{1+2d}}\frac{1}{2}\sum_{t=1}^{n} u_t^2$ *is dominated by* $\frac{1}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1}u_t$ *as* $m \to \infty$.

Remark 18 *As in Phillips, Wu and Yu (2011), Theorem 3 suggests that a confidence interval (CI) for ρ can be constructed as*

$$\left\{\hat{\rho}_j \pm \frac{\hat{\rho}_j^2 - 1}{\hat{\rho}_j^n} C_a\right\},\ j \in \{a, b\}, \tag{30}$$

where C_a is the critical value for the two-tailed test with the significance level α and $C_{0.1} = 6.315, C_{0.05} = 12.7, C_{0.01} = 63.65674$.

Remark 19 *The Cauchy limiting distribution also holds when we have weakly/strongly dependent errors in Model (23). For example, suppose u_t is weakly dependent with $\sum_{j=0}^{\infty} |c_j| < \infty$, and $\sum_{j=0}^{\infty} c_j \neq 0$, $y_0 = o_p(n^{1/2})$ and $E|\epsilon_t|^{\beta+\varepsilon} < \infty$ for some $\beta > 2$ and $\varepsilon > 0$. With the sequential asymptotic, we have*

$$\begin{aligned}
\frac{1}{2c}\frac{n}{m} e^{cm} (\hat{\rho}_a - \rho) &\Rightarrow \frac{e^{-cm}\int_0^1 J_{cm}(s)dW(s) + e^{-cm}\frac{1}{2}\left(1 - \frac{\upsilon}{\lambda^2}\right)}{2ce^{-2cm}\int_0^1 J_{cm}^2(s)ds}, \text{ as } n \to \infty \text{ with fixed } m \\
&= \frac{e^{-cm}\int_0^m \widetilde{J}_c(s)d\widetilde{W}(s)}{2ce^{-2cm}\int_0^m \widetilde{J}_c^2(s)ds} + O_p(e^{-cm}) \\
&\Rightarrow C, \text{ as } m \to \infty.
\end{aligned} \tag{31}$$

The first convergence follows from Theorem 1 of Phillips (1987), where $\upsilon = \sigma^2 \sum_{j=0}^{\infty} c_j^2$ and $\lambda = \sigma \sum_{j=0}^{\infty} c_j$. The second convergence follows from the martingale convergence theorem.

Remark 20 *Suppose that $\rho_n = 1 + c/n^\alpha$ with $\alpha \in (0, 1)$, $c > 0$, and $u_t = \epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. According to Theorem 4.3 of PM (2007a),*

$$\rho_n^{-n}/n^\alpha \sum_{t=1}^n y_{t-1}u_t \Rightarrow \omega_0\eta_0,\quad \rho_n^{-2n}/n^{2\alpha}\sum_{t=1}^n y_{t-1}^2 \Rightarrow \eta_0^2,$$

where ω_0 and η_0 are independent $N(0, \sigma^2/2c)$ random variables. In our model, we have $\rho = \rho_{n,m} = 1 + cm/n$ and anti-persistent errors. Under the sequential asymptotic scheme, we have

$$\frac{e^{-cm}}{m}\frac{1}{n^{1+2d}}\frac{1}{\zeta^2}\sum_{t=1}^n y_{t-1}u_t \Rightarrow \omega_d\eta_d,\quad 2ce^{-2cm}\frac{1}{n^{2+2d}}\frac{1}{\zeta^2}\sum_{t=1}^n y_{t-1}^2 \Rightarrow \eta_d^2, \tag{32}$$

where ω_d and η_d are independent $N(0, H\Gamma(2H)/2c)$ random variables. We complement the results of PM and PMG to the model with anti-persistent errors.

Remark 21 *When u_t is strongly dependent, using the similar arguments in proving Theorem 3, we can obtain the results of (29) and (32). In this case, the assumption $m = \delta \ln n$ with $\delta > -\frac{2d}{c}$, which is used to eliminate $\frac{1}{n^{1+2d}} \frac{1}{2} \sum_{t=1}^{n} u_t^2$ as $m \to \infty$, is not needed.*

2.4 Monte Carlo Studies

In this section, we design several Monte Carlo experiments to evaluate the precision of the derived asymptotic distribution in finite samples. In all experiments, we simulate data from the following data generating process (DGP):

$$y_t = \mu_n + \rho y_{t-1} + u_t, t = 1, 2, ..., n, \tag{33}$$

where $\rho = \left(1 + \frac{cm}{n}\right)$, $y_0 = 0, c > 0, \mu_n = \mu/n^{\vartheta}$, $u_t = (1 - L)^d \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0, 1)$. We consider the following parameter settings:

$$\begin{aligned} (n, m) &\in \{(100, 10), (500, 15), (1000, 20)\}, \\ d &\in \{-0.45, -0.4, -0.3, -0.2, -0.1, -0.01\}, \\ c &\in \{0.5, 1\}, \mu = 1, \vartheta = \frac{1}{2} - d + 0.1. \end{aligned} \tag{34}$$

The number of replications is always set at 10,000.

Under the parameter settings (34), we first obtain the LS estimates $\hat{\rho}_a$ and $\hat{\rho}_b$, and then apply the Cauchy distribution to construct the 95% CI (CI$_a$ and CI$_b$) based on (30) for $\rho_{n,m}$. We calculate the empirical coverage of the true value ρ, i.e., $\frac{1}{10000} \sum_{l=1}^{10000} 1\left(\rho_L^{(l)} \leqslant \rho \leqslant \rho_U^{(l)}\right)$, where $\rho_L^{(l)}$ and $\rho_U^{(l)}$ are the two bounds of the CI in the l^{th} replication, and $1(\cdot)$ is the indicator function.

Tables 5 reports the empirical coverage of 95% CIs for alternative parameter settings in (34). With $n = 100$, $m = 10$ and $c = 0.5$, there is an obvious over coverage problem for both CI$_a$ and CI$_b$. This problem is less severe as c increases to 1 or as both m and n increase. Moreover, the CIs have good finite sample performance when c is relatively large and d is between -0.01 and -0.3. When $c = 1$, it can be seen that both CI$_a$ and CI$_b$ provide the empirical coverage which is close to the nominal coverage 95%. Finally, the empirical coverage obtained from CI$_a$ and CI$_b$ are similar.

Table 5: Empirical coverage of 95% CI of ρ

	$(n=100, m=10)$				$(n=500, m=15)$				$(n=1000, m=20)$			
	$c=0.5$		$c=1$		$c=0.5$		$c=1$		$c=0.5$		$c=1$	
d	CI_a	CI_b	CI_a	CI_b	CI_a	CI_b	CI_a	CI_b	CI_a	CI_b	CI_a	CI_b
-0.45	.995	.995	.928	.92	.905	.889	.923	.917	.915	.905	.922	.917
-0.4	.996	.995	.933	.924	.917	.903	.928	.924	.925	.915	.929	.925
-0.3	.996	.994	.945	.937	.934	.923	.939	.933	.936	.928	.937	.935
-0.2	.995	.995	.948	.943	.949	.939	.944	.941	.944	.937	.947	.942
-0.1	.991	.992	.947	.943	.95	.946	.95	.945	.948	.943	.950	.947
-0.01	.988	.99	.951	.947	.952	.946	.952	.949	.950	.946	.952	.950

2.5 An Empirical Study

To highlight the usefulness of the proposed model and the derived limiting distribution in practice, we now conduct an empirical study of a rational bubble based on Model (23) and the asymptotic theory in Theorem 3. The standard no-arbitrage condition suggests that

$$P_t = \frac{1}{1+r_f} E_t \left[P_{t+1} + D_{t+1}\right], \tag{35}$$

where P_t, r_f, D_t and E_t denote the price of asset, the discount rate, the dividend, and the expectation based on information at time t, respectively. Equation (35) can be solved by forward substitutions, giving rise to the following expressions:

$$\begin{aligned} P_t &= P_t^f + B_t, && (36) \\ P_t^f &= \sum_{i=1}^{\infty} \left(\frac{1}{1+r_f}\right)^i E_t\left(D_{t+i}\right), && (37) \\ B_t &= \frac{1}{1+r_f} E_t\left(B_{t+1}\right). && (38) \end{aligned}$$

Equation (36) expresses price as a sum of two components: the fundamental price P_t^f which summarizes all the expected future discounted dividend and a bubble component B_t which is not related to the fundamentals.

If the transversality condition is imposed, then $B_t = 0$ and hence, $P_t = P_t^f$. Note that B_t is an explosive process since $(1+r_f) > 1$. Therefore, when P_t^f is not explosive, testing the existence of a bubble is equivalent to examining the explosiveness in P_t. That is why

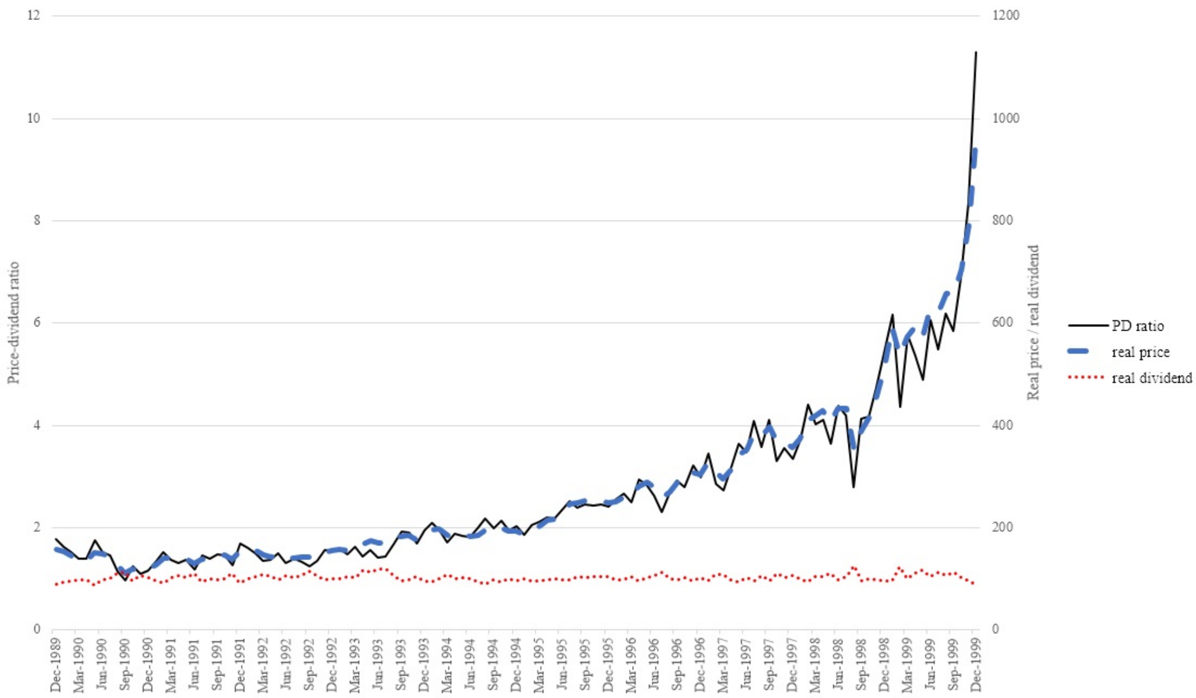


Figure 2: Price-dividend ratio in NASDAQ from December 1989 to December 1999

in the literature looking for an explosive behavior in the price-dividend ratio (P_t/D_t) has been widely used; see, for example, Phillips, Shi and Yu (2015a, 2015b).

Our paper studies the price-dividend ratio in the NASDAQ composite index, we obtain the data set from Phillips, Wu and Yu (2011), which contains the monthly real price and real dividend series from February 1973 to June 2005. We then construct the price-dividend (PD) ratio based on the two time series. After obtaining the PD ratio, we focus on the sample period from December 1989 to December 1999.

In Figure 2, the PD ratio, the real price, and the real dividend are plotted in the black solid line, the blue dash line, and the red dotted line, respectively. We fit Model (23) with and without intercept to the PD ratio by LS, and then estimate the memory parameter (d) in the residuals by the LW method of Robinson (1994). The point estimate ("estimate" should be "estimates") of the intercept ($\hat{\mu}$), the AR coefficient ($\hat{\rho}$), and the memory parameter ($\hat{d}$) are reported in Panel A of Table 6. We use the subscript a and b to denote the LS estimate without and with intercept, respectively. Since the estimates of the AR coefficient are greater than 1 and $\hat{d} \in (-0.5, 0)$, Model (23) is relevant and the asymptotic theory developed in Theorem 3 is applicable. We then use the Cauchy limiting distribution to form the 95% CI of ρ which is reported in Panel A of Table 6. As the 95% CI excludes the unity, suggesting that there is strong evidence of explosiveness in the PD ratio and, hence, strong evidence of the presence of a bubble. In Panel B, we report the empirical results based on a subsample of the NASDAQ index, namely, January 1993 to December 1999. We continue to find that $\hat{\rho} > 1$, $\hat{d} \in (-0.5, 0)$, and that the 95% CI suggests the strong evidence of the presence of a bubble in the subsample.

Table 6: Empirical results for the NASDAQ Index

Panel A: Sample Period: December 1989 to December 1999, $n = 120$							
	$\hat{d}_a$	$\hat{\rho}_a$	95% CI$_a$	$\hat{d}_b$	$\hat{\mu}$	$\hat{\rho}_b$	95% CI$_b$
P_t/D_t	-0.084	1.0437	$[1.0370, 1.0504]$	-0.060	-0.1445	1.0862	$[1.0860, 1.0863]$
Panel B: Sample Period: January 1993 to December 1999, $n = 83$							
	$\hat{d}_a$	$\hat{\rho}_a$	95% CI$_a$	$\hat{d}_b$	$\hat{\mu}$	$\hat{\rho}_b$	95% CI$_b$
P_t/D_t	-0.079	1.0478	$[1.0220, 1.0736]$	-0.066	-0.1865	1.0969	$[1.0957, 1.0981]$

2.6 Conclusion

In this paper, we have made two contributions to the rapidly growing literature on explosive time series. First, we show that in empirical data, it is very plausible that we may have to use a mildly explosive model with anti-persistent errors to describe the movement of financial assets. Second, we show that, when anti-persistent errors are in a first-order autoregression with a mildly explosive root, the Cauchy limiting distribution remains valid for the LS estimate. To develop the limiting distribution, we following PMG's setup by assuming the autoregressive parameter is $\rho_{n,m} = 1 + \frac{cm}{n}$ and by adopting a sequential limit with $n \to \infty$ followed by $m \to \infty$. When the errors are anti-persistent, an extra rate condition $m = \delta \ln n$ with $\delta > -\frac{2d}{c}$ is needed.

We also discuss how to obtain a feasible confidence interval for the AR coefficient. Empirical coverage of CI based on the Cauchy limiting distribution is presented in the Monte Carlo studies, suggesting that the limiting distribution works well in finite samples. Finally, an empirical study of a rational bubble in the NASDAQ index is provided, highlighting the usefulness of the proposed model and the derived asymptotic theory.

3 Testing for Rational Bubbles under Strongly Dependent Errors

3.1 Introduction

The standard no-arbitrage condition implies that

$$P_t = \frac{1}{1+R} E_t \left(P_{t+1} + D_{t+1} \right), \tag{39}$$

where R, E_t, P_t, and D_t denote the discount rate, the expectation based on information at time t, asset price, and the fundamentals (such as the dividend for a stock or the rent from a house) at time t, respectively. Solving (39) by forward substitution, we can express $P_t = P_t^F + B_t$ where $P_t^F = \sum_{i=1}^{\infty} \left(\frac{1}{1+R}\right)^i E_t \left(D_{t+i}\right)$ is the fundamental price and $B_t = \frac{1}{1+R} E_t(B_{t+1})$ is the bubble component. Note that B_t is not related to the fundamentals. If P_t is the price of a stock, P_t^F is determined by the sum of the discounted dividends. Suppose that the transversality condition is satisfied, namely, $\lim_{T\to\infty} (1+R)^{-T} E_t P_{t+T} = 0$. This condition implies $B_t = 0$ and hence, $P_t = P_t^F$. When the transversality condition is not satisfied, $B_t \neq 0$. In this case, B_t is an explosive process since $R > 0$ and hence, $1 + R > 1$. The explosiveness in B_t also makes P_t an explosive process even when P_t^F is not explosive. This is how a rational bubble is related to explosiveness in time series. In practice, empirical studies often verify the explosiveness of the price-fundamental ratio for the purpose of bubble detection; see, for instance, Phillips et al. (2015a) (hereinafter PSY) and Pedersen and Schütte (2017). A natural approach to the bubble detection is to employ a right-tailed unit root test, popularized by Diba and Grossman (1988), Phillips et al. (2011) (hereinafter PWY) and Phillips and Yu (2011). PSY (2015a, b) extend the work of PWY to detect multiple bubbles. Harvey et al. (2016, 2019a, 2019b) extend it to account for heteroskedastic errors and Pedersen and Schütte (2017) for weakly dependent errors.

Considers the following simple first-order autoregressive (AR) model

$$y_t = \rho y_{t-1} + \epsilon_t,\ y_0 = O_p(1), \epsilon_t \overset{iid}{\sim} (0, \sigma^2), t = 1, ..., n. \tag{40}$$

Under the null hypothesis, y_t is a unit root process (i.e., $\rho = 1$). Under the alternative hypothesis, y_t displays explosiveness (i.e., $\rho > 1$), suggesting why one would use the right-tailed unit root test.

This paper focuses on model (40) with a simple extension, where errors are assumed to follow a strongly dependent process. The phenomenon of strong dependence is widespread in economic and financial time series. Cheung (1993) and Baillie et al. (1996) find empirical evidence of strong dependence in exchange rates. Christensen and Nielsen (2007), Andersen et al. (2003) and Ohanissian et al. (2008) show empirical evidence of strong dependence in volatility of stock returns and exchange rate returns. In addition, empirical studies obtained by Gil-Alana et al. (2014) and Barros et al. (2014) suggest strong dependence in housing prices in US cities. More recently, Chevillon and Mavroeidis (2017) show statistical learning can generate strong dependence and find empirical evidence of strong dependence in the US monthly CPI inflation rates.

Consider the following model

$$\begin{aligned} y_t &= y_{t-1} + u_t,\ t = 1, ..., n, \\ u_t &= (1-L)^{-d}\epsilon_t, d \in (0, 0.5), \epsilon_t \overset{iid}{\sim} (0, \sigma^2), \end{aligned} \tag{41}$$

where L is the lag operator with $(1-L)^{-d}$ defined as

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j \text{ and } L^j\epsilon_t = \epsilon_{t-j}.$$

Denote $\psi(k)$ the k^{th} order autocovariance function of u_t, namely $\psi(k) := E(u_t u_{t-k})$. For u_t to be strongly dependent, we have $\sum_{k=-\infty}^{\infty} |\psi(k)| = \infty$. It can be shown that when $d \in (0, 0.5)$, $\sum_{k=-\infty}^{\infty} |\psi(k)| = \infty$, suggesting u_t is indeed strongly dependent. If $d = 0$ in (41), model (41) becomes model (40) with $\rho = 1$. In the time series literature, we often say $u_t \sim I(d)$, fractionally integrated of order d. Since the first difference of y_t is $I(d)$, $y_t \sim I(\lambda)$ with $\lambda = 1 + d$.

Although y_t in (41) is a unit root process with strongly dependent errors, it is plausible to see an explosive trajectory in y_t. To see why this is the case, we can express $y_t = \sum_{i=1}^{t} u_i + y_0$. Since u_t is strongly dependent, a positive realization of the error term is likely to generate a long stream of positive errors due to strong dependence. Since y_t is the cumulative sum of errors, a long stream of positive errors will generate an upward trend that looks like an explosive process.

Suppose that we use the least squares (LS) method to estimate the AR(1) coefficient when data come from in (41) and then construct the conventional t statistic. Sowell (1990) shows that the t statistic diverges with the sample size. Therefore, applying a traditional right-tailed unit root test tends to reject the unit root null hypothesis when the sample size is large. In the context of rational bubble detection, due to the diverging type I error, applying a unit root test that ignores the strong dependence in u_t (i.e. assuming $d = 0$) tends to conclude explosiveness in y_t, thereby incorrectly detecting a rational bubble in model (41) when there is no bubble.

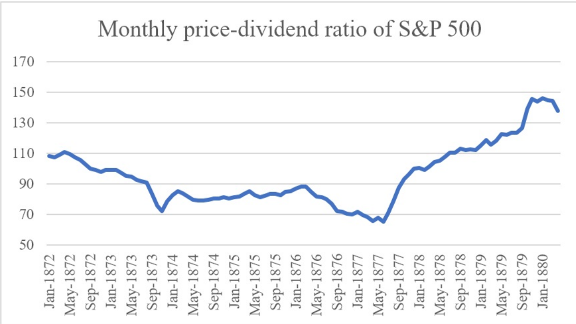
To showcase the empirical relevance of this problem, Figure 3 plots the monthly price-dividend ratio of S&P 500.[8] In particular, we consider six sampling periods: (a) November 1874 to October 1879; (b) June 1882 to May 1887; (c) May 1940 to February 1946; (d) June 1948 to November 1955; (e) October 1979 to March 1987 and (f) May 1988 to April 1998. It is noteworthy that each sampling period contains a trajectory in which the market experiences exuberance with a rising price-dividend ratio. Under the assumption that the true model is (40), that is, the errors are not strongly dependent, we perform an LS regression with an intercept and calculate the Dickey-Fuller t statistic for each sample (denoted by DF_n).

Table 7 Right-tailed unit root test for the price-dividend ratio

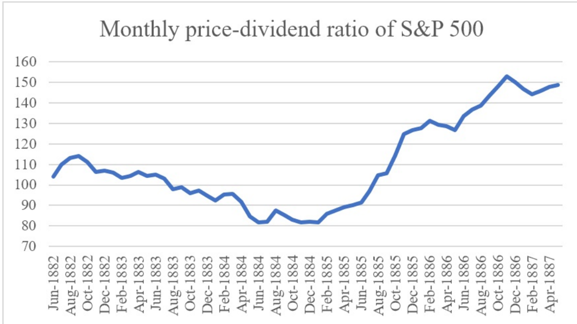
Sampling Period	DF_n	$\hat{d}$
(a) Jan 1872 to Feb 1880	1.35***	0.24***
(b) Jun 1882 to May 1887	0.66***	0.32***
(c) May 1940 to Feb 1946	1.38***	0.34***
(d) Jun 1948 to Nov 1955	1.70***	0.29***
(e) May 1980 to July 1987	1.43***	0.20**
(f) May 1988 to Apr 1998	3.76***	0.24***

NOTES: "*", "***" and "***" denote the 90%, 95% and 99% level of significance of the right-tailed test for $\rho > 1$ and $d > 0$, respectively. For the right-tailed unit root test, following PWY we use 95% and 99% critical values for the alternatives ($cv^{90\%} = -0.44$, $cv^{95\%} = -0.08$, and $cv^{99\%} = 0.6$). The right-tailed tests for d apply the limit theorem of ELW estimator to obtain the critical values.

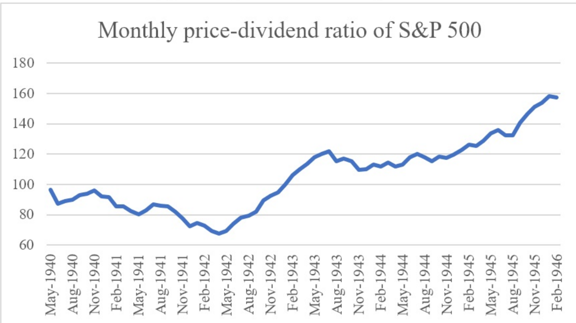
[8] The price-dividend ratio was used in PSY (2015a).



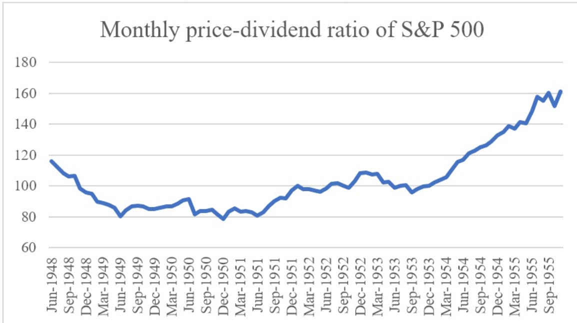
(a) January 1872 to February 1880



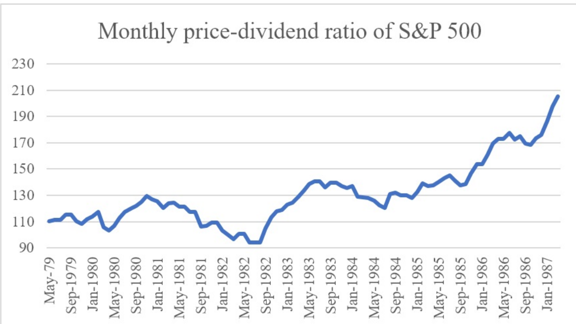
(b) June 1882 to May 1887



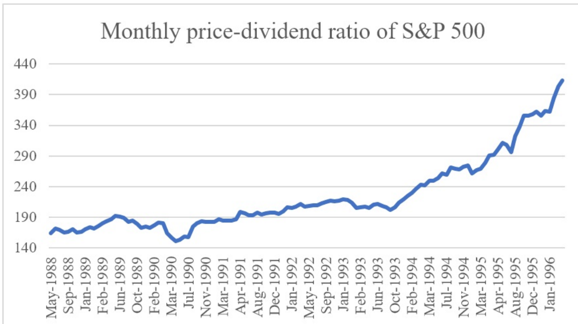
(c) May 1940 to February 1946



(d) June 1948 to November 1955



(e) May 1979 to March 1987



(f) May 1988 to April 1998

Figure 3: Monthly price-dividend ratio of S&P 500

Table 7 reports DF_n and it is noteworthy that we reject the unit root hypothesis at the 99% confidence level critical value for each of the sampling period. Therefore, the right-tailed unit root tests give a very strong evidence of the existence of a rational bubblen in each period. In fact, adopting a form of recursive unit root test, PSY (2015a) also find bubbles during similar periods.

However, if we assume that the time series are fractionally integrated as in model (41), we can estimate λ and d. In particular, we apply the exact local Whittle method of Shimotsu and Phillips (2005) to estimate d and test i.i.d. hypothesis of ϵ_t against the strongly dependent alternative.[9] Table 7 shows that positive estimates $\hat{d}$ for all three sampling periods are found. Moreover, for each sampling period, the i.i.d. hypothesis is rejected under 95% confidence level, suggesting strong evidence of strong dependence in u_t. Therefore, it is possible that the true model is (41) and that the divergent t statistic leads to the rejection of the unit root null hypothesis in favor of an explosive alternative. In other words, these rational bubbles can be spurious.

Motivated by the empirical evidence of strongly dependent errors and their implication for bubble detection, this paper proposes a method to address the spurious explosiveness problem in detecting rational bubbles when a time series model has strongly dependent errors. We construct a heteroskedasticity-autocorrelation robust (HAR) test statistic, which converges to a proper distribution under no bubble assumption but diverges when the underlying model has an explosive or a mildly explosive root. Therefore, we can distinguish between an explosive and unit root time series even when the error process is strongly dependent. After a bubble is detected, a new estimator is proposed to consistently timestamp its origination and termination dates.

The remainder of this paper is organized as follows. Section 3.2 provides a brief review of the traditional right-tailed unit root test for bubble detection and the estimation method to timestamp bubble origination and termination dates. Section 3.3 introduces our model with strongly dependent errors, proposes the new test, and derives the asymptotic

[9] We first estimate the fractionally integrated order (λ) of y_t. Then we subtract 1 from the estimate of λ to obtain $\hat{d}$. The confidence interval obtained is based on the asymptotic of the ELW estimate where $\sqrt{m}(\hat{d} - d) \xrightarrow{d} N(0, 1/4)$, with $m = n^{\delta}$. We set $\delta = 0.65$ in all applications in this paper.

theory under the null hypothesis of unit root. Section 3.4 examines the asymptotic properties of the proposed test statistic under two explosive alternatives. Section 3.5 proposes new estimators of bubble origination and termination dates based on the new test statistic. Monte Carlo simulation studies are carried out in Section 3.6 to study the finite sample performance of the proposed test and estimators. Section 3.7 provides an empirical study using the S&P 500 index. Finally, Section 3.8 concludes the paper. A discussion of the sup statistics, detailed graphs for the empirical study and the proofs of the main results in the paper are provided in the Appendix. We use the following notations throughout the paper: $\xrightarrow{p}$, $\xrightarrow{d}$, $\xrightarrow{as}$, $\Rightarrow$, $\overset{a}{\sim}$ and $\overset{iid}{\sim}$ denote convergence in probability, convergence in distribution, almost sure convergence, weak convergence, asymptotic equivalence, and independent and identically distributed, respectively.

3.2 A Brief Review of Literature

In this section, we briefly review the traditional right-tailed unit root test statistic for bubble detection and the estimation method to timestamp bubble origination and termination dates. Consider model (40). Suppose that we perform an LS regression with an intercept from the full sample and obtain our LS estimator $\hat{\rho}_n$ of ρ. Denote the t statistic $DF_n = (\hat{\rho}_n - 1)/se(\hat{\rho}_n)$, where $se(\hat{\rho}_n)$ is the standard error of $\hat{\rho}_n$. Under the assumption that $\rho = 1$, following Phillips (1987a), we have

$$DF_n \Longrightarrow DF_\infty := \frac{\int_0^1 \tilde{W}(s)dW(s)}{\left(\int_0^1 \tilde{W}(s)^2 ds\right)^{1/2}}, \text{ (as } n \to \infty), \tag{42}$$

where $W(r)$ is the standard Brownian motion and $\tilde{W}(r) = W(r) - \frac{1}{r}\int_0^r W(s)ds$ is the demeaned Brownian motion. To implement a right-tailed unit root test, we can obtain the $(1-\beta)\%$ critical value as the $(1-\beta)$ percentile of DF_∞ and reject the null hypothesis when DF_n is greater than the critical value.

In practice, a bubble usually starts not from the first observation of the full sample, but from the middle of the sample, say at $\tau_e = \lfloor nr_e \rfloor$ where $\lfloor \cdot \rfloor$ denotes the integer part of its argument and $r_e \in (0,1)$. Moreover, a bubble usually does not last forever, but collapses later in the sample, say at $\tau_f = \lfloor nr_f \rfloor$. The collapse of a bubble typically corresponds to a market correction. If a bubble emerges and collapses in the sample, Phillips and Yu (2009) show that $DF_n \to -\infty$, suggesting that the traditional right-tailed unit root test statistic, if calculated from the full sample, cannot reject the unit root null hypothesis.

To identify the bubble in the full sample, PWY (2011) propose a sup statistic based on recursive regressions. The regression in the first recursion is

$$y_t = \hat{\mu} + \hat{\rho}_\tau y_{t-1} + \hat{u}_t, \tag{43}$$

where $t = 1, ..., \tau_0 (:= \lfloor nr_0 \rfloor)$, $\hat{\mu}$, $\hat{\rho}_\tau$, and $\hat{u}_t$ are the intercept estimator, the AR coefficient estimator and the LS residuals, respectively, and $0 < r_0 < r_e$. Subsequent regressions employ this originating data set supplemented by successive observations giving a sample of size $\tau = \lfloor nr \rfloor$ for $r_0 \leqslant r \leqslant 1$. Let DF_τ denote the DF t statistic based on the first τ observations of the sample, that is,

$$DF_\tau := \frac{\hat{\rho}_\tau - 1}{s_\tau}, \tag{44}$$

where $s_\tau = \left(\frac{\frac{1}{\tau}\sum_{t=1}^{\tau} \hat{u}_t^2}{\sum_{t=1}^{\tau} y_{t-1}^2 - \frac{1}{\tau}\left(\sum_{t=1}^{\tau} y_{t-1}\right)^2} \right)^{1/2}$ is the standard error of $\hat{\rho}_\tau$. The test statistic proposed by PWY is $\sup_{\tau \in [\tau_0, n]} DF_\tau$, Under the null hypothesis, PWY show

$$SDF := \sup_{\tau \in [\tau_0, n]} DF_\tau \Longrightarrow \sup_{r \in [r_0, 1]} \frac{\int_0^r \tilde{W}(s) dW(s)}{\left(\int_0^r \tilde{W}(s)^2 ds\right)^{1/2}}, \text{ (as } n \to \infty).$$

If SDF takes a value larger than the right-tailed critical value, the null unit root hypothesis is rejected in favor of the explosive alternative. In this case, the evidence of a bubble is found.

After a bubble is detected, one may want to estimate bubble origination and conclusion dates, that is, r_e and r_f. Assume the model under the alternative hypothesis is given by

$$\begin{aligned} y_t &= y_{t-1} 1\{t < \tau_e\} + \rho_n y_{t-1} 1\{\tau_e \leqslant t \leqslant n\} \\ &\quad + \left(\sum_{k=\tau_f+1}^{t} \epsilon_k + y_{\tau_f}^* \right) 1\{t > \tau_f\} + \epsilon_t \, 1\{t \leqslant \tau_f\}, \\ \rho_n &= 1 + \frac{c}{n^\alpha}, \;\; c > 0, \;\; \alpha \in (0,1), \epsilon_t \overset{iid}{\sim} (0, \sigma^2). \end{aligned} \tag{45}$$

Model (45) has two structural breaks. Before the first break (i.e. $t < \tau_e$), y_t follows a unit root process. After the first break but before the second break (i.e. $\tau_e \leqslant t \leqslant \tau_f$), it follows a mildly explosive process with a root above 1 taking the form $\rho_n = 1 + \frac{c}{n^\alpha}$. At $\tau_f + 1$, the bubble terminates with a crash to $y^*_{\tau_f}$ which is in the neighborbood of the fundamental value prior to the emergence of the bubble. Here τ_e and τ_f are the true absolute bubble origination and termination dates. Since $\tau_e = \lfloor nr_e \rfloor$ and $\tau_f = \lfloor nr_f \rfloor$, r_e and r_f are the true fractional bubble origination and termination dates.

PWY (2011) introduce the estimator of r_e as

$$\hat{r}_e^{PWY} = \inf_{r \geqslant r_0} \{r : DF_\tau > cv_n\}. \tag{46}$$

Conditional on finding some originating date $\hat{r}_e^{PWY}$ of a bubble, PWY introduce the estimator of r_f as

$$\hat{r}_f^{PWY} = \inf_{s \geqslant \hat{r}_e + \frac{\gamma \ln(n)}{n}} \{s : DF_s < cv_n\}. \tag{47}$$

In (46) and (47), cv_n is a critical value function that increases with the sample size. Provided cv_n goes to infinity at a slower rate than $n^{1-\alpha/2}$, Phillips and Yu (2009) showed that $\hat{r}_e^{PWY} \xrightarrow{p} r_e$ and $\hat{r}_f^{PWY} \xrightarrow{p} r_f$ under some general regularity conditions. In the empirical applications of PWY, cv_n is set to be proportional to $\ln \ln n$.

3.3 Model, New Test and Asymptotic Null Distribution

Motivated by the empirical studies in section 3.1, we now consider the following model

$$\begin{aligned} y_t &= \rho_n y_{t-1} + u_t, \; t = 1, ..., n, \\ u_t &= (1-L)^{-d} \epsilon_t, \; \epsilon_t \overset{iid}{\sim} (0, \sigma^2), \; d \in [0, 0.5), \\ y_0 &= o_p(n^{1/2+d}). \end{aligned} \tag{48}$$

Model (48) is different from model (40) in that u_t can be strongly dependent. We first consider the asymptotic property of the traditional Dickey-Fuller t test when $\rho_n = 1$.

3.3.1 Asymptotic null distribution of DF_τ

Lemma 5 *Assume the true data generation process is model (48) with $\rho_n = 1$ and $d \in (0, 0.5)$. Suppose that DF_τ is constructed from the empirical regression (43) based on the* *first τ observations. Then, for any $r \in (0,1]$, as $n \to \infty$,*

$$DF_{\tau} = O_p\left(n^d\right). \tag{49}$$

Lemma 5 indicates a serious implication for bubble detection using the traditional method. Namely, the t statistic DF_{τ} diverges with the sample size. When $r = 1$, $DF_n = DF_{\tau} = O_p\left(n^d\right)$. This divergence leads to excessive rejection of the null hypothesis when the sample size is large and conclude with a spurious bubble.

Remark 22 *To detect the presence of a bubble, PWY and PSY propose to use SDF and $GSDF$ defined by*

$$SDF(\tau_0) = \sup_{\tau \in [\tau_0, n]} DF_{\tau},$$
$$GSDF(\tau_0) = \sup_{\tau_2 \in [\tau_0, n], \tau_1 \in [0, \tau_2 - \tau_0]} DF_{\tau_1}^{\tau_2},$$

where $\tau_0 = \lfloor nr_0 \rfloor$ is the minimum data window and $DF_{\tau_1}^{\tau_2}$ is the t statistic based on the observations from $\tau_1 = \lfloor nr_1 \rfloor$ to $\tau_2 = \lfloor nr_2 \rfloor$.

As Lemma 5 holds uniformly for $r \in (0,1]$, under model (48) with $\rho_n = 1$ and $d \in (0, 0.5)$, we have

$$\begin{aligned} SDF(\tau_0) &= O_p\left(n^d\right), \\ GSDF(\tau_0) &= O_p\left(n^d\right). \end{aligned}$$

Both statistics can lead to spurious rational bubble detection as they diverge to infinity.

Remark 23 *Similar to PSY (2014), our model (48) can be generalized to have an asymptotically negligible intercept. In this case,*

$$y_t = \mu_n + \rho_n y_{t-1} + u_t, \tag{50}$$

where $\mu_n = O(n^{-\theta})$, with $\theta > 1/2 - d$. It can be shown that, since μ_n is asymptotically negligible, the result in Lemma 5 remains valid.

Remark 24 *Note that if the k^{th} order augmented Dickey-Fuller test is used, the same result as in (49) can be obtained.*

Remark 25 *If the LS regression is carried out without intercept and* $r = 1$*, the result of Lemma 5 coincides with that of Sowell (1990) (see Theorem 4 in Sowell, 1990) when* $d \in (0, 0.5)$.

Remark 26 *Homm and Breitung (2012) use the CUSUM statistic* (CS_t) *to detect rational bubble. They show that if* $CS_t = \frac{1}{\hat{\sigma}} \sum_{j=T+1}^{t} \Delta y_t$ *with* $\Delta y_t = y_t - y_{t-1}$*, we have* $n^{-1/2} CS_{t=\lfloor nk \rfloor} \Longrightarrow W(k) - W(1)$ *with* $k > 1$ *under the assumption that* u_t *is an i.i.d. error. Clearly, under model (48) with* $d \in (0, 0.5)$*,* $n^{-1/2} CS_{t=\lfloor nk \rfloor}$ *is not properly normalized and has order* $O_p\left(n^d\right)$*. Therefore, the CUSUM statistic will also share the spurious bubble detection problem under model (48) with* $d \in (0, 0.5)$.

3.3.2 New test statistic

The failure of standard t statistic stems from estimating the variance of u_t by the average squared residuals $\frac{1}{\tau} \sum_{t=1}^{\tau} \hat{u}_t^2$. As this estimator does not provide a proper normalization, it results in the divergence of DF_τ. In this paper, we use a properly self-normalized statistic that converges to a proper distribution for $d \in [0, 0.5)$. To design the new statistic, noting that as u_t is potentially strongly dependent, we propose to estimate the variance of u_t by using $\hat{\Omega}_{HAR} = \sum_{j=-\tau+1}^{\tau} K\left(\frac{j}{M}\right) \hat{\gamma}_j$, where $K(\cdot)$ is a kernel function with bandwidth M and $\hat{\gamma}_j = \frac{1}{\tau} \sum_{t=j+1}^{\tau} \Delta y_t \Delta y_{t-j}$ is the j^{th} order sample autocovariance.

Based on $\hat{\Omega}_{HAR}$, we can define the new t statistic as

$$DF_{\tau,HAR} = \frac{\hat{\rho}_\tau - 1}{s_{\tau,HAR}}, \tag{51}$$

where the robust standard error is defined as

$$s_{\tau,HAR} = \sqrt{\frac{\hat{\Omega}_{HAR}}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2}} \text{ with } \bar{y}_t = y_t - \frac{1}{\tau} \sum_{t=1}^{\tau} y_{t-1}.$$

In addition, we select the bandwidth by letting $M = b \times \tau$, where $b \in (0, 1]$, so that the bandwidth is the same as the sample size τ in the regression window. This approach is popularized by Kiefer and Vogelsang (2002a, 2002b, 2005), Bunzel et al. (2002) and Vogelsang (2003). The test statistic based on $\hat{\Omega}$ is heteroskedasticity-autocorrelation robust (HAR). Our test also shares the same spirit as the test proposed in Sun (2004), where the HAR test statistic with a fix-b approach is used to tackle the problem of spurious co-integration.

Theorem 4 *Suppose* $M = b\tau$, *and* $K(x) = K_B(x)$ *is the Bartlett kernel function with* $K_B(x) = (1 - |x|)\mathbf{1}(|x| \leqslant 1)$. *Under model (48), for* $r \in (0, 1]$, *as* $n \to \infty$, $DF_{\tau,HAR}$ *has the following fixed-b asymptotic distribution,*

$$DF_{\tau,HAR} \Longrightarrow \begin{cases} F_{r,0} := \dfrac{b^{1/2} r \int_0^r \tilde{W}(s) dW(s)}{\left[2 \int_0^r \tilde{W}(s)^2 ds \left(\int_0^r W(p)^2 dp - \int_0^{(1-b)r} W(p) W(p+br) dp\right)\right]^{1/2}} & \textit{for } d = 0 \\ F_{r,d} := \dfrac{b^{1/2} \left[\frac{r}{2}\left(B^H(r)\right)^2 - \left(\int_0^r B^H(s) ds\right) B^H(r)\right]}{\left[2 \int_0^r \left(\tilde{B}^H(s)\right)^2 ds \left(\int_0^r B^H(p)^2 - \int_0^{(1-b)r} B^H(p) B^H(p+br) dp\right)\right]^{1/2}} & \textit{for } d \in (0, 0.5) \end{cases} \tag{52}$$

where $B^H(t)$ *is a fractional Brownian motion (fBm) with the Hurst parameter* $H = 1/2 + d$, *and* $\tilde{B}^H(r) = B^H(r) - \frac{1}{r}\int_0^r B^H(s) ds$ *is the demeaned fBm.*

Unlike DF_τ, Theorem 4 shows that the HAR test statistics converge to proper limit distributions for both $d = 0$ and $d \in (0, 0.5)$, therefore, it does not share the diverging size problem as in Lemma 5 and the spurious rational bubble conclusion can be avoided asymptotically.

Although the test statistic is well normalized within the range of d that we are interested, it should be be noted that the limit distribution of $DF_{\tau,HAR}$ does not have a uniform expression for any $d \in [0, 0.5)$. This is because the centered LS estimator $\hat{\rho}_\tau - 1$ has a component $-n^{-1-2d}\left(\sum_{t=1}^{\tau} u_t^2\right)$. When $d > 0$, this term is asymptotically negligible. However, when $d = 0$, this term cannot be ignored and shows up at the limit. As d is generally unknown and needed to be estimated, to use the asymptotic distribution, one needs to first obtain a consistent estimator $\hat{d}$, and determine whether $F_{r,0}$ or $F_{r,d}|_{d=\hat{d}}$ is used.

In our unreported simulation, the critical values obtained from the limit distribution of $DF_{\tau,HAR}$ does not provide a satisfactory performance in size especially when d is closed to zero. We discover that the bad size performance is due to the following two reasons. Firstly, when $d > 0$ but reasonably closed to zero, the component $n^{-1-2d}\left(\sum_{t=1}^{\tau} u_t^2\right)$ converges in probability to zero in a slowly, therefore $F_{r,d}$ does not provide a good finite sample approximation. Secondly, when $d = 0$, it possible that we may end up applying

$F_{r,d}|_{d=\hat{d}}$ with $\hat{d} > 0$, therefore, the correct distribution $F_{r,0}$ is not used for the construction of critical values.

This motivates us to design a modified statistic $\widetilde{DF}_{\tau,HAR}$ which has a uniform limit expression and a satisfactory size performance. We now consider the following test statistic

$$\widetilde{DF}_{\tau,HAR} = \frac{\tilde{\rho}_\tau - 1}{s_{\tau,HAR}} \tag{53}$$

where $\tilde{\rho}_\tau = \hat{\rho}_\tau + \frac{\frac{1}{2}\sum_{t=1}^{\tau} \Delta y_t^2}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2}$ and $\Delta y_t = y_t - y_{t-1}$.

Theorem 5 *Under the same set of assumption in Theorem 4, for $r \in (0,1]$, as $n \to \infty$, we have*

$$\widetilde{DF}_{\tau,HAR} \Longrightarrow F_{r,d}, \textit{ for } d \in [0, 0.5). \tag{54}$$

where $F_{r,d}$ is defined in (52).

From Theorem 5, we can see the new statistic has a uniform expression for $d \in [0, 0.5)$. It is noteworthily that $B^H(r) = W(r)$ when the $H = 1/2$. We achieve this uniform expression by simply algebraically removing the component $n^{-1-2d}\left(\sum_{t=1}^{\tau} u_t^2\right)$ in the centered LS estimator $\hat{\rho}_\tau - 1$. By doing so, the slowly converging to zero behavior of $n^{-1-2d}\left(\sum_{t=1}^{\tau} u_t^2\right)$ is no longer relevant and the limit of $\widetilde{DF}_{\tau,HAR}$ does not have an abrupt shift at $d = 0$. We have the following corollary for the full sample HAR statistic $\widetilde{DF}_{n,HAR}$.

Corollary 1 *Under the assumption of Theorem 4, if $d \in [0, 0.5)$, as $n \to \infty$, $\widetilde{DF}_{n,HAR}$ has the following limit distribution*

$$\begin{aligned} \widetilde{DF}_{n,HAR} &\Longrightarrow F_{1,d} \\ &:= \frac{b^{1/2}\left[\frac{1}{2}\left(B^H(1)\right)^2 - \left(\int_0^1 B^H(s)ds\right) B^H(r)\right]}{\left[2\int_0^1 \left(\tilde{B}^H(s)\right)^2 ds \left(\int_0^1 B^H(p)^2 - \int_0^{(1-b)} B^H(p) B^H(p+br)dp\right)\right]^{1/2}}. \end{aligned}$$

To carry out right-tailed unit root tests based on the full sample, we can adopt the test statistic $\widetilde{DF}_{n,HAR}$ and its $(1-\beta)\times 100\%$ critical value $cv_{HAR}^{(1-\beta)\%}(d)$, which is defined by

$$\Pr\left(F_{1,d} > cv_{HAR}^{(1-\beta)\%}(d)\right) = \beta. \tag{55}$$

Moreover, the memory parameter (d) that appears in the critical value can be consistently estimated (e.g. using the ELW method). Therefore, a feasible critical value can be obtained.

Remark 27 *The limit distributions in Theorems 28 and 4 apply when the error term* u_t *follows a stationary ARFIMA(*p, d, q*) process with* $d \in (0, 0.5)$*. Indeed,* $n^{-1-2d}\left(\sum_{t=1}^{\tau} u_t^2\right)$ *vanishes with* $d > 0$ *as* $n \to \infty$*. It washes out the variance term which is dependent on the specific form of the ARFIMA(*p, d, q*) process.*

Remark 28 *In fact, other kernel functions (*$K_2(.)$*) can also produce a fixed-b asymptotic distribution of* $\widetilde{DF}_{\tau,HAR}$*. Suppose* $\hat{\Omega}_{HAR} = \sum_{j=-\tau+1}^{\tau} K_2\left(\frac{j}{M}\right)\hat{\gamma}_j$*,* $K_2(\cdot) \geqslant 0$*,* $K_2(x) = K_2(-x)$*, with* $K_2(\cdot)$ *being a twice differentiable kernel function, we can easily show as* $n \to \infty$*,*

$$\widetilde{DF}_{\tau,HAR} \Longrightarrow \widetilde{F}_{r,d} := \frac{\frac{br^{3/2}}{2}\left(B^H(r)\right)^2 - br^{1/2}\left(\int_0^r B^H(s)ds\right)B^H(r)}{\left(\left(\int_0^r \tilde{B}^H(s)^2 ds\right)\int_0^r\int_0^r -K_2''\left(\frac{p-q}{br}\right)B^H(p)B^H(q)dpdq\right)^{1/2}}, \text{ for } d \in [0, 0.5)$$

where $K_2''(.)$ *is the second derivative of* $K_2(\cdot)$*.*

Remark 29 *As in PWY, a sup test statistic (e.g.* $\sup_{\tau\in[\tau_0,n]}\widetilde{DF}_{\tau,HAR}$*) can be defined and we can also obtain the corresponding asymptotic distribution. This recursive formulation can identify a rational bubble when the time series has a mildly explosive root. We discuss the construction of this recursive statistic and its test for explosiveness in Appendix C.1.*

3.4 Alternative Hypothesis and Asymptotic Theory

To study the asymptotic behavior of the proposed test statistic under the alternative hypothesis, following the literature we consider two popular explosive models. The first alternative adopts the local-to-unit-root framework of Phillips (1987b) to study the locally explosive time series; see Harvey et al. (2016, 2019a, and 2019b). The advantage of using the locally explosive model is that it facilitates the computation of local power.

The second alternative is the mildly explosive model of Phillips and Magdalinos (2007). It assumes that the AR parameter has a greater deviation from the unit root than the local-to-unit-root model; see PWY, PSY and Phillips and Yu (2011). Under this explosive alternative, a consistent test can be obtained.

3.4.1 Locally explosive model

We first consider the alternative hypothesis with the following locally explosive setting:

$$\begin{aligned} y_t &= (y_{t-1} + u_t)\, 1\{t < \tau_e\} + (\rho_n y_{t-1} + u_t)\, 1\{\tau_e \leqslant t \leqslant n\}, t = 1, ..., n, \\ u_t &= (1-L)^{-d}\epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0, \sigma^2),\ d \in [0, 0.5), \\ \rho_n &= 1 + c/n,\ c > 0, \\ y_0 &= o_p(n^{1/2+d}). \end{aligned} \tag{56}$$

In model (56), y_t has a unit root before τ_e. It becomes mildly explosive after τ_e. That is, there is a structural break at τ_e. During both periods, the error term in the AR model has strong dependence with the same memory parameter d.

We now consider the asymptotic behavior of the HAR test statistic $\widetilde{DF}_{\tau,HAR}$.

Theorem 6 *Under model (56), for any $r > r_e$, as $n \to \infty$,*

$$\widetilde{DF}_{\tau,HAR} \Longrightarrow F^c_{r,d} := \frac{\left(\frac{\left(\frac{1}{2}C_{r,d} - \frac{1}{r}A_{r,d}B^H(r)\right)r}{B_{r,d} - \frac{1}{r}A^2_{r,d}} + cr\right)\left(B_{r,d} - \frac{1}{r}A^2_{r,d}\right)^{1/2}}{\left[\frac{2}{b}\left(\int_0^r G_{r_e,c}(d,p)^2 dp - \int_0^{(1-b)r} G_{r_e,c}(d,p) G_{r_e,c}(d,p+br) dp\right)\right]^{1/2}}, \tag{57}$$

where

$$\begin{aligned} A_{r,d} &= \int_0^r \left(e^{(x-r_e)c} B^H(r_e) + \int_{r_e}^x e^{(x-s)c} dB^H(s)\right) dx, \\ B_{r,d} &= \int_0^r \left(e^{(x-r_e)c} B^H(r_e) + \int_{r_e}^x e^{(x-s)c} dB^H(s)\right)^2 dx, \\ C_{r,d} &= \left(e^{(r-r_e)c} B^H(r_e) + \int_{r_e}^r e^{(r-s)c} dB^H(s)\right)^2, \\ G_{r_e,c}(p) &= B^H(p) - cA_{p,d} - \int_0^{r_e} B^H(p) dp. \end{aligned}$$

The limit distribution in Theorem 6 depends on the non-centralized parameter c. This parameter departs the distribution of $F^c_{r,d}$ from $F_{r,d}$. One can directly verify that if $c = 0$, $F^c_{r,d} = F_{r,d}$ from (54). Since both $F^c_{r,d}$ and $F_{r,d}$ are of $O_p(1)$, one may use them to compute the local power of the proposed test.

3.4.2 Mildly explosive model

We then consider the alternative hypothesis with the following mildly explosive setting:

$$\begin{aligned} y_t &= (y_{t-1}+u_t)\,1\{t<\tau_e\}+(\rho_n y_{t-1}+u_t)\,1\{\tau_e\leqslant t\leqslant n\},\ y_0=o_p(n^{1/2+d_1}), \\ u_t &= \begin{cases} (1-L)^{-d_1}\epsilon_t & \text{if } t<\tau_e, \\ (1-L)^{-d_2}\epsilon_t & \text{if } \tau_e\leqslant t\leqslant n, \end{cases} \quad \text{and } \epsilon_t \overset{iid}{\sim} (0,\sigma^2),\ d_1,d_2\in[0,0.5). \end{aligned} \tag{58}$$

where

$$\rho_n = 1+\frac{c}{n^\alpha},\ c>0,\ \alpha\in\left(0,\min\left\{\frac{1/2-d_1}{1/2-d_2},1\right\}\right). \tag{59}$$

In model (58), y_t has a unit root before τ_e. It becomes mildly explosive after τ_e. That is, there is a structural break at τ_e. During both periods, the error term in the AR model has strong dependence with memory parameters d_1 prior to the break and d_2 after the break. As $0<\alpha<1$, we obtain a higher degree of explosiveness than the local-to-unit-root explosiveness considered earlier. The upper bound on α (i.e. $\frac{1/2-d_1}{1/2-d_2}$) is needed as we have to ensure that the explosive observations dominate asymptotically the unit root observations with long memory errors.

Theorem 7 *Under model (56) with (59), as $n\to\infty$,*

$$DF_n \xrightarrow{p} \infty,$$

$$\widetilde{DF}_{n,HAR} \xrightarrow{p} \infty\ .$$

Theorem 7 first shows that DF_n diverges to infinity under the mildly explosive assumption. Combining with the result in Lemma 5, it means that the divergence of DF_n can be either from a strong dependent error or a mildly explosive process. So the rejection of null hypothesis does not solely come from the explosiveness and hence produce a difficulty for empirical researcher to identify a rational bubble. On the other hand, apply the robust statistic will not share the same problem as $\widetilde{DF}_{n,HAR}$ only diverges under the alternative assumption. Naturally, for any $(1-\beta)\%$ confidence level critical value $cv_{HAR}^{(1-\beta)\%}(d)$, we have $\Pr(\widetilde{DF}_{n,HAR} > cv_{HAR}^{(1-\beta)\%}(d)) \to 1$ under model (56) with (59). Therefore, we have a consistent test for this explosive alternative.

3.5 Estimation of Bubble Origination and Termination Dates

In this section, we discuss the estimation of the bubble origination and termination dates. Following PWY, we consider the following model:

$$\begin{aligned}
y_t &= (y_{t-1}+u_t)\,1\{t<\tau_e\} + (\rho_n y_{t-1}+u_t)\,1\{\tau_e \leqslant t \leqslant \tau_f\} + \left(\sum_{k=\tau_f+1}^{t} u_k + y_{\tau_f}^*\right) 1\{t>\tau_f\}, \\
\rho_n &= 1+\frac{c}{n^\alpha},\ c>0,\ \alpha\in\left(0, \min\left\{\frac{1/2-d_1}{1/2-d_2}, 1\right\}\right), \\
u_t &= (1-L)^{-d_t}\epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0,\sigma^2), \qquad (60)\\
d_t &= d_1 \text{ for } t\in[1,\tau_e)\cup[\tau_f+1,n],\ d_t=d_2 \text{ for } t\in[\tau_e,\tau_f],\ \tau_e=\lfloor nr_e\rfloor,\ \tau_f=\lfloor nr_f\rfloor, \\
y_{\tau_f}^* &= y_{\tau_e}+y^*, \text{ and } y^*=O_p(1).
\end{aligned}$$

Once again, we require $\alpha < \frac{1/2-d_1}{1/2-d_2}$.

Theorem 8 *Under model (60) with $\tau = \lfloor nr \rfloor$, $DF_{\tau,HAR}$ has the following asymptotic behaviour:*

$$\begin{aligned}
\widetilde{DF}_{\tau,HAR} &\xrightarrow{p} \infty \text{ if } \tau\in[\tau_e,\tau_f], \\
\widetilde{DF}_{\tau,HAR} &\xrightarrow{p} -\infty \text{ if } \tau\in[\tau_f+1,n].
\end{aligned} \qquad (61)$$

The estimators of r_e and r_f are defined as

$$\begin{aligned}
\hat{r}_e^{HAR} &= \inf_{r\geqslant r_0}\{r : \widetilde{DF}_{\tau,HAR} > cv_{n,HAR}\}, \\
\hat{r}_f^{HAR} &= \inf_{r>\hat{r}_e+\gamma\ln(n)/n}\{r : \widetilde{DF}_{\tau,HAR} < cv_{n,HAR}\}.
\end{aligned} \qquad (62)$$

If $r_e \geqslant r_0$ and the critical value $cv_{n,HAR}$ satisfies the following condition

$$\frac{1}{cv_{n,HAR}} + \frac{cv_{n,HAR}}{n^{(1-\alpha)/2}} \to 0, \qquad (63)$$

then, as $n\to\infty$, we have

$$\hat{r}_e^{HAR} \xrightarrow{p} r_e \text{ and } \hat{r}_f^{HAR} \xrightarrow{p} r_f.$$

Intuitively, $\lfloor n\hat{r}_e^{HAR}\rfloor$ represents the first observation when $\widetilde{DF}_{\tau,HAR} > cv_{n,HAR}$ and, after a bubble is deemed to have emerged and lasts longer than $\gamma\ln(n)/n$, $\lfloor n\hat{r}_f^{HAR}\rfloor$ represents the first observation when $\widetilde{DF}_{\tau,HAR} < cv_{n,HAR}$. Note that we require a bubble to have a minimum duration $\gamma\ln(n)/n$ where γ is a frequency dependent parameter.

Theorem 8 implies that $\widetilde{DF}_{\tau,HAR}$ provides a consistent test when data do not contain any observations after the bubble collapses. Otherwise, it diverges to negative infinity. Therefore, if the full sample is used to do the right-tailed unit root test, one should avoid including observations after a bubble collapses. Otherwise, the test cannot detect the presence of a bubble. This is known as Evans's critique (Evans, 1991).

For the consistent estimation of the bubble origination and termination dates, we require that the critical value $cv_{n,HAR}$ grows to infinity at a rate slower than $n^{(1-\alpha)/2}$, which is different from the rate obtained in PWY.

PWY propose to estimate r_e using $\hat{r}_e^{PWY}$ in (3.6.3) with cv_n increasing at the $\ln \ln n$ rate. However, in our model (60) with $d_1 \in (0, 0.5)$, we can easily obtain

$$\hat{r}_e^{PWY} \xrightarrow{p} r_0 \leqslant r_e.$$

Indeed, for $r < r_e$, $DF_{\tau,HAR}$ diverges at the rate n^{d_1} (as shown in Lemma 5), which is faster than $\ln \ln n$. Therefore, we can expect $\hat{r}_e^{PWY}$ with cv_n increasing at the $\ln \ln n$ rate to be inconsistent when $d > 0$ and $r_0 < r_e$. In the Monte Carlo simulations presented in section 3.6, we show that $\hat{r}_e^{PWY}$ with cv_n increasing at the $\ln \ln n$ rate tends to lead to a too early estimation of the bubble origination and termination dates. In contrast, $\hat{r}_e^{HAR}$ is a consistent estimator if $\ln \ln n$ is the growth rate adopted for $cv_{n,HAR}$.

Remark 30 *For model (60), as we have a negative divergence when full sample is used to calculate $\widetilde{DF}_{n,HAR}$, the sup statistic (e.g. $\sup_{\tau\in[\tau_0,n]} \widetilde{DF}_{\tau,HAR}$) can be employed, as the sup operator avoids the negative divergence part elicited from observations $\{y_t\}_{t\in[\tau_f+1,n]}$. We discuss the asymptotic property of the sup statistic in Appendix C.1.*

3.6 Monte Carlo Studies

In this section, we design some Monte Carlo experiments to study the size and power of our proposed test for bubble detection and our estimators of bubble origination and termination dates in finite samples. We use the normalized partial sum of $u_t = (1-L)^{-d}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} (0,1)$ to approximate a fractional Brownian motion.[10] This approximation allows

[10]To be more precise, we use $\frac{1}{\sigma\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \epsilon_t$ and $\frac{1}{\zeta n^{1/2+d}} \sum_{t=1}^{\lfloor nr \rfloor} u_t$ to approximate $B(r)$ and $B^H(r)$, respectively. See Equation (94) for the definition of ζ.

us to simulate DF_∞, $F_{r,0}$ and $F_{r,d}$ to obtain the critical values. The number of replications is always set to 2,500.

3.6.1 Empirical size

To investigate the empirical size of our test statistic, we perform a small-scale Monte Carlo study. In particular, we consider the following DGP,

$$\begin{aligned} y_t &= y_{t-1} + u_t,\ t = 1, ..., n, \\ u_t &= (1-L)^{-d}\epsilon_t,\ \epsilon_t \overset{iid}{\sim} N(0,1), \end{aligned} \tag{64}$$

with the following parameter settings: $d \in \{0, 0.05, 0.1, ..., 0.45\}$, $y_0 = 0$, and $n \in \{100, 500\}$.

Table 8 Empirical size of the right-tailed unit root test with 95% confidence level

$n=100$	d									
Test statistic	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
DF_n	0.05	0.09	0.14	0.20	0.26	0.31	0.36	0.41	0.45	0.48
$DF_{n,HAR}$	0.13	0.13	0.11	0.09	0.07	0.05	0.04	0.03	0.03	0.04
$\widetilde{DF}_{n,HAR}$	0.04	0.04	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06
$n=500$										
DF_n	0.05	0.10	0.18	0.27	0.34	0.40	0.42	0.49	0.53	0.56
$DF_{n,HAR}$	0.15	0.12	0.07	0.03	0.02	0.03	0.03	0.04	0.04	0.05
$\widetilde{DF}_{n,HAR}$	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Under these parameter settings, we perform a right-tailed unit test and calculate DF_n, $DF_{n,HAR}$ and $\widetilde{DF}_{n,HAR}$. For the standard right-tailed unit root test, we reject the null hypothesis when DF_n is greater than the 95% asymptotic critical value $cv^{95\%} = -0.08$.

For the HAR tests, we adopt $DF_{n,HAR}$ and $\widetilde{DF}_{n,HAR}$ and the corresponding 95% critical value. Moreover, we let $b = 0.05$ to calculate $\hat{\Omega}_{HAR}$[11].Note that as d is unknown, we obtain $\hat{d}$ using the ELW method proposed by Shimotsu and Phillips (2005) to get a feasible inference. For the tests with $DF_{n,HAR}$, the critical value 95% critical value is

[11]We choose this value of b because in extensive simulations, we find that for any $b \geqslant 0.05$, the test can deliver a empirical size which is close to its nominal value, while for the power analysis, a lower value of b yields a higher power.

obtained from the limit distribution in Theorem 4, we select $F_{1,0}$ if $\hat{d}=0$, and select $F_{1,\hat{d}}$ if $\hat{d}>0$. Finally, numerical simulation allows us to generate the 95% critical values based on $\Pr\left(F_{r,0} > cv^{95\%}_{DF_{n,HAR}}(0)\right) = 0.95$ if $\hat{d}=0$ and $\Pr\left(F_{r,\hat{d}} > cv^{95\%}_{DF_{n,HAR}}(\hat{d})\right) = 0.95$ if $\hat{d}>0$. For $\widetilde{DF}_{n,HAR}$, we generate the 95% critical value based on (55) for $\hat{d}\in[0,0.5)$.

Table 8 reports the empirical size of the 5% level right-tailed unit root test based on DF_n, $DF_{n,HAR}$ and $\widetilde{DF}_{n,HAR}$. Several observations can be made from Table 8. First, the standard unit root test only has a decent performance when $d=0$, it has a divergent size when $d>0$. For example, it can be seen that when $d=0.3$ and $n=500$, the test rejects the null hypothesis about 40% of the time, which is far above the nominal rate of 5%. These simulation results are consistent with the prediction of the asymptotic theory in Sowell (1990) and Lemma 5 and suggest that standard unit root test very often produce a spurious rational bubble detection under DGP (64) when $d>0$. Secondly, when $DF_{n,HAR}$ is adopted, while we do not see a divergent empirical size, the empirical size performance is not well controlled for a wide range of d. Finally, our modified statistic has a good performace across different value of d. In fact, the empirical size under all value of d is within 1% difference of the nominal size.

3.6.2 Power

To investigate the power of our test under finite sample, we design a Monte Carlo study based on model (58) with the following parameter settings: $n=100$, $y_0=100$, $r_e\in\{0.6,0.8\}$, $d_1=d_2=d\in\{0,0.05,0.1,...,0.45\}$, $\rho_n=1+c/n^{\alpha}$, $c=1$, and $\alpha=0.6$. Under these parameter settings, we perform right-tailed unit root tests with DF_n and $\widetilde{DF}_{n,HAR}$. We use the same procedures as in section 3.6.1 to perform tests with 95% confidence level.

Table 9 Finite sample power of the HAR test

$n=100$		d									
		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$r_e=0.6$	DF_n	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
	$\widetilde{DF}_{n,HAR}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98
$r_e=0.8$	DF_n	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.94
	$\widetilde{DF}_{n,HAR}$	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.93	0.89

Table 9 reports the empirical rejection rates under different parameter settings. Several observations can be made from Table 9. First, for both tests, higher power comes with a smaller r_e. This is expected as a lower value of r_e extends the duration of an explosive path and it leads to an easier bubble detection. Secondly, we see that while the difference in Finite sample power of tests with DF_n and $\widetilde{DF}_{n,HAR}$ is small, we see that finite sample power of DF_n is higher. This can be explained by the higher divergence rate of DF_n. Therefore, it can be seen that while $\widetilde{DF}_{n,HAR}$ do not provide a spurious explosiveness conclusion, a smaller power is a price for such a robust conclusion.

3.6.3 Real time bubble detection and estimation of origination and termination dates

To study the finite sample performance of the proposed estimators of bubble origination and termination dates, we design a Monte Carlo experiment based on model (60) with the following parameter settings: $n \in \{100, 150\}$, $y_0 = 100$, $c = 1$, $\alpha = 0.6$, $d_1 = d_2 = d \in \{0, 0.05, 0.1, ..., 0.45\}$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $y^*_{\tau_f} = y_{\tau_e}$, $r_e = 0.6$, $r_f = 0.8$, $r_0 = 0.4$, and $\gamma \ln(n)/n = 0.1$.

To obtain $\hat{r}_e^{HAR}$ and $\hat{r}_f^{HAR}$, we first calculate $\{DF_\tau, \widetilde{DF}_{\tau,HAR}\}_{\tau=\lfloor nr_0 \rfloor}^n$. Second, we let τ_0 be the minimum estimation window and obtain $\{\hat{d}_\tau\}_{\tau=\tau_0}^n$ using the ELW method based on the recursive data window $\{\{y_t\}_{t=1}^\tau\}_{\tau=\tau_0}^n$. Third, we specify the following critical value function $cv_{n,HAR}$,

$$cv_{n,HAR} = cv_{n,HAR}^{97\%}\left(\hat{d}_\tau\right) + \ln(\ln ns)/100, \tag{65}$$

where $s \in (0, 1]$ is proportional to the sample size.

Note that the critical value function is constructed as the sum of 97% critical value function under the memory parameter estimate $\hat{d}_\tau$ and a diverging factor $\ln(\ln ns)/100$, while this diverging factor guarantees that $cv_{n,HAR}$ satisfies condition (63) and therefore ensure the consistency of our bubble origination and terminate date estimators, it has a small impact in our finite sample simulation as $\ln(\ln ns)/100$ is between 0.01 and 0.015 while $cv_{n,HAR}^{97\%}\left(\hat{d}_\tau\right)$ has a much greater magnitude.

We also follow PWY and specify

$$cv_n = \ln(\ln ns)/100. \quad (66)$$

It can be seen that $cv_{n,HAR}$ is analogus to cv_n, as by construction cv_n is between the 95% and 99% right-tailed test critical value utilized the Dickey-Fuller-Phillips distribution in (42).

By combining (46), (47), (62), (65), and (66), we can calculate $\hat{r}_e^{HAR}, \hat{r}_f^{HAR}, \hat{r}_e^{PWY}$ and $\hat{r}_f^{PWY}$.

Table 10 Finite sample performance of $\hat{r}_e^{HAR}, \hat{r}_f^{HAR}, \hat{r}_e^{PWY}$ and $\hat{r}_f^{PWY}$

$r_e = 0.6$		d									
$r_f = 0.8$		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$n = 100$	$\hat{r}_e^{HAR}$	0.60	0.60	0.60	0.61	0.61	0.61	0.62	0.63	0.63	0.63
	$\hat{r}_f^{HAR}$	0.80	0.79	0.79	0.79	0.79	0.79	0.79	0.80	0.80	0.79
	$\hat{r}_e^{PWY}$	0.61	0.61	0.60	0.60	0.59	0.58	0.57	0.56	0.55	0.54
	$\hat{r}_f^{PWY}$	0.81	0.81	0.80	0.80	0.79	0.78	0.78	0.77	0.77	0.77
$n = 150$	$\hat{r}_e^{HAR}$	0.59	0.60	0.60	0.60	0.61	0.62	0.62	0.62	0.63	0.63
	$\hat{r}_f^{HAR}$	0.78	0.78	0.78	0.78	0.78	0.78	0.79	0.79	0.79	0.78
	$\hat{r}_e^{PWY}$	0.61	0.60	0.59	0.58	0.57	0.57	0.55	0.54	0.53	0.52
	$\hat{r}_f^{PWY}$	0.80	0.80	0.79	0.78	0.78	0.77	0.76	0.76	0.75	0.75

Table 10 reports the average of $\hat{r}_e^{HAR}$, $\hat{r}_f^{HAR}, \hat{r}_e$ and $\hat{r}_f^{HAR}$ across 2,500 replications[12]. With different values of d, our estimators of r_e and r_f are reasonably close to the true values of 0.6 and 0.8, respectively, and it has a stable performance across various values of d. When $d = 0$, $\hat{r}_e^{PWY}$ and $\hat{r}_f^{PWY}$ perform very well to estimate r_e and r_f. When $d \geqslant 0.2$, we can see a too early detection, as $\hat{r}_e^{PWY}$ is smaller than r_e on average, and the severeness of too early detection gets worse as d increases. Indeed, when $d \in (0, 0.5)$ with the critical value increasing at the rate $\ln\ln n$ and $r_0 < r_e$, $\hat{r}_e^{PWY}$ is an inconsistent estimator, which converges in probability to r_0. Interestingly, when $d > 0$, $\hat{r}_f^{PWY}$ is also inaccurate, because when $\hat{r}_e^{PWY}$ is close to r_0, the region between r_0 and r_e has a random wandering and non-standardized test statistic, as the test statistic occasionally falls below the too small critical value function (as $\ln\ln n$ is dominated by n^d) in this region. In this

[12] We only report the average of the the estimators when they are well defined.

experiment, the inconsistent estimator $\hat{r}_e^{PWY}$ also induces an inaccurate estimate of r_f in finite sample. Interestingly, two tests seem to give a similar estimate when d is small, the simulation exercise therefore suggests that the very different estimate of r_e and r_f can be a signal of the presence of a strongly dependent error.

The Monte Carlo simulation exercise to estimate r_e and r_f can also be seen as a real-time bubble detection exercise, as we adopt the first-time crossing principle to mark the time when the sequence of test statistic $\widetilde{DF}_{\tau,HAR}$ surpasses the 97% critical value function $cv_{n,HAR}^{97\%}\left(\hat{d}_\tau\right)$. Given the current data ($\{y_t\}_{t=1}^{\tau_0}$ in our simulations), we can initialize the calculation of HAR test statistic $\widetilde{DF}_{\tau_0,HAR}$, $cv_{n,HAR}^{97\%}\left(\hat{d}_{\tau_0}\right)$. As we update the HAR test statistic $\widetilde{DF}_{\tau,HAR}$ and $cv_{n,HAR}^{97\%}\left(\hat{d}_\tau\right)$ when new data become available, we can detect a rational bubble in real time.

3.7 Empirical Studies

To highlight the usefulness of our HAR test statistics, we conduct an empirical study using the same time series as that was used to obtain Table 1. We implement our HAR test statistics $\widetilde{DF}_{n,HAR}$ and use the critical values under 95% and 99% confidence level (also report the 90% confidence level) when performing the right-tailed unit root tests.

Table 11 Test for an explosive alternative

Sampling Period	$\hat{d}$	$\widetilde{DF}_{n,HAR}$	$cv_{HAR}^{90\%}\left(\hat{d}\right)$	$cv_{HAR}^{95\%}\left(\hat{d}\right)$	$cv_{HAR}^{99\%}\left(\hat{d}\right)$
(a) Jan 1872 to Feb 1880	0.24	1.25**	0.72	0.93	1.28
(b) Jun 1882 to May 1887	0.32	0.62	0.80	1.02	1.37
(c) May 1940 to Feb 1946	0.34	0.89*	0.78	0.99	1.38
(d) June 1948 to Nov 1955	0.29	1.54***	0.75	0.96	1.35
(e) Mar 1980 to July 1987	0.20	0.83*	0.67	0.89	1.26
(f) May 1988 to Apr 1998	0.24	1.20**	0.73	0.93	1.28

NOTES: "*", "***" and "***" denote the 90%, 95% and 99% level of significance for testing the existence of an mildly explosive epsiode.

Table 11 reports the HAR test statistic $\widetilde{DF}_{n,HAR}$ and the 90%, 95% and 99% confidence level critical values ($cv_{n,HAR}^{90\%}\left(\hat{d}\right)$, $cv_{n,HAR}^{95\%}\left(\hat{d}\right)$ and $cv_{n,HAR}^{99\%}\left(\hat{d}\right)$) for the six sampling periods. In Table 7, it is clear that the standard test statistics DF_n is greater than its 99% critical value for each sampling period, indicating a strong evidence for existence of rational bubble.

From Table 11, we can see that the test for sampling period (a) fails to reject the null under $cv^{90\%}_{n,HAR}\left(\hat{d}\right)$, while for sampling period (c) and (e), the tests reject the no bubble assumption at 90% confidence level. For the other sampling periods, our test finds strong evidence of the presence of explosiveness based on 95% confidence level[13]. If the conventional 95% confidence level is adopted, only sampling period (a), (d) and (f) are considered having a rational bubble.

Moreover, as discussed in Section 3.5 and Section 3.6.3, a strongly dependent error term will make the bubble origination date estimate inaccurate, this has a important implication on real time bubble detection. To showcase the empirical consequence, we use the sampling periods where we have concluded the existence of explosive episode in Table 5 to do a pseudo real time test. To implement the method, we use 60 monthly observations to initialize the estimation and calculate a sequence of $\widetilde{DF}_{\tau,HAR}$, $cv^{97\%}_{n,HAR}\left(\hat{d}_{\tau}\right)$ and $cv_{n,HAR}$ recursively. Finally, we estimate $\hat{r}^{PWY}_{e}$, $\hat{r}^{PWY}_{f}$, $\hat{r}^{HAR}_{e}$ and $\hat{r}^{HAR}_{f}$ as in our previous Monte Carlo simulation in Section 3.6.3, and we assume the minimum bubble duration to be 6 months. Through this exercise, we can estimate when an explosive episode starts and ends and hence the length of the episode in the stock market.

Table 12 Estimation for an explosive episode

Sampling period	$\hat{r}^{PWY}_{e}$	$\hat{r}^{PWY}_{f}$	Duration	$\hat{r}^{HAR}_{e}$	$\hat{r}^{HAR}_{f}$	Duration
(a') Jan 1872 to May 1880	July 1879	May 1880	10 mo	Oct 1879	Apr 1880	6 mo
(d') Jun 1948 to Feb 1957	Nov 1954	Feb 1957	26 mo	Dec 1954	Feb 1956	14 mo
(f') Jan 1989 to Dec 1998	Sep 1995	–	$>$ 40 mo	Feb 1997	Oct 1997	8 mo

Table 12 reports the estimates of bubble origination date ($\hat{r}^{PWY}_{e}$, $\hat{r}^{HAR}_{e}$), termination date ($\hat{r}^{PWY}_{f}$, $\hat{r}^{HAR}_{f}$) and duration (in months) for the corresponding sampling periods. The graphs which plot the estimates, sequence of statistics and critical values are documented in Appendix C.1.

Some conclusions can be drawn from Table 12. Firstly, using the PWY method, we always have an earlier bubble origination date estimate ($\hat{r}^{PWY}_{e}$) than HAR estimate ($\hat{r}^{HAR}_{e}$)

[13] In Appendix 1, a robustness checks for the rational bubble conclusion using the sup statistics was carried out to confirm the explosiveness.

in the 4 sampling periods, this is particularly obvious in sampling period (e') and (f') where the $\hat{r}_e$ using PWY method and our method are drastically different. Secondly, the bubble termination date estimate from the two methods also have a large difference. In sampling period (d') while our method estimates that the explosive episode ends in February 1956, PWY method predicts it ends at a year later. Finally, while the PWY method concludes that rational bubble can exist in the stock market for more than 2 years, our HAR method concludes that the longest explosive episode only last slightly longer than a year.

From these empirical applications, we has two interesting takeaways which are sharply different from PWY. Firstly, our HAR test concludes that it is less frequent that we have an explosive episode in US stock market. Secondly, when we do have an explosive episode, the duration of this exuberance period is shorter than PWY's estimate.

3.8 Conclusion

This paper introduces a new test and dating algorithm for the purpose of bubble detection. We motivate our test by showing empirical evidence that an autoregressive model may have strongly dependent errors. Because strongly dependent errors produce divergent Dickey-Fuller t statistics, the use of the traditional right-tailed unit root test statistics, such as the PWY statistic, spuriously detects a rational bubble. Not surprisingly, the PWY method also gives inaccurate estimators of the bubble origination and termination dates.

To avoid the spurious bubble detection, we propose a heteroskedasticity autocorrelation robust (HAR) test statistic. The idea behind our test is to use a properly self-normalized estimator of the standard error of the LS estimator of the AR(1) coefficient. We obtain the limit distribution of the proposed test statistic.

Based on a sequence of proposed test statistic, we then introduce new estimators to timestamp the bubble origination and termination dates based on the first-time-crossing principle. We show that the proposed estimators consistently estimate the bubble origination and termination dates when the true data generate process switches from a unit root model to a mild explosive model with a crash at the end of the explosive period and then switch back to a unit root model.

We have designed several Monte Carlo experiments to study the finite sample properties of the proposed test and estimator. Via simulated data, we first show that the traditional unit root test identifies too many bubbles when in an AR model with strongly dependent errors. We also show that the PWY estimator tends to have a too early estimation on the bubble origination and termination dates. Via the same simulated data, we then show that the proposed HAR statistic provides a test with well-controlled size and power in finite samples. The proposed estimator also leads to much better finite sample performance than the PWY estimators for the bubble origination and termination dates.

Our proposed test and estimators are applied to the data of S&P 500 monthly price-dividend ratio. According to the new test, one of the rational bubbles ((b) June 1882 to May 1887) identified by the traditional unit root test are not robust against strongly dependent errors, and some rational bubbles ((c) May 1940 to February 1946 and (e) March 1980 to July 1987) only are only detected at a weak confidence level. However, the proposed robust test re-confirm the explosiveness of these time series ((a) January 1872 to February 1880, (d) June 1948 to November 1955 and (f) May 1988 to April 1998). Using the PWY method, we notice that the estimate of bubble origination date are always earlier, and the bubble duration is noticeably longer than our robust estimate. Based on our theory and simulation result, we believe our estimate is more accurate, and the empirical application suggests that the explosive episode in stock market should be reasonably shorter than the PWY estimate.

While in this paper we have not addressed the issue of multiple bubbles, we should point out that our test statistic and the estimators can be extended to deal with the multiple bubbles in the same ways as in PSY (2015a, 2015b). The idea is to replace DF_τ with $\widetilde{DF}_{\tau,HAR}$. Such an extension will be investigated in a future study.

4 Reference

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. Journal of Finance, 54(4), 1361-1395.

Aït-Sahalia, Y., Jacod, J. (2014). High-Frequency Financial Econometrics. Princeton University Press.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and Forecasting Realized Volatility. Econometrica, 71(2), 579–625.

Anderson, T.W. (1959). On asymptotic distribution of estimates of parameters of stochastic difference equations. Annals of Mathematical Statistics, 30(3), 676–687.

Arvanitis, S. and Magdalinos, T. (2018). Mildly explosive autoregression under stationary conditional heteroskedasticity. Journal of Time Series Analysis, 39(6), 892-908.

Bai, J., Ng, S. (2005). Tests for Skewness, Kurtosis, and Normality for Time Series Data. Journal of Business & Economic Statistics, 23(1), 49-60.

Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 74(1),3-30.

Bao, Y., Ullah, A., & Wang, Y. (2017). Distribution of the mean reversion estimator in the Ornstein–Uhlenbeck process. Econometric Reviews, 36(6-9), 1039-1056.

Barros, C., Gil-Alana, L., & Payne, J. (2013). Tests of convergence and long memory behavior in US housing prices by state. Journal of Housing Research, 23(1), 73-87.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., & Taylor, R. L. (1991). Bootstrapping unstable first-order autoregressive processes. Annals of Statistics, 19(2), 1098-1101.

Billingsley, P. (1968) Convergence of probability measures. Wiley, New York.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. Annals of Statistics 16(4), 1709–1722.

Chang, J., Cheng, G., & Yao, Y., (2019). A Power One Test for Unit Roots Based on Sample Autocovariances, London School of Economics and Political Science, Working Paper.

Chang, J., Hall, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation, Biometrika 102 (1), 203-214.

Cheung, Y W. (1993). Long memory in foreign-exchange rates. Journal of Business and Economic Statistics, 11(1): 93-101.

Chevillon, G., & Mavroeidis, S. (2017). Learning can generate long memory. Journal of econometrics, 198(1), 1-9.

Christensen, B. J., & Nielsen, M. Ø. (2007). The effect of long memory in volatility on stock market fluctuations. The Review of Economics and Statistics, 89(4), 684-700.

Diba, B. & Grossman, H., (1988). Explosive Rational Bubbles in Stock Prices? American Economic Review, 78(3), 520-30.

EI Machkouri, M., Es-Sebaiy, K. and Ouknine, Y. (2016). Least squares estimator for non-ergodic Ornstein–Uhlenbeck processes driven by Gaussian processes. Journal of the Korean Statistical Society, 45(3), 329–341.

Evans, G.W. (1991). Pitfalls in Testing for Explosive Bubbles in Asset Prices. The American Economic Review, 81, 922–30.

Gatheral, J., Jaisson, T. and Rosenbaum, M. (2018). Volatility is rough. Quantitative Finance, 18(6), 933–949.

Gil-Alana, L. A., Barros, C., & Peypoch, N. (2014). Long memory and fractional integration in the housing price series of London and Paris. Applied Economics, 46(27), 3377-3388.

Giraitis, L., Koul, H. L. and Surgailis, D. (2012). Large sample inference for long memory processes. World Scientific Publishing Company.

Hall, P. (2013). The bootstrap and Edgeworth expansion: Springer Science & Business Media.

Hansen, B.E. (1999). The grid bootstrap and the autoregressive model. Review of Economics and Statistics, 81(4), 594-607.

Harvey, D. I., Leybourne, S. J., & Zu, Y. (2019a). Testing explosive bubbles with time-varying volatility. Econometric Reviews, 38(10), 1131-1151.

Harvey, D. I., Leybourne, S. J., & Zu, Y. (2019b). Sign-based Unit Root Tests for Explosive Financial Bubbles in the Presence of Nonstationary Volatility. Econometric Theory, 1-48.

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

Jiang, J., Wang, X., & Yu, J. (2018). New Distribution Theory for the Estimation of Structural Break Point in Mean. Journal of Econometrics, 205, 156-176.

Jiang, J., Wang, X., & Yu, J. (2020). In-fill Asymptotic Theory for Structural Break Point in Autoregression: A Unified Theory, Singapore Management University, Working Paper

Kiefer, N. M., & Vogelsang, T. J. (2002a). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. Econometric Theory, 18(6), 1350-1366.

Kiefer, N. M., & Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica, 70(5), 2093-2095.

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

Kiefer, N. M., Vogelsang, T. J., & Bunzel, H. (2000). Simple robust testing of regression hypotheses. Econometrica, 68(3), 695-714.

Kruse, R., & Wegener, C. (2019). Explosive behaviour and long memory with an application to European bond yield spreads. Scottish Journal of Political Economy, 66(1), 139-153.

Kwiatkowski, D., Phillips, P.C.B. , Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, Journal of Econometrics 54 (1-3), 159-178.

Li, J., Xiu, D. (2016), Generalized Method of Integrated Moments for High-Frequency Data. Econometrica, 84, 1613–1633.

Madan, D. B., Carr, P. P., & Chang, E. C. (1998). The variance gamma process and option pricing. Review of Finance, 2(1), 79-105.

Madan, D., & Seneta, E. (1990). The Variance Gamma (V.G.) Model for Share Market Returns. The Journal of Business, 63(4), 511-524.

Magdalinos, T. (2012). Mildly explosive autoregression under weak and strong dependence. Journal of Econometrics, 169(2), 179–187.

Mikusheva, A. (2007). Uniform inference in autoregressive models. Econometrica, 75(5), 1411-1452.

Mikusheva, A. (2015). Second order expansion of t-statistic in autoregressive models. Econometric Theory, 31(3), 426-448.

Ohanissian, A., Russell, J. R., & Tsay, R. S. (2008). True or spurious long memory? A new test. Journal of Business & Economic Statistics, 26(2), 161-175.

Park, J. (2003). Weak unit roots. Working Paper, Rice University.

Park, J. Y. (2003). Bootstrap unit root tests. Econometrica, 71(6), 1845-1895.

Pedersen, T. W. and Schütte, E.C.M. (2017), Testing for Explosive Bubbles in the Presence of Autocorrelated Innovations, CREATES Working Paper 2017-9, Aarhus University.

Perron, P. (1991). A continuous-time approximation to the unstable first-order autoregressive process: the case without an intercept. Econometrica, 59(1), 211-236.

Phillips, P. C. B. (1977). Approximations to Some Finite Sample Distributions Associated with a First-Order Stochastic Difference Equation. Econometrica, 45(2), 463–485.

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

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

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

Phillips, P. C. B. (2012). Folklore theorems, implicit maps, and indirect inference. Econometrica, 80(1), 425-454.

Phillips, P. C. B. & Yu, J. (2011). Dating the Timeline of Financial Bubbles During the Subprime Crisis. Quantitative Economics 2, 455-491.

Phillips, P. C. B., and Magdalinos, T. (2007a). Limit theory for moderate deviations from a unit root. Journal of Econometrics, 136(1), 115–130.

Phillips, P. C. B., and Magdalinos, T. (2007b). Limit theory for moderate deviations from a unit root under weak dependence. In: Phillips, G.D.A., Tzavalis, E. (Eds.), The Refinement of Econometric Estimation and Test Procedures. Cambridge University Press, UK.

Phillips, P. C. B., Magdalinos, T. and Giraitis, L. (2010). Smoothing local-to-moderate unit root theory. Journal of Econometrics, 158(2), 274–279.

Phillips, P. C. B., Shi, S. and Yu, J., (2014). Specification Sensitivity in Right-Tailed Unit Root Testing for Explosive Behaviour. Oxford Bulletin of Economics and Statistics, 76(3), 315-333.

Phillips, P. C. B., Shi, S. and Yu, J., (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. International Economic Review, 56(4), 1043–1078.

Phillips, P. C. B., Shi, S. and Yu, J., (2015b). Testing for Multiple Bubbles: Limit Theory of Real Time Detector. International Economic Review, 56(4), 1079-1134.

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 (1), 201–226.

Robinson, P. M. (1994). Semiparametric analysis of long-memory time series. Annals of Statistics, 22(1): 515–539.

Shimotsu, K., & Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. The Annals of Statistics, 33(4), 1890-1933.

Skorokhod, A.V. (1965) Studies in the Theory of Random Processes. Addison-Wesley.

Sowell, F. (1990). The fractional unit root distribution. Econometrica, 58(2), 495-505.

Sun, Y. (2004). A convergent t statistic in spurious regressions. Econometric Theory, 20(5), 943-962.

Tanaka, K. (1983). Asymptotic Expansions Associated with the AR(1) Model with Unknown Mean. Econometrica, 51(4), 1221-1231.

Tang, Chen Yong, & Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. Journal of Econometrics, 149(1), 65-81.

Vogelsang, T. J. (2003). Testing in GMM models without truncation. In Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later (pp. 199-233). Emerald Group Publishing Limited.

Wang, X., & Yu, J. (2016). Double asymptotics for explosive continuous-time models. Journal of Econometrics, 193(1), 35-53.

Wang, X., Xiao, W. and Yu, J. (2019). Estimation and Inference of Fractional Continuous-Time Model with Discrete-Sampled Data. SMU Economics and Statistics Working Paper Series, Paper No. 17-2019.

White, J. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. Annals of Mathematical Statistics, 29(4), 1188–1197.

Xiao, W. and Yu, J. (2019a). Asymptotic Theory for Estimating Drift Parameters in the Fractional Vasicek Model. Econometric Theory, 35(1), 198–231.

Xiao, W. and Yu, J. (2019b). Asymptotic Theory for Rough Fractiona Vasicek Models. Economics Letters, 177, 26-29.

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

Yu, J. (2014). Econometric Analysis of Continuous Time Models: A Survey of Peter Phillips's Work and Some New Results. Econometric Theory, 30(4), 737-774.

Zhou, Q., & Yu, J. (2010). Asymptotic distributions of the least squares estimator for diffusion processes. Working Paper 02-2010, Centre for Financial Econometrics, Singapore Management University.

Zhou, Q., & Yu, J. (2015). Asymptotic theory for linear diffusions under alternative sampling schemes. Economics Letters, 128, 1-5.

A Appendix to Chapter 1

A.1 Proof of Lemma 1 and Remark 4

Proof: Before proving Lemma 1, we define Υ_1, Υ_2 and Υ_3 in Lemma 1 by

$$\begin{aligned}
\Upsilon_1 :=\; & \frac{\exp(2c) - 4\exp(c) + 2c + 3}{2c^3} b^2 + \frac{2b}{c}\int_0^1 (\exp(rc) - 1) J_c(r) dr \\
& + \int_0^1 J_c^2(r) dr + \frac{\exp(2c) - 2\exp(c) + 1}{c^2} b\gamma_0 + 2\gamma_0 \int_0^1 \exp(rc) J_c(r) + \gamma_0^2 \frac{\exp(2c) - 1}{2c}; \\
\Upsilon_2 :=\; & \frac{\exp(c) - c - 1}{c^2} b + \int_0^1 J_c(r) dr + \frac{\exp(c) - 1}{c}\gamma_0; \\
\Upsilon_3 :=\; & \frac{2b}{c}\int_0^1 (\exp(rc) - 1) J_c(r) dr + \int_0^1 J_c(r) dW(r) + \gamma_0 \int_0^1 \exp(rc) dW(r); \\
J_c(r) :=\; & \int_0^r \exp(c(r-s)) dW(s);\ \gamma_0 := \frac{y_0}{\sigma_\psi \sqrt{N}};\ b := \left(\mu + \frac{\sigma i \psi'(0)}{\kappa}\right)\frac{\sqrt{-c\kappa}}{\sigma_\psi};\ c := -\kappa N.
\end{aligned}$$

Proof of Lemma 1 and Remark 4 can be done in the same way as in Zhou and Yu (2010). The only difference is that in Zhou and Yu (2010) $L(t) = W(t)$. If we divide Equation (7) by $\sigma_\psi \lambda_h$, and let $x_t = y_t / (\sigma_\psi \lambda_h)$, then we have $x_t = \rho_h x_{t-1} + \breve{g}_h + u_t$, where $\breve{g}_h = \frac{g_h}{\sigma_\psi \lambda_h}$. Under the in-fill scheme, we have

$$\frac{1}{T^2}\sum_{t=1}^{T} x_{t-1}^2 \Rightarrow \Upsilon_1,\ \frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t \Rightarrow \Upsilon_2,\ \frac{1}{T}\sum_{t=1}^{T} x_{t-1} u_t \Rightarrow \Upsilon_3. \tag{67}$$

Let $S(T,\kappa) = \frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^T y_{t-1}\epsilon_t - \frac{1}{\hat{\sigma} T}\sum_{t=1}^T y_{t-1} \frac{1}{\hat{\sigma} T}\sum_{t=1}^T \epsilon_t$, and $R(T,\kappa) = \frac{1}{\hat{\sigma}^2 T^2}\sum_{t=1}^T y_{t-1}^2 - \left(\frac{1}{\hat{\sigma} T^{\frac{3}{2}}}\sum_{t=1}^T y_{t-1}\right)^2$, where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T (y_t - \hat{g}_h - \hat{\rho}_h(\kappa) y_{t-1})^2$. By construction, it can be seen that

$$T(\hat{\rho}_h(\kappa) - \rho_h(\kappa)) = \frac{S(T,\kappa)}{R(T,\kappa)} \text{ and } t\,(Y,\rho,T) = \frac{S(T,\kappa)}{\sqrt{R(T,\kappa)}}.$$

Hence,

$$T(\hat{\rho}_h(\kappa) - \rho_h(\kappa)) = \frac{\frac{1}{T}\sum_{t=1}^T x_{t-1}u_t - \frac{1}{\sqrt{T}}\sum_{t=1}^T \epsilon_t \frac{1}{T^{3/2}}\sum_{t=1}^T x_t}{\frac{1}{T^2}\sum_{t=1}^T x_{t-1}^2 - \left(\frac{1}{T^{3/2}}\sum_{t=1}^T x_{t-1}\right)^2}.$$

Letting $\varsigma_h(\cdot) = -\ln(\cdot)/h$, we have

$$\hat{\kappa}_h - \kappa = \varsigma_h(\widehat{\rho}_h(\kappa)) - \varsigma_h(\rho_h(\kappa)) = \varsigma_h'(\widetilde{\rho}_h(\kappa))(\widehat{\rho}_h(\kappa) - \rho_h(\kappa)),$$

where $\widetilde{\rho}_h(\kappa)$ is a value between $\widehat{\rho}_h(\kappa)$ and $\rho_h(\kappa)$. Therefore, we can write

$$\frac{T}{\varsigma_h'(\rho_h(\kappa))}(\hat{\kappa}_h-\kappa)=\left(1+\frac{\varsigma_h'(\tilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right)T(\widehat{\rho}_h(\kappa)-\rho_h(\kappa)).$$

This implies

$$z\left(Y,\kappa,h\right)=h\varsigma_h'(\rho_h(\kappa))\left(1+\frac{\varsigma_h'(\tilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right)z\left(Y,\rho,T\right). \tag{68}$$

Since $\hat{\kappa}_h=\frac{-\ln(\hat{\rho}_h(\kappa))}{h}$, applying the generalized Delta method and using the relationship in (68), $Th=N$, $\left(1+\frac{\varsigma_h'(\tilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right)\to_p 1$, and $h\varsigma_h'(\widetilde{\rho}_h(\kappa))\to_p -1$, we obtain the limiting result $z\left(Y,\kappa,h\right)\Rightarrow-\frac{\Upsilon_3-\Upsilon_2\int_0^1 dW(r)}{\Upsilon_1-\Upsilon_2^2}$.

For $t\left(Y,\rho,T\right)$, we have

$$\begin{aligned}
t\left(Y,\rho,T\right) &= \frac{\sum_{t=1}^T y_{t-1}\epsilon_t-\frac{1}{T}\sum_{t=1}^T y_{t-1}\frac{1}{T}\sum_{t=1}^T\epsilon_t}{\sqrt{\hat{\sigma}^2\left(\sum_{t=1}^T y_{t-1}^2-\frac{1}{T}\left(\sum_{t=1}^T y_{t-1}\right)^2\right)}}\\
&= \frac{\frac{1}{\hat{\sigma}^2T}\sum_{t=1}^T y_{t-1}\epsilon_t-\frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^T y_{t-1}\frac{1}{\hat{\sigma}\sqrt{T}}\sum_{t=1}^T\epsilon_t}{\sqrt{\frac{1}{\hat{\sigma}^2T^2}\sum_{t=1}^T y_{t-1}^2-\left(\frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^T y_{t-1}\right)^2}}\\
&= \frac{\sigma_\psi\lambda_h}{\hat{\sigma}_c\sqrt{h}}\left[\frac{\frac{1}{T}\sum_{t=1}^T x_{t-1}u_t-\frac{1}{T^{3/2}}\sum_{t=1}^T x_{t-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T u_t}{\sqrt{\frac{1}{T^2}\sum_{t=1}^T x_{t-1}^2-\left(\frac{1}{T^{3/2}}\sum_{t=1}^T x_{t-1}\right)^2}}\right].
\end{aligned}$$

By Lemma 2, $\frac{\sigma_\psi\lambda_h}{\hat{\sigma}_c\sqrt{h}}\to_p 1$. Applying results in (67), we can obtain the limit of $t\left(Y,\rho,T\right)$.

To show the limit of $t(Y,\kappa,h)$, similar to (68), we have

$$t\left(Y,\kappa,h\right)=\varsigma_h'(\rho_h)h\left(1+\frac{\varsigma_h'(\widetilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right)t\left(Y,\rho,T\right).$$

Later, we will show that $\frac{\varsigma_h'(\widetilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}$ is $o_p(1)$, and $\varsigma_h'(\rho_h)h\to-1$. Hence, $t\left(Y,\kappa,h\right)=-t\left(Y,\rho,T\right)+o_p(1)$ under the in-fill scheme, giving the result in Remark 4. ■

A.2 Proof of Lemma 2

Before proving Lemma 2, we need the following lemma to show that we can obtain a consistent estimator of $\breve{g}_h$ at the rate of $h^{-1/2}$.

Lemma 6 *For Model (7), let $\hat{\breve{g}}_h$ be the LS estimator of $\breve{g}_h$. Then under the in-fill scheme, for any $\kappa \geqslant 0$, we have*

$$h^{-1/2}(\hat{\breve{g}}_h - \breve{g}_h) \Rightarrow \frac{1}{\sqrt{N}} \frac{\Upsilon_1 \eta - \Upsilon_2 \Upsilon_3}{\Upsilon_1 - \Upsilon_2^2},$$

where $\eta \sim i.i.d.N(0,1)$.

Proof: Note that

$$\hat{g}_h = \frac{\sum_{t=1}^T y_t \sum_{t=1}^T y_{t-1}^2 - \sum_{t=1}^T y_{t-1} \sum_{t=1}^T y_{t-1} y_t}{T \sum_{t=1}^T y_{t-1}^2 - \left(\sum_{t=1}^T y_{t-1}\right)^2}.$$

Using (7) and let $\hat{\breve{g}}_h = \frac{\hat{g}_h}{\sigma_\psi \lambda_h}$, we have

$$\hat{\breve{g}}_h - \breve{g}_h = \frac{\sum_{t=1}^T x_{t-1}^2 \sum_{t=1}^T u_t - \sum_{t=1}^T x_{t-1} \sum_{t=1}^T x_{t-1} u_t}{T \sum_{t=1}^T x_{t-1}^2 - \left(\sum_{t=1}^T x_{t-1}\right)^2}.$$

Therefore, we have

$$T^{1/2}(\hat{\breve{g}}_h - \breve{g}_h) = \left[\frac{\frac{1}{T^2}\sum_{t=1}^T x_{t-1}^2 \frac{1}{\sqrt{T}} \sum_{t=1}^T u_t - \frac{1}{T^{3/2}} \sum_{t=1}^T x_{t-1} \frac{1}{T} \sum_{t=1}^T x_{t-1} u_t}{\frac{1}{T^2}\sum_{t=1}^T x_{t-1}^2 - \left(\frac{1}{T^{3/2}} \sum_{t=1}^T x_{t-1}\right)^2}\right].$$

Note that $T = N/h$. Using (67), we therefore establish the result in Lemma 6. ■

We now prove Lemma 2.

Proof: Let the LS residual be $e_t = y_t - \hat{g}_h - \hat{\rho}_h y_{t-1}$ and

$$\begin{aligned}
\hat{\sigma}_c^2 &= \frac{1}{Th}\sum_{t=1}^T e_t^2 = \frac{1}{Th}\sum_{t=1}^T (\epsilon_t + (g_h - \hat{g}_h) + (\rho_h(\kappa) - \hat{\rho}_h) y_{t-1})^2 \\
&= \frac{1}{Th}\sum_{t=1}^T \epsilon_t^2 + \frac{1}{Th}\sum_{t=1}^T (g_h - \hat{g}_h)^2 + (\rho_h(\kappa) - \hat{\rho}_h)^2 \frac{1}{Th}\sum_{t=1}^T y_{t-1}^2 \\
&\quad + 2(g_h - \hat{g}_h)\frac{1}{Th}\sum_{t=1}^T \epsilon_t + 2(g_h - \hat{g}_h)(\rho_h(\kappa) - \hat{\rho}_h)\frac{1}{Th}\sum_{t=1}^T y_{t-1} \\
&\quad + 2(\rho_h(\kappa) - \hat{\rho}_h)\frac{1}{Th}\sum_{t=1}^T y_{t-1}\epsilon_t. \qquad (69)
\end{aligned}$$

We now investigate the five terms on the right-hand side of (69) one-by-one.

$$\frac{1}{Th}\sum_{t=1}^T \epsilon_t^2 = \frac{1}{Th}\sigma_\psi^2 \lambda_h^2 \sum_{t=1}^T u_t^2 \to_p \sigma_\psi^2,$$

$$\frac{1}{Th}\sum_{t=1}^{T}(g_h-\hat{g}_h)^2=\frac{(g_h-\hat{g}_h)^2}{h}=\frac{\sigma_\psi^2\lambda_h^2(\hat{\breve{g}}_h-\breve{g}_h)^2}{h}=O_p(h)=o_p(1),\quad \text{(by Lemma 6)}$$

$$\begin{aligned}
&(\rho_h(\kappa)-\hat{\rho}_h)^2\frac{1}{Th}\sum_{t=1}^{T}y_{t-1}^2\\
=&\left(\frac{\sum_{t=1}^{T}y_{t-1}\epsilon_t-\frac{1}{T}\sum_{t=1}^{T}y_{t-1}\epsilon_t}{\sum_{t=1}^{T}y_{t-1}^2-\frac{1}{T}(\sum_{t=1}^{T}y_{t-1})^2}\right)^2\frac{1}{Th}\sum_{t=1}^{T}y_{t-1}^2\\
=&\frac{\sigma_\psi^2\lambda_h^2}{Th}\left(\frac{\sum_{t=1}^{T}x_{t-1}u_t-\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t}{\sum_{t=1}^{T}y_{t-1}^2-\frac{1}{T}(\sum_{t=1}^{T}x_{t-1})^2}\right)^2\sum_{t=1}^{T}x_{t-1}^2\\
=&\frac{\sigma_\psi^2\lambda_h^2}{Th}\frac{\left(\sum_{t=1}^{T}x_{t-1}u_t\right)^2-2\frac{1}{T}\left(\sum_{t=1}^{T}x_{t-1}u_t\right)^2+\frac{1}{T^2}\left(\sum_{t=1}^{T}x_{t-1}u_t\right)^2}{\sum_{t=1}^{T}x_{t-1}^2-2\frac{1}{T}\left(\sum_{t=1}^{T}x_{t-1}\right)^2+\frac{1}{T^2}\frac{\left(\sum_{t=1}^{T}x_{t-1}\right)^4}{\sum_{t=1}^{T}x_{t-1}^2}}\\
=&\frac{\sigma_\psi^2\lambda_h^2}{Th}\frac{\left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t\right)^2-\frac{2}{T}\left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t\right)^2+\frac{1}{T^2}\left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t\right)^2}{\frac{1}{T^2}\sum_{t=1}^{T}x_{t-1}^2-2\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\right)^2+\frac{\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\right)^4}{\frac{1}{T^2}\sum_{t=1}^{T}x_{t-1}^2}}=O_p(T^{-1}),
\end{aligned}$$

$$(g_h-\hat{g}_h)\frac{1}{Th}\sum_{t=1}^{T}\epsilon_t=h^{-1/2}(g_h-\hat{g}_h)\sigma_\psi\frac{\lambda_h}{\sqrt{h}}\frac{1}{T}\sum_{t=1}^{T}u_t=O_p(h^{1/2})o_p(1)=o_p(1).$$

$$\begin{aligned}
(g_h-\hat{g}_h)(\rho_h(\kappa)-\hat{\rho}_h)\frac{1}{Th}\sum_{t=1}^{T}y_{t-1}&=O_p(h)O_p(T^{-1})\sigma_\psi\frac{\lambda_h}{h}\frac{1}{T}\sum_{t=1}^{T}x_{t-1}\\
&=O_p(1)\sigma_\psi\lambda_h\frac{1}{T^2}\sum_{t=1}^{T}x_{t-1}=o_p(1).
\end{aligned}$$

And finally,

$$(\rho_h(\kappa)-\hat{\rho}_h)\frac{1}{Th}\sum_{t=1}^{T}y_{t-1}\epsilon_t=(\rho_h(\kappa)-\hat{\rho}_h)\sigma_\psi^2\frac{\lambda_h^2}{h}\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t=O_p(T^{-1}).$$

Thus,

$$\begin{aligned}
\frac{\hat{\sigma}_c^2}{\sigma_\psi^2}-1 =& \frac{\lambda_h^2}{h}\frac{1}{T}\sum_{t=1}^{T}u_t^2-1+\frac{1}{\sigma_\psi^2}\frac{1}{Th}\sum_{t=1}^{T}(g_h-\hat{g}_h)^2+(\rho_h(\kappa)-\hat{\rho}_h)^2\frac{\lambda_h^2}{h}\frac{1}{\sigma_\psi^2}\frac{1}{T}\sum_{t=1}^{T}x_{t-1}^2\\
&+\frac{2}{\sigma_\psi}(g_h-\hat{g}_h)\frac{\lambda_h}{\sqrt{h}}\frac{1}{T}\sum_{t=1}^{T}u_t+\frac{2}{\sigma_\psi}(g_h-\hat{g}_h)(\rho_h(\kappa)-\hat{\rho}_h)\frac{1}{Th}\sum_{t=1}^{T}x_{t-1}\\
&+2(\rho_h(\kappa)-\hat{\rho}_h)\frac{\lambda_h}{\sqrt{h}}\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t.
\end{aligned}$$

Clearly, all terms on the right-hand side converge to zero in probability when $h \to 0$ and N is fixed. ■

A.3 Proof of Theorem 1 and Remark 7

Before proving Theorem 1 and Remark 7, we need some notations. Define $\epsilon_t^* = \hat{\sigma}_c \lambda_h u_t^*$ and a pair of statistics $(S^*(T,\kappa), R^*(T,\kappa))$ by

$$\begin{aligned}&(S^*(T,\kappa), R^*(T,\kappa))\\ =&\left(\frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^{T} y_{t-1}^* \epsilon_t^* - \frac{1}{\hat{\sigma} T}\sum_{t=1}^{T} y_{t-1}^* \frac{1}{\hat{\sigma} T}\sum_{t=1}^{T} \epsilon_t^*\ ,\ \frac{1}{\hat{\sigma}^2 T^2}\sum_{t=1}^{T} y_{t-1}^{*2} - \left(\frac{1}{\hat{\sigma} T^{\frac{3}{2}}}\sum_{t=1}^{T} y_{t-1}^*\right)^2\right).\end{aligned}$$

By construction, we have $z(Y^*,\rho,T) = S^*(T,\kappa)/R^*(T,\kappa)$ and $t(Y^*,\rho,T) = S^*(T,\kappa)/\sqrt{R^*(T,\kappa)}$. The ideas here is to show the asymptotic closeness of $z(Y^*,\kappa,T)$ and $z(Y,\kappa,T)$ uniformly in κ. We first restate Lemma 2 and Lemma 12 in Mikusheva (2007) which are used in our proof.

Lemma 7 (Lemma 2 and Lemma 12 in Mikusheva (2007)) *Under Model (7), let $S_j = \sum_{t=1}^{j} u_i$ be the partial sums. We can construct a sequence of processes $w_T(t) = \frac{1}{\sqrt{T}} S_{\lfloor Tt \rfloor}$ and a sequence of Brownian motions $\varsigma_T(t)$ on a common probability space, such that for every $\varepsilon > 0$, we have $\sup_{0 \leqslant t < 1} |w_T(t) - \varsigma_T(t)| = o_{as}(T^{-1/2+1/r+\varepsilon})$.*

Suppose that bootstrap error term $\{u_t^*\}_{t=1}^T$ drawn from our resampling method in Section 1.4, we can construct a process $\eta_T(t) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tt \rfloor} u_t^*$ and $\varsigma_T(t)$ on a common probability space such that as $T \to \infty$, $\sup_{0 \leqslant t < 1} |\eta_T(t) - \varsigma_T(t)| = o_p(T^{-\delta})$ for some $\delta > 0$. Thus, we have

$$\sup_{0 \leqslant t < 1} |\eta_T(t) - w_T(t)| = o_p(T^{-\delta}) \text{ for some } \delta > 0. \tag{70}$$

We now introduce the following Lemma which shows that, for every $\kappa \in K$, various bootstrap moments and statistics are close to their finite sample counterparts.

Lemma 8 *Suppose $\kappa_0 \in K$, where K is a compact set in the positive half-line, then for every $\varepsilon > 0$ and $\delta > 0$, we have*

1. $\lim_{h\to 0}\sup_{\kappa\in K}\Pr\left\{\left|\left(\tilde{g}_h - g_h\right)/\sigma_\psi\lambda_h\right| > \varepsilon\right\} = 0$;

2. $\sup_{\kappa\in K}\sup_t\left|\frac{1}{\hat{\sigma}}\left(\frac{y_t}{\sqrt{T}} - \frac{y_t^*}{\sqrt{T}}\right)\right| = o_p(T^{-\delta})$;

3. $\sup_{\kappa\in K}\sup_t\left|\frac{y_t}{\hat{\sigma}\sqrt{T}}\right| = O_p(1)$;

4. $\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \eta_T\left(\frac{t}{T}\right)u_t - \frac{1}{\sqrt{T}}\sum_{t=1}^T w_T\left(\frac{t}{T}\right)u_t^*\right| = o_p(T^{-\delta})$;

5. $\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}T}\sum_{t=1}^T y_{t-1}\epsilon_t - \frac{1}{\hat{\sigma}T}\sum_{t=1}^T y_{t-1}^*\epsilon_t^*\right| = o_p(T^{-\delta})$;

6. $\sup_{\kappa\in K}\left|\frac{1}{T^2\hat{\sigma}^2}\sum_{t=1}^T y_{t-1}^2 - \frac{1}{T^2\hat{\sigma}^2}\sum_{t=1}^T y_{t-1}^{*2}\right| = o_p(T^{-\delta})$;

7. $\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^T \epsilon_t\frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1} - \frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^T \epsilon_t^*\frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1}^*\right| = o_p(T^{-\delta})$;

8. $\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^T y_{t-1} - \frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^T y_{t-1}^*\right| = o_p(T^{-\delta})$;

9. $\lim_{h\to 0}\sup_{\kappa\in K}\Pr\{|z(Y,\rho,T) - z(Y^*,\rho,T)| > \varepsilon\} = 0$ *and* $\lim_{h\to 0}\sup_{\kappa\in K}\Pr\{|z(Y,\kappa,T) - z(Y^*,\kappa,T)| > \varepsilon\} = 0$.

Proof: Since u_t is i.i.d. with zero mean and unit variance, the Lindeberg–Lévy CLT Central Limit Theorem applies. Hence,

$$1.\ \frac{\tilde{g}_h - g_h}{\sigma_\psi\lambda_h} = \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^T u_t = O_p(T^{-1/2}).$$

$$
\begin{aligned}
2.\ \frac{1}{\hat{\sigma}}\frac{y_t}{\sqrt{T}} &= \frac{1}{\hat{\sigma}}\frac{1}{\sqrt{T}}\left[\sum_{i=1}^t \rho_h^{t-i}\epsilon_i + \rho_h^t y_0 + g_h\sum_{i=1}^t \rho_h^i\right] \\
&= \frac{\sigma_\psi\lambda_h}{\hat{\sigma}}\frac{1}{\sqrt{T}}\sum_{i=1}^t \rho_h^{t-i}u_i + \frac{\rho_h^t y_0}{\hat{\sigma}\sqrt{T}} + \frac{g_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^t \rho_h^i \\
&= \frac{\sigma_\psi\lambda_h}{\hat{\sigma}}\sum_{i=1}^t \rho_h^{t-i}\left[\eta_T\left(\frac{i}{T}\right) - \eta_T\left(\frac{i-1}{T}\right)\right] + \frac{\rho_h^t y_0}{\hat{\sigma}\sqrt{T}} + \frac{g_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^t \rho_h^i \\
&= \frac{\sigma_\psi\lambda_h}{\hat{\sigma}}\left[\sum_{i=1}^t \left(\rho_h^{t-i} - \rho_h^{t-i-1}\right)\eta_T\left(\frac{i}{T}\right) + \eta_T\left(\frac{t}{T}\right) + \rho_h^t\eta_T\left(\frac{0}{T}\right)\right] + \frac{\rho_h^t y_0}{\hat{\sigma}\sqrt{T}} + \frac{g_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^t \rho_h^i \\
&= \frac{\sigma_\psi\lambda_h}{\hat{\sigma}}\left[(\rho_h - 1)\sum_{i=1}^t \rho_h^{t-i-1}\eta_T\left(\frac{i}{T}\right) + \eta_T\left(\frac{t}{T}\right)\right] + \frac{\rho_h^t y_0}{\hat{\sigma}\sqrt{T}} + \frac{g_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^t \rho_h^i.
\end{aligned}
$$

$$\frac{1}{\hat{\sigma}}\frac{y_t^*}{\sqrt{T}} = \frac{\hat{\sigma}_c \lambda_h}{\hat{\sigma}}\left[(\rho_h - 1)\sum_{i=1}^{t}\rho_h^{t-i-1} w_T\left(\frac{i}{T}\right) + w_T\left(\frac{t}{T}\right)\right] + \frac{\rho_h^t y_0}{\hat{\sigma}\sqrt{T}} + \frac{\tilde{g}_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^{t}\rho_h^i.$$

Note that by Lemma 2 and the continuous mapping theorem, when N is fixed and $h \to 0$, we have $T \to \infty$, $\frac{\sigma_\psi \lambda_h}{\hat{\sigma}} \to_p 1$, and $\frac{\hat{\sigma}_c \lambda_h}{\hat{\sigma}} \to_p 1$. Hence,

$$\begin{aligned}
& \sup_{\kappa \in K}\sup_t \left|\frac{1}{\hat{\sigma}}\left(\frac{y_{t-1}}{\sqrt{T}} - \frac{y_{t-1}^*}{\sqrt{T}}\right)\right| \\
= \; & \sup_{\kappa \in K}\sup_t (1 + o_p(1)) \left| \begin{array}{c} (\rho_h - 1)\sum_{i=1}^{t}\rho_h^{t-i-1}\left(\eta_T\left(\frac{i}{T}\right) - w_T\left(\frac{i}{T}\right)\right) \\ + \eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right) + \frac{g_h - \tilde{g}_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^{t}\rho_h^i \end{array} \right| \\
\leqslant \; & \sup_{\kappa \in K}\left[(1 + o_p(1))\sup_t \left(\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|\left|\frac{\rho_h - 1}{\rho_h}\sum_{i=1}^{t}\rho_h^{t-i} + 1\right|\right)\right] + \sup_{\kappa \in K}\sup_t \left|\frac{g_h - \tilde{g}_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^{t}\rho_h^i\right| \\
\leqslant \; & (1 + o_p(1))\sup_t \left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| \sup_{\kappa \in K}\left|\frac{\rho_h - 1}{\rho_h}\frac{1 - \rho_h^t}{1 - \rho_h} + 1\right| + \sup_{\kappa \in K}\left|\frac{g_h - \tilde{g}_h}{\hat{\sigma}_c\sqrt{N}}\frac{\rho_h(1 - \rho_h^T)}{1 - \rho_h}\right| \\
\leqslant \; & (1 + o_p(1))\sup_t \left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| \sup_{\kappa \in K}\left|\frac{1}{\rho_h} + 1\right| + \sup_{\kappa \in K}|g_h - \tilde{g}_h|\frac{C_{\rho,1}}{\hat{\sigma}_c\sqrt{N}} \\
\leqslant \; & (1 + o_p(1))\sup_t \left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|(C_{\rho,2} + 1) + \sup_{\kappa \in K}|\frac{g_h - \tilde{g}_h}{\sigma_\psi \lambda_h}/|\frac{\sigma_\psi \lambda_h C_{\rho,1}}{\hat{\sigma}_c\sqrt{N}} = o_p(T^{-\delta}),
\end{aligned}$$

where both $C_{\rho,1}$ and $C_{\rho,2}$ depend on ρ_h.

3.
$$\begin{aligned}
& \sup_{\kappa \in K}\sup_t \left|\frac{1}{\hat{\sigma}}\frac{y_t}{\sqrt{T}}\right| \\
= \; & \sup_{\kappa \in K}\sup_t \left|\frac{\sigma \lambda_h}{\hat{\sigma}}\left[(\rho_h - 1)\sum_{i=1}^{t}\rho_h^{t-i-1}\eta_T\left(\frac{i}{T}\right) + \eta_T\left(\frac{t}{T}\right)\right] + \frac{\rho_h^t y_0}{\hat{\sigma}\sqrt{T}} + \frac{g_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^{t}\rho_h^i\right| \\
\leqslant \; & \sup_{\kappa \in K}\left[(1 + o_p(1))\left(\frac{\rho_h - 1}{\rho_h}\sum_{i=1}^{t}\rho_h^{t-j} + 1\right)\right]\sup_t |\eta_T(t)| + \sup_{\kappa \in K}\left|\frac{y_0}{\hat{\sigma}_c\sqrt{N}}\right| + \frac{C_\rho}{\hat{\sigma}_c\sqrt{N}} = O_p(1),
\end{aligned}$$

where C_ρ depends on ρ_h.

4. See Lemma 4c) in Mikusheva (2007).

5. Let $\breve{g}_h = \frac{g_h}{\sigma_\psi \lambda_h}$. Then, we have

$$
\begin{aligned}
& \frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^T y_{t-1}\epsilon_t \\
= & \frac{1}{\hat{\sigma}^2 T}\left(y_T \sum_{t=1}^T \epsilon_t - \sum_{t=1}^T (y_t - y_{t-1}) \sum_{k=0}^t \epsilon_k \right) \\
= & \frac{1}{\hat{\sigma}^2 T}\left(y_T \sum_{t=1}^T \epsilon_t - \sum_{t=2}^T (y_t - y_{t-1}) \sum_{k=0}^t \epsilon_k - (y_1 - y_0)\epsilon_1 \right) \\
= & \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\left(\frac{x_T}{\sqrt{T}} \sum_{t=1}^T \frac{u_t}{\sqrt{T}} - \sum_{t=2}^T \frac{\breve{g}_h + (\rho_h - 1)x_t + u_t}{\sqrt{T}} \sum_{k=0}^t \frac{z_k}{\sqrt{T}} \right) - \frac{[g_h + (\rho_h - 1)y_0 + \epsilon_1]\epsilon_1}{\hat{\sigma}^2 T} \\
= & \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\left(\frac{x_T}{\sqrt{T}}\eta_T(1) - \sum_{t=2}^T \frac{\breve{g}_h + (\rho_h - 1)x_{t-1} + u_t}{\sqrt{T}} \eta_T\left(\frac{t}{T}\right) \right) - \frac{\sigma_\psi \lambda_h (\rho_h - 1) y_0 u_1}{\hat{\sigma}^2 T} \\
& - \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\left(\frac{u_1}{\sqrt{T}}\right)^2 - \frac{g_h \epsilon_1}{\hat{\sigma}^2 T}.
\end{aligned}
$$

Similarly, denoting $\breve{\tilde{g}}_h = \frac{\tilde{g}_h}{\sigma_c \lambda_h}$, we have

$$
\begin{aligned}
\frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^T y_{t-1}^*\epsilon_t^* = & \frac{\lambda_h^2}{h}\left(\frac{x_T^*}{\sqrt{T}} w_T(1) - \sum_{t=2}^T \frac{\breve{\tilde{g}}_h + (\rho_h - 1)x_{t-1}^* + u_t^*}{\sqrt{T}} w_T\left(\frac{t}{T}\right) \right) \\
& - \frac{\hat{\sigma}_c \lambda_h (\rho_h - 1) y_0 u_1^*}{\hat{\sigma}^2 T} - \frac{\lambda_h^2}{h}\left(\frac{u_1^*}{\sqrt{T}}\right)^2 - \frac{\tilde{g}_h \epsilon_1^*}{\hat{\sigma}^2 T}.
\end{aligned}
$$

Hence,

$$
\frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^T y_{t-1}\epsilon_t - \frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^T y_{t-1}^*\epsilon_t^* = A + B + C + D + E + F + G,
$$

where

$$
\begin{aligned}
A = & \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\frac{x_T}{\sqrt{T}}\eta_T(1) - \frac{\lambda_h^2}{h}\frac{x_T^*}{\sqrt{T}} w_T(1),\ B = \frac{\lambda_h^2}{h}\frac{\breve{\tilde{g}}_h}{\sqrt{T}}\sum_{t=1}^T w_T\left(\frac{t}{T}\right) - \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\frac{\breve{g}_h}{\sqrt{T}}\sum_{t=1}^T \eta_T\left(\frac{t}{T}\right), \\
C = & \frac{(\rho_h - 1)\lambda_h^2}{h}\sum_{t=2}^T \frac{x_{t-1}^*}{\sqrt{T}} w_T\left(\frac{t}{T}\right) - \frac{(\rho_h - 1)\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^T \frac{x_{t-1}}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right), \\
D = & \frac{\lambda_h^2}{h}\sum_{t=2}^T \frac{u_t^*}{\sqrt{T}} w_T\left(\frac{t}{T}\right) - \frac{\sigma^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^T \frac{u_t}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right),\ E = \frac{(\rho_h - 1)\hat{\sigma}_c \lambda_h}{\hat{\sigma}^2 T} y_0 u_1^* - \frac{(\rho_h - 1)\sigma_\psi \lambda_h}{\hat{\sigma}^2 T} y_0 u_1, \\
F = & \frac{\lambda_h^2}{h}\left(\frac{z_1^*}{\sqrt{T}}\right)^2 - \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h}\left(\frac{z_1}{\sqrt{T}}\right)^2,\ G = \frac{\tilde{g}_h \epsilon_1^*}{\hat{\sigma}^2 T} - \frac{g_h \epsilon_1}{\hat{\sigma}^2 T}.
\end{aligned}
$$

We now examine these terms one-by-one.

$$\begin{aligned}
\sup_{\kappa\in K}\sup_t |A| &= \sup_{\kappa\in K}\sup_t \left|\frac{\sigma_\psi^2\lambda_h^2}{\hat\sigma_c^2 h}\frac{x_T}{\sqrt{T}}\eta_T(1) - \frac{\lambda_h^2}{h}\frac{x_T^*}{\sqrt{T}}w_T(1)\right| \\
&= \sup_{\kappa\in K}\sup_t \left|(1+o_p(1))\left(\frac{x_T}{\sqrt{T}}\eta_T(1) - \frac{x_T^*}{\sqrt{T}}w_T(1)\right)\right| \\
&= \sup_{\kappa\in K}\sup_t \left|(1+o_p(1))\left(\frac{x_T}{\sqrt{T}}(\eta_T(1) - w_T(1)) + \left(\frac{x_T}{\sqrt{T}} - \frac{x_T^*}{\sqrt{T}}\right)w_T(1)\right)\right| \\
&\leqslant (1+o_p(1))\left[\sup_{\kappa\in K}\left|\frac{x_T}{\sqrt{T}}\right|\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| + \sup_t\left|w_T(t)\left(\frac{x_t}{\sqrt{T}} - \frac{x_t^*}{\sqrt{T}}\right)\right|\right] \\
&= o_p(T^{-\delta}).
\end{aligned}$$

$$\begin{aligned}
\sup_{\kappa\in K}\sup_t |B| &= \sup_{\kappa\in K}\sup_t \left|\frac{\lambda_h^2}{h}\frac{\check{\tilde{g}}_h}{\sqrt{T}}\sum_{t=1}^T w_T\left(\frac{t}{T}\right) - \frac{\sigma_\psi^2\lambda_h^2}{\hat\sigma_c^2 h}\frac{\check g_h}{\sqrt{T}}\sum_{t=1}^T \eta_T\left(\frac{t}{T}\right)\right| \\
&= \sup_{\kappa\in K}\sup_t \left|(1+o_p(1))\frac{\check g_h}{\sqrt{T}}\left(\sum_{t=1}^T\left(w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right)\right)\right| \\
&\leqslant (1+o_p(1))\sup_{\kappa\in K}\left|\frac{g_h T}{\sigma_\psi\lambda_h\sqrt{T}}\right|\sup_t\left|w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| \\
&= (1+o_p(1))\sup_{\kappa\in K}\left|\frac{(\mu\kappa + \sigma i\psi'(0))\sqrt{N}}{\sigma\psi''(0)}\right|\sup_t\left|w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| \\
&= o_p(T^{-\delta}).
\end{aligned}$$

$$\begin{aligned}
\sup_{\kappa\in K}\sup_t |C| &= \sup_{\kappa\in K}\sup_t \left|\frac{(\rho_h-1)\lambda_h^2}{h}\sum_{t=2}^T\frac{x_{t-1}^*}{\sqrt{T}}w_T\left(\frac{t}{T}\right) - \frac{(\rho_h-1)\sigma_\psi^2\lambda_h^2}{\hat\sigma_c^2 h}\sum_{t=2}^T\frac{x_{t-1}}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right)\right| \\
&= (1+o_p(1))\sup_{\kappa\in K}|\rho_h - 1|\sup_t\sum_{t=2}^T\left|\begin{array}{l}\frac{x_{t-1}^*}{\sqrt{T}}\left(w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right) \\ +\eta_T\left(\frac{t}{T}\right)\left(\frac{x_{t-1}^*}{\sqrt{T}} - \frac{x_{t-1}}{\sqrt{T}}\right)\end{array}\right| \\
&\leqslant \sup_{\kappa\in K}|-\kappa h + o(h^2)|T\left(\begin{array}{c}\sup_t\left|\frac{x_{t-1}^*}{\sqrt{T}}\right|\sup_t\left|w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| \\ +\sup_t\left|\eta_T\left(\frac{t}{T}\right)\right|\sup_t\left|\frac{x_{t-1}^*}{\sqrt{T}} - \frac{x_{t-1}}{\sqrt{T}}\right|\end{array}\right) \\
&\leqslant CN\left(\sup_t\left|\frac{x_{t-1}^*}{\sqrt{T}}\right|\sup_t\left|w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| + \sup_t\left|\eta_T\left(\frac{t}{T}\right)\right|\sup_t\left|\frac{x_{t-1}^*}{\sqrt{T}} - \frac{x_{t-1}}{\sqrt{T}}\right|\right) \\
&= o_p(T^{-\delta}).
\end{aligned}$$

$$
\begin{aligned}
\sup_{\kappa\in K}\sup_t |D| &= \sup_{\kappa\in K}\sup_t \left|\frac{\lambda_h^2}{h}\sum_{t=2}^{T}\frac{u_t^*}{\sqrt{T}}w_T\left(\frac{t}{T}\right) - \frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^{T}\frac{u_t}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right)\right| \\
&= \sup_t\left|(1+o_p(1))\left(\sum_{t=2}^{T}\frac{u_t^*}{\sqrt{T}}w_T\left(\frac{t}{T}\right) - \sum_{t=2}^{T}\frac{u_t}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right)\right)\right| \\
&= \sup_t\left|(1+o_p(1))\left(\sum_{t=2}^{T}\frac{u_t^*}{\sqrt{T}}\left(w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right) + \sum_{t=2}^{T}\left(\frac{u_t^*-u_t}{\sqrt{T}}\right)\eta_T\left(\frac{t}{T}\right)\right)\right| \\
&\leqslant \sup_t\left|w_T\left(\frac{t}{T}\right)\right|\sup_t\left|w_T\left(\frac{t}{T}\right)-\eta_T\left(\frac{t}{T}\right)\right| + \sup_t\left|w_T\left(\frac{t}{T}\right)-\eta_T\left(\frac{t}{T}\right)\right|\sup_t\left|\eta_T\left(\frac{t}{T}\right)\right| \\
&= o_p(T^{-\delta}).
\end{aligned}
$$

$$
\begin{aligned}
\sup_{\kappa\in K}\sup_t |E| &= \sup_{\kappa\in K}\sup_t\left|\frac{(\rho_h-1)\hat{\sigma}_c\lambda_h}{\hat{\sigma}^2 T}y_0u_1^* - \frac{(\rho_h-1)\sigma_\psi\lambda_h}{\hat{\sigma}^2 T}y_0u_1\right| \\
&= \sup_{\kappa\in K}\sup_t\left|\kappa h\frac{1}{\sigma}\frac{\lambda_h}{\sqrt{h}}\frac{y_0}{\sqrt{hT}}\left[\frac{u_1}{\sqrt{T}}\right] - \kappa h\frac{1}{\sigma}\frac{\lambda_h}{\sqrt{h}}\frac{y_0}{\sqrt{hT}}\left[\frac{u_1^*}{\sqrt{T}}\right]\right| + o_p(h) \\
&\leqslant C_\kappa h\frac{1}{\sigma}\frac{|y_0|}{\sqrt{N}}\left[\sup_t\left|\eta_T\left(\frac{t}{T}\right)\right| + \sup_t\left|w_T\left(\frac{t}{T}\right)\right|\right] + o_p(1) = o_p(h),
\end{aligned}
$$

where C_κ depends on κ.

$$
\begin{aligned}
\sup_{\kappa\in K}\sup_t |F| &= \sup_{\kappa\in K}\sup_t\left|\frac{\lambda_h^2}{h}\left(\frac{u_1^*}{\sqrt{T}}\right)^2 - \frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\left(\frac{u_1}{\sqrt{T}}\right)^2\right| \\
&= \sup_{\kappa\in K}\sup_t\left|(1+o_p(h))\left(\left(\frac{u_1^*}{\sqrt{T}}\right)^2 - \left(\frac{u_1}{\sqrt{T}}\right)^2\right)\right| \\
&= \sup_{\kappa\in K}\sup_t\left|(1+o_p(h))\left[\left(\frac{u_1^*}{\sqrt{T}}\right) - \left(\frac{u_1}{\sqrt{T}}\right)\right]\left[\left(\frac{u_1^*}{\sqrt{T}}\right) + \left(\frac{u_1}{\sqrt{T}}\right)\right]\right| \\
&\leqslant (1+o_p(h))\sup_t\left|w_T\left(\frac{t}{T}\right)-\eta_T\left(\frac{t}{T}\right)\right|\left(\sup_t\left|w_T\left(\frac{t}{T}\right)\right| + \sup_t\left|\eta_T\left(\frac{t}{T}\right)\right|\right) \\
&= o_p(T^{-\delta}).
\end{aligned}
$$

$$
\sup_{\kappa\in K}\sup_t |G| \leqslant \sup_{\kappa\in K}\left|(1+o_p(1))\frac{g_h}{\hat{\sigma}_c^2 N}\right|\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| = o_p(T^{-\delta}).
$$

Thus, we have established item 5.

For item 6,

$$\begin{aligned}
\sup_{\kappa\in K}\left|\frac{\sum_{t=1}^{T}y_{t-1}^{2}}{T\hat{\sigma}^{2}}-\frac{\sum_{t=1}^{T}y_{t-1}^{*2}}{T\hat{\sigma}^{2}}\right| &= \sup_{\kappa\in K}\left|\frac{1}{T\hat{\sigma}^{2}}\sum_{t=2}^{T}y_{t-1}^{2}-\frac{1}{T\hat{\sigma}^{2}}\sum_{t=2}^{T}y_{t-1}^{*2}+\frac{1}{T\hat{\sigma}^{2}}y_{0}^{2}-\frac{1}{T\hat{\sigma}^{2}}y_{0}^{2}\right| \\
&= \sup_{\kappa\in K}\left|\frac{\sigma_{\psi}^{2}\lambda_{h}^{2}}{\hat{\sigma}_{c}^{2}h}\sum_{t=2}^{T}\left(\frac{x_{t-1}}{\sqrt{T}}\right)^{2}-\frac{\lambda_{h}^{2}}{h}\sum_{t=2}^{T}\left(\frac{x_{t-1}^{*}}{\sqrt{T}}\right)^{2}\right| \\
&= \sup_{\kappa\in K}\left|(1+o_{p}(h))\left[\sum_{t=2}^{T}\left(\frac{x_{t-1}}{\sqrt{T}}\right)^{2}-\sum_{t=2}^{T}\left(\frac{x_{t-1}^{*}}{\sqrt{T}}\right)^{2}\right]\right| \\
&\leqslant (1+o_{p}(h))\sup_{t}\left|\frac{x_{t}}{\sqrt{T}}-\frac{x_{t}^{*}}{\sqrt{T}}\right|\left(\sup_{t}\left|\frac{x_{t}}{\sqrt{T}}\right|+\sup_{t}\left|\frac{x_{t}^{*}}{\sqrt{T}}\right|\right) \\
&= o_{p}(T^{-\delta}).
\end{aligned}$$

For item 7,

$$\begin{aligned}
&\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}^{2}\sqrt{T}}\sum_{t=1}^{T}\epsilon_{t}\frac{1}{T^{3/2}}\sum_{t=1}^{T}y_{t-1}-\frac{1}{\hat{\sigma}^{2}\sqrt{T}}\sum_{t=1}^{T}\epsilon_{t}^{*}\frac{1}{T^{3/2}}\sum_{t=1}^{T}y_{t-1}^{*}\right| \\
=\;& \sup_{\kappa\in K}\left|\frac{\sigma_{\psi}^{2}\lambda_{h}^{2}}{\hat{\sigma}_{c}h}\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\sum_{t=1}^{T}\frac{u_{t}}{\sqrt{T}}-\frac{\lambda_{h}^{2}}{h}\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\sum_{t=1}^{T}\frac{u_{t}^{*}}{\sqrt{T}}\right. \\
&\left.+\frac{\sigma_{\psi}^{2}\lambda_{h}^{2}}{\hat{\sigma}_{c}^{2}h}\sum_{t=1}^{T}\frac{u_{t}^{*}}{\sqrt{T}}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}-\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}^{*}\right)\right| \\
\leqslant\;& \sup_{\kappa\in K}\sup_{t}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{x_{t-1}}{\sqrt{T}}\right|\sup_{t}\left|\eta_{T}\left(\frac{t}{T}\right)-w_{T}\left(\frac{t}{T}\right)\right| \\
&+\frac{1}{T}\sup_{t}\left|w_{T}\left(\frac{t}{T}\right)\right|\sup_{t}\left|\eta_{T}\left(\frac{t}{T}\right)-w_{T}\left(\frac{t}{T}\right)\right|+o_{a.s.}(1)=o_{p}(T^{-\delta}).
\end{aligned}$$

For item 8,

$$\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^{T}y_{t-1}-\frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^{T}y_{t-1}^{*}\right|=\sup_{\kappa\in K}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{\hat{\sigma}}\left(\frac{y_{t-1}}{\sqrt{T}}-\frac{y_{t-1}^{*}}{\sqrt{T}}\right)\right]\right|=o_{p}(T^{-\delta}).$$

For item 9,

$$\begin{aligned}
&\sup_{\kappa\in K}\Pr\{|z\left(Y,\rho,T\right)-z\left(Y^{*},\rho,T\right)|>\epsilon\} \\
=\;& \sup_{\kappa\in K}\Pr\left\{\left|\frac{S(T,\kappa)}{R(T,\kappa)}-\frac{S^{*}(T,\kappa)}{R^{*}(T,\kappa)}\right|>\epsilon\right\} \\
\leqslant\;& \sup_{\kappa\in K}\Pr\left\{\left|C\left(|S(T,\kappa)-S^{*}(T,\kappa)|+|R(T,\kappa)-R^{*}(T,\kappa)|\right)\right|>\epsilon\right\}\to_{p}0.
\end{aligned}$$

From the relationship of $z\left(Y,\rho,T\right)$ and $z\left(Y,\kappa,h\right)$, the closeness of $z\left(Y,\rho,T\right)$ and

$z(Y^*, \rho, T)$ implies the closeness of $z(Y, \kappa, h)$ and $z(Y^*, \kappa, h)$.

$$
\begin{aligned}
& \sup_{\kappa \in K} \Pr\{|z(Y, \kappa, h) - z(Y^*, \kappa, h)| > \epsilon\} \\
= \; & \sup_{\kappa \in K} \Pr\left\{ \left| \varsigma_h'(\rho_h(\kappa)) h \left(1 + \frac{\varsigma_h'(\tilde{\rho}_h(\kappa)) - \varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right) z(Y, \rho, T) \right. \right. \\
& \left. \left. - \varsigma_h'(\rho_h) h \left(1 + \frac{\varsigma_h'(\tilde{\rho}_h^*(\kappa)) - \varsigma_h'(\rho_h)}{\varsigma_h'(\rho_h)}\right) z(Y^*, \rho, T) \right| > \epsilon \right\} \\
= \; & \sup_{\kappa \in K} \Pr\{(1 + o_p(1)) |z(Y, \rho, T) - z(Y^*, \rho, T)|\} \to_p 0.
\end{aligned}
$$

The last step is due to Theorem 1 in Phillips (2012) as the sequence $\{\varsigma_h'(\rho_h(\kappa))\}$ is asymptotically locally relatively equicontinuous in ρ. Since $\hat{\rho}_h - \rho_h = O_p(T^{-1})$, let a shrinking neighborhood denoted by

$$
B_\delta^h = \left\{ \hat{\rho}_h : |\hat{\rho}_h - \rho_h| < \frac{\delta}{T^a} \right\},
$$

where $\delta > 0$ and $a \in (0, 1)$. Note that for any ρ_h in B_δ^h, we have:

$$
\frac{\varsigma_h'(\hat{\rho}_h) - \varsigma_h'(\rho_h)}{\varsigma_h'(\rho_h)} = -\frac{\frac{1}{h\rho_h} - \frac{1}{h\hat{\rho}_h}}{\frac{1}{h\rho_h}} = \frac{\rho_h - \hat{\rho}_h}{\hat{\rho}_h} \leqslant \frac{\delta}{T^a(\rho_h + o_p(1))} \to_p 0.
$$

Now we are in the position to show Theorem 1, that is,

$$
\sup_{\kappa \in K} \sup_x |\Pr\{z(Y, \kappa, h) < x\} - \Pr{}^*\{z(Y^*, \kappa, h) < x | Y\}| \to 0;
$$

$$
\inf_{\kappa \in K} \Pr\{\kappa_0 \in CI_q\} \to x_2 - x_2 = q. \tag{71}
$$

Since $S^*(T, \kappa)$ and $R^*(T, \kappa)$ are jointly uniformly continuous by Assumption 4, this implies that $z(Y^*, \rho, T)$ is uniformly continuous in the following sense (see Lemma 2 in Mikusheva (2007)),

$$
\lim_{h \to 0} \sup_{\kappa \in K} \sup_x |\Pr\{z(Y, \rho, T) < x | \kappa\} - \Pr{}^*\{z(Y^*, \rho, T) < x | \kappa, Y\}| = 0.
$$

Similarly, for $\Pr(z(Y, \kappa, h) < x | \kappa)$ and $\Pr{}^*\{z(Y^*, \kappa, T) < x | \kappa, Y\}$, we have

$$
\begin{aligned}
\Pr(z(Y, \kappa, h) < x | \kappa) &= \Pr\left(z(Y, \rho, T) < x \frac{1}{\varsigma_h(\rho_h) h} \left(1 + \frac{\varsigma_h'(\rho_h) - \varsigma_h'(\rho_h)}{\varsigma_h'(\rho_h)}\right)^{-1} | \kappa \right) \\
&= \Pr(z(Y, \rho, T) < -x\rho_h + o_p(1) | \kappa) \\
&= \Pr(z(Y, \rho, T) < -x + o_p(1) | \kappa),
\end{aligned}
$$

and

$$\Pr{}^*\{z\left(Y^*,\kappa,T\right)<x|\kappa,Y\}=\Pr{}^*\left(z\left(Y^*,\rho,T\right)<-x+o_p(1)|\kappa,Y\right).$$

Thus, as $h\to 0$, we have

$$\sup_{\kappa\in K}\sup_{x}|\Pr\{z\left(Y,\kappa,h\right)<x\}-\Pr{}^*\{z\left(Y^*,\kappa,h\right)<x|Y\}|\to 0.$$

The final claim (71) is a direct result of Lemma 1 in Mikusheva (2007). The result in Remark 7 is established based on the same argument and is omitted. ■

A.4 Proof of Theorem 2

Before we prove Theorem 2, we need to introduce three lemmas. All three lemmas rely on the probabilistic embedding of the partial sum process in an expanded probability space. For details about the embedding, see Park (2003).

Lemma 9 (Park (2003), Lemma 3.5(a)) *Assume that z_j are i.i.d. random variable with mean 0 and variance σ_z^2, and $E|z_j|^r<\infty$ for some $r\geqslant 8$. Let $N(t)=W(1+t)-W(1)$, and $M(t)$ be a Brownian motion which is independent on W. Then*

$$\frac{1}{\sqrt{T}\sigma_z}\sum_{t=1}^{T}u_t=W(1)+\frac{1}{T^{1/4}}M(V)+\frac{1}{\sqrt{T}}N(V)+o_p(T^{-1/2}),$$

where $\mathcal{B}=(W,V,U)$ is a Brownian motion with variance matrix Σ as

$$\Sigma=\begin{bmatrix}1 & \mu_3/3\sigma_z^3 & \mu_3/\sigma_z^3\\ \mu_3/3\sigma_z^3 & \varrho/\sigma_z^4 & (\mu_4-3\sigma_z^4+3\varrho)/6\sigma_z^4\\ \mu_3/\sigma_z^3 & (\mu_4-3\sigma_z^4+3\varrho)/6\sigma_z^4 & (\mu_4-\sigma_z^4)/\sigma_z^4\end{bmatrix}.$$

Here, $\mu_3=Ez_j^3$, $\mu_4=Ez_j^4$, $\varrho=E(\tau_j-\sigma_z^2)^2$. We define τ_j implicitly by Skorohod's embedding scheme (Skorohod, 1965) such that on an extended probability space, we have the distribution equivalence given by

$$\left\{\frac{1}{\sqrt{T}\sigma_z}\sum_{i=1}^{j}z_i\right\}_{j=1}^{T}\overset{d}{=}\left\{W\left(\frac{1}{T\sigma_z^2}\sum_{i=1}^{j}\tau_i\right)\right\},$$

where $\left(\frac{1}{T\sigma_z^2}\sum_{i=1}^{j}\tau_i\right)$ is known as the stopping time.

Lemma 10 (Mikusheva (2015), Theorem 1) *Suppose $c \leqslant 0$ and z_j satisfies the assumption in Lemma 9. Let $\tilde{x}_t = \sum_{j=1}^{t} \exp\left(c\left(\frac{t-j}{T}\right)\right) z_j$, and z_j is an i.i.d. random variable with mean 0 and variance 1. Then we have the following results:*

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{t-1}u_t &= \int_0^1 J_c(r)dW(r) + \frac{1}{T^{1/4}}J_c(1)M(V) \\
&\quad + \frac{1}{\sqrt{T}}\left(-c\int_0^1\int_0^r e^{c(r-s)}J_c(s)dV(s)dW(r) + J_c(1)N(V) + \frac{1}{2}M^2(V) - \frac{1}{2}U\right) \\
&\quad + o_p(T^{-1/2}).
\end{aligned}$$

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T}\tilde{x}_t^2 &= \int_0^1 J_c^2(r)dr - \frac{2c}{\sqrt{T}}\int_0^1 J_c(r)\int_0^r e^{c(r-s)}J_c(s)dV(s)dr \\
&= -\frac{1}{\sqrt{T}}\int_0^1 J_c^2(r)dV(r) + \frac{1}{\sqrt{T}}J_c^2(1)V - \frac{2\mu_3}{3\sqrt{T}}\int_0^1 J_c(r)dr + o_p(T^{-1/2}).
\end{aligned}$$

$$\begin{aligned}
\frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_t &= \int_0^1 J_c(r)dr - \frac{c}{\sqrt{T}}\int_0^1\int_0^r e^{c(r-s)}J_c(s)dV(s)dr - \frac{1}{\sqrt{T}}\int_0^1 J_c(r)dV(r) \\
&= +\frac{1}{\sqrt{T}}J_c(1)V - \frac{\mu_3}{3\sqrt{T}} + o_p(T^{-1/2}).
\end{aligned}$$

Lemma 11 *Under model (4), if $\kappa \geqslant 0$, then we have*

1. $\frac{1}{T}\sum_{t=1}^{T} x_t z_{t+1} = \Upsilon_3 + \frac{1}{T^{1/4}}R_{3,T^{-1/4}} + \frac{1}{T^{1/2}}R_{3,T^{-1/2}} + o_p(T^{-1/2})$;
2. $\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 = \Upsilon_1 + \frac{1}{T^{1/2}}R_{1,T^{-1/2}} + o_p(T^{-1/2})$;
3. $\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t = \Upsilon_2 + \frac{1}{T^{1/2}}R_{2,T^{-1/2}} + o_p(T^{-1/2})$,

where

$$\begin{aligned}
R_{3,T^{-1/4}} &= J_c(1)N(V) + \frac{b}{c}M(V); \\
R_{3/T^{-1/2}} &= -c\int_0^1\int_0^r e^{c(r-s)}J_c(s)dV(s)dW(r) + \left(J_c(1) + \frac{b}{c}\right)N(V) + \frac{1}{2}M^2(V) - \frac{1}{2}U; \\
R_{2,T^{-1/2}} &= -c\int_0^1\int_0^r e^{r(c-s)}J_c(s)dV(s)dr - \int_0^1 J_c(r)dV(r) + J_c(1)V - \frac{\mu_3}{3}; \\
R_{1,T^{-1/2}} &= -2c\int_0^1 J_c(r)\int_0^r e^{c(r-s)}J_c(s)dV(s)dr - \int_0^1 J_c^2(r)dV(r) + J_c^2(1)V \\
&\quad + 2b\int_0^1 (e^{rc}-1)\int_0^r e^{c(r-s)}J_c(s)dV(s)dr - 2\frac{\mu_3}{3}\int_0^1 J_c(r)dr.
\end{aligned}$$

Proof: By backward substitutions, we can write x_t as

$$\begin{aligned} x_t &= \sum_{j=1}^{t} e^{(t-j)c/T} z_j + \frac{b}{\sqrt{T}} \frac{e^{ct/T}-1}{e^{c/T}-1} + e^{ct/T} x_0 + o_p(T^{-1/2}) \\ &= \tilde{x}_t + \frac{b}{\sqrt{T}} \frac{e^{ct/T}-1}{e^{c/T}-1} + e^{ct/T} x_0 + o_p(T^{-1/2}). \end{aligned} \tag{72}$$

This expression allows us to evaluate the asymptotic behavior of $\frac{1}{T}\sum_{t=1}^{T} x_t z_{t+1}$, $\frac{1}{T^2}\sum_{t=1}^{T} x_t^2$ and $\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t$.

We now show the first claim in Lemma 11.

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} x_t z_{t+1} &= \frac{1}{T}\sum_{t=1}^{T} z_{t+1} \sum_{j=1}^{t} e^{c\left(\frac{t-j}{T}\right)} z_j + \frac{1}{T}\sum_{t=1}^{T} \frac{b}{\sqrt{T}} \frac{e^{tc/T}-1}{e^{c/T}-1} z_{t+1} + \frac{x_0}{T}\sum_{t=1}^{T} e^{tc/T} z_{t+1} \\ &= \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_t z_{t+1} + \frac{1}{T}\sum_{t=1}^{T} \frac{b}{\sqrt{T}} \frac{e^{tc/T}-1}{e^{c/T}-1} z_{t+1} + \frac{x_0}{T}\sum_{t=1}^{T} e^{tc/T} z_{t+1}. \end{aligned}$$

The approximation to the first term is given in Lemma 10(1). For the second term, we have

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \frac{b}{\sqrt{T}} \frac{e^{tc/T}-1}{e^{c/T}-1} z_{t+1} &= \frac{b}{T(e^{c/T}-1)} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (e^{tc/T}-1) z_{t+1} \\ &= \frac{b}{c}\frac{1}{\sqrt{T}} \sum_{t=1}^{T} e^{ct/T} z_{t+1} - \frac{b}{c}\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{t+1} + o(T^{-1}) \\ &= \frac{b}{c}\int_0^1 e^{rc} dW(r) + \frac{b}{c}\left(W(1) + \frac{1}{T^{1/4}} M(V) + \frac{1}{\sqrt{T}} N(V)\right) + o_p(T^{-1/2}), \end{aligned}$$

where the last equality is due to Lemma 9. For the third term, we have

$$\frac{x_0}{T}\sum_{t=1}^{T} e^{tc/T} z_{t+1} = \frac{x_0}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e^{tc/T} z_{t+1} = \frac{y_0}{\sigma_\psi \sqrt{N}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e^{tc/T} z_{t+1} = \gamma_0 \int_0^1 e^{rc} dW(r) + o_p(T^{-1/2}).$$

To show the second claim of Lemma 11, note that

$$\begin{aligned} \frac{1}{T^2}\sum_{t=1}^{T} x_t^2 &= \frac{1}{T^2}\sum_{t=1}^{T} \tilde{x}_t^2 + \frac{1}{T^2}\sum_{t=1}^{T} \frac{b^2}{T}\frac{(e^{tc/T}-1)^2}{(e^{c/T}-1)^2} + \frac{1}{T^2}\sum_{t=1}^{T} \frac{2b}{\sqrt{T}}\frac{e^{tc/T}-1}{e^{c/T}-1} \sum_{t=1}^{T}\sum_{j=0}^{t} e^{(t-j)c/T} z_j \\ &= +\frac{1}{T^2}\sum_{t=1}^{T} \frac{2b}{\sqrt{T}}\frac{e^{tc/T}-1}{e^{c/T}-1} e^{tc/T} x_0 + \frac{1}{T^2}\sum_{t=1}^{T} e^{tc/T} x_0 \sum_{j=0}^{t} e^{(t-j)c/T} z_j + \frac{1}{T^2}\sum_{t=1}^{T} e^{2tc/T} x_0^2. \end{aligned}$$

The first term is approximated by using Lemma 10. For the second term, as in Zhou and Yu (2015), we can write

$$\frac{1}{T^2}\sum_{t=1}^{T} \frac{b^2}{T}\frac{(e^{tc/T}-1)^2}{(e^{c/T}-1)^2} = \frac{e^{2c} - 4e^{c} + 2c + 3}{2c^3} b^2 + O(T^{-1}).$$

For the third term, we have

$$\begin{aligned}
&= \frac{1}{T^2}\sum_{t=1}^{T}\frac{2b}{\sqrt{T}}\frac{e^{tc/T}-1}{e^{c/T}-1}\sum_{t=1}^{T}\sum_{j=0}^{t}e^{(t-j)c/T}z_j \\
&= \frac{2b}{T(e^{c/T}-1)}\frac{1}{T}\sum_{t=1}^{T}(e^{ct/T}-1)\frac{1}{\sqrt{T}}\sum_{j=1}^{t}e^{(t-j)c/T}z_j \\
&= \frac{2b}{c}\frac{1}{T}\sum_{t=1}^{T}(e^{ct/T}-1)\frac{1}{\sqrt{T}}\sum_{j=1}^{t}e^{(t-j)c/T}z_j + O_p(T^{-1}) \\
&= \frac{2b}{c}\int_0^1 (e^{cr}-1)J_c(r)dr + \frac{2b}{c}\frac{1}{T}\sum_{t=1}^{T}(e^{ct/T}-1)\frac{c}{\sqrt{T}}\int_0^{t/T} e^{c(t/T-s)}J_c(s)dV(s) + o_p(T^{-1/2}) \\
&= \frac{2b}{c}\int_0^1 (e^{cr}-1)J_c(r)dr + \frac{2b}{\sqrt{T}}\int_0^1 (e^{cr}-1)\int_0^r e^{c(r-s)}J_c(s)dV(s)dr + o_p(T^{-1/2}).
\end{aligned}$$

Finally, for the last three terms, we have:

$$\frac{1}{T^2}\sum_{t=1}^{T}\frac{2b}{\sqrt{T}}\frac{e^{tc/T}-1}{e^{c/T}-1}e^{tc/T}x_0 = \frac{e^{2c}-2e^c+1}{c^2}b\gamma_0 + O(T^{-1}),$$

$$\frac{1}{T^2}\sum_{t=1}^{T}e^{tc/T}x_0\sum_{j=0}^{t}e^{(t-j)c/T}z_j = 2\gamma_0\int_0^1 e^{rc}J_c(r)dr + O_p(T^{-1}),$$

$$\frac{1}{T^2}\sum_{t=1}^{T}e^{2tc/T}x_0^2 = \gamma_0^2\frac{e^{2c}-1}{2c} + O(T^{-1}).$$

For the last claim, we have

$$\begin{aligned}
\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_t &= \frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_t + \frac{T^{-2}b}{c^{t/T}-1}\left(\sum_{t=1}^{T}e^{tc/T}-T\right) + \frac{1}{T^{3/2}}\sum_{t=1}^{T}e^{ct/T}x_0 + O_p(T^{-1}) \\
&= \frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_t + \frac{b(e^{c(T+1)/T}-e^{c/T})}{T^2(e^{c/T}-1)^2} - \frac{b}{T(e^{c/T}-1)} + \frac{e^c-1}{c}\gamma_0 + O_p(T^{-1}) \\
&= \int_0^1 J_c(r)dr - \frac{c}{\sqrt{T}}\int_0^1\int_0^r e^{c(r-s)}J_c(s)dV(s)dr - \frac{1}{\sqrt{T}}\int_0^1 J_c(r)dV(r) \\
&\quad + \frac{1}{\sqrt{T}}J_c(1)V - \frac{\mu_3}{3\sqrt{T}} + \frac{e^c-c-1}{c^2}b + \frac{e^c-1}{c}\gamma_0 + o_p(T^{-1/2}).
\end{aligned}$$

By summing all three terms, we obtain the results in Lemma 11. ■

Now we are in the position to prove Theorem 2.

Proof: To show the probabilistic expansion, we rewrite $z(Y,\rho,T)$ as:

$$z(Y,\rho,T) = \frac{\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t - \frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t}{\frac{1}{T^2}\sum_{t=1}^{T}x_{t-1}^2 - \left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\right)^2}.$$

For the numerator and the denominator, after applying Lemma 11, we obtain

$$\begin{aligned}
&\Upsilon_3 - \Upsilon_2 W(1) + \frac{1}{T^{1/4}}(R_{3,T^{-1/4}} - M(V)\Upsilon_2)\\
&+\frac{1}{T^{1/2}}\left(R_{3,T^{-1/2}} - N(V)\Upsilon_2 - R_{2,T^{-1/2}}W(1)\right)\\
&-\frac{1}{T^{3/4}}R_{2,T^{-1/2}}M(V) - \frac{1}{T}R_{2,T^{-1/2}}N(V) + o_p(T^{-1/2}),
\end{aligned}$$

and

$$\Upsilon_1 - \Upsilon_2^2 + \frac{1}{T^{1/2}}(R_{1,T^{-1/2}} - 2R_{2,T^{-1/2}}) - \frac{1}{T}R_{2,T^{-1/2}} + o_p(T^{-1/2}).$$

Expanding $z\left(Y, \rho, T\right)$ around the in-fill limit by the Taylor series expansion, we obtain

$$\begin{aligned}
z\left(Y, \rho, T\right) =& \frac{\Upsilon_3 - \Upsilon_2 W(1)}{\Upsilon_1 - \Upsilon_2^2} + \frac{1}{T^{1/4}}\frac{R_{3,T^{-1/4}} - M(V)\Upsilon_2}{\Upsilon_1 - \Upsilon_2^2}\\
&+\frac{1}{T^{1/2}}\left(\begin{array}{c}\frac{R_{3,T^{-1/2}} - N(V)\Upsilon_2 - R_{2,T^{-1/2}}W(1)}{\Upsilon_1 - \Upsilon_2^2}\\ -\frac{\Upsilon_3 - \Upsilon_2 W(1)}{(\Upsilon_1 - \Upsilon_2^2)^2}\left(R_{1,T^{-1/2}} - 2R_{2,T^{-1/2}}\right)\end{array}\right) + o_p\left(T^{-1/2}\right)\\
=& z^{y_0}\left(\rho, \theta\right) + T^{-1/4}\tilde{A} + T^{-1/2}\tilde{B} + o_p(T^{-1/2}),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{A} &= \frac{R_{3,T^{-1/4}} - M(V)\Upsilon_2}{\Upsilon_1 - \Upsilon_2^2},\\
\tilde{B} &= \frac{R_{3,T^{-1/2}} - N(V)\Upsilon_2 - R_{2,T^{-1/2}}W(1)}{\Upsilon_1 - \Upsilon_2^2} - \frac{\Upsilon_3 - \Upsilon_2 W(1)}{(\Upsilon_1 - \Upsilon_2^2)^2}\left(R_{1,T^{-1/2}} - 2R_{2,T^{-1/2}}\right).
\end{aligned}$$

The expansion of $z\left(Y, \kappa, h\right)$ can be obtained from (68) and the Taylor series expansion of $h\varsigma_h'(\rho_h(\kappa)) = -\exp(\kappa h)$.

Finally, for the last claim in Theorem 2, following Theorem 3 in Mikusheva (2015), we can easily show that the difference between the distribution of the coefficient-based statistic and the bootstrap statistic is of order $o(T^{-1/2})$. ■

Before proving Proposition 1, we need the following lemma.

Lemma 12 (In-fill distribution under Model (15)) *Under Model (16), as $h \to 0$, we have*

$$z(Y, \kappa, h) \Rightarrow \bar{z}^{y_0}(\kappa, \theta).$$

Proof: From the discrete-time model (16), letting $x_t = y_t/\lambda_h$, we have

$$x_t = \rho_h(\kappa)x_{t-1} + \breve{g}_h + \sigma_t u_t, \tag{73}$$

where $\breve{g}_h = g_h/\lambda_h$. Letting $h \to 0$ and applying Donsker's theorem, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor sT \rfloor} \sigma_t u_t \Rightarrow \psi''(0)\int_0^s \omega(s)dW(s) = \bar{\sigma}_\psi W_\omega(s), \tag{74}$$

where $W_\omega(s) = \bar{\sigma}_\psi^{-1}\psi''(0)\int_0^s \omega(s)dW(s)$, and $\bar{\sigma}_\psi^2 = \psi''(0)^2\int_0^1 \omega(r)^2 dr$. Applying the continuous mapping theorem and following the proof of Lemma 1, we have

$$\begin{aligned}
\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t &\Rightarrow \frac{e^c - c - 1}{c^2}\bar{b} + \bar{\sigma}_\psi \int_0^1 J_c^\omega(r)dr + \frac{e^c - 1}{c}\bar{\gamma} := \bar{\Upsilon}_2; \\
\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 &\Rightarrow \frac{e^{2c} + 4e^c + 2c + 3}{2c^3}\bar{b}^2 + \bar{\sigma}_\psi \int_0^1 \left(J_c^\omega(r)\right)^2 dr + \frac{2\bar{b}\bar{\sigma}_\psi}{c}\int_0^1 e^{rc-1}J_c^\omega(r)dr \\
&\quad + \frac{e^{2c} - 2e^c + 1}{c^2}\bar{b}\bar{\gamma}_0 + 2\bar{\gamma}_0\bar{\sigma}_\psi \int_0^1 e^{rc}J_c^\omega(r)dr + \bar{\gamma}_0^2\frac{e^{2c} - 1}{2c} := \bar{\Upsilon}_1; \qquad (75) \\
\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\sigma_t u_t &\Rightarrow \frac{2\bar{b}\bar{\sigma}_\psi}{c}\int_0^1 (e^{rc} - 1)dW_\omega(r) + \bar{\sigma}_\psi^2 \int_0^1 J_c^\omega(r)dW_\omega(r) + \bar{\gamma}\bar{\sigma}_\psi \int_0^1 dW_\omega(r) = \bar{\Upsilon}_3,
\end{aligned}$$

where $J_c^\omega(r) = \int_0^r e^{c(r-s)}dW_\omega(s)$, $\bar{b} = \mu\sqrt{c\kappa}$ and $\bar{\gamma}_0 = \frac{y_0}{\sqrt{N}}$. Eventually, we can obtain the in-fill distribution $\bar{z}^{y_0}(\kappa, \theta) = -\frac{\bar{\Upsilon}_3 - \bar{\Upsilon}_2\psi''(0)\int_0^s \omega(s)dW(s)}{\bar{\Upsilon}_1 - \bar{\Upsilon}_2^2}$. ■

We are in the position to prove Proposition 1

Proof: We only scratch the proof for the sake of brevity. Let e_t be the LS residual. Since $\frac{1}{Th}\sum_{t=1}^{\lfloor sT \rfloor} e_t^2 \Rightarrow_p \psi''(0)^2\int_0^s \omega(r)^2 dr$, the scaled sum of squared residuals is a consistent estimator of the integrated variance $\psi''(0)^2\int_0^s \omega(r)^2 dr$. Since the partial sum $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor sT \rfloor}\frac{e_t}{\sqrt{h}}z_t^*$ is Gaussian with the covariance kernel $M_T(s \wedge t)$ and $M_T(s) = \frac{1}{Th}\sum_{t=1}^{\lfloor sT \rfloor} e_t^2 \Rightarrow_p \psi''(0)^2\int_0^s \omega(r)^2 dr = \bar{\sigma}_\psi^2$ and $\psi''(0)\int_0^s \omega(r)^2 dr$ is the kernel function of the transformed Brownian motion $\bar{\sigma}_\psi W_\omega(s) = \bar{\sigma}_\psi W(\omega(s))$, it implies the weak convergence in probability, that is,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor sT \rfloor}\frac{e_t}{\sqrt{h}}z_t^* \overset{p}{\Rightarrow} \bar{\sigma}_\psi W_\omega(s). \tag{76}$$

Since $\breve{g}_h$ can be consistently estimated (conditional on κ) as $\hat{\breve{g}}_h - \breve{g}_h = \frac{1}{T}\sum_{t=1}^{\lfloor sT \rfloor}\sigma_t u_t = o_p(1)$ from (74), analogous to (73), the bootstrap data generating process can be written as

$$x_t^* = \rho_h(\kappa) x_{t-1}^* + \breve{g}_h + e_t h^{-1/2} z_t^* + o_p(1). \tag{77}$$

Applying the continuous mapping theorem, we obtain the analogous results to those in (75). Eventually we have

$$z(Y^*, \kappa, h) \overset{p}{\Rightarrow} \bar{z}^{y_0}(\kappa, \theta). \tag{78}$$

The convergence in (78) implies that $\Pr^*(z(Y^*, \kappa, h) < x|\kappa, Y^*) \to \Pr(\bar{z}^{y_0}(\kappa, \theta) < x|\kappa)$ uniformly in probability and that $\Pr^*(z(Y^*, \kappa, h) < x|\kappa, Y) \to_d U[0, 1]$ (since $\Pr(\bar{z}^{y_0}(\kappa, \theta) < x|\kappa)$ is a cumulative distribution function). From the definition of BCI, we have

$$\begin{aligned} CI_q^B &= \{\kappa \in R : c_T^*(x_1|\kappa) \leqslant z(Y^*, \kappa, h) \leqslant c_T^*(x_2|\kappa)\} \\ &= \{\kappa \in R : x_1 \leqslant \Pr{}^*(z(Y^*, \kappa, h) < x|\kappa, Y) \leqslant x_2\}. \end{aligned}$$

As $\Pr^*(z(Y^*, \kappa, h) < x|\kappa, Y) \to_d U[0, 1]$, we obtain the desired result. ■

B Appendix to Chapter 2

Lemma 13 (Lemma 2.3 in El Machkouri, Es-Sebaiy and Ouknine (2016)) *Suppose we have the following stochastic differential equation:*

$$dX(t) = cX(t)dt + dG(t), X(0) = X_0 = 0,$$

where $G(t)$ is a Gaussian process and $c > 0$. Further assume the following two assumptions hold for $G = (G(t), t \geqslant 0)$.

1. *The process G has Hölder continuous paths of order $\delta \in (0, 1]$;*

2. *For every $t \geqslant 0$, $E(G^2(t)) \leqslant ct^{2\gamma}$ for some positive constants c and γ.*

Then, for every $t \geqslant 0$, we have

$$\frac{1}{2}X^2(t) = c\int_0^t X^2(s)ds + cZ(t)\int_0^t e^{cs}dG(s) + R(t),$$

where

$$Z(t) = \int_0^t e^{-cs}G(s)ds,$$

$$R(t) = \frac{1}{2}G^2(t) - c\int_0^t G^2(s)ds + c^2\int_0^t\int_0^s e^{-c(s-r)}G(s)G(r)drds.$$

Proof of Lemma 4 Throughout the proof, we assume $n \to \infty$ with m fixed. By backward substitutions, we can write

$$y_{\lfloor nr\rfloor} = \frac{1-\rho_{n,m}^{\lfloor nr\rfloor}}{1-\rho_{n,m}}\mu_n + \rho_{n,m}^{\lfloor nr\rfloor}y_0 + \sum_{j=1}^{\lfloor nr\rfloor}\rho_{n,m}^{\lfloor nr\rfloor-j}u_j.$$

Note that $\rho_{n,m} = \exp(\frac{cm}{n}) + R_\rho$, with $R_\rho = -\sum_{k=2}^{\infty}\left(\frac{cm}{n}\right)^k/k! = O(n^{-2})$. Applying the binomial expansion, we have

$$\begin{aligned}
\rho_{n,m}^{\lfloor nr\rfloor} &= \left(\exp\left(\frac{cm}{n}\right) + R_\rho\right)^{\lfloor nr\rfloor} \\
&= \sum_{k=0}^{\lfloor nr\rfloor}\binom{\lfloor nr\rfloor}{k}\exp\left(\frac{cm}{n}\right)^{\lfloor nr\rfloor-k}R_\rho^k \\
&= \exp\left(\frac{cm}{n}\right)^{\lfloor nr\rfloor} + \sum_{k=1}^{\lfloor nr\rfloor}\binom{\lfloor nr\rfloor}{k}\exp\left(\frac{cm}{n}\right)^{\lfloor nr\rfloor-k}R_\rho^k.
\end{aligned}$$

We will show for any $k \geqslant 1$,

$$\binom{\lfloor nr \rfloor}{k} \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - k} R_\rho^k \to 0. \tag{79}$$

To do so, note that $\binom{\lfloor nr \rfloor}{k} = O(n^k)$, $\exp(\frac{cm}{n})^{\lfloor nr \rfloor - k} = O(1)$, and $R_\rho^k = \exp(k \ln R_\rho) = \exp\left(k \ln O(n^{-2})\right) = \exp\left(-2k \ln(O(n))\right)$. Hence,

$$\binom{nr}{k} \exp\left(\frac{cm}{n}\right)^{nr-k} R_\rho^k = O[n^k \exp(-2k \ln(O(n)))].$$

Moreover,

$$\ln[n^k \exp(-2k \ln(n))] = k \ln(n) - 2k \ln(n) = -k \ln(n) \to -\infty.$$

This proves (79).

Letting $k^* = \arg\max_{k \in \{2,\ldots,n\}} \binom{nr}{k} \exp\left(\frac{cm}{n}\right)^{nr-k} R_\rho^k$, we have

$$\sum_{k=2}^{\lfloor nr \rfloor} \binom{\lfloor nr \rfloor}{k} \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - k} R_\rho^k = O[n^{1+k^*} \exp(-2k^* \ln(O(n)))] \to 0, \tag{80}$$

because

$$\begin{aligned} \ln[n^{1+k^*} \exp(-2k^* \ln(n))] &= (k^* + 1 - 2k^*) \ln(n) \\ &= (1 - k^*) \ln n \to -\infty \text{ since } k^* \geqslant 2. \end{aligned}$$

From (79) and (80), we have

$$\sum_{k=1}^{\lfloor nr \rfloor} \binom{\lfloor nr \rfloor}{k} \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - k} R_\rho^k \to 0.$$

So $\rho_{n,m}^{\lfloor nr \rfloor} = \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor} + o(1)$. Since $\lfloor nr \rfloor / nr \to 1$, we can write

$$y_{\lfloor nr \rfloor} = \frac{1 - \exp(cmr)}{-cm/n} \mu_n + (\exp(cmr) + o(1)) y_0 + \sum_{j=1}^{\lfloor nr \rfloor} \rho_{n,m}^{\lfloor nr \rfloor - j} u_j + o(1). \tag{81}$$

For the third term in (81), we can show that $\rho_{n,m}^{\lfloor nr \rfloor - j} = \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - j} + o(1)$ which allows us to express

$$\begin{aligned} \sum_{j=1}^{\lfloor nr \rfloor} \rho_{n,m}^{\lfloor nr \rfloor - j} u_j &= \sum_{j=1}^{\lfloor nr \rfloor} \left(\exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - j} + o(1)\right) u_j \\ &= \sum_{j=1}^{\lfloor nr \rfloor} \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - j} u_j + o(1) \sum_{j=1}^{\lfloor nr \rfloor} u_j \\ &= \sum_{j=1}^{\lfloor nr \rfloor} \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - j} u_j + o_p(n^{1/2+d}). \end{aligned}$$

We obtain the third equality by using (94) where $\sum_{j=1}^{\lfloor nr \rfloor} u_j = O_p(n^{1/2+d})$.

Eventually, we can rewrite (81) as

$$y_{\lfloor nr \rfloor} = nr\mu_n + \exp(cmr) y_0 + \sum_{j=1}^{\lfloor nr \rfloor} \exp\left(\frac{cm}{n}\right)^{\lfloor nr \rfloor - j} u_j + o_p(n^{1/2+d}). \tag{82}$$

Let $X_n(r) = \frac{1}{n^{1/2+d}\varsigma} S_{\lfloor nr \rfloor}$ with $S_{\lfloor nr \rfloor} = \sum_{j=1}^{\lfloor nr \rfloor} u_j$. Recall that under Model (25), $y_0 = o_p(n^{1/2+d})$, $\mu_n = \mu/n^{\vartheta}$ with $\vartheta > 1/2 - d$. The first two terms in (82) vanish as $n \to \infty$. If we multiply both sides in (82) by $n^{-1/2-d}$, we have

$$\begin{aligned} n^{-1/2-d} y_{\lfloor nr \rfloor} &= \varsigma \sum_{j=1}^{\lfloor nr \rfloor} e^{(\lfloor nr \rfloor - j)cm/n} \int_{(j-1)/n}^{j/n} dX_n(s) + o_p(1) \\ &= \varsigma \sum_{j=1}^{\lfloor nr \rfloor} \int_{(j-1)/n}^{j/n} e^{(r-s)cm} dX_n(s) + o_p(1) \\ &= \varsigma \int_0^r e^{(r-s)cm} dX_n(s) + o_p(1) \\ &\Rightarrow \varsigma \int_0^r e^{(r-s)cm} dB^H(s) := \varsigma J_{cm}^H(r). \end{aligned}$$

We have applied Lemma 16 with the continuous mapping theorem (Billingsley, 1968, p. 30) to obtain the last result.

For the terms involving $\sum_{t=1}^{n} y_t$ and $\sum_{t=1}^{n} y_{t-1}^2$, note that we can write

$$n^{-3/2-d} \sum_{t=1}^{n} y_t = \frac{1}{n} \sum_{t=1}^{n} \left(n^{-1/2-d} y_t\right),$$

$$n^{-2-2d} \sum_{t=1}^{n} y_t^2 = \frac{1}{n} \sum_{t=1}^{n} \left(n^{-1/2-d} y_t\right)^2.$$

By applying the continuous mapping theorem, we obtain the second claim and the third claim in Lemma 4.

For the last claim, after squaring y_t and summing over t, we have

$$\begin{aligned} \sum_{t=1}^{n} y_t^2 &= \rho_{n,m}^2 \sum_{t=1}^{n} y_{t-1}^2 + 2\rho_{n,m} \sum_{t=1}^{n} y_{t-1} u_t + \sum_{t=1}^{n} u_t^2 \\ &\quad + n\mu_n^2 + 2\mu_n \rho_{n,m} \sum_{t=1}^{n} y_{t-1} + 2\mu_n \sum_{t=1}^{n} u_t, \end{aligned}$$

which leads to

$$
\begin{aligned}
y_n^2 &= \frac{2cm}{n}\sum_{t=1}^{n} y_{t-1}^2 + 2\rho_{n,m}\sum_{t=1}^{n} y_{t-1}u_t + \sum_{t=1}^{n} u_t^2 \\
&\quad + \frac{(cm)^2}{n^2}\sum_{t=1}^{n} y_{t-1}^2 + n\mu_n^2 + 2\mu_n\rho_{n,m}\sum_{t=1}^{n} y_{t-1} + 2\mu_n\sum_{t=1}^{n} u_t.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
2\rho_{n,m}\sum_{t=1}^{n} y_{t-1}u_t &= y_n^2 - \frac{2cm}{n}\sum_{t=1}^{n} y_{t-1}^2 - \sum_{t=1}^{n} u_t^2 - \frac{(cm)^2}{n^2}\sum_{t=1}^{n} y_{t-1}^2 \\
&\quad - n\mu_n^2 - 2\mu_n\rho_{n,m}\sum_{t=1}^{n} y_{t-1} - 2\mu_n\sum_{t=1}^{n} u_t,
\end{aligned}
$$

$$
\begin{aligned}
\frac{2}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1}u_t &= \frac{1}{n^{1+2d}}y_n^2 - \frac{2cm}{n^{2+2d}}\sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n^{2d}}\frac{1}{n}\sum_{t=1}^{n} u_t^2 \\
&\quad - \frac{n\mu_n^2}{n^{1+2d}} - 2\frac{\mu_n}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1} - 2\frac{\mu_n}{n^{1+2d}}\sum_{t=1}^{n} u_t + o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{2}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1}u_t + \frac{1}{n^{2d}}\frac{1}{n}\sum_{t=1}^{n} u_t^2 &= \frac{1}{n^{1+2d}}y_n^2 - \frac{2cm}{n^{2+2d}}\sum_{t=1}^{n} y_{t-1}^2 - \frac{n\mu_n^2}{n^{1+2d}} \\
&\quad - 2\frac{\mu_n}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1} - 2\frac{\mu_n}{n^{1+2d}}\sum_{t=1}^{n} u_t + o_p(1),
\end{aligned}
$$

as $\rho_{n,m} \to 1$, and $\frac{(cm)^2}{n^2}\sum_{t=1}^{n} y_{t-1}^2 = O_p(n^{2d})$ is dominated by $\frac{2cm}{n}\sum_{t=1}^{n} y_{t-1}^2 = O_p(n^{1+2d})$ when m is fixed.

Note that $\frac{1}{n}\sum_{t=1}^{n} u_t^2 \overset{as}{\to} E[u_t^2]$ by the ergodic theorem and

$$
\frac{n\mu_n^2}{n^{1+2d}} = \frac{n^{-2d}}{n^{2\theta}}\mu^2 < \frac{n^{1-2d}}{n^{2\theta}}\mu^2 = \left(\frac{n^{1/2-d}}{n^{\vartheta}}\right)^2 \mu^2 \to 0,
$$

$$
\frac{\mu_n}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1} = \mu\frac{1}{n^{\vartheta}}\frac{n^{3/2+d}}{n^{1+2d}}\left(\frac{1}{n^{3/2+d}}\sum_{t=1}^{n} y_{t-1}\right) = \mu\frac{n^{1/2-d}}{n^{\vartheta}}O_p(1) = o_p(1),
$$

$$
\begin{aligned}
\frac{\mu_n}{n^{1+2d}}\sum_{t=1}^{n} u_t &= \mu\frac{1}{n^{\vartheta}}\frac{n^{1/2+d}}{n^{1+2d}}\left(\frac{1}{n^{1/2+d}}\sum_{t=1}^{n} u_t\right) \\
&= \mu n^{-1/2-d-\vartheta}O_p(1) = o_p(1) \text{ since } \vartheta > 0, \text{ and } d < 1/2.
\end{aligned}
$$

These results lead to

$$\frac{2}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1}u_t + \frac{1}{n^{2d}}\frac{1}{n}\sum_{t=1}^{n} u_t^2 \Rightarrow \varsigma^2\left[\left(J_{cm}^H(1)\right)^2 - 2cm\int_0^1 \left(J_{cm}^H(r)\right)^2 dr\right].$$

So we have

$$\begin{aligned}\frac{1}{n^{1+2d}}\sum_{t=1}^{n} y_{t-1}u_t + \frac{1}{n^{2d}}\frac{1}{2n}\sum_{t=1}^{n} u_t^2 \quad &\Rightarrow \quad \varsigma^2\left[\frac{1}{2}\left(J_{cm}^H(1)\right)^2 - cm\int_0^1 (J_{cm}^H(r))^2 dr\right] \\ &= \quad \varsigma^2\left[cmZ(1)\int_0^1 e^{cms}dB^H(s) + R(1)\right],\end{aligned}$$

where the last step follows from Lemma 13. This completes the proof of Lemma 4.

To analyze the asymptotics when $m \to \infty$, we introduce the following lemma, which documents some results of distributional equivalence. By the self-similarity property of fBM, we have $B^H\left(\frac{t}{m}\right) \overset{d}{=} \left(\frac{1}{m}\right)^H B^H(t)$. Let $\widetilde{B}^H(t) := m^H B^H\left(\frac{t}{m}\right)$.

Lemma 14 *Applying the self-similarity property of fBM, we can obtain the following:*

1. $\int_0^1 J_{cm}^H(r)drB^H(1) \overset{d}{=} \frac{1}{m^{2H+1}}\int_0^m \widetilde{J}_c^H(s)ds\widetilde{B}^H(m)$;
2. $\left(\int_0^1 J_{cm}^H(r)dr\right)^2 \overset{d}{=} \frac{1}{m^{2H+2}}\left(\int_0^m \widetilde{J}_c^H(s)ds\right)^2$;
3. $\int_0^1 (J_{cm}^H(r))^2 dr \overset{d}{=} \frac{1}{m^{2H+1}}\int_0^m \left(\widetilde{J}_c^H(s)\right)^2 ds$;
4. $cmZ(1)\int_0^1 e^{cms}dB^H(s) + R(1) \overset{d}{=} \frac{1}{m^{2H}}\left(c\tilde{Z}(m)\int_0^m e^{cs}d\widetilde{B}^H(s) + \widetilde{R}(m)\right)$,

where

$$\begin{aligned}\widetilde{J}_c^H(r) &= \int_0^r e^{c(r-s)}d\widetilde{B}^H(s), \\ \tilde{Z}(m) &= \int_0^m e^{-cs}\widetilde{B}^H(s)ds, \\ \widetilde{R}(m) &= \frac{1}{2}\left(\widetilde{B}^H(m)\right)^2 - c\int_0^m \left(\widetilde{B}^H(s)\right)^2 ds + c^2\int_0^m\int_0^r e^{c(r-s)}\widetilde{B}^H(r)\widetilde{B}^H(s)drds.\end{aligned}$$

Proof of Lemma 14

Lemma 14 obtains the distributional representations of different functionals of fBM. We can prove the lemma by obtaining the distributional representations of all the components on the left-hand sides of the four equations in Lemma 14. That is, we only need to show the following results are correct:

1. $Z(1) \stackrel{d}{=} \frac{1}{m^{H+1}}\tilde{Z}(m)$;

2. $\int_0^1 e^{cms} dB^H(s) \stackrel{d}{=} \frac{1}{m^H}\int_0^m e^{cs} d\widetilde{B}^H(s)$;

3. $\int_0^1 \left(B^H(s)\right)^2 ds \stackrel{d}{=} \frac{1}{m^{2H+1}}\int_0^m \left(\widetilde{B}^H(s)\right)^2 ds$;

4. $\int_0^1 J^H_{cm}(s) ds \stackrel{d}{=} \frac{1}{m^{H+1}}\int_0^m \widetilde{J}^H_c(s) ds$;

5. $\int_0^1 (J^H_{cm}(s))^2 ds \stackrel{d}{=} \frac{1}{m^{2H+1}}\int_0^m \left(\widetilde{J}^H_c(s)\right)^2 ds$;

6. $m^2\int_0^1\int_0^s e^{cm(r-s)} B^H(r)B^H(s) dr ds \stackrel{d}{=} \frac{1}{m^{2H}}\int_0^m\int_0^s e^{c(r-s)}\widetilde{B}^H(r)\widetilde{B}^H(s) dr ds.$

As the steps to prove the above results are similar, we shall only prove the last two claims. For the fifth claim, we have

$$\begin{aligned}
\int_0^1 \left(J^H_{cm}(r)\right)^2 dr &= \int_0^1 \left(\int_0^r e^{cm(r-s)} dB^H(s)\right)^2 dr \\
&= \int_0^1 e^{2cmr}\left(\int_0^r e^{-cms} dB^H(s)\right)^2 dr \\
&= \int_0^1 e^{2cmr}\left(\int_0^{mr} e^{-cv} dB^H\left(\frac{v}{m}\right)\right)^2 dr \\
&= \frac{1}{m^{2H}}\int_0^1 e^{2cmr}\left(\int_0^{mr} e^{-cv} d\left(m^H B^H\left(\frac{v}{m}\right)\right)\right)^2 dr \\
&= \frac{1}{m^{2H}}\int_0^m e^{2cu}\left(\int_0^u e^{-cv} d\widetilde{B}^H(v)\right)^2 d\left(\frac{u}{m}\right) \\
&= \frac{1}{m^{2H+1}}\int_0^m \left(\int_0^u e^{c(u-v)} d\widetilde{B}^H(v)\right)^2 du \\
&= \frac{1}{m^{2H+1}}\int_0^m \left(\widetilde{J}^H_c(u)\right)^2 du.
\end{aligned}$$

For the sixth result, we have

$$\begin{aligned}
m^2\int_0^1\int_0^s e^{cm(r-s)} B^H(r)B^H(s) dr ds &= m^2\int_0^1 e^{-cms}\left(\int_0^s e^{cmr} B^H(r) dr\right) B^H(s) ds \\
&= m^2\int_0^1 e^{-cms}\left(\int_0^{ms} e^{cr} B^H\left(\frac{r}{m}\right) d\left(\frac{r}{m}\right)\right) B^H(s) ds \\
&= \frac{m}{m^H}\int_0^m e^{-cv}\left(\int_0^{ms} e^{cr}\widetilde{B}^H(r) dr\right) B^H\left(\frac{v}{m}\right) d\left(\frac{v}{m}\right) \\
&= \frac{1}{m^{2H}}\int_0^m e^{-cv}\left(\int_0^v e^{cr}\widetilde{B}^H(r) dr\right)\widetilde{B}^H(v) dv \\
&= \frac{1}{m^{2H}}\int_0^m\int_0^v e^{c(r-v)}\widetilde{B}^H(r)\widetilde{B}^H(v) dr dv.
\end{aligned}$$

Proof of Theorem 3

To avoid confusion, we now refer $n \to \infty$ with m fixed as the "fix-m asymptotics", and $n \to \infty$ followed by $m \to \infty$ as the "sequential asymptotics".

From (26) and (27), we can have the following expressions for the normalized centered LS estimates

$$\begin{aligned} \frac{e^{cm}}{m} n(\hat{\rho}_a - \rho) &= \frac{e^{cm}}{m} n \left(\frac{\sum_{t=1}^{n} y_{t-1}u_t + \frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2} - \frac{\frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2} \right) \\ &= \frac{e^{cm}}{m} n \frac{\sum_{t=1}^{n} y_{t-1}u_t + \frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2} - \frac{e^{cm}}{m} B_n^a \\ &:= \frac{e^{cm}}{m} A_n^a - \frac{e^{cm}}{m} B_n^a, \end{aligned} \tag{83}$$

where $A_n^a = n\frac{\sum_{t=1}^{n} y_{t-1}u_t + \frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2}$ and $B_n^a = \frac{n}{2}\sum_{t=1}^{n} u_t^2 / \sum_{t=1}^{n} y_{t-1}^2$.

Similarly, we can express

$$\begin{aligned} \frac{e^{cm}}{m} n(\hat{\rho}_b - \rho) &= \frac{e^{cm}}{m} n \left(\begin{array}{c} \frac{\sum_{t=1}^{n} y_{t-1}u_t - \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\sum_{t=1}^{n} u_t + \frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^{n} y_{t-1}\right)^2} \\ - \frac{\frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^{n} y_{t-1}\right)^2} \end{array} \right) \\ &:= \frac{e^{cm}}{m} A_n^b - \frac{e^{cm}}{m} B_n^b, \end{aligned} \tag{84}$$

where

$$A_n^b = n\frac{\sum_{t=1}^{n} y_{t-1}u_t - \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\sum_{t=1}^{n} u_t + \frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^{n} y_{t-1}\right)^2},\ B_n^b = n\frac{\frac{1}{2}\sum_{t=1}^{n} u_t^2}{\sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^{n} y_{t-1}\right)^2}.$$

Since the proofs of the sequential asympotics for $\frac{e^{cm}}{m} n(\hat{\rho}_a - \rho)$ are very similar to those for $\frac{e^{cm}}{m} n(\hat{\rho}_b - \rho)$, we shall only prove the later. In fact, the only difference between the two estimates is the extra terms induced by the inclusion of an intercept in the LS regression. As we proceed, we will see the extra terms vanish in the sequential asympotics.

We first show the sequential limit of $\frac{e^{cm}}{m} A_n^b$ in (84). Applying Lemma 4 and Lemma

14, as $n \to \infty$ with fixed m,

$$\begin{aligned}
\frac{e^{cm}}{m} A_n^b \quad &\Rightarrow \quad \frac{e^{cm}}{m} \frac{cmZ(1)\int_0^1 e^{cms} dB^H(s) + R(1) - \int_0^1 J_{cm}^H(r) dr B^H(1)}{\int_0^1 (J_{cm}^H(r))^2 dr - \left(\int_0^1 J_{cm}^H(r) dr\right)^2} \\
&\stackrel{d}{=} \quad \frac{e^{cm}}{m} \frac{\frac{1}{m^{2H}}\left(c\tilde{Z}(m)\int_0^m e^{-cr}\widetilde{B}^H(r)dr + \widetilde{R}(m) - \frac{1}{m}\int_0^m J_c^H(s) ds \tilde{B}^H(m)\right)}{\frac{1}{m^{2H+1}}\left(\int_0^m \tilde{J}_c^H(s)^2 ds - \frac{1}{m}\left(\int_0^m \tilde{J}_c^H(s) ds\right)^2\right)} \\
&= \quad e^{cm} \frac{c\tilde{Z}(m)\int_0^m e^{-cr}\widetilde{B}^H(r)dr + \widetilde{R}(m) - \frac{1}{m}\int_0^m J_c^H(s) ds \tilde{B}^H(m)}{\int_0^m \tilde{J}_c^H(s)^2 ds - \frac{1}{m}\left(\int_0^m \tilde{J}_c^H(s) ds\right)^2}. \qquad (85)
\end{aligned}$$

For the sake of notational simplicity, we now introduce the following process with $m \geqslant 0$,

$$\xi(m) = \int_0^m e^{-cr} d\widetilde{B}^H(r), \qquad (86)$$

where the integral is interpreted in the Young sense.

From Lemma 2.1 of El Machkouri, Es-Sebaiy and Ouknine (2016), we obtain a well-defined limit $\tilde{Z}(\infty) = \int_0^\infty e^{-cr}\widetilde{B}^H(r)dr$. As $m \to \infty$, we have

$$\tilde{Z}(m) \stackrel{as}{\to} \tilde{Z}(\infty) \text{ and } \xi(m) \stackrel{as}{\to} \xi(\infty) = c\tilde{Z}(\infty). \qquad (87)$$

These two results are similar to those obtained by the martingale convergence theorem used in PMG when $m \to \infty$.

By the definition of the Young integral, we obtain $\widetilde{B}^H(0) = 0$. By the definition of $\tilde{Z}(m)$, we have

$$\xi(m) = e^{-cm}\widetilde{B}^H(m) + c\int_0^m e^{-cr}\widetilde{B}^H(r)dr = e^{-cm}\widetilde{B}^H(m) + c\tilde{Z}(m),$$

$$\tilde{J}_c^H(r) = \int_0^r e^{c(r-s)} d\widetilde{B}^H(s) = e^{cr}\int_0^r e^{-cs} d\widetilde{B}^H(s) = e^{cr}\xi(r).$$

So we can express (85) as

$$\begin{aligned}
&e^{cm} \frac{\left(c\tilde{Z}(m)\int_0^m e^{cs} d\tilde{B}^H(s) + \tilde{R}(m)\right) - e^{cs}\xi(s) ds \frac{1}{m}\tilde{B}^H(m)}{\int_0^m e^{2cs}\xi^2(s) ds - \frac{1}{m}\left(e^{-cm}\int_0^m e^{cs}\xi(s) ds\right)^2} \\
= \quad & \frac{e^{-cm}\left[\left(c\tilde{Z}(m)\int_0^m e^{cs} d\tilde{B}^H(s) + \tilde{R}(m)\right) - e^{cs}\xi(s) ds \frac{1}{m}\tilde{B}^H(m)\right]}{e^{-2cm}\left[\int_0^m e^{2cs}\xi^2(s) ds - \frac{1}{m}\left(e^{-cm}\int_0^m e^{cs}\xi(s) ds\right)^2\right]} \\
= \quad & \frac{e^{-cm}\left(c\tilde{Z}(m)\int_0^m e^{cs} d\tilde{B}^H(s) + \tilde{R}(m)\right) - \varphi_1'}{e^{-2cm}\left[\int_0^m e^{2cs}\xi^2(s) ds\right] - \varphi_2'}, \qquad (88)
\end{aligned}$$

where $\varphi_1' = \left(e^{-cm}\int_0^m e^{cs}\xi(s)ds\right)\left(\frac{1}{m}\tilde{B}^H(m)\right) := \varphi_{1a}' \times \varphi_{1b}'$, $\varphi_2' = \frac{1}{m}\left(e^{-cm}\int_0^m e^{cs}\xi(s)ds\right)^2$.

For the term $\int_0^m e^{cs}\xi(s)ds$, Propostion 3.1 in El Machkouri, Es-Sebaiy and Ouknine (2016) shows it is a Gaussian process with a diverging variance. Therefore, applying (87) and L'Hospital's rule, we have, as $m \to \infty$,

$$\begin{aligned}\lim_{m\to\infty} \varphi_{1a}' &= \lim_{m\to\infty} e^{-cm}\int_0^m e^{cs}\xi(s)ds \\ &= \lim_{m\to\infty} \frac{e^{cm}\xi(m)}{ce^{cm}} = \frac{\xi(\infty)}{c} = \tilde{Z}(\infty)\,. \end{aligned} \tag{89}$$

Since $E\left(\frac{1}{m}\tilde{B}^H(m)\right) = 0$, $Var\left(\frac{1}{m}\tilde{B}^H(m)\right) = \frac{m^{2H}}{m^2} \to 0$, as $m \to \infty$,

$$\varphi_{1b}' = \frac{1}{m}\tilde{B}^H(m) \xrightarrow{p} 0.$$

Moreover, the continuous mapping theorem and Equation (89) imply that, as $m \to \infty$,

$$\left(e^{-cm}\int_0^m e^{cs}\xi(s)ds\right)^2 \xrightarrow{as} \tilde{Z}^2(\infty).$$

As $\frac{1}{m} \to 0$, $\varphi_2' \xrightarrow{p} 0$. Note that φ_1' and φ_2' are the extra terms due to the inclusion of the intercept in the LS regression. As they vanish, $\frac{e^{cm}}{m}A_n^a$ and $\frac{e^{cm}}{m}A_n^b$ are asymptotically equivalent in the sequential asymptotics. Therefore, as $m \to \infty$, we can write (88) as,

$$\frac{e^{-cm}\left(c\tilde{Z}(m)\int_0^m e^{cs}d\tilde{B}^H(s) + \tilde{R}(m)\right)}{e^{-2cm}\int_0^m e^{2cs}\xi^2(s)ds} + o_p(1) \tag{90}$$

To derive the sequential limit of $\frac{e^{cm}}{m}A_n^b$, we need the following lemma.

Lemma 15 *let ω and η be two independent standard normal random variables. Then, as $m \to \infty$, we obtain:*

1. $e^{-2cm}\int_0^m e^{2cs}\xi^2(s)ds \xrightarrow{as} \frac{c}{2}\tilde{Z}^2(\infty)$;
2. $c\tilde{Z}(m)\left(e^{-cm}\int_0^m e^{cs}d\widetilde{B}^H(s)\right) \Rightarrow c\tilde{Z}(\infty)\sqrt{\frac{H\Gamma(2H)}{c^{2H}}}\eta$;
3. $\xi(m) \xrightarrow{as} \xi(\infty) = \sqrt{\frac{H\Gamma(2H)}{c^{2H}}}\omega$;
4. $e^{-cm}\widetilde{R}(m) \xrightarrow{p} 0$.

The first result is immediate after applying (87) and L'Hospital's rule. The last three results can be obtained by applying Lemma 2.1, Lemma 2.2 and Lemma 2.4 of El Machkouri, Es-Sebaiy and Ouknine (2016). Hence, as $m \to \infty$ and using $\xi(\infty) = c\tilde{Z}(\infty)$, we have

$$\begin{aligned}\frac{e^{-cm}\left(c\tilde{Z}(m)\int_0^m e^{cs}d\tilde{B}^H(s) + \tilde{R}(m)\right)}{e^{-2cm}\int_0^m e^{2cs}\xi^2(s)ds} + o_p(1) &\Rightarrow \frac{c\tilde{Z}(\infty)\sqrt{\frac{H\Gamma(2H)}{c^{2H}}}\eta}{\frac{c}{2}\tilde{Z}^2(\infty)} \\ &= 2c \times \frac{\eta}{\omega} = 2c \times C, \qquad (91)\end{aligned}$$

where C is the standard Cauchy variate.

We now analyze the sequential limit of $\frac{e^{cm}}{m}B_n^b$ in (84). A standard calculation shows

$$\begin{aligned}\frac{e^{cm}}{m}B_n^b &= \frac{e^{cm}}{m}n\frac{\frac{1}{2}\sum_{t=1}^n u_t^2}{\sum_{t=1}^n y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n y_{t-1}\right)^2} \\ &= \frac{e^{cm}}{m}\frac{n^{-1-2d}\frac{1}{2}\sum_{t=1}^n u_t^2}{n^{-2-2d}\left(\sum_{t=1}^n y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^n y_{t-1}\right)^2\right)} \\ &= \frac{e^{-cm}}{e^{-2cm}}\frac{m^{2H}}{m^{2H+1}}\frac{O_p(n^{-2d})}{O_p\left(\frac{e^{2cm}}{m^{2H+1}}\right)} \text{ as } n \to \infty \\ &= O_p\left(\frac{m^{2H}n^{-2d}}{e^{cm}}\right).\end{aligned}$$

The third equality is established by $\frac{1}{n}\sum_{t=1}^n u_t^2 = O_{as}(1)$, Lemma 4, Lemma 14 and Lemma 17. The assumption $m = \delta \ln n$, with $\delta > -\frac{2d}{c}$ implies that $\frac{m^{2H}n^{-2d}}{e^{cm}} \to 0$. To see this,

$$\begin{aligned}\ln\frac{m^{2H}n^{-2d}}{e^{cm}} &= 2H\ln m - 2d\ln n - cm \\ &= 2H(\ln\delta + \ln\ln n) - 2d\ln n - c\delta\ln n \\ &= -(c\delta + 2d)\ln n + 2H(\ln\delta + \ln\ln n) \\ &\to -\infty.\end{aligned}$$

Hence,

$$\frac{e^{cm}}{m}B_n^b = o_p(1). \qquad (92)$$

This suggests that when $n \to \infty$ followed by $m \to \infty$ and when $m = \delta\ln n$ with $\delta > -\frac{2d}{c}$, $\frac{1}{n^{1+2d}}\frac{1}{2}\sum_{t=1}^n u_t^2$ is dominated by $\frac{1}{n^{1+2d}}\sum_{t=1}^n y_{t-1}u_t$.

Equations (84), (91) and (92) imply that the sequential limit of $\frac{1}{2c}\frac{e^{cm}}{m}n(\hat{\rho}_b - \rho)$ is the standard Cauchy random variable C.

C Appendix to Chapter 3

C.1 Sup statistic

In Remark 29 and 30, we mentioned that a version of sup statistics can be employed. We now construct the sup statistic $\widetilde{SDF}_{HAR}(\tau_0)$ which fits our testing purpose. Let

$$\widetilde{SDF}_{HAR}(\tau_0) = \sup_{\tau\in[\tau_0,n]} \frac{\tilde{\rho}_\tau - 1}{s_{\tau,HAR}}$$

where $\tilde{\rho}_\tau$ is defined in (53), $s_{\tau,HAR} = \sqrt{\frac{\tilde{\Omega}_{HAR}}{\sum_{t=1}^{\tau}\tilde{y}_{t-1}}}, \tilde{\Omega}_{HAR} = \sum_{j=-n+1}^{n} K_B\left(\frac{j}{M}\right)\hat{\gamma}_j$, with $\hat{\gamma}_j$ being the jth order sample covariance and $\tau_0 = \lfloor nr_0 \rfloor$ is the minimum data window.

We proceed to discuss the limit of $\widetilde{SDF}_{HAR}(\tau_0)$ under the null or alternative hypotheses in the following theorem.

Theorem 9 *Suppose $M = \lfloor b\tau \rfloor$ and $K_B(x)$ is the Bartlett kernel function.*
Under model (48), as $n \to \infty$,

$$\widetilde{SDF}_{HAR}(\tau_0) \Rightarrow \sup_{r\in[r_0,1]} \frac{b^{1/2}\left[\frac{r}{2}(B^H(r)^2) - \left(\int_0^r B^H(s)dsB^H(r)\right)\right]}{\left[2r\left(\int_0^r \tilde{B}^H(s)^2ds\right)\left(\int_0^1 B^H(p)^2dp - \int_0^{1-b} B^H(p)B^H(p+br)\right)dp\right]^{1/2}}. \quad (93)$$

Under model (56), as $n \to \infty$,

$$\widetilde{SDF}_{HAR}(\tau_0) \Rightarrow \sup_{r\in[r_0,1]} \frac{b^{1/2}\left(\frac{r}{2}C_{r,d} - A_{r,d}B^H(r) + B_{r,d}cr - cA_{r,d}^2\right)}{\left[2r\left(B_{r,d} - \frac{1}{r}A_{r,d}^2\right)\left(\int_0^1 G_{r_e,c}(p)^2dp - \int_0^{(1-b)} G_{r_e,c}(p)G_{r_e,c}(p+br)dp\right)\right]^{1/2}}.$$

Under model (56) with (59), furthermore $\tau_0 < \tau_f$, as $n \to \infty$,

$$\widetilde{SDF}_{HAR}(\tau_0) \xrightarrow{p} \infty.$$

Theorem 9 documents the asymptotic behaviour of under the null hypothesis, locally and mildly explosive alternatives. Under the null hypothesis, as expected, the sup statistic converges to a stable distribution; under the local alternative, the limit of test statistic gives us the expression for asymptotical local power; and under the mildly explosive alternative, the divergent behaviour for the test statistic implies a consistent detection for an explosive episode when the data experience a sharp collapse.

The asymptotic distribution (93) and a consistent estimation of d will allow us to obtain the $(1-\beta)\%$ confidence level critical value $scv_{HAR}^{(1-\beta)\%}\left(\hat{d}\right)$.

C.1.1 Size of the sup test

To investigate the empirical size of the sup statistic $\widetilde{SDF}_{HAR}(\tau_0)$, we perform a Monte Carlo study based on DGP (64), and we let $y_0 = 0$, $d \in \{0, 0.05, ..., 0.45\}$, ,. To calculate sup statistics, as in Section 3.6.1, we let $b = 0.05$ to calculate $\tilde{\Omega}_{HAR}$. Also, we need to choose the minimum data window. based on extensive simulations, we find that the following rule of thumb gives a satisfactory power and size performance in finite sample: $r_0 = 0.01 + 4.9/\sqrt{n}$. So $r_0 \approx 0.5$ if $n = 100$ and $r_0 \approx 0.23$ if $n = 500$. For comparison purpose, we report both the empirical size of $SDF(\tau_0)$ and $\widetilde{SDF}_{HAR}(\tau_0)$ based on the 95% confidence level right-tailed unit root test.

Table 13 Empirical size of the right tailed unit root tests (95% confidence level)

$n = 100, r_0 = 0.50$					d					
Test statistics	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$SDF(\tau_0)$	0.04	0.08	0.15	0.23	0.32	0.41	0.50	0.56	0.63	0.68
$\widetilde{SDF}_{HAR}(\tau_0)$	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06
$n = 500, r_0 = 0.23$					d					
$SDF(\tau_0)$	0.04	0.15	0.32	0.49	0.64	0.75	0.83	0.87	0.90	0.93
$\widetilde{SDF}_{HAR}(\tau_0)$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Table 13 documents the empirical size of the right tailed unit root test. The finding indeed echoes the result in Table 8, where we see that $SDF(\tau_0)$ has a severe oversize problem when d becomes large, and $\widetilde{SDF}_{HAR}(\tau_0)$ has a empirical size which is reasonably close to the nominal level.

C.1.2 Power of the sup test

To illustrate the power of the sup test, we construct Monte Carlo studies based on model (60) with the following parameter settings: $y_0 = 100$, $n = 100$, $c = 1$, $\alpha = 0.6$, $r_e = 0.6$, $r_f \in \{0.7, 0.75, 0.85\}$ and $d \in \{0, 0.05, 0.1, ..., 0.45\}$. Similar to Table 7, we report the power of $SDF(\tau_0)$ and $\widetilde{SDF}_{HAR}(\tau_0)$ based on the 95% confidence level tests.

Table 14 Power of the right tailed unit root tests (95% confidence level)

$n = 100, r_0 = 0.5$		d									
	Test statistic	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$r_e = 0.6, r_f = 0.7$	$SDF(\tau_0)$	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.96	0.95	0.94
	$\widetilde{SDF}_{HAR}(\tau_0)$	1.00	1.00	1.00	0.99	0.97	0.94	0.90	0.85	0.80	0.75
$r_e = 0.6, r_f = 0.75$	$SDF(\tau_0)$	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.97
	$\widetilde{SDF}_{HAR}(\tau_0)$	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.91	0.87	0.82
$r_e = 0.6, r_f = 0.8$	$SDF(\tau_0)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98
	$\widetilde{SDF}_{HAR}(\tau_0)$	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.95	0.92	0.87

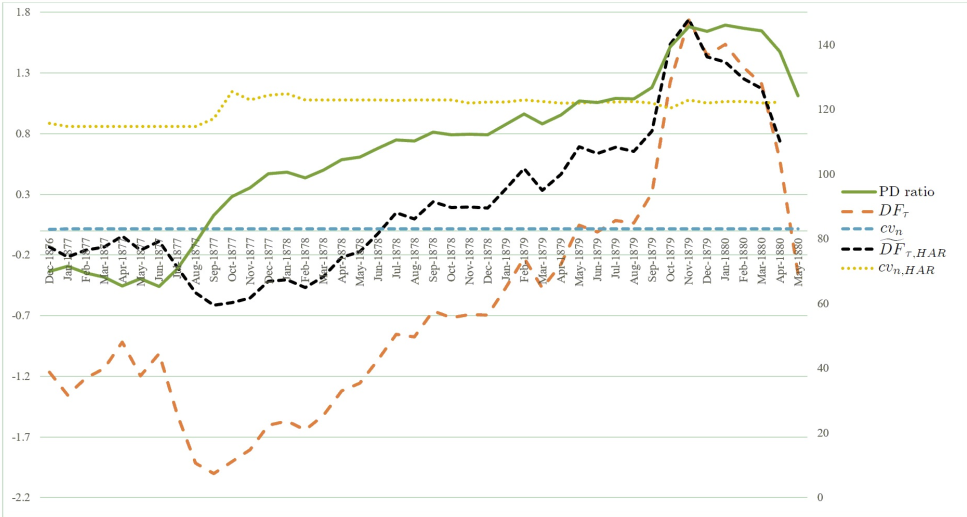
C.1.3 Sup tests for empirical studies in Table 7

To verify the robustness of the rational bubble conclusion section 3.7, we perform tests using $\widetilde{SDF}_{HAR}(\tau_0)$ based on the following time periods in the S&P 500 monthly data : January 1872 to June 1880, June 1948 to September 1956 and January 1989 to August 1997. In the following applications, we let $r_0 = 0.01 + 4.9/\sqrt{n}$ and $b = 0.05$ to calculate $\widetilde{SDF}_{HAR}(\tau_0)$.

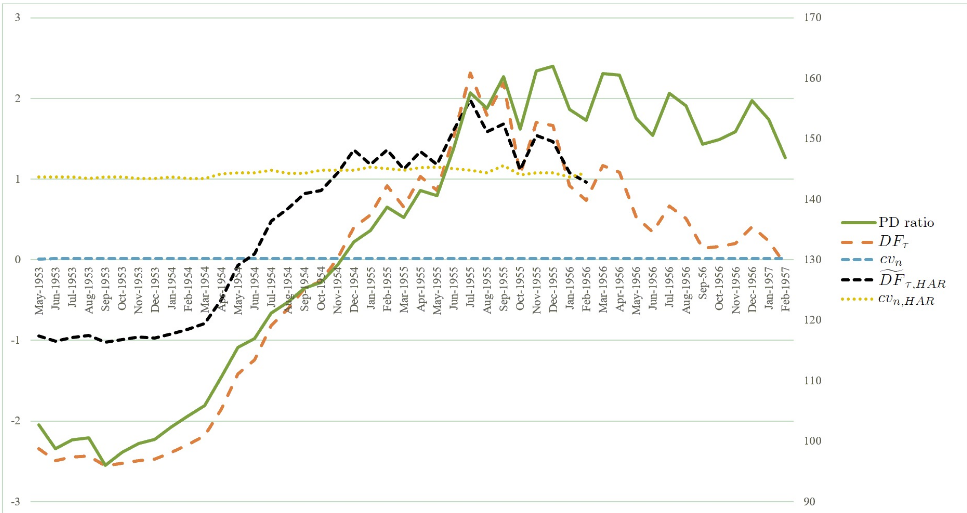
Table 15 Sup tests for the explosiveness in S&P 500

Sampling period	$\hat{d}$	$\widetilde{SDF}_{HAR}(\tau_0)$	$scv^{90\%}_{n,HAR}\left(\hat{d}\right)$	$scv^{95\%}_{n,HAR}\left(\hat{d}\right)$	$scv^{99\%}_{n,HAR}\left(\hat{d}\right)$
Jan 1872 to Jun 1880	0.32	1.742**	1.291	1.465	1.927
Jun 1948 to Sep 1956	0.3	1.980***	1.311	1.507	1.859
Jan 1989 to Aug 1997	0.26	1.303*	1.301	1.471	1.868

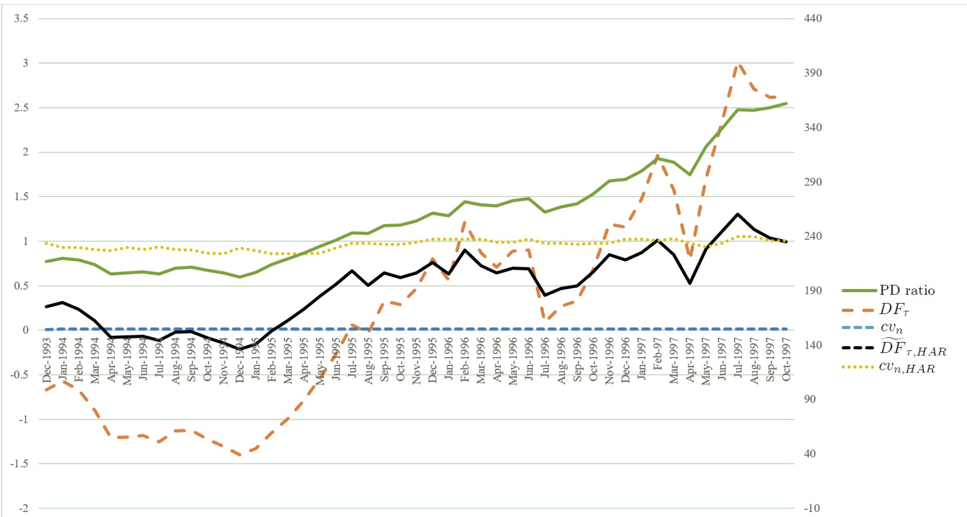
From Table 15, we see that all the conclusions of rational bubble from Section 3.7 are robust at least at 90% confidence level.



(a') Jan 1872 to May 1880



(d') Jun 1948 to Feb 1957



(f') Jan 1989 to Dec 1998

Figure 4: Estimation of the bubble origination and termination dates

C.1.4 Estimation for an explosive episode

In Figure 4, the green, brown, blue, black and yellow lines represent the price dividend ratio of the S&P 500 index, DF_τ, cv_n, $\widetilde{DF}_{\tau,HAR}$ and $cv_{n,HAR}$ respectively. While the left axis shows the sequence of test statistics and critical values, the right axis shows the values of price-dividend ratio. We use the first-time crossing principle to calculate the result in Table 12. Note that for sampling period (f'), $\hat{r}_f^{PWY}$ is not well defined, we assume the explosive episode does not end at the end of the sampling period.

C.2 Proofs of main results

Before we prove the main results in the paper, it is useful to list the following lemmas.

Lemma 16 (Corollary 4.4.1 in Giraitis et al. (2012)) *Suppose* $u_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$, *and* $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$. *Assume* $c_j \overset{a}{\sim} \gamma j^{-1+d}$ *with* $d \in (0, 0.5)$, γ *being a constant, as* $j \to \infty$,

Then we have

$$n^{-\left(\frac{1}{2}+d\right)} \sum_{t=1}^{\lfloor nr \rfloor} u_t \Rightarrow \varsigma B^H(r), \tag{94}$$

in $\mathcal{D}[0,1]$ *with the uniform metric, where* $H = \frac{1}{2} + d$, $\varsigma = \sqrt{\sigma^2 \gamma^2 \frac{\mathcal{B}(d,1-2d)}{d(1+2d)}}$ *with* $\mathcal{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, $B^H(r)$ *being an fBm with Hurst parameter* H.

Note that Lemma 16 is general enough to include the fractional intergraded process where $u_t = (1-L)^d \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$. As one can show $u_t = (1-L)^d \epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$, with $c_j \overset{a}{\sim} \frac{j^{-1+d}}{\Gamma(d)}$.

Lemma 17 *Suppose that we have the following data generating process:*

$$\begin{aligned} y_t &= y_{t-1} + u_t,\ y_0 = o_p(n^{1/2+d}), \\ (1-L)^d u_t &= \epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0, \sigma^2),\ d \in [0, 0.5) \end{aligned}$$

Let $\tau = \lfloor nr \rfloor$ *with* $r \in (0, 1]$, *as* $n \to \infty$, *we have*

$$\frac{1}{n^{1+2d}} \sum_{t=1}^{\tau} y_{t-1} u_t \Longrightarrow \begin{cases} \frac{\sigma_d^2}{2} \left[(W(r))^2 - r \right] & \text{if } d = 0, \\ \frac{\sigma_d^2}{2} \left[\left(B^H(r) \right)^2 \right] & \text{if } d \in (0, 0.5) \end{cases}, \tag{95}$$

$$\frac{1}{n^{3/2+d}}\sum_{t=1}^{\tau} y_{t-1} \Longrightarrow \sigma_d \int_0^r B^H(s)ds, \tag{96}$$

$$\frac{1}{n^{2+2d}}\sum_{t=1}^{\tau} y_{t-1}^2 \Longrightarrow \sigma_d^2 \int_0^r B^H(s)^2 ds, \tag{97}$$

where $\sigma_d = \sigma$ *if* $d = 0$, *and* $\sigma_d = \varsigma = \sqrt{\sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}}$.

Suppose empirical regression (43) with observations indexed by $t = 1, ..., \tau$ *is applied, for* $r \in (0, 1]$, *as* $n \to \infty$, *we have*

$$\tau(\hat{\rho}_\tau - 1) \Longrightarrow \begin{cases} \frac{\frac{r}{2}\left((W(r))^2 - r\right)^2 - \left(\int_0^r W(s)ds\right)W(r)}{\int_0^r \left(\tilde{W}(s)\right)^2 ds} & \text{if } d = 0, \\ \frac{\frac{r}{2}\left(B^H(r)\right)^2 - \left(\int_0^r B^H(s)ds\right)B^H(r)}{\int_0^r \left(\tilde{B}^H(s)\right)^2 ds} & \text{if } d \in (0, 0.5) \end{cases}. \tag{98}$$

Furthermore, let $\tilde{\rho}_\tau = \hat{\rho}_\tau + \frac{\frac{1}{2}\sum_{t=1}^{\tau} \triangle y_t^2}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2}$, *we have the following limit*

$$\tau(\tilde{\rho}_\tau - 1) \Longrightarrow \frac{\frac{r}{2}\left(B^H(r)\right)^2 - \left(\int_0^r B^H(s)ds\right)B^H(r)}{\int_0^r \left(\tilde{B}^H(s)\right)^2 ds}, \text{for } d \in [0, 0.5). \tag{99}$$

Proof of Lemma 17.

Note that when $d = 0$, the error term is just an i.i.d. process and the results (95), (96), (97) and (98) are well known in the literature. Therefore, we shall only prove the claims under $d \in (0, 0.5)$.

For the first claim, note that by standard calculation, we have $\sum_{t=1}^{\tau} y_{t-1}u_t = \frac{1}{2}\left(y_{\lfloor nr\rfloor}^2 - y_0^2 - \sum_{t=1}^{\tau} u_t^2\right)$ upon normalization, we have

$$\begin{aligned} n^{-1-2d}\sum_{t=1}^{\tau} y_{t-1}u_t &= \frac{1}{2}\left[\left(n^{-1/2-d}y_{\lfloor nr\rfloor}\right)^2 - \frac{1}{n^{2d}}\left(\frac{1}{n}\sum_{t=1}^{\tau} u_t^2\right)\right] + o_p(1) \\ &\Longrightarrow \frac{\varsigma^2}{2}\left[\left(B^H(r)\right)^2\right], \end{aligned}$$

where we have the last result since $n^{-1/2-d}y_{\lfloor nr\rfloor} \Rightarrow \varsigma B^H(r)$ (from Lemma 16), $\frac{1}{n}\sum_{t=1}^{\tau} u_t^2 = \frac{\tau}{n}\frac{1}{\tau}\sum_{t=1}^{\tau} u_t^2$, $\frac{1}{\tau}\sum_{t=1}^{\tau} u_t^2 \overset{a.s.}{\to} E[u_t^2]$ (by ergodic theorem) and $\frac{\tau}{n} \to r$. These imply $\frac{1}{n^{2d}}\left(\frac{1}{n}\sum_{t=1}^{\tau} u_t^2\right) \overset{p}{\to} 0$, and we obtain (95).

For the second claim, since $\sum_{t=1}^{\tau} y_{t-1} = \sum_{t=1}^{\tau}\left(\sum_{i=1}^{t-1} u_i + y_0\right)$, upon the correct normalization, applying the result of Lemma 16 and continuous mapping theorem (CMT), it is straightforward to see $\frac{1}{n^{3/2+d}}\sum_{t=1}^{\tau} y_{t-1} = \frac{1}{n}\sum_{t=1}^{\tau}\left(\frac{1}{n^{1/2+d}}\sum_{i=1}^{t-1} u_i\right) + o_p(1) \Longrightarrow \sigma_d \int_0^r B^H(s)ds$.

Applying the similar argument, we can establish the third claim by expressing $\frac{1}{n^{2+2d}}\sum_{t=1}^{\tau} y_{t-1}^2 = \frac{1}{n}\sum_{t=1}^{\tau}\left(\frac{1}{n^{1/2+d}}\sum_{i=1}^{\tau-1} u_i\right)^2 + o_p(1)\frac{1}{n^{2+2d}}\sum_{t=1}^{\tau} y_{t-1}^2 \Longrightarrow \sigma_d^2 \int_0^r B^H(s)^2 ds$.

For the normalized centered LS estimator $\tau(\hat{\rho}_\tau - 1)$, note that we can write

$$
\begin{aligned}
\tau(\hat{\rho}_\tau - 1) &= \frac{\tau}{n}\frac{n^{-1-2d}}{n^{-2-2d}}\left[\frac{\sum_{t=1}^{\tau} y_{t-1}u_t - \frac{1}{\tau}\sum_{t=1}^{\tau} y_{t-1}\sum_{t=1}^{\tau} u_t}{\sum_{t=1}^{\tau} y_{t-1}^2 - \frac{1}{\tau}\left(\sum_{t=1}^{\tau} y_{t-1}\right)^2}\right] \\
&= \frac{\tau}{n}\frac{n^{-1-2d}\sum_{t=1}^{\tau} y_{t-1}u_t - \frac{n}{\tau}n^{-3/2-d}\sum_{t=1}^{\tau} y_{t-1} n^{-1/2-d}\sum_{t=1}^{\tau} u_t}{n^{-2-2d}\sum_{t=1}^{\tau} y_{t-1}^2 - \frac{n}{\tau}\left(n^{-3/2-d}\sum_{t=1}^{\tau} y_{t-1}\right)^2} \\
&\Longrightarrow \frac{\frac{r}{2}\left[\left(B^H(r)\right)^2\right] - \int_0^r B^H(s)dsB^H(r)}{\int_0^r B^H(s)^2 ds - \frac{1}{r}\left(\int_0^r B^H(s)ds\right)^2},
\end{aligned}
\tag{100}
$$

where we obtain the last result after applying (94), (96) and (97).

As expression (100) is equivalent to (98) when $d \in (0, 0.5)$, we have shown (98).

To show (99), note that we can express

$$
\begin{aligned}
\tilde{\rho}_\tau - 1 &= \hat{\rho}_\tau - 1 + \frac{\frac{1}{2}\sum_{t=1}^{\tau} \triangle y_t^2}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2} \\
&= \frac{\sum_{t=1}^{\tau} y_{t-1}u_t - \frac{1}{\tau}\sum_{t=1}^{\tau} y_{t-1}\sum_{t=1}^{\tau} u_t}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2} + \frac{\frac{1}{2}\sum_{t=1}^{\tau} \triangle y_t^2}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2} \\
&= \frac{D_\tau}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2} - \frac{\frac{1}{\tau}\sum_{t=1}^{\tau} y_{t-1}\sum_{t=1}^{\tau} u_t}{\sum_{t=1}^{\tau} \bar{y}_{t-1}^2},
\end{aligned}
$$

where $D_\tau = \sum_{t=1}^{\tau} y_{t-1}u_t + \frac{1}{2}\sum_{t=1}^{\tau} \triangle y_t^2$.

Note that we can express $\sum_{t=1}^{\tau} y_{t-1}u_t = \frac{1}{2}\left(y_\tau^2 - y_0^2\right) - \frac{1}{2}\sum_{t=1}^{\tau} u_t^2$. Under the assumption that $\rho_n = 1$, $\triangle y_t = u_t$, this makes $\sum_{t=1}^{\tau} y_{t-1}u_t + \frac{1}{2}\sum_{t=1}^{\tau} \triangle y_t^2 = \frac{1}{2}\left(y_\tau^2 - y_0^2\right)$. And after normalization, we have

$$
n^{-1-2d} D_\tau \Longrightarrow \frac{1}{2}B^H(r)^2. \tag{101}
$$

Eventually,

$$
\begin{aligned}
\tau(\tilde{\rho}_\tau - 1) &= \frac{\tau}{n}\frac{n^{-1-2d}D_\tau - \frac{n}{\tau}n^{-3/2-d}\sum_{t=1}^{\tau} y_{t-1} n^{-1/2-d}\sum_{t=1}^{\tau} u_t}{n^{-2-2d}\sum_{t=1}^{\tau} y_{t-1}^2 - \frac{n}{\tau}\left(n^{-3/2-d}\sum_{t=1}^{\tau} y_{t-1}\right)^2} \\
&\Longrightarrow \frac{\frac{r}{2}\left[\left(B^H(r)\right)^2\right] - \int_0^r B^H(s)dsB^H(r)}{\int_0^r B^H(s)^2 ds - \frac{1}{r}\left(\int_0^r B^H(s)ds\right)^2}, \text{ for } d \in [0, 0.5).
\end{aligned}
$$

Proof of Lemma 5

We proceed to show the divergence of DF_τ under $d \in (0, 0.5)$. For the test statistics $DF_\tau = \frac{\hat{\rho}_\tau - 1}{s_\tau}$, note that by definition $s_\tau^2 = \frac{\frac{1}{\tau}\sum_{t=1}^{\tau} \hat{u}_t^2}{\sum_{t=1}^{\tau} y_{t-1}^2 - \frac{1}{\tau}\left(\sum_{t=1}^{\tau} y_{t-1}\right)^2}$.

As the LS residuals $\hat{u}_t = u_t + (1-\hat{\rho}_\tau)y_{t-1} - \hat{\mu}$ involves the intercept estimator $\hat{\mu}$, we first study the asymptotic of $\hat{\mu}$. From standard calculation, we can express

$$\hat{\mu} = \frac{\sum_{t=1}^{\tau} y_{t-1}^2 \sum_{t=1}^{\tau} u_t - \sum_{t=1}^{\tau} y_{t-1}u_t \sum_{t=1}^{\tau} y_{t-1}}{\frac{\tau}{n}\sum_{t=1}^{\tau} y_{t-1}^2 - \frac{1}{n}\left(\sum_{t=1}^{\tau} y_{t-1}\right)^2}.$$

Upon normalization, we have

$$\begin{aligned} n^{1/2-d}\hat{\mu} &= \frac{\frac{1}{n^{2+2d}}\sum_{t=1}^{\tau} y_{t-1}^2 \frac{1}{n^{1/2+d}}\sum_{t=1}^{\tau} u_t - \frac{1}{n^{1+2d}}\sum_{t=1}^{\tau} y_{t-1}u_t \frac{1}{n^{3/2+d}}\sum_{t=1}^{\tau} y_{t-1}}{\frac{\tau}{n}\frac{1}{n^{2+2d}}\sum_{t=1}^{\tau} y_{t-1}^2 - \left(\frac{1}{n^{3/2+d}}\sum_{t=1}^{\tau} y_{t-1}\right)^2} \\ &\Longrightarrow \frac{\sigma_d^3 \int_0^r B^H(s)^2 ds B^H(r) - \frac{\sigma_d^3}{2} B^H(r)^2 \int_0^r B^H(s)ds}{r\sigma_d^2 \int_0^r B^H(s)^2 ds - \sigma_d^2\left(\int_0^r B^H(s)ds\right)^2}. \qquad (102) \end{aligned}$$

This implies $\hat{\mu} = O_p(n^{-1/2+d}) = o_p(1)$.

For the average squared residuals $\frac{1}{\tau}\sum_{t=1}^{\tau} \hat{u}_t^2$, we can write

$$\begin{aligned} \frac{1}{\tau}\sum_{t=1}^{\tau} \hat{u}_t^2 &= \frac{1}{\tau}\sum_{t=1}^{\tau} (u_t + (1-\hat{\rho}_\tau)y_{t-1} - \hat{\mu})^2 \\ &= \frac{1}{\tau}\sum_{t=1}^{\tau} u_t^2 + \frac{2(1-\hat{\rho}_\tau)}{\tau}\sum_{t=1}^{\tau} y_{t-1}u_t + \frac{(1-\hat{\rho}_\tau)^2}{\tau}\sum_{t=1}^{\tau} y_{t-1}^2 \\ &\quad -2\hat{\mu}\frac{1}{\tau}\sum_{t=1}^{\tau} u_t - 2\frac{(1-\hat{\rho}_\tau)\hat{\mu}}{\tau}\sum_{t=1}^{\tau} y_{t-1} + \frac{\hat{\mu}^2}{\tau}\sum_{t=1}^{\tau} 1. \\ &= \frac{1}{\tau}\sum_{t=1}^{\tau} u_t^2 + \frac{2\tau(1-\hat{\rho}_\tau)}{\tau^2}\sum_{t=1}^{\tau} y_{t-1}u_t + \frac{(\tau(1-\hat{\rho}_\tau))^2}{\tau^3}\sum_{t=1}^{\tau} y_{t-1}^2 \\ &\quad -2\tau(1-\hat{\rho}_\tau)\hat{\mu}\left(\frac{1}{\tau^2}\sum_{t=1}^{\tau} y_{t-1}\right) + \hat{\mu}^2. \end{aligned}$$

From Lemma 17, $\tau(1-\hat{\rho}_\tau) = O_p(1)$, $\tau^{-2}\sum_{t=1}^{\tau} y_{t-1}u_t = O_p(n^{2d-1}) = o_p(1)$, $\tau^{-3}\sum_{t=1}^{\tau} y_{t-1}^2 = O_p(n^{2d-1}) = o_p(1)$, $\tau^{-2}\sum_{t=1}^{\tau} y_{t-1} = O_p(n^{d-1/2})$ and from (102), $\hat{\mu}^2 = O_p(n^{-1+2d}) = o_p(1)$, we can express $\frac{1}{\tau}\sum_{t=1}^{\tau} \hat{u}_t^2 = \frac{1}{\tau}\sum_{t=1}^{\tau} u_t^2 + o_p(1) \xrightarrow{p} E[u_t^2]$ by ergodic theorem.

Note that applying Lemma 17, we know the limit of the sum of demeaned y_t,

$$n^{-(2+2d)}\left[\sum_{t=1}^{\tau}y_{t-1}^2-\frac{1}{\tau}\left(\sum_{t=1}^{\tau}y_{t-1}\right)^2\right] = n^{-(2+2d)}\sum_{t=1}^{\tau}y_{t-1}^2-\frac{\left(n^{-3/2-d}\sum_{t=1}^{\tau}y_{t-1}\right)^2}{\frac{\tau}{n}}$$
$$\Longrightarrow \sigma_d^2\left(\int_0^r\left(B^H(s)\right)^2ds-\frac{1}{r}\left(\int_0^r B^H(s)ds\right)^2\right) \quad (103)$$

Therefore it is straightforward to show

$$\begin{aligned}
& n^{-d}DF_\tau \\
= & \frac{n(\hat{\rho}_\tau-1)}{(n^{2+2d}s_\tau^2)^{1/2}} \\
= & \frac{\frac{n}{\tau}\tau(\hat{\rho}_\tau-1)}{(n^{2+2d}s_\tau^2)^{1/2}} \\
\Longrightarrow & \frac{1}{r}\frac{\frac{r}{2}\left[\left(B^H(r)\right)^2\right]-\int_0^r B^H(s)dsB^H(r)}{\int_0^r B^H(s)^2ds-\frac{1}{r}\left(\int_0^r B^H(s)ds\right)^2}\times\left(\frac{\sigma_d^2\left(\int_0^r\left(B^H(s)\right)^2ds-\frac{1}{r}\left(\int_0^r B^H(s)ds\right)^2\right)}{E[u_t^2]}\right)^{1/2}.
\end{aligned}$$

This implies the test statistic DF_τ is of $O_p(n^d)$ and completes the proof of Lemma 5.

Proof of Theorem 4, Theorem 5 and Corollary 1

Note that we can express

$$DF_{\tau,HAR}=\frac{(\hat{\rho}_\tau-1)}{s_{\tau,HAR}}=\frac{\tau\left(\hat{\rho}_\tau-1\right)}{\left(\tau^2s_{\tau,HAR}^2\right)^{1/2}} \quad (104)$$

and

$$\widetilde{DF}_{\tau,HAR}=\frac{\tau\left(\tilde{\rho}_\tau-1\right)}{\left(\tau^2s_{\tau,HAR}^2\right)^{1/2}}. \quad (105)$$

To show the limit we first study the denominator of (104) and (105). Note that $s_{\tau,HAR}^2=\frac{\hat{\Omega}_{HAR}}{\sum_{t=1}^{\tau}\bar{y}_{t-1}^2}$. For $\hat{\Omega}_{HAR}$, letting $K_{i,j}=K\left(\frac{i-j}{b\tau}\right)$ and $S_t=\sum_{i=1}^{t}\Delta y_i$ allow us to write

$$
\begin{aligned}
\hat{\Omega}_{HAR} &= \sum_{j=-\tau+1}^{\tau} K\left(\frac{j}{b\tau}\right)\hat{\gamma}_j \\
&= \frac{1}{\tau}\sum_{i=1}^{\tau}\sum_{i=1}^{\tau}\Delta y_i K_{i,j}\Delta y_j \\
&= \frac{1}{\tau}\sum_{i=1}^{\tau-1}\frac{1}{\tau}\sum_{j=1}^{\tau-1}\tau^2\left[(K_{i,j}-K_{i,j+1})-(K_{i+1,j}-K_{i+1,j+1})\right]\frac{1}{\sqrt{\tau}}\hat{S}_i\frac{1}{\sqrt{\tau}}\hat{S}_j \\
&= \frac{1}{\tau}\sum_{i=1}^{\tau-1}\frac{1}{\tau}\sum_{j=1}^{\tau-1}\tau^2 D_\tau\left(\frac{i-j}{b\tau}\right)\frac{1}{\sqrt{\tau}}S_i\frac{1}{\sqrt{\tau}}S_j \qquad (106)
\end{aligned}
$$

where $D_\tau\left(\frac{i-j}{b\tau}\right) = (K_{i,j}-K_{i,j+1})-(K_{i+1,j}-K_{i+1,j+1})$. The last equality follows from Equation (A.1) in Kiefer and Vogelsang (2002a) .

From straightforward calculation, we can show $D_\tau\left(\frac{i-j}{b\tau}\right) = \begin{cases} \frac{2}{b\tau} & \text{if } |i-j|=0, \\ -\frac{1}{b\tau} & \text{if } |i-j|=\lfloor b\tau\rfloor, \\ 0 & \text{otherwise.}\end{cases}$

This implies

$$
\begin{aligned}
\hat{\Omega}_{HAR} &= \sum_{i=1}^{\tau-1}\sum_{j=1}^{\tau-1} D_\tau\left(\frac{i-j}{b\tau}\right)\frac{1}{\sqrt{\tau}}S_i\frac{1}{\sqrt{\tau}}S_j \\
&= \frac{2}{b\tau}\sum_{i=1}^{\tau-1}\left(\frac{1}{\sqrt{\tau}}S_i\right)^2 - \frac{2}{b\tau}\sum_{i=1}^{\tau-\lfloor b\tau\rfloor-1}\left(\frac{1}{\sqrt{\tau}}S_i\right)\left(\frac{1}{\sqrt{\tau}}S_{i+\lfloor b\tau\rfloor}\right) \\
&= \frac{2}{b}\frac{n}{\lfloor nr\rfloor}\frac{1}{n}\sum_{i=1}^{\tau-1}\left(\frac{1}{\sqrt{\tau}}S_i\right)^2 - \frac{2}{b}\frac{n}{\lfloor nr\rfloor}\frac{1}{n}\sum_{i=1}^{\tau-\lfloor b\tau\rfloor-1}\left(\frac{1}{\sqrt{\tau}}S_i\right)\left(\frac{1}{\sqrt{\tau}}S_{i+\lfloor b\tau\rfloor}\right) \quad (107)
\end{aligned}
$$

Thus, with $i = \lfloor np\rfloor$ and under the assumption $\rho_n = 1$, we have $S_i = \sum_{j=1}^{i}\Delta y_j = \sum_{j=1}^{\lfloor np\rfloor} u_j$, it implies that

$$
\begin{aligned}
&\frac{1}{n^d}\frac{1}{\sqrt{\tau}}S_{\lfloor np\rfloor} \\
&= \left(\frac{n}{\tau}\right)^{1/2}\frac{1}{n^{1/2+d}}\sum_{t=1}^{\lfloor np\rfloor}u_i \Longrightarrow \frac{\sigma_d}{r^{1/2}}B^H(p). \qquad (108)
\end{aligned}
$$

Therefore we have,

$$\begin{aligned}
& \frac{1}{n^{2d}}\hat{\Omega}_{HAC} \\
= & \frac{2n}{b\tau}\frac{1}{n}\sum_{i=1}^{\tau-1}\left(\frac{1}{n^d}\frac{1}{\sqrt{\tau}}S_{\lfloor np\rfloor}\right)^2 - \frac{2n}{b\tau}\frac{1}{n}\sum_{i=1}^{\tau-\lfloor b\tau\rfloor-1}\left(\frac{1}{n^d}\frac{1}{\sqrt{\tau}}S_i\right)\left(\frac{1}{n^d}\frac{1}{\sqrt{\tau}}S_{i+\lfloor b\tau\rfloor}\right) \\
\Longrightarrow & \frac{2}{br}\int_0^r\left(\frac{\sigma_d}{r^{1/2}}B^H(p)\right)^2 dp - \frac{2}{br}\int_0^{(1-b)r}\frac{\sigma_d^2}{r}B^H(p)B^H(p+br)dp \qquad (109) \\
= & \frac{2\sigma_d^2}{br^2}\left(\int_0^r B^H(p)^2dp - \int_0^{(1-b)r}B^H(p)B^H(p+br)dp\right) \qquad (110)
\end{aligned}$$

where we have applied (108) and continuous mapping theorem to obtain the limit (109).

Combining (103) and (109), upon normalization, we can show

$$\begin{aligned}
\tau^2 s^2_{\tau,HAR} &= \left(\frac{\tau}{n}\right)^2\frac{\frac{1}{n^{2d}}\hat{\Omega}_{HAR}}{\frac{1}{n^{2+2d}}\left(\sum_{t=1}^{\tau}y_{t-1}^2 - \tau^{-1}\left(\sum_{t=1}^{\tau}y_{t-1}\right)^2\right)} \\
&\Longrightarrow \frac{2\left(\int_0^r B^H(p)^2dp - \int_0^{(1-b)r}B^H(p)B^H(p+br)dp\right)}{b\int_0^r\tilde{B}^H(s)^2ds} \qquad (111)
\end{aligned}$$

We now proceed to show the limit of $DF_{\tau,HAR}$.

Suppose $d=0$, we have

$$\begin{aligned}
& DF_{\tau,HAR} \\
= & \frac{\tau\left(\hat{\rho}_\tau - 1\right)}{\left(\tau^2 s^2_{\tau,HAR}\right)^{1/2}} \\
= & \frac{\lfloor nr\rfloor}{nr}r\frac{n\left(\hat{\rho}_\tau-1\right)}{\left(\tau^2 s^2_{\tau,HAR}\right)^{1/2}} \\
\Longrightarrow & \frac{r\int_0^r\tilde{W}(s)dW(s)}{\int_0^r\tilde{W}(s)^2ds}\left(\frac{b\int_0^r\tilde{W}(s)^2ds}{2\left(\int_0^r W(p)^2dp - \int_0^{(1-b)r}W(p)W(p+br)dp\right)}\right)^{1/2} \\
= & \frac{b^{1/2}r\int_0^r\tilde{W}(s)dW(s)}{\left[2\int_0^r\tilde{W}(s)^2ds\left(\int_0^r W(p)^2dp - \int_0^{(1-b)r}W(p)B^H(p+br)dp\right)\right]^{1/2}},
\end{aligned}$$

where we use the well known result that $n\left(\hat{\rho}_\tau-1\right) \Longrightarrow \int_0^r\tilde{W}(s)dW(s)/\int_0^r\tilde{W}(s)^2ds$ and (111) with $H=1/2$.

For $d \in (0, 0.5)$, similarly, we can express

$$DF_{\tau,HAR} = \frac{\tau\left(\hat{\rho}_\tau - 1\right)}{\left(\tau^2 s^2_{\tau,HAR}\right)^{1/2}}$$

$$\Longrightarrow \frac{\frac{r}{2}\left(B^H(r)\right)^2 - \left(\int_0^r B^H(s)ds\right)B^H(r)}{\int_0^r \left(\tilde{B}^H(s)\right)^2 ds}\left(\frac{b\int_0^r \tilde{B}^H(s)^2 ds}{2\left(\int_0^r B^H(p)^2dp - \int_0^{(1-b)r} B^H(p)B^H(p+br)dp\right)}\right)^{1/2}$$

$$= \frac{\frac{rb^{1/2}}{2}\left(B^H(r)\right)^2 - b^{1/2}\left(\int_0^r B^H(s)ds\right)B^H(r)}{\left[2\int_0^r \left(\tilde{B}^H(s)\right)^2 ds\left(\int_0^r B^H(p)^2dp - \int_0^{(1-b)r} B^H(p)B^H(p+br)dp\right)\right]^{1/2}}, \tag{112}$$

where we obtain the limit from using (100) and (111).

To show the limit of $\widetilde{DF}_{\tau,HAR}$, using the similar steps to show (112) with the use of (99) and (111), we have

$$\widetilde{DF}_{\tau,HAR} = \frac{\tau\left(\tilde{\rho}_\tau - 1\right)}{\left(\tau^2 s^2_{\tau,HAR}\right)^{1/2}}$$

$$\Longrightarrow \frac{\frac{rb^{1/2}}{2}\left(B^H(r)\right)^2 - b^{1/2}\left(\int_0^r B^H(s)ds\right)B^H(r)}{\left[2\int_0^r \left(\tilde{B}^H(s)\right)^2 ds\left(\int_0^r B^H(p)^2dp - \int_0^{(1-b)r} B^H(p)B^H(p+br)dp\right)\right]^{1/2}}.$$

Corollary 1 is a special case under Theorem 5 with $r = 1$.

This completes the proof of Theorem 4, Theorem 5 and Corollary 1.

Proof of Remark 28

Let $\hat{\Omega}_{HAR} = \sum_{j=-\tau+1}^{\tau} K\left(\frac{j}{M}\right)\hat{\gamma}_j$ as in Theorem 5. Furthernore, $K(.) \geqslant 0$, $K(x) = K(-x)$ and $K(\cdot)$ is a twice differentiable function. As in (106), we can express $\hat{\Omega}_{HAR} = \frac{1}{\tau}\sum_{i=1}^{\tau-1}\frac{1}{\tau}\sum_{j=1}^{\tau-1}\tau^2 D_\tau\left(\frac{i-j}{b\tau}\right)\frac{1}{\sqrt{\tau}}S_i\frac{1}{\sqrt{\tau}}S_j$, and following the steps in proving Theorem 4 in Sun (2004) we can show $\lim_{n\to\infty}\tau^2 D_\tau\left(\frac{i-j}{b\tau}\right) = -\frac{1}{b^2r^2}K''\left(\frac{p-q}{br}\right)$, given $(i/n, j/n) \to (p, q)$.

Combining (106), (108) and applying continuous mapping theorem, we have

$$\frac{1}{n^{2d}}\hat{\Omega}_{HAR} = \frac{1}{\tau}\sum_{i==1}^{\tau-1}\frac{1}{\tau}\sum_{j=1}^{\tau-1} D_\tau\left(\frac{i-j}{b\tau}\right)\frac{1}{n^d}\frac{1}{\sqrt{\tau}}\hat{S}_i\frac{1}{n^d}\frac{1}{\sqrt{\tau}}\hat{S}_j$$

$$\Longrightarrow -\frac{\sigma_d^2}{b^2r^3}\int_0^r\int_0^r K''\left(\frac{p-q}{br}\right)B^H(p)B^H(q)dpdq. \tag{113}$$

Since $s^2_{\tau,HAR} = \frac{\hat{\Omega}_{HAR}}{\sum_{t=1}^{\tau} y_{t-1}^2 - \tau^{-1}\left(\sum_{t=1}^{\tau} y_{t-1}\right)^2}$, combining (103) and (113) gives us

$$\tau^2 s^2_{\tau,adj} \Longrightarrow \frac{\int_0^r \int_0^r -K''\left(\frac{p-q}{br}\right) B^H(p) B^H(q) dp dq}{b^2 r \int_0^r \tilde{B}^H(s)^2 ds}. \tag{114}$$

Finally, $\widetilde{DF}_{\tau,HAR} = \frac{\tau(\tilde{\rho}_\tau - 1)}{\left(\tau^2 s^2_{\tau,HAR}\right)^{1/2}}$, from (99), (114) and standard calculation yields

$$\widetilde{DF}_{\tau,HAR} \Longrightarrow \frac{\frac{br^{3/2}}{2}\left(B^H(r)\right)^2 - br^{1/2}\left(\int_0^r B^H(s) ds\right) B^H(r)}{\left(\left(\int_0^r \tilde{B}^H(s)^2 ds\right) \int_0^r \int_0^r -K''\left(\frac{p-q}{br}\right) B^H(p) B^H(q) dp dq\right)^{1/2}}.$$

This completes the proof of Remark 28.

Before we proof Theorem 6, it is useful to introduce the following lemma.

To study the limit distributions of $DF_{\tau,HAR}$ and $DF^{Bartlett}_{\tau,HAR}$, we first introduce the following lemma.

Lemma 18 *Under the local alternative model (56), let $\tau = \lfloor nr \rfloor$ with $r \in (r_e, 1]$, as $n \to \infty$,*

1. $\frac{1}{n^{1/2+d}} y_\tau \Rightarrow \sigma_d \left(e^{(r-r_e)c} B^H(r_e) + \int_{r_e}^r e^{(r-s)c} dB^H(s) \right)$;
2. $\frac{1}{n^{3/2+d}} \sum_{t=1}^{\tau} y_{t-1} \Rightarrow \sigma_d A_{r,d}$;
3. $\frac{1}{n^{2+2d}} \sum_{t=1}^{\tau} y_{t-1}^2 \Rightarrow \sigma_d^2 B_{r,d}$;
4. $\frac{1}{n^{1+2d}} \left(\sum_{t=1}^{\tau} y_{t-1} u_t + \frac{1}{2} \sum_{t=1}^{\tau} \Delta y_t^2 \right) \Rightarrow \frac{\sigma_d^2}{2} C_{r,d}$;
5. $n(\tilde{\rho}_\tau - \rho_c) \Rightarrow X_c(r, d)$;
6. $n(\tilde{\rho}_\tau - 1) \Rightarrow X_c(r, d) + c$;

where

$$
\begin{aligned}
A_{r,d} &= \int_0^r \left(e^{(x-r_e)c} B^H(r_e) + \int_{r_e}^x e^{(x-s)c} dB^H(s) \right) dx, \\
B_{r,d} &= \int_0^r \left(e^{(x-r_e)c} B^H(r_e) + \int_{r_e}^x e^{(x-s)c} dB^H(s) \right)^2 dx, \\
C_{r,d} &= \left(e^{(r-r_e)c} B^H(r_e) + \int_{r_e}^r e^{(r-s)c)} dB^H(s) \right)^2 - B^H(r_e)^2, \\
X_c(r,d) &= \frac{\frac{1}{2} C_{r,d} - \frac{1}{r} A_{r,d} B^H(r)}{B_{r,d} - \frac{1}{r} A_{r,d}^2}, \\
Y_c(r,d) &: = \frac{B_{r,d} B^H(r) - \frac{1}{2} C_{r,d} A_{r,d}}{r \left(B_{r,d} - A_{r,d}^2 \right)}.
\end{aligned}
$$

C.2.1 Proof of Lemma 18

1. To show the first claim, note that by backward substitution, we can express $y_{\lfloor nr \rfloor}$ as

$$
y_{\lfloor nr \rfloor} = \rho_n^{\lfloor nr \rfloor - (\lfloor nr_e \rfloor - 1)} y_{\lfloor nr_e \rfloor - 1} + \sum_{j=\lfloor r_e \rfloor}^{\lfloor nr \rfloor} \left(1 + \frac{c}{n} \right)^{\lfloor nr \rfloor - j} u_j.
$$

Note that $\rho_n^{\lfloor nr \rfloor - (\lfloor nr_e \rfloor - 1)} = (1 + c/n)^{\lfloor nr \rfloor - (\lfloor nr_e \rfloor - 1)} = \exp(c (\lfloor nr \rfloor - (\lfloor nr_e \rfloor - 1)) / n) + o(1) = \exp((r - r_e)c) + o(1)$. Therefore we have

$$
\begin{aligned}
\frac{1}{n^{1/2+d}} y_{\lfloor nr \rfloor} &= \exp((r - r_e)c) \frac{1}{n^{1/2+d}} y_{\lfloor nr_e \rfloor - 1} + \frac{1}{n^{1/2+d}} \sum_{j=\lfloor r_e \rfloor}^{\lfloor nr \rfloor} \left(1 + \frac{c}{n} \right)^{\lfloor nr \rfloor - j} u_j \\
&\Rightarrow \sigma_d \left(e^{(r-r_e)c} B^H(r_e) + \int_{r_e}^r e^{(r-s)c} dB^H(s) \right), \qquad (115)
\end{aligned}
$$

where we obtain the limit by applying Lemma 16 and continuous mapping theorem (add footnote here to cite my OBES paper).

For the second and the third claims, note that we can express $\frac{1}{n^{2/3+d}} \sum_{t=1}^{\tau} y_{t-1} = \frac{1}{n} \sum_{t=1}^{\tau} \left(\frac{1}{n^{1/2+d}} y_{t-1} \right)$ and $\frac{1}{n^{2+2d}} \sum_{t=1}^{\tau} y_{t-1}^2 = \frac{1}{n} \sum_{t=1}^{\tau} \left(\frac{1}{n^{1/2+d}} y_{t-1} \right)^2$, (115) and CMT will yiled the results.

For the fourth claim, we first study the second component $\frac{1}{2} \sum_{t=1}^{\tau} \Delta y_t^2$. Since

$$
\Delta y_t = \begin{cases} u_t & \text{if } t < \tau_e, \\ \frac{c}{n} y_{t-1} + u_t & \text{otherwise} \end{cases},
$$

we can express

$$
\begin{aligned}
\sum_{t=1}^{\tau} \Delta y_t^2 &= \sum_{t=1}^{\tau_e - 1} u_t^2 + \sum_{t=\tau_e}^{\tau} \left(\frac{c}{n} y_{t-1} + u_t \right)^2 \\
&= \sum_{t=1}^{\tau} u_t^2 + \frac{c^2}{n^2} \sum_{t=\tau_e}^{\tau} y_{t-1}^2 + \frac{2c}{n} \sum_{t=\tau_e}^{\tau} y_{t-1} \\
&= \sum_{t=1}^{\tau} u_t^2 + \frac{c^2}{n^2} n^{2+2d} \left(\frac{1}{n^{2+2d}} \sum_{t=\tau_e}^{\tau} y_{t-1}^2 \right) + \frac{2c}{n} n^{3/2+d} \left(\frac{1}{n^{3/2+d}} \sum_{t=\tau_e}^{\tau} y_{t-1} \right) \quad (116)
\end{aligned}
$$

For the term $\frac{1}{n^{2+2d}} \sum_{t=\tau_e}^{\tau} y_{t-1}^2$, as $n \to \infty$, we have

$$
\begin{aligned}
\frac{1}{n^{2+2d}} \sum_{t=\tau_e}^{\tau} y_{t-1}^2 &= \frac{1}{n^{2+2d}} \sum_{t=1}^{\tau} y_{t-1}^2 - \frac{1}{n^{2+2d}} \sum_{t=1}^{\tau_e - 1} y_{t-1}^2 \\
&\Rightarrow \sigma_d^2 B_{r,d} - \sigma_d^2 \int_0^{r_e} B^H(s)^2 ds, \quad (117)
\end{aligned}
$$

where we have applied the results Lemma 18.3 and (97) to obtain the limit.

For the term $\frac{1}{n^{3/2+d}} \sum_{t=\tau_e}^{\tau} y_{t-1}$, similarly, using Lemma 18.2 and (96), we have the limit

$$
\frac{1}{n^{3/2+d}} \sum_{t=\tau_e}^{\tau} y_{t-1} \Rightarrow \sigma_d A_{r,d} - \sigma_d \int_0^{r_e} B^H(s) ds. \quad (118)
$$

Combining (116), (117) and (118), it can be seen that,

$$
\sum_{t=1}^{\tau} \Delta y_t^2 = \sum_{t=1}^{\tau} u_t^2 + R_{1,n}, \; R_{1,n} = O_p(n^{1/2+d}).
$$

This implies that, upon normalization, we can express

$$
\begin{aligned}
& \frac{1}{n^{1+2d}} \left(\sum_{t=1}^{\tau} y_{t-1} u_t + \frac{1}{2} \sum_{t=1}^{\tau} \Delta y_t^2 \right) \\
= & \frac{1}{n^{1+2d}} \left(\sum_{t=1}^{\tau} y_{t-1} u_t + \frac{1}{2} \sum_{t=1}^{\tau} u_t^2 + R_{1,n} \right) \\
= & \frac{1}{n^{1+2d}} \left(\sum_{t=1}^{\lfloor nr_e \rfloor - 1} y_{t-1} u_t + \frac{1}{2} \sum_{t=1}^{\lfloor nr_e \rfloor - 1} u_t^2 \right) + \frac{1}{n^{1+2d}} \left(\sum_{t=\lfloor nr_e \rfloor}^{\tau} y_{t-1} u_t + \frac{1}{2} \sum_{t=\lfloor nr_e \rfloor}^{\tau} u_t^2 \right) + \frac{R_{1,n}}{n^{1+2d}} \quad (119)
\end{aligned}
$$

For the first component in (119), applying (101) we have

$$
\frac{1}{n^{1+2d}} \left(\sum_{t=1}^{\lfloor nr_e \rfloor - 1} y_{t-1} u_t + \frac{1}{2} \sum_{t=1}^{\lfloor nr_e \rfloor - 1} u_t^2 \right) \Rightarrow \frac{\sigma_d^2}{2} \left[(B^H(r_e))^2 \right].
$$

For the second component $\frac{1}{n^{1+2d}}\left(\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}u_t + \frac{1}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2\right)$, as from standard calculation we can write

$$\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}u_t = \frac{1}{2\rho_n}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_t^2 - \frac{\rho_n}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}^2 - \frac{1}{2\rho_n}\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2,$$

we have

$$\begin{aligned}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}u_t + \frac{1}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2 &= \frac{1}{2\rho_n}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_t^2 - \frac{\rho_n}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}^2 - \frac{1}{2\rho_n}\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2 + \frac{1}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2 \\ &= \frac{1}{2\rho_n}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_t^2 - \frac{\rho_n}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}^2 + \frac{1}{2}\left(1-\frac{1}{\rho_n}\right)\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2.\end{aligned}$$

As $\rho_n = 1 + o(1)$, we can express

$$\frac{1}{n^{1+2d}}\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}u_t = \frac{1}{n^{1+2d}}\frac{1}{2}\left[y_\tau^2 - y_{\lfloor nr_e\rfloor-1}^2\right] - \frac{1}{n^{1+2d}}\frac{1}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2 + o_p(1).$$

This makes

$$\begin{aligned}\frac{1}{n^{1+2d}}\left(\sum_{t=\lfloor nr_e\rfloor}^{\tau} y_{t-1}u_t + \frac{1}{2}\sum_{t=\lfloor nr_e\rfloor}^{\tau} u_t^2\right) &= \frac{1}{n^{1+2d}}\frac{1}{2}\left[y_\tau^2 - y_{\lfloor nr_e\rfloor-1}^2\right] + o_p(1) \\ &\Rightarrow \frac{\sigma_d^2}{2}\left[\left(e^{(r-r_e)c}B^H(r_e) + \int_{r_e}^{r} e^{(r-s)c)}dB^H(s)\right)^2 - (B^H(r_e))^2\right]\end{aligned}$$

where we obtain the limit by applying Lemma 18.1 and Lemma 16.

The last term $R_{1,n}/n^{1+2d}$ in (119) vanishes as $\infty \to \infty$, as $R_{1,n} = O_p(n^{1/2+d})$. So we have shown the fourth claim.

To show the fifth claim, note that

$$\begin{aligned}\tilde{\rho}_\tau - \rho_n &= (\hat{\rho}_\tau - \rho_n) + \frac{\frac{1}{2}\sum_{t=1}^{\tau}\Delta y_t^2}{\sum_{t=1}^{\tau}\bar{y}_{t-1}^2} \\ &= \frac{\sum_{t=1}^{\tau}\bar{y}_{t-1}(y_t - \rho_n y_{t-1}) + \frac{1}{2}\sum_{t=1}^{\tau}\Delta y_t^2}{\sum_{t=1}^{\tau}\bar{y}_{t-1}^2}. \qquad (121)\end{aligned}$$

We now analyse the numerator in (121).

$$
\begin{aligned}
&\sum_{t=1}^{\tau} \bar{y}_{t-1}\left(y_t - \rho_n y_{t-1}\right) + \frac{1}{2}\sum_{t=1}^{\tau} \Delta y_t^2 \\
= &\sum_{t=1}^{\tau_e-1} \bar{y}_{t-1}\left(y_t - \rho_n y_{t-1}\right) + \sum_{t=\tau_e}^{\tau} \bar{y}_{t-1}\left(y_t - \rho_n y_{t-1}\right) + \frac{1}{2}\sum_{t=1}^{\tau} u_t^2 + R_{1,n} \\
= &\sum_{t=1}^{\tau_e-1} \bar{y}_{t-1}\left(y_{t-1} + u_t - \rho_c y_{t-1}\right) + \sum_{t=\tau_e}^{\tau} \bar{y}_{t-1} u_t + \frac{1}{2}\sum_{t=1}^{\tau} u_t^2 + R_{1,n} \\
= &\left(\sum_{t=1}^{\tau} \bar{y}_{t-1} u_t + \frac{1}{2}\sum_{t=1}^{\tau} u_t^2\right) - \frac{c}{n}\sum_{t=1}^{\tau_e-1} \bar{y}_{t-1} + R_{1,n}. \qquad (122)
\end{aligned}
$$

Upon normalization, the first component in (122)

$$
\begin{aligned}
&\frac{1}{n^{1+2d}}\left(\sum_{t=1}^{\tau} \bar{y}_{t-1} u_t + \frac{1}{2}\sum_{t=1}^{\tau} u_t^2\right) \\
= &\frac{1}{n^{1+2d}}\left(\sum_{t=1}^{\tau} y_{t-1} u_t + \frac{1}{2}\sum_{t=1}^{\tau} u_t^2 - \frac{1}{\tau}\sum_{t=1}^{\tau} y_{t-1}\sum_{t=1}^{\tau} u_t\right) \\
= &\frac{1}{n^{1+2d}}\left(\sum_{t=1}^{\tau} y_{t-1} u_t + \frac{1}{2}\sum_{t=1}^{\tau} u_t^2\right) - \frac{n}{\tau}\frac{1}{n^{3/2+d}}\sum_{t=1}^{\tau} y_{t-1}\frac{1}{n^{1/2+d}}\sum_{t=1}^{\tau} u_t \\
\Rightarrow &\frac{1}{2}\sigma_d^2 C_{r,d} - \frac{1}{r}\sigma_d^2 A_{r,d} B^H(r), \qquad (123)
\end{aligned}
$$

1. where we have applied (120), Lemma 18 and Lemma 16 to obtain the limit.

For the second component, note that

$$
\begin{aligned}
\frac{c}{n}\sum_{t=1}^{\tau_e-1} \bar{y}_{t-1} &= \frac{c}{n}\sum_{t=1}^{\tau_e-1} y_{t-1} - \frac{c}{n}\frac{1}{\tau}\sum_{t=1}^{\tau_e-1}\sum_{j=1}^{\tau} y_{j-1} \\
&= \frac{c}{n}\sum_{t=1}^{\tau_e-1} y_{t-1} - \frac{c}{n}\frac{\tau_e-1}{\tau}\sum_{j=1}^{\tau} y_{j-1}.
\end{aligned}
$$

After normalization, we have

$$
\begin{aligned}
\frac{1}{n^{1/2+d}}\left(\frac{c}{n}\sum_{t=1}^{\tau_e-1} y_{t-1} - \frac{c}{n}\frac{\tau_e-1}{\tau}\sum_{j=1}^{\tau} y_{j-1}\right) &= \frac{c}{n^{3/2+d}}\sum_{t=1}^{\tau_e-1} y_{t-1} - \frac{\tau_e-1}{\tau}\frac{c}{n^{3/2+d}}\sum_{j=1}^{\tau} y_{j-1} \\
&\Rightarrow c\sigma_d\int_0^{r_e} B^H(s)ds - c r_e \sigma_d A_{r,d}
\end{aligned}
$$

where we have obtain the limit by using Lemma 17 and 18.2.

Therefore, the first component in (122) is $O_p(n^{1+2d})$ dominates $\frac{c}{n}\sum_{t=1}^{\tau_e-1}\bar{y}_{t-1} = O_p(n^{1/2+d})$ and the third component $R_{1,n} = O_p(n^{1/2+d})$, consequently, we have

$$\frac{1}{n^{1+2d}}\left(\sum_{t=1}^{\tau}\bar{y}_{t-1}\left(y_t-\rho_n y_{t-1}\right)+\frac{1}{2}\sum_{t=1}^{\tau}\Delta y_t^2\right) \Rightarrow \frac{1}{2}\sigma_d^2 C_{r,d} - \frac{1}{r}\sigma_d^2 A_{r,d}B^H(r).$$

For the denominator in (121), applying Lemma 18.2 and 18.3 we have

$$\begin{aligned}\frac{1}{n^{2+2d}}\sum_{t=1}^{\tau}\bar{y}_{t-1}^2 &= \frac{1}{n^{2+2d}}\left[\sum_{t=1}^{\tau}y_{t-1}^2-\frac{1}{\tau}\left(\sum_{t=1}^{\tau}y_{t-1}\right)^2\right]\\ &\Rightarrow \sigma_d^2\left(B_{r,d}-\frac{1}{r}A_{r,d}^2\right).\end{aligned} \tag{124}$$

Eventually, we have

$$\begin{aligned} n(\tilde{\rho}_\tau-\rho_n) &= \frac{\frac{1}{n^{1+2d}}\left(\sum_{t=1}^{\tau}\bar{y}_{t-1}\left(y_t-\rho_n y_{t-1}\right)+\frac{1}{2}\sum_{t=1}^{\tau}\Delta y_t^2\right)}{\frac{1}{n^{2+2d}}\sum_{t=1}^{\tau}\bar{y}_{t-1}^2}\\ &\Rightarrow \frac{\frac{1}{2}C_{r,d}-\frac{1}{r}A_{r,d}B^H(r)}{B_{r,d}-\frac{1}{r}A_{r,d}^2} := X_c(r,d)\ .\end{aligned}$$

For the sixth claim,

As $n(\tilde{\rho}_\tau-1) = n(\tilde{\rho}_\tau-\rho_n)+n(\rho_n-1) = n(\tilde{\rho}_\tau-\rho_n)+c$, we have

$$n(\hat{\rho}_\tau-1)\Rightarrow X_c(r,d)+c.$$

This completes the proof of Lemma 18.

Proof of Theorem 6

Recall that from (107), we have

$$\hat{\Omega}_{HAR} = \frac{2}{b}\frac{n}{\lfloor nr\rfloor}\frac{1}{n}\sum_{i=1}^{\tau-1}\left(\frac{1}{\sqrt{\tau}}S_i\right)^2-\frac{2}{b}\frac{n}{\lfloor nr\rfloor}\frac{1}{n}\sum_{i=1}^{\tau-\lfloor b\tau\rfloor-1}\left(\frac{1}{\sqrt{\tau}}S_i\right)\left(\frac{1}{\sqrt{\tau}}S_{i+\lfloor b\tau\rfloor}\right) \tag{125}$$

with $S_{\lfloor np\rfloor} = \sum_{i=1}^{\lfloor np\rfloor}\Delta y_i$.

For the partial sum $S_{\lfloor np\rfloor} = \sum_{i=1}^{\lfloor np\rfloor}\Delta y_i$, we can express

$$
\begin{aligned}
S_{\lfloor np \rfloor} &= \sum_{i=1}^{\tau_e - 1} \Delta y_i + \sum_{i=\tau_e}^{\lfloor np \rfloor} \Delta y_i \\
&= \sum_{i=1}^{\tau_e - 1} u_i + \frac{c}{n} \sum_{i=\tau_e}^{\lfloor np \rfloor} y_{i-1} + \sum_{i=\tau_e}^{\lfloor np \rfloor} u_i \\
&= \sum_{i=1}^{\lfloor np \rfloor} u_i + \frac{c}{n} \sum_{i=1}^{\lfloor np \rfloor} y_{i-1} - \frac{c}{n} \sum_{i=1}^{\tau_e - 1} y_{i-1}.
\end{aligned}
$$

Upon normalization, we have

$$
\begin{aligned}
& \frac{1}{n^{1/2+d}} \hat{S}_{\lfloor np \rfloor} \\
= & \frac{1}{n^{1/2+d}} \sum_{i=1}^{\lfloor np \rfloor} u_i + \frac{c}{n^{3/2+d}} \sum_{i=1}^{\lfloor np \rfloor} y_{i-1} - \frac{c}{n^{3/2+d}} \sum_{i=1}^{\tau_e - 1} y_{i-1} \\
\Rightarrow & \sigma_d \left(B^H(p) + cA_{p,d} - \int_0^{r_e} B^H(p) dp \right) := \sigma_d G_{r_e,c}(p). \qquad (126)
\end{aligned}
$$

So, combining (125) and (126), as $n \to \infty$, we have

$$
\frac{1}{n^{2d}} \hat{\Omega}_{HAR} \Rightarrow \frac{2\sigma_d^2}{br^2} \left(\int_0^r G_{r_e,c}(d,p)^2 dp - \int_0^{(1-b)r} G_{r_e,c}(d,p) G_{r_e,c}(d,p+br) dp \right).
$$

With (124),

$$
\begin{aligned}
\tau^2 s_{\tau,HAR}^2 &= \left(\frac{\tau}{n}\right)^2 \frac{\frac{1}{n^{2d}} \hat{\Omega}_{HAR}}{\frac{1}{n^{2+2d}} \left(\sum_{t=1}^{\tau} y_{t-1}^2 - \tau^{-1} \left(\sum_{t=1}^{\tau} y_{t-1} \right)^2 \right)} \\
&\Rightarrow \frac{\frac{2}{b} \left(\int_0^r G_{r_e,c}(d,p)^2 dp - \int_0^{(1-b)r} G_{r_e,c}(d,p) G_{r_e,c}(d,p+br) dp \right)}{B_{r,d} - \frac{1}{r} A_{r,d}^2}. \qquad (127)
\end{aligned}
$$

Eventually, for our test statistic $\widetilde{DF}_{\tau,HAR}$,

$$
\begin{aligned}
\widetilde{DF}_{\tau,HAR} &= \frac{\lfloor nr \rfloor}{nr} \frac{nr(\tilde{\rho}_\tau - 1)}{\left(\tau^2 s_{\tau,HAR}^2\right)^{1/2}} \\
&\Longrightarrow \frac{\frac{\left(\frac{1}{2} C_{r,d} - \frac{1}{r} A_{r,d} B^H(r)\right) r}{B_{r,d} - \frac{1}{r} A_{r,d}^2} + cr}{\sqrt{\frac{\frac{2}{b} \left(\int_0^r G_{r_e,c}(d,p)^2 dp - \int_0^{(1-b)r} G_{r_e,c}(d,p) G_{r_e,c}(d,p+br) dp \right)}{B_{r,d} - \frac{1}{r} A_{r,d}^2}}} \\
&= \frac{\left(\frac{\left(\frac{1}{2} C_{r,d} - \frac{1}{r} A_{r,d} B^H(r)\right) r}{B_{r,d} - \frac{1}{r} A_{r,d}^2} + cr \right) \left(B_{r,d} - \frac{1}{r} A_{r,d}^2 \right)^{1/2}}{\left[\frac{2}{b} \left(\int_0^r G_{r_e,c}(d,p)^2 dp - \int_0^{(1-b)r} G_{r_e,c}(d,p) G_{r_e,c}(d,p+br) dp \right) \right]^{1/2}},
\end{aligned}
$$

where we obtain the above limit from noticing $\frac{\lfloor nr \rfloor}{nr} \to 1$, applying Lemma18.6 and (127). This complete the proof of Theorem 6.

C.2.2 Proof of Theorem 7 and Theorem 8

Since the proofs of Theorem 7 and Theorem 8 are similar, we shall only prove the latter. It is useful to list the following lemmas (Lemma 19 to 27) which study various sample moments.

Lemma 19 *Let $B = [\tau_e, \tau_f]$ be the bubble period, $N_0 \in [1, \tau_e)$ and $N_1 = [\tau_f + 1, n]$ are the normal market period before and after the bubble period respectively. Under the dgp (60), with $t = \lfloor nr \rfloor$, we have the following asymptotic approximation :*

1. *For $t \in N_0$,*

$$y_t \overset{a}{\sim} n^{1/2+d_1} \sigma_{d_1} B^{H_1}(r_e).$$

2. *For $t \in B$,*

$$y_t \overset{a}{\sim} \rho_n^{(t-\tau_e)} n^{1/2+d_1} \sigma_{d_1} B^{H_1}(r_e).$$

3. *For $t \in N_1$,*

$$y_{\lfloor nr \rfloor} \overset{a}{\sim} \begin{cases} n^{1/2+d_1} \sigma_{d_1} B^{H_1}(r_e) & if \;\; d_1 > d_2, \\ n^{1/2+d_2} \sigma_{d_2} \left(B^{H_2}(r) - B^{H_2}(r_f) \right) & if \;\; d_1 < d_2, \\ n^{1/2+d_1} \left[\sigma_{d_2} \left(B^{H_2}(r) - B^{H_2}(r_f) \right) + \sigma_{d_1} B^{H_1}(r_e) \right] & if \;\; d_1 = d_2 \end{cases}$$

where $\sigma_{d_1} = \sqrt{\sigma^2 \frac{\Gamma(1-2d_1)}{\Gamma(d_1)\Gamma(1-d_1)}}$, $\sigma_{d_2} = \sqrt{\sigma^2 \frac{\Gamma(1-2d_2)}{\Gamma(d_2)\Gamma(1-d_2)}}$, $B^{H_1}(r)$ and $B^{H_2}(r)$ are fractional Brownian motion with Hurst parameter $H_1 = 1/2 + d_1$ and $H_2 = 1/2 + d_2$, respectively.

Proof.

1. From Lemma 16, we have $\frac{1}{n^{1/2+d_1}} y_{\lfloor nr \rfloor} \Longrightarrow \sigma_{d_1} B^{H_1}(r_e)$.

2. For $t \in B$, we have

$$y_t = \rho_n^{t-\tau_e+1} y_{\tau_e-1} + \sum_{j=0}^{t-\tau_e} \rho_n^j u_{t-j}$$

Pre-multiplying the both term by $\rho_n^{-(t-\tau_e)}$, we have

$$\rho_n^{-(t-\tau_e)} y_t = \rho_n y_{\tau_e-1} + \sum_{j=0}^{t-\tau_e} \rho_n^{-j} u_{t-j} \tag{128}$$

From Lemma 3 in Magdalinos (2012), the second term in (128) is of order $O_p(n^{(1/2+d_2)\alpha})$, and the first term is of order $O_p\left(\rho_n n^{1/2+d_1}\right)$. The order of $\rho_n^{-(t-\tau_e)} y_t$ depends on the ratio $\frac{n^{1/2+d_1}}{n^{(1/2+d_2)\alpha}}$.

As $\rho_n y_{\tau_e-1}$ is asymptotically dominant in (128) and $\rho_n \to 1$, we have

$$\rho_n^{-(t-\tau_e)} \frac{1}{n^{1/2+d_1}} y_t \overset{a}{\sim} \frac{\rho_n}{n^{1/2+d_1}} y_{\tau_e-1} \overset{a}{\sim} \sigma_{d_1} B^{H_1}(r_e).$$

3. For $t \in N_1$, we have

$$y_{\lfloor nr \rfloor} = \sum_{k=\tau_f+1}^{\lfloor nr \rfloor} u_k + y_{\tau_f}^* = \sum_{k=\tau_f+1}^{\lfloor nr \rfloor} u_k + y_{\tau_e} + y^*$$

$$y_{\lfloor nr \rfloor} = \sum_{k=\tau_f+1}^{\lfloor nr \rfloor} u_k + y_{\tau_f}^* = \sum_{k=1}^{\lfloor nr \rfloor} u_k - \sum_{k=1}^{\tau_f} u_k + y_{\tau_e} + y^*$$

Note that $y_{\tau_e} \overset{a}{\sim} n^{1/2+d_1} \sigma_{d_1} B^{H_1}(r_e)$, we need to compare the order of $\sum_{k=\tau_f+1}^{\lfloor nr \rfloor} u_k$ and y_{τ_e}.

Suppose that $d_1 = d_2$, we have

$$y_{\lfloor nr \rfloor} \overset{a}{\sim} n^{1/2+d_1} \left[\sigma_{d_2} \left(B^{H_2}(r) - B^{H_2}(r_f)\right) + \sigma_{d_1} B^{H_1}(r_e)\right].$$

Suppose that $d_1 > d_2$, we have

$$y_{\lfloor nr \rfloor} \overset{a}{\sim} n^{1/2+d_1} \sigma_{d_1} B^{H_1}(r_e).$$

Suppose that $d_1 < d_2$, we have

$$y_{\lfloor nr \rfloor} \overset{a}{\sim} n^{1/2+d_2} \sigma_{d_2} \left(B^{H_2}(r) - B^{H_2}(r_f)\right).$$

This completes the proof of Lemma 19.

Lemma 20 *For the sample average,*

1. *For* $\tau \in B$,
$$\frac{1}{\tau}\sum_{j=1}^{\tau} y_j \overset{a}{\sim} n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)$$

2. *For* $\tau \in N_1$,
$$\frac{1}{\tau}\sum_{j=1}^{\tau} y_j \overset{a}{\sim} n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e).$$

Proof. For $\tau \in B$, we have

$$\frac{1}{\tau}\sum_{j=1}^{\tau} y_j = \frac{1}{\tau}\sum_{j=1}^{\tau_e-1} y_j + \frac{1}{\tau}\sum_{j=\tau_e}^{\tau} y_j$$

The first term is

$$\begin{aligned}\frac{1}{\tau}\sum_{j=1}^{\tau_e-1} y_j &= n^{1/2+d_1}\frac{\tau_e}{\tau}\left(\frac{1}{\tau_e}\sum_{j=1}^{\tau_e-1}\frac{1}{n^{1/2+d_1}}y_j\right)\\ &\overset{a}{\sim} n^{1/2+d_1}\frac{r_e}{r}\sigma_{d_1}\int_0^{r_e} B^{H_1}(s)ds, \qquad (129)\end{aligned}$$

where have applied Lemma 19.1 and continuous mapping theorem to obtain (129).

For the second term,

$$\begin{aligned}&\frac{1}{\tau}\sum_{j=\tau_e}^{\tau} y_j\\ \overset{a}{\sim}\ & \frac{1}{\tau}\sum_{j=\tau_e}^{\tau}\rho_n^{(j-\tau_e)}n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\\ =\ & n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\frac{1}{\tau}\sum_{j=\tau_e}^{\tau}\rho_n^{j-\tau_e}\\ =\ & \frac{n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)}{\tau}\frac{\rho_n^{\tau-\tau_e+1}-1}{\rho_n-1}\\ =\ & \frac{n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\left[\left(\rho_n^{\tau-\tau_e}\rho_n\right)n^{\alpha}-n^{\alpha}\right]}{\lfloor nr\rfloor c}\\ \overset{a}{\sim}\ & n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e) \qquad (130)\end{aligned}$$

where we obtain the asymptotic equivalence because $\rho_n^{\tau-\tau_e} n^\alpha$ asymptotically dominates n^α.

Comparing (130) with (129), as the second term is of higher order in all three cases, therefore, we have the results in Lemma 20.1.

For $\tau \in N_1$,

$$\frac{1}{\tau}\sum_{j=1}^{\tau} y_j = \frac{1}{\tau}\sum_{j=1}^{\tau_e-1} y_j + \frac{1}{\tau}\sum_{j=\tau_e}^{\tau_f} y_j + \frac{1}{\tau}\sum_{j=\tau_f+1}^{\tau} y_j.$$

For the first term, similar to (129), we have

$$\frac{1}{\tau}\sum_{j=1}^{\tau_e-1} y_j \overset{a}{\sim} n^{1/2+d_1}\frac{r_e}{r}\sigma_{d_1}\int_0^{r_e} B^{H_1}(s)ds.$$

For the second term, similar to (130), we have

$$\frac{1}{\tau}\sum_{j=\tau_e}^{\tau_f} y_j \overset{a}{\sim} n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e).$$

For the last term, using Lemma 19.3, we have

$$\frac{1}{\tau}\sum_{j=\tau_f+1}^{\tau} y_j = \frac{\tau-\tau_f}{\tau}n^{1/2+d_1}O_p(1) \overset{a}{\sim} O_p(n^{1/2+d_1}).$$

Similar to the proof in Lemma 20, the second term has the highest order, therefore we obtain the result in Lemma 20.2.

Lemma 21 *Define the centered quantity* $\bar{y}_t = y_t - \frac{1}{\tau}\sum_{j=1}^{\tau} y_{j-1}$.

1. *For* $\tau \in B$,

   if $t \in N_0$
   $$\bar{y}_t \overset{a}{\sim} -n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e), \tag{131}$$

   if $t \in B$,
   $$\bar{y}_t \overset{a}{\sim} \left(\rho_n^{(t-\tau_e)} - \frac{n^\alpha}{nrc}\rho_n^{\tau-\tau_e}\right) n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e). \tag{132}$$

2. *For* $\tau \in N_1$,

   if $t \in N_0$,
   $$\bar{y}_t \overset{a}{\sim} -n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e), \tag{133}$$

$if\ t \in B$

$$\bar{y}_t \overset{a}{\sim} \left(\rho_n^{(t-\tau_e)} - \frac{n^{\alpha}}{nrc}\rho_n^{\tau-\tau_e}\right) n^{1/2+d_1}\sigma_{d_1} B^{H_1}(r_e), \tag{134}$$

$if\ t \in N_1$

$$\bar{y}_t \overset{a}{\sim} -n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1} B^{H_1}(r_e). \tag{135}$$

Proof.

1. Suppose $\tau \in B$.

$$\bar{y}_t = y_t - \frac{1}{\tau}\sum_{j=1}^{\tau} y_{j-1}$$

If $t \in N_0$, from Lemma 19.1, $y_t = O_p(n^{1/2+d_1})$, for the second term, $\frac{1}{\tau}\sum_{j=1}^{\tau} y_{j-1}$, following Lemma 20.1, we can obtain $\frac{1}{\tau}\sum_{j=1}^{\tau} y_{j-1} \overset{a}{\sim} n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1} B^{H_1}(r_e)$.

Therefore if $t \in N_0$, the second term has a higher order and we obtain

$$\bar{y}_t \overset{a}{\sim} -n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1} B^{H_1}(r_e).$$

If $t \in B$, from Lemma 19.2,

$$\bar{y}_t \overset{a}{\sim} \rho_n^{(t-\tau_e)} n^{1/2+d_1}\sigma_d B_u^H(r_e) - n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1} B^{H_1}(r_e).$$

2. Suppose that $\tau \in N_1$.

If $t \in N_0$,

$$\bar{y}_t = y_t - \frac{1}{\tau}\sum_{j=1}^{\tau} y_{j-1}$$

Similar to the proof in Lemma 21.1, as y_t is asymptotically dominated by the latter term, we can express

$$\bar{y}_t \overset{a}{\sim} -n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1} B^{H_1}(r_e).$$

If $t \in B$,

$$\bar{y}_t \overset{a}{\sim} \left(\rho_n^{(t-\tau_e)} - \rho_n^{\tau_f-\tau_e}\frac{n^\alpha}{nrc}\right) n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e).$$

If $t \in N_1$, components in y_t will be dominated by the components in $\frac{1}{\tau}\sum_{j=1}^{\tau} y_{j-1}$, eventually, following the proof of Lemma 21.1, we have

$$\bar{y}_t \overset{a}{\sim} -n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e).$$

Lemma 22 *The sample variance terms involving $\bar{y}_t$ behave as follows.*

1. *If $\tau \in B$,*

$$\sum_{j=1}^{\tau} \bar{y}_{j-1}^2 \overset{a}{\sim} n^{1+2d_1+\alpha}\frac{\rho_n^{2(\tau-\tau_e)}}{2c}\sigma_{d_1}^2 B^{H_1}(r_e)^2.$$

2. *if $\tau \in N_1$,*

$$\sum_{j=1}^{\tau} \bar{y}_{j-1}^2 \overset{a}{\sim} n^{1+\alpha+2d_1}\frac{\rho_n^{2(\tau_f-\tau_e)}}{2c}\sigma_{d_1}^2 B^{H_1}(r_e)^2.$$

Proof.

1. For $\tau \in B$

$$\sum_{j=1}^{\tau} \bar{y}_{j-1}^2 = \sum_{j=1}^{\tau_e-1} \bar{y}_{j-1}^2 + \sum_{j=\tau_e}^{\tau} \bar{y}_{j-1}^2. \tag{136}$$

For the first term in (136),

$$\begin{aligned}\sum_{j=1}^{\tau_e-1} \bar{y}_{j-1}^2 &\overset{a}{\sim} \sum_{j=1}^{\tau_e-1}\left(-n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\right)^2 \\ &= \frac{(\tau_e-1)}{n}n^{2(\alpha+d_1)}\rho_n^{2(\tau-\tau_e)}\frac{1}{r^2c^2}\sigma_{d_1}^2 B^{H_1}(r_e)^2 \\ &\overset{a}{\sim} \frac{r_e}{r^2c^2}n^{2(\alpha+d_1)}\rho_n^{2(\tau-\tau_e)}\sigma_{d_1}^2 B^{H_1}(r_e)^2.\end{aligned}$$

For the second term in (136),

$$
\begin{aligned}
\sum_{j=\tau_e}^{\tau} \bar{y}_{j-1}^2 &\overset{a}{\sim} \sum_{j=\tau_e}^{\tau} \left[\left(\rho_n^{(j-\tau_e)} - \frac{n^\alpha}{nrc} \rho_n^{\tau-\tau_e} \right) n^{1/2+d_1} \sigma_{d_1} B^{H_1}(r_e) \right]^2 \\
&= n^{1+2d_1} \sigma_{d_1}^2 B^{H_1}(r_e)^2 \sum_{j=\tau_e}^{\tau} \left(\rho_n^{(j-\tau_e)} - \frac{n^\alpha}{nrc} \rho_n^{\tau-\tau_e} \right)^2 \\
&= n^{1+2d_1} \sigma_{d_1}^2 B^{H_1}(r_e)^2 \sum_{j=\tau_e}^{\tau} \left(\rho_n^{2(j-\tau_e)} - 2\rho_n^{(j-\tau_e)} \frac{n^\alpha}{nrc} \rho_n^{\tau-\tau_e} + \frac{n^{2\alpha}}{n^2 r^2 c^2} \rho_n^{2(\tau-\tau_e)} \right) \\
&= n^{1+2d_1} \sigma_{d_1}^2 B^{H_1}(r_e)^2 \left[\frac{n^\alpha \rho_n^{2(\tau-\tau_e)}}{2c} - 2 \frac{n^{2\alpha-1} \rho_n^{2(\tau-\tau_e)}}{nrc} + \frac{r - r_e + 1/n}{r^2 c^2} n^{2\alpha-1} \rho_n^{2(\tau-\tau_e)} \right] \\
&\overset{a}{\sim} n^{1+2d_1+\alpha} \sigma_{d_1}^2 B^{H_1}(r_e)^2 \frac{\rho_n^{2(\tau-\tau_e)}}{2c} \text{ (since } \alpha > 2\alpha - 1).
\end{aligned}
$$

Since $1 + 2d_1 + \alpha > 2(\alpha + d_1)$, $\sum_{j=\tau_e}^{\tau} \bar{y}_{j-1}^2$ dominates $\sum_{j=1}^{\tau_e-1} \bar{y}_{j-1}^2$ asymptotically, and we have

$$
\sum_{j=1}^{\tau} \bar{y}_{j-1}^2 \overset{a}{\sim} n^{1+2d_1+\alpha} \sigma_{d_1}^2 B^{H_1}(r_e)^2 \frac{\rho_n^{2(\tau-\tau_e)}}{2c}.
$$

2. For $\tau_1 \in N_0$ and $\tau_2 \in N_1$

$$
\sum_{j=\tau_1}^{\tau_2} \bar{y}_{j-1}^2 = \sum_{j=\tau_1}^{\tau_e-1} \bar{y}_{j-1}^2 + \sum_{j=\tau_e}^{\tau_f} \bar{y}_{j-1}^2 + \sum_{j=\tau_f+1}^{\tau} \bar{y}_{j-1}^2. \tag{137}
$$

For the first term in (137), we have

$$
\sum_{j=1}^{\tau_e-1} \bar{y}_{j-1}^2 \overset{a}{\sim} \frac{r_e}{r^2 c^2} n^{2(\alpha+d_1)} \rho_n^{2\left(\tau_f-\tau_e\right)} \sigma_{d_1}^2 B^{H_1}(r_e)^2.
$$

For the second term, we have

$$
\sum_{j=\tau_e}^{\tau_f} \bar{y}_{j-1}^2 \overset{a}{\sim} n^{1+\alpha+2d_1} \sigma_{d_1}^2 B^{H_1}(r_e)^2 \frac{\rho_n^{2(\tau_f-\tau_e)}}{2c}.
$$

For the third term,

$$\sum_{j=\tau_f+1}^{\tau} \bar{y}_{j-1}^2 \overset{a}{\sim} \sum_{j=\tau_f+1}^{\tau} \left(-n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\right)^2$$
$$= \frac{\tau-\tau_f}{n}n^{2(\alpha+d_1)}\rho_n^{2\left(\tau_f-\tau_e\right)}\frac{1}{r^2c^2}\sigma_{d_1}^2 B^{H_1}(r_e)^2$$
$$\overset{a}{\sim} \frac{(r-r_f)}{r^2c^2}n^{2(\alpha+d_1)}\rho_n^{2\left(\tau_f-\tau_e\right)}\sigma_{d_1}^2 B^{H_1}(r_e)^2.$$

As $1+\alpha+2d_1 > 2(\alpha+d_1)$, the middle term dominates, and we have

$$\sum_{j=1}^{\tau} \bar{y}_{j-1}^2 \overset{a}{\sim} n^{1+\alpha+2d_1}\sigma_{d_1}^2 B^{H_1}(r_e)^2\frac{\rho_n^{2(\tau_f-\tau_e)}}{2c}.$$

This ends the proof of Lemma 22.

Lemma 23 *The sample variances of $\bar{y}_t$ and u_t behave as follows:*

1. *For $\tau \in B$,*
$$\sum_{j=1}^{\tau} \bar{y}_{j-1}u_j \overset{a}{\sim} \rho_n^{\tau-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}(r-r_e)^{(1/2+d_2)\alpha}\sigma_{d_1}B^{H_1}(r_e)\Phi. \tag{138}$$

2. *For $\tau \in N_1$,*
$$\sum_{j=1}^{\tau} \bar{y}_{j-1}u_j \overset{a}{\sim} \rho_n^{\tau-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}(r_f-r_e)^{(1/2+d_2)\alpha}\sigma_{d_1}B^{H_1}(r_e)\Phi, \tag{139}$$

where Φ is a normally distribution random variable with a finite variance.

1. For $\tau \in B$,
$$\sum_{j=1}^{\tau_2} \bar{y}_{j-1}u_j = \sum_{j=1}^{\tau_e-1} \bar{y}_{j-1}u_j + \sum_{j=\tau_e}^{\tau} \bar{y}_{j-1}u_j. \tag{140}$$

The first term in (140) is

$$\sum_{j=1}^{\tau_e-1} \bar{y}_{j-1}u_j \overset{a}{\sim} \sum_{j=1}^{\tau_e-1}\left(-n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\right)u_j$$
$$= \left(-n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\right)\sum_{j=1}^{\tau_e-1}u_j$$
$$= \frac{-n^{\alpha+d_1-1/2+d_1+1/2}}{rc}\rho_n^{\tau-\tau_e}\sigma_{d_1}B^{H_1}(r_e)\frac{1}{n^{1/2+d_1}}\sum_{j=1}^{\tau_e-1}u_j$$
$$\overset{a}{\sim} -\frac{n^{\alpha+2d_1}}{rc}\rho_n^{\tau-\tau_e}\sigma_{d_1}^2 B^{H_1}(r_e)^2. \tag{141}$$

For the second term in (140),

$$
\begin{aligned}
&\sum_{j=\tau_e}^{\tau} \bar{y}_{j-1}u_j \\
\overset{a}{\sim}\ & \sum_{j=\tau_e}^{\tau} \left[\left(\rho_n^{(j-\tau_e)} - \rho_n^{\tau-\tau_e}\frac{n^\alpha}{nrc}\right) n^{1/2+d_1}\sigma_d B_1^H(r_e)\right] u_j \\
=\ & n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\left[\rho_n^{\tau-\tau_e}\sum_{j=\tau_e}^{\tau}\rho_n^{-(\tau-j+1)}u_j - \rho_n^{\tau-\tau_e}\frac{n^\alpha}{nrc}\sum_{j=\tau_e}^{\tau}u_j\right] \\
=\ & n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\left[\begin{array}{c}\rho_n^{\tau-\tau_e}\frac{(\tau-\tau_e+1)^{(1/2+d_2)\alpha}}{(\tau-\tau_e+1)^{(1/2+d_2)\alpha}}\sum_{j=\tau_e}^{\tau}\rho_n^{-(\tau-j+1)}u_j \\ -\rho_n^{\tau-\tau_e}\frac{n^{\alpha-1/2+d_2}}{rc}\frac{1}{n^{1/2+d_2}}\sum_{j=\tau_e}^{\tau}u_j\end{array}\right] \\
\overset{a}{\sim}\ & n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\left[\rho_n^{\tau-\tau_e}\left[n\left(r-r_e\right)\right]^{(1/2+d_2)\alpha}\Phi - \rho_n^{\tau-\tau_e}\frac{n^{\alpha-1/2+d_2}}{rc}\sigma_{d_2}\left(B^{H_2}(r)-B^{H_2}(r_e)\right)\right] \\
\overset{a}{\sim}\ & n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\rho_n^{\tau-\tau_e}\left[n\left(r-r_e\right)\right]^{(1/2+d_2)\alpha}\Phi. \\
=\ & \rho_n^{\tau-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}\left(r-r_e\right)^{(1/2+d_2)\alpha}\sigma_{d_1}B^{H_1}(r_e)\Phi.
\end{aligned} \tag{142}
$$

We have applied Lemma 3 in Magdalinos(2012) to acquire the second asymptotic equivalence and since $(1/2+d_2)\alpha-(\alpha-1/2+d_2)=(1/2-d_2)(1-\alpha)>0$, we have the last asymptotic equivalence.

The assumption that $\frac{1/2-d_1}{1/2-d_2}>\alpha$ implies $\frac{1}{2}(1+\alpha)+d_1+d_2\alpha>\alpha+2d_1$, so the $\sum_{j=\tau_e}^{\tau}\bar{y}_{j-1}u_j$ asymptotically dominates $\sum_{j=1}^{\tau_e-1}\bar{y}_{j-1}u_j$, eventually, we have

$$\sum_{j=1}^{\tau}\bar{y}_{j-1}u_j \overset{a}{\sim} \rho_n^{\tau-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}\left(r-r_e\right)^{(1/2+d_2)\alpha}\sigma_{d_1}B_1^H(r_e)\Phi.$$

2. For $\tau_1\in N_0$ and $\tau_2\in N_1$,

$$\sum_{j=\tau_1}^{\tau_2}\bar{y}_{j-1}u_j = \sum_{j=\tau_1}^{\tau_e-1}\bar{y}_{j-1}u_j + \sum_{j=\tau_d}^{\tau_f}\bar{y}_{j-1}u_j + \sum_{j=\tau_f+1}^{\tau_2}\bar{y}_{j-1}u_j.$$

As in (141), the first term is

$$\sum_{j=\tau_1}^{\tau_e-1}\bar{y}_{j-1}u_j \overset{a}{\sim} -\frac{n^{\alpha+2d_1}}{rc}\rho_n^{\tau-\tau_e}\sigma_{d_1}^2B^{H_1}(r_e)^2.$$

As in (142), the second term is

$$\sum_{j=\tau_e}^{\tau_f}\bar{y}_{j-1}u_j \overset{a}{\sim} \rho_n^{\tau_f-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}\left(r-r_e\right)^{(1/2+d_2)\alpha}\sigma_{d_1}B^{H_1}(r_e)\Phi.$$

The third term is

$$\begin{aligned}
\sum_{j=\tau_f+1}^{\tau} \bar{y}_{j-1} u_j &\overset{a}{\sim} \sum_{j=\tau_f+1}^{\tau} \left(-n^{\alpha+d_1-1/2} \rho_n^{\tau-\tau_e} \frac{1}{rc} \sigma_{d_1} B^{H_1}(r_e) \right) u_j \\
&= -n^{\alpha+d_1-1/2} \rho_n^{\tau-\tau_e} \frac{1}{rc} \sigma_{d_1} B^{H_1}(r_e) \sum_{j=\tau_f+1}^{\tau} u_j \\
&= -n^{\alpha+d_1-1/2} \rho_n^{\tau-\tau_e} \frac{1}{rc} \sigma_d B^{H_1}(r_e) n^{1/2+d_1} \left(B^{H_1}(r) - B^{H_1}(r_f) \right) \\
&\overset{a}{\sim} -n^{\alpha+2d_1} \rho_n^{\tau-\tau_e} \frac{1}{rc} \sigma_d B^{H_1}(r_e) \left(B^{H_1}(r) - B^{H_1}(r_f) \right).
\end{aligned}$$

Note that $\frac{1/2-d_1}{1/2-d_2} > \alpha$ implies $\frac{1}{2} + \frac{1}{2}\alpha + d_1 + d_2\alpha > \alpha + 2d_1$. So the second term dominates the first and third term and we have

$$\sum_{j=\tau_f+1}^{\tau_2} \bar{y}_{j-1} u_j \overset{a}{\sim} \rho_n^{\tau-\tau_e} n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha} (r - r_e)^{(1/2+d_2)\alpha} \sigma_{d_1} B^{H_1}(r_e) \Phi.$$

So we have completed the proof of Lemma 23.

Lemma 24 *The sample covariances of $\bar{y}_{j-1}$ and $y_j - \rho_n y_{j-1}$ behave as follows:*

1. *For $\tau \in B$,*

$$\sum_{j=1}^{\tau} \bar{y}_{j-1}(y_j - \rho_n y_{j-1}) \overset{a}{\sim} \rho_n^{\tau-\tau_e} n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha} (r - r_e)^{(1/2+d_2)\alpha} \sigma_{d_1} B^{H_1}(r_e) \Phi.$$

2. *For $\tau \in N_1$,*

$$\sum_{j=1}^{\tau} \bar{y}_{j-1}(y_j - \rho_n y_{j-1}) \overset{a}{\sim} -\rho_n^{2\left(\tau_f-\tau_e\right)} n^{1+2d_1} \sigma_d^2 \left(B^{H_1}(r_e) \right)^2.$$

Proof.

1. Note that we can separate $\sum_{j=1}^{\tau} \bar{y}_{j-1}(y_j - \rho_n y_{j-1})$ into two parts such that

$$\begin{aligned}
\sum_{j=1}^{\tau} \bar{y}_{j-1}(y_j - \rho_n y_{j-1}) &= \sum_{j=1}^{\tau_e-1} \bar{y}_{j-1}(y_j - \rho_n y_{j-1}) + \sum_{j=\tau_e}^{\tau} \bar{y}_{j-1}(y_j - \rho_n y_{j-1}) \\
&= \sum_{j=1}^{\tau_e-1} \bar{y}_{j-1} \left[(1-\rho_n) y_{j-1} + u_t \right] + \sum_{j=\tau_e}^{\tau} \bar{y}_{j-1} u_t \\
&= \sum_{j=1}^{\tau} \bar{y}_{j-1} u_t - \frac{c}{n^\alpha} \sum_{j=1}^{\tau_e-1} \bar{y}_{j-1} y_{j-1}.
\end{aligned}$$

From (138), $\sum_{j=1}^{\tau}\bar{y}_{j-1}u_t \overset{a}{\sim} \rho_n^{\tau-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}(r-r_e)^{(1/2+d_2)\alpha}\sigma_{d_1}B^{H_1}(r_e)\Phi$, for the second term, applying (131) gives us

$$
\begin{aligned}
\frac{c}{n^\alpha}\sum_{j=1}^{\tau_e-1}\bar{y}_{j-1}y_{j-1} &\overset{a}{\sim} \frac{c}{n^\alpha}\sum_{j=1}^{\tau_e-1}\left(-n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\right)y_{j-1} \\
&= \left(-n^{d_1-1/2}\rho_n^{\tau-\tau_e}\frac{1}{r}\sigma_{d_1}B^{H_1}(r_e)\right)\tau_e\left(\frac{1}{\tau_e}\sum_{j=1}^{\tau_e-1}y_{j-1}\right) \\
&\overset{a}{\sim} \left(-n^{d_1-1/2}\frac{1}{r}\rho_n^{\tau-\tau_e}\sigma_{d_1}B^{H_1}(r_e)\right)\tau_e\left(n^{1/2+d_1}\sigma_{d_1}\int_0^{r_e}B^{H_1}(s)ds\right) \\
&= -n^{2d_1+1}\rho_n^{\tau-\tau_e}\frac{r_e}{r}\sigma_{d_1}^2B^{H_1}(r_e)\int_0^{r_e}B^{H_1}(s)ds. \qquad (143)
\end{aligned}
$$

Under the assumption $\alpha < \frac{1/2+d_1}{1/2+d_2}$, we can show $\frac{1}{2}(1+\alpha)+d_1+d_2\alpha > 2d_1+1$ thus $\sum_{j=1}^{\tau}\bar{y}_{j-1}u_t$ asymptotically dominates $-\frac{c}{n^\alpha}\sum_{j=1}^{\tau_e-1}\bar{y}_{j-1}y_{j-1}$, and we have

$$
\sum_{j=1}^{\tau}\bar{y}_{j-1}(y_j-\rho_n y_{j-1}) \overset{a}{\sim} \rho_n^{\tau-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}(r-r_e)^{(1/2+d_2)\alpha}\sigma_{d_1}B^{H_1}(r_e)\Phi.
$$

2. For $\tau \in N_1$, we can express

$$
\begin{aligned}
&\sum_{j=1}^{\tau}\bar{y}_{j-1}(y_j-\rho_n y_{j-1}) \\
=\;& \sum_{j=1}^{\tau_e-1}\bar{y}_{j-1}(y_j-\rho_n y_{j-1}) + \sum_{j=\tau_e}^{\tau_f}\bar{y}_{j-1}(y_j-\rho_n y_{j-1}) + \bar{y}_{\tau_f}(y_{\tau_f}-\rho_n y_{\tau_f}) \\
&+ \sum_{j=\tau_f+2}^{\tau}\bar{y}_{j-1}(y_j-\rho_n y_{j-1}) \\
=\;& \sum_{j=1}^{\tau_e-1}\bar{y}_{j-1}\left[-\frac{c}{n^\alpha}y_{j-1}+u_t\right] + \sum_{j=\tau_e}^{\tau_f}\bar{y}_{j-1}u_t + \bar{y}_{\tau_f}(y_{\tau_f+1}-\rho_n y_{\tau_f}) \\
&+ \sum_{j=\tau_f+2}^{\tau}\bar{y}_{j-1}\left[-\frac{c}{n^\alpha}y_{j-1}+u_t\right] \\
=\;& \sum_{j=1}^{\tau}\bar{y}_{j-1}u_t - \frac{c}{n^\alpha}\sum_{j=1}^{\tau_e-1}\bar{y}_{j-1}y_{j-1} - \frac{c}{n^\alpha}\sum_{j=\tau_f+2}^{\tau}\bar{y}_{j-1}y_{j-1} - \rho_n\bar{y}_{\tau_f}y_{\tau_f}.
\end{aligned}
$$

For the first term, similar to (138), we have

$$
\sum_{j=1}^{\tau}\bar{y}_{j-1}u_t \overset{a}{\sim} \rho_n^{\tau-\tau_e}n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}(r_f-r_e)^{(1/2+d_2)\alpha}\sigma_{d_1}B^{H_1}(r_e)\Phi,
$$

for the second term, following the step to obtain (143), we have

$$\frac{c}{n^{\alpha}}\sum_{j=1}^{\tau_e-1}\bar{y}_{j-1}y_{j-1} \overset{a}{\sim} -n^{2d_1+1}\rho_n^{\tau_f-\tau_e}\frac{r_e}{r}\sigma_{d_1}^2 B^{H_1}\int_0^{r_e}B^{H_1}(s)ds,$$

for the third term,

$$\begin{aligned}
\frac{c}{n^{\alpha}}\sum_{j=\tau_f+2}^{\tau}\bar{y}_{j-1}y_{j-1} &\overset{a}{\sim} \frac{c}{n^{\alpha}}\sum_{j=\tau_f+2}^{\tau}\left(-n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\right)y_{j-1}\\
&= -n^{d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\sum_{j=\tau_f+2}^{\tau}y_{j-1}\\
&= -n^{d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\left(\tau-\tau_f-1\right)\frac{n^{1/2+d_1}}{\tau-\tau_f-1}\sum_{j=\tau_f+2}^{\tau}\frac{1}{n^{1/2+d_1}}y_{j-1}\\
&\overset{a}{\sim} -n^{2d_1+1}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}^2 B^{H_1}(r_e)^2\left(r-r_f\right)\int_{r_f}^{r}B^{H_1}(s)ds.
\end{aligned}$$

For the last term, from Lemma 19.2 and (134)

$$\begin{aligned}
\rho_n\bar{y}_{\tau_f}y_{\tau_f} &\overset{a}{\sim} \bar{y}_{\tau_f}y_{\tau_f}\ (\text{as } \rho_n\to 1)\\
&\overset{a}{\sim} \left(\rho_n^{\tau_f-\tau_e}-\rho_n^{\tau_f-\tau_e}\frac{n^{\alpha}}{nrc}\right)n^{1/2+d_1}\sigma_{d_1}B^{H_1}\left(\rho_n^{\tau_f-\tau_e}n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\right)\\
&= \rho_n^{2\left(\tau_f-\tau_e\right)}n^{1+2d_1}\sigma_{d_1}^2\left(B^{H_1}(r_e)\right)^2.
\end{aligned}$$

Note that the last component dominates the previous terms as $\rho_n^{2\left(\tau_f-\tau_e\right)}$ is overwhelming. Finally, we have

$$\sum_{j=1}^{\tau}\bar{y}_{j-1}(y_j-\rho_n y_{j-1}) \overset{a}{\sim} -\rho_n^{2\left(\tau_f-\tau_e\right)}n^{1+2d_1}\sigma_{d_1}^2\left(B^{H_1}(r_e)\right)^2.$$

This ends the proof of Lemma 24.

Lemma 25 *For the sum of squared difference $\sum_{t=1}^{\tau}\Delta y_t^2$, we have the following limit results:*

1. *When $\tau\in B$,*

$$\sum_{t=1}^{\tau}\Delta y_t^2 = O_p(n^{1+2d_1-\alpha}\rho_n^{2(\tau_f-\tau_e)}). \tag{144}$$

2. *When* $\tau \in N_1$,

$$\sum_{t=1}^{\tau} \Delta y_t^2 = O_p(n^{1+2d_1} \rho_n^{2(\tau_f - \tau_e)}). \tag{145}$$

Proof.

For Δy_t, note that it has a different expression for different period,

$$\Delta y_t = \begin{cases} u_t & if \;\; t < \tau_e, \\ (\rho_n - 1) y_{t-1} + u_t & if \;\; \tau_e \leqslant t \leqslant \tau_f, \\ y_{\tau_e} + y^* + u_{\tau_f+1} - y_{\tau_f} & if \;\; t = \tau_f + 1, \\ u_t & if \;\; t > \tau_f + 1. \end{cases} . \tag{146}$$

Note that,

$$\sum_{t=1}^{\tau} \Delta y_t^2 = \sum_{t=1}^{\tau_e - 1} \Delta y_t^2 + \sum_{t=\tau_e}^{\tau_f} \Delta y_t^2 + \Delta y_{\tau_f+1}^2 + \sum_{t=\tau_f+2}^{\tau} \Delta y_t^2. \tag{147}$$

And using (146), we can write

$$\begin{aligned} & \sum_{t=1}^{\tau} \Delta y_t^2 \\ = & \sum_{t=1}^{\tau_e-1} u_t^2 + \sum_{t=\tau_e}^{\tau_f} ((\rho_n - 1) y_{t-1} + u_t)^2 + \left(y_{\tau_e} + y^* + u_{\tau_f+1} - y_{\tau_f}\right)^2 + \sum_{t=\tau_f+2}^{\tau} u_t^2 \\ = & \sum_{t=1}^{\tau_e-1} u_t^2 + \frac{c^2}{n^{2\alpha}} \sum_{t=\tau_e}^{\tau_f} y_{t-1}^2 + \frac{2c}{n^{\alpha}} \sum_{t=\tau_e}^{\tau_f} y_{t-1} u_t + \sum_{t=\tau_e}^{\tau_f} u_t^2 \\ & + \left(y_{\tau_e} + y^* - y_{\tau_f}\right)^2 + 2\left(y_{\tau_e} + y^* - y_{\tau_f}\right) u_{\tau_f+1} + u_{\tau_f+1}^2 + \sum_{t=\tau_f+2}^{\tau} u_t^2 \\ = & \sum_{t=1}^{\tau} u_t^2 + \frac{c^2}{n^{2\alpha}} \sum_{t=\tau_e}^{\tau_f} y_{t-1}^2 + \frac{2c}{n^{\alpha}} \sum_{t=\tau_e}^{\tau_f} y_{t-1} u_t + \left(y_{\tau_e} + y^* - y_{\tau_f}\right)^2 + 2\left(y_{\tau_e} + y^* - y_{\tau_f}\right) u_{\tau_f+1} \end{aligned} \tag{148}$$

We now proceed to compare the stochastic order of different component in (148). By ergodic theorem $\sum_{t=1}^{\tau} u_t^2 = O_p(n)$. For the second term, applying Lemma 20, we have

$$\frac{c^2}{n^{2\alpha}}\sum_{t=\tau_e}^{\tau_f} y_{t-1}^2 \overset{a}{\sim} \frac{c^2}{n^{2\alpha}}\left(n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\right)^2\sum_{t=\tau_e}^{\tau_f}\rho_n^{2(t-\tau_e)} \tag{149}$$

$$= \frac{c^2}{n^{2\alpha}}\left(n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)\right)^2\frac{n^{\alpha}\rho_n^{2(\tau_f-\tau_e)}}{2c}$$

$$= \frac{c}{2}n^{1+2d_1-\alpha}\rho_n^{2(\tau_f-\tau_e)}\sigma_{d_1}^2 B^{H_1}(r_e)^2$$

$$= O_p(n^{1+2d_1-\alpha}\rho_n^{2(\tau_f-\tau_e)}). \tag{150}$$

Suppose $\tau\in B$, we do not have the term in (147), and (150) yields (144).

For the third term, note that

$$y_{\tau_e}+y^*-y_{\tau_f} \overset{a}{\sim} n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)+O_p(1)-\rho_n^{(\tau_f-\tau_e)}n^{1/2+d_1}\sigma_{d_1}B^{H_1}(r_e)$$

$$= O_p(\rho_n^{(\tau_f-\tau_e)}n^{1/2+d_1}).$$

It implies $\left(y_{\tau_e}+y^*-y_{\tau_f}\right)^2 = O_p(\rho_n^{2(\tau_f-\tau_e)}n^{1+2d_1})$.

Finally,

$$2\left(y_{\tau_e}+y^*-y_{\tau_f}\right)u_{\tau_f+1} = O_p(\rho_n^{(\tau_f-\tau_e)}n^{1/2+d_1})\times O_p(1)$$

$$= O_p(\rho_n^{(\tau_f-\tau_e)}n^{1/2+d_1}).$$

Therefore, the third term dominates the other term as $n\to\infty$, and we have (145). This completes the proof.

Lemma 26 *For the LS estimator $\hat{\rho}_\tau$, we have the following asymptotic results:*

1. *When $\tau\in B$,*

$$n\left(\tilde{\rho}_\tau-1\right)\xrightarrow{p}\infty.$$

2. *When $\tau\in N_1$,*

$$n\left(\tilde{\rho}_\tau-1\right)\xrightarrow{p}-\infty.$$

Proof.

By our definition, $\tilde{\rho}_\tau=\hat{\rho}_\tau+\frac{\frac{1}{2}\sum_{j=1}^{\tau}\Delta y_t^2}{\sum_{j=1}^{\tau}\bar{y}_{j-1}^2}$, while from Lemma 22 and Lemma 25, it is clear that $\frac{1}{2}\sum_{j=1}^{\tau}\Delta y_t^2$ is at most $O_p(n^{1+2d_1}\rho_n^{2(\tau_f-\tau_e)})$, and $\sum_{j=1}^{\tau}\bar{y}_{j-1}^2=O_p(n^{1+2d_1+\alpha}\rho_n^{2(\tau_f-\tau_e)})$

for $\tau \in B \cup N_1$. So the term $\frac{\frac{1}{2}\sum_{j=1}^{\tau}\Delta y_t^2}{\sum_{j=1}^{\tau}\bar{y}_{j-1}^2} = o_p(1)$, it means we only need to study the asymptotic properties of $\hat{\rho}_\tau$.

We first focus on the centered statistics $\hat{\rho}_\tau - \rho_n = \frac{\sum_{j=1}^{\tau}\bar{y}_{j-1}(y_j - \rho_n y_{j-1})}{\sum_{j=1}^{\tau}\bar{y}_{j-1}^2}$.

1. When $\tau \in B$,

$$\begin{aligned}\frac{\sum_{j=1}^{\tau}\bar{y}_{j-1}(y_j - \rho_n y_{j-1})}{\sum_{j=1}^{\tau}\bar{y}_{j-1}^2} &\overset{a}{\sim} \frac{\rho_n^{\tau-\tau_e} n^{\frac{1}{2}(1+\alpha)+d_1+d_2\alpha}\left(r-r_e\right)^{(1/2+d_2)\alpha}\sigma_{d_1} B^{H_1}(r_e)\Phi}{n^{1+2d_1+\alpha}\frac{\rho_n^{2(\tau-\tau_e)}}{2c}\sigma_{d_1}^2 B^{H_1}(r_e)^2} \\ &= \frac{2c\left(r-r_e\right)^{(1/2+d_2)\alpha}\Phi}{\rho_n^{\tau-\tau_e} n^{1/2+d_1+(1/2-d_2)\alpha}\sigma_{d_1} B^{H_1}(r_e)} \\ &= O_p(n^{-\alpha}\rho_n^{-(\tau-\tau_e)}), \end{aligned} \tag{151}$$

where we have applied Lemma 22.1 and 24.1 to obtain the asymptotic equivalence. Note that 151 also implies $\tilde{\rho}_\tau - \rho_n = O_p(n^{-\alpha}\rho_n^{-(\tau-\tau_e)})$.

As $n(\tilde{\rho}_\tau - 1) = n(\rho_n - 1) + n(\tilde{\rho}_\tau - \rho_n)$,

$$\begin{aligned} n(\rho_n - 1) + n(\tilde{\rho}_\tau - \rho_n) &= n^{1-\alpha}c + O_p(n^{1-\alpha}\rho_n^{-(\tau-\tau_e)}) \\ &= n^{1-\alpha}c + o_p(1) \to \infty. \end{aligned} \tag{152}$$

2. When $\tau \in N_1$,

$$\begin{aligned}\frac{\sum_{j=1}^{\tau}\bar{y}_{j-1}(y_j - \rho_n y_{j-1})}{\sum_{j=1}^{\tau}\bar{y}_{j-1}^2} &\overset{a}{\sim} \frac{-\rho_n^{2\left(\tau_f-\tau_e\right)} n^{1+2d_1}\sigma_{d_1}^2\left(B^{H_1}(r_e)\right)^2}{n^{1+\alpha+2d_1}\frac{\rho_n^{2(\tau_f-\tau_e)}}{2c}\sigma_{d_1}^2 B^{H_1}(r_e)^2} \\ &= -n^{-\alpha}2c. \end{aligned} \tag{153}$$

Similarly, (153) also gives the order of $\tilde{\rho}_\tau - \rho_n$.

As $n(\tilde{\rho}_\tau - 1) = n(\rho_n - 1) + n(\tilde{\rho}_\tau - \rho_n)$,

$$\begin{aligned} n(\rho_n - 1) + n(\tilde{\rho}_\tau - \rho_n) &= n^{1-\alpha}c - n\left(n^{-\alpha}2c\right) \\ &= -n^{1-\alpha}c \to \infty. \end{aligned} \tag{154}$$

This completes the proof of Lemma 26.

Lemma 27 *Under model (60), we have the following asymptotic result.*

$$\hat{\Omega}_{HAR} = O_p(n^{2d_1}\rho_n^{2(\tau-\tau_e)}).$$

Proof. From (106), we have

$$\hat{\Omega}_{HAR} = \frac{1}{\tau}\sum_{i=1}^{\tau-1}\frac{1}{\tau}\sum_{j=1}^{\tau-1}\tau^2 D_\tau\left(\frac{i-j}{\tau}\right)\frac{1}{\sqrt{\tau}}\hat{S}_i\frac{1}{\sqrt{\tau}}\hat{S}_j. \tag{155}$$

To study the order of Ω_{HAR}, we only need to study the limit of $\frac{1}{\sqrt{\tau}}\hat{S}_i$.

Suppose $\tau \in B$,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\hat{S}_\tau &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\tau}(\bar{y}_i - \hat{\rho}_\tau\bar{y}_{i-1}) \\
&= \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{\tau_e-1}(\bar{y}_i - \hat{\rho}_\tau\bar{y}_{i-1}) + \sum_{i=\tau_e}^{\tau}(\bar{y}_i - \hat{\rho}_\tau\bar{y}_{i-1})\right] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{\tau_e-1}(u_i - (\hat{\rho}_\tau - 1)\bar{y}_{i-1}) + \frac{1}{\sqrt{n}}\sum_{i=\tau_e}^{\tau}(u_i - (\hat{\rho}_\tau - \rho_n)\bar{y}_{i-1}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{\tau}u_i - (\hat{\rho}_\tau - 1)\frac{1}{\sqrt{n}}\sum_{i=1}^{\tau_e-1}\bar{y}_{i-1} - (\hat{\rho}_\tau - \rho_n)\frac{1}{\sqrt{n}}\sum_{i=\tau_e}^{\tau}\bar{y}_{j-1}.
\end{aligned} \tag{156}$$

We now compare the order of the three terms. It is clear that $\frac{1}{\sqrt{n}}\sum_{i=1}^{\tau}u_i = O_p(n^{\max\{d_1,d_2\}})$. For the second term $(\hat{\rho}_\tau - 1)\frac{1}{\sqrt{n}}\sum_{i=1}^{\tau_e-1}\bar{y}_{i-1}$, note that $(\hat{\rho}_\tau - 1) \overset{a}{\sim} \frac{c}{n^\alpha}$ and

$$\begin{aligned}
\sqrt{n}\frac{1}{n}\sum_{i=1}^{\tau_e-1}\bar{y}_{i-1} &= \sqrt{n}\frac{\lfloor nr\rfloor}{n}\frac{1}{\tau}\sum_{i=1}^{\tau_e-1}\bar{y}_{i-1} \\
&= \sqrt{n}\frac{\lfloor nr\rfloor}{n}\left(\frac{1}{\tau}\sum_{i=1}^{\tau_e-1}y_{i-1} - \frac{\tau_e-1}{\tau}\frac{1}{\tau}\sum_{i=1}^{\tau}y_i\right) \\
&= O(\sqrt{n})\left(O_p(n^{1/2+d_1}) - O_p\left(n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\right)\right) \\
&= O_p\left(n^{\alpha+d_1}\rho_n^{\tau-\tau_e}\right)
\end{aligned}$$

This makes $(\hat{\rho}_\tau - 1)\frac{1}{\sqrt{n}}\sum_{i=1}^{\tau_e-1}\bar{y}_{i-1} = \frac{c}{n^\alpha}O_p\left(n^{\alpha+d_1}\rho_n^{\tau-\tau_e}\right) = O_p\left(n^{d_1}\rho_n^{\tau-\tau_e}\right)$. For the last term, note that from (152), we have $(\hat{\rho}_\tau - \rho_n) = O_p\left(\frac{1}{n^\alpha\rho_n^{\tau-\tau_e}}\right)$, and

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=\tau_e}^{\tau}\bar{y}_{j-1} &= r\sqrt{n}\frac{\lfloor nr\rfloor}{nr}\frac{1}{\tau}\sum_{i=\tau_e}^{\tau}\bar{y}_{j-1} \\
&= r\sqrt{n}\frac{\lfloor nr\rfloor}{nr}\frac{1}{\tau}\sum_{i=\tau_e}^{\tau}\left(y_{j-1} - \frac{1}{\tau}\sum_{i=1}^{\tau}y_j\right) \\
&= r\sqrt{n}\frac{\lfloor nr\rfloor}{nr}\left(\frac{1}{\tau}\sum_{i=\tau_e}^{\tau}y_{j-1} - \frac{\tau-\tau_e+1}{\tau}\frac{1}{\tau}\sum_{i=1}^{\tau}y_j\right).
\end{aligned}$$

Note that from Lemma 20.1 and (130) we have

$$\frac{1}{\tau}\sum_{j=\tau_e}^{\tau} y_{j-1} = O_p\left(n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\right),$$

and

$$\frac{1}{\tau}\sum_{j=1}^{\tau} y_j = O_p\left(n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\right).$$

So we have

$$\frac{1}{\sqrt{n}}\sum_{i=\tau_e}^{\tau} \bar{y}_{j-1} = O\left(n^{1/2}\right) O_p\left(n^{\alpha+d_1-1/2}\rho_n^{\tau-\tau_e}\right) = O_p\left(n^{\alpha+d_1}\rho_n^{\tau-\tau_e}\right).$$

This implies

$$(\hat{\rho}_\tau - \rho_n)\frac{1}{\sqrt{n}}\sum_{i=\tau_e}^{\tau} \bar{y}_{j-1} = O_p\left(\frac{1}{n^\alpha \rho_n^{\tau-\tau_e}}\right) O_p\left(n^{\alpha+d_1}\rho_n^{\tau-\tau_e}\right) = O_p\left(n^{d_1}\right). \quad (157)$$

Comparing the order of three terms in (156), we obtain

$$\frac{1}{\sqrt{n}}\hat{S}_\tau \overset{a}{\sim} -(\hat{\rho}_\tau - 1)\frac{1}{\sqrt{n}}\sum_{i=1}^{\tau_e-1} \bar{y}_{i-1} = O_p\left(n^{d_1}\rho_n^{\tau-\tau_e}\right). \quad (158)$$

Then (155) implies $\hat{\Omega}_{HAR} = O_p(n^{2d_1}\rho_n^{2(\tau-\tau_e)})$.

Suppose $\tau \in N_1$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\hat{S}_\tau \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{\tau}(\bar{y}_i - \hat{\rho}_\tau \bar{y}_{i-1}) \\
= \;& \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{\tau_e-1}(\bar{y}_i - \hat{\rho}_\tau \bar{y}_{i-1}) + \sum_{i=\tau_e}^{\tau_f}(\bar{y}_i - \hat{\rho}_\tau \bar{y}_{i-1}) + \sum_{i=\tau_f+1}^{\tau}(\bar{y}_i - \hat{\rho}_\tau \bar{y}_{i-1})\right] \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{\tau_e-1}(u_i - (\hat{\rho}_\tau - 1)\bar{y}_{i-1}) + \frac{1}{\sqrt{n}}\sum_{i=\tau_e}^{\tau_f}(u_i - (\hat{\rho}_\tau - \rho_n)\bar{y}_{i-1}) + \frac{1}{\sqrt{n}}\sum_{i=\tau_f+1}^{\tau}(u_i - (\hat{\rho}_\tau - 1)\bar{y}_{i-1}) \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{\tau} u_i + (\hat{\rho}_\tau - 1)\frac{1}{\sqrt{n}}\sum_{i=1}^{\tau_e-1}\bar{y}_{i-1} - (\hat{\rho}_\tau - \rho_n)\frac{1}{\sqrt{n}}\sum_{i=\tau_e}^{\tau}\bar{y}_{j-1} + (\hat{\rho}_\tau - 1)\frac{1}{\sqrt{n}}\sum_{i=\tau_f+1}^{\tau}\bar{y}_{i-1}. \quad (159)
\end{aligned}$$

Note that the order of the first three terms in (159) are $O_p(n^{d_1})$, $O_p\left(n^{d_1}\rho_n^{\tau-\tau_e}\right)$, and $O_p\left(n^{d_1}\right)$, respectively. For the last term, we have

$$
\begin{aligned}
(\hat{\rho}_\tau - 1)\frac{1}{\sqrt{n}}\sum_{i=\tau_f+1}^{\tau}\bar{y}_{i-1} &\overset{a}{\sim} \frac{c}{n^\alpha}\frac{1}{\sqrt{n}}\sum_{i=\tau_f+1}^{\tau}\left(-n^{\alpha+d_1-1/2}\rho_n^{\tau_f-\tau_e}\frac{1}{rc}\sigma_{d_1}B^{H_1}(r_e)\right) \\
&= O_p(n^{d_1}\rho_n^{\tau_f-\tau_e}),
\end{aligned}
$$

where we have applied (135) and (154) to obtain the asymptotic equivalence.

As $\tau \in N_1$, $\tau_f < \tau$, eventually we have the same expression as (158) and it implies $\Omega_{HAR} = O_p(n^{2d_1}\rho_n^{2(\tau-\tau_e)})$. This ends the proof of Lemma 27.

Proof of Theorem 8

Note that $\widetilde{DF}_{\tau,HAR} = \left(\frac{\sum_{i=1}^{\tau}\bar{y}_{i-1}^2}{\hat{\Omega}_{HAR}}\right)^{1/2}(\tilde{\rho}_\tau - 1)$.

Suppose that $\tau \in B$, applying the results in Lemma 22.1, (152) and Lemma 27 , we obtain

$$
\begin{aligned}
\left(\frac{\sum_{i=1}^{\tau}\bar{y}_{i-1}^2}{\hat{\Omega}_{HAR}}\right)^{1/2}(\tilde{\rho}_\tau - 1) &= O_p\left(\frac{n^{1+\alpha+2d_1}\rho_n^{2(\tau-\tau_e)}}{n^{2d_1}\rho_n^{2(\tau-\tau_e)}}\right)^{1/2}\frac{c}{n^\alpha} \\
&= O_p\left(n^{\frac{1-\alpha}{2}}\right) \to \infty.
\end{aligned}
$$

Suppose that $\tau \in N_{1,}$applying the results in Lemma 22.1, (154) and Lemma 27, we have

$$
\begin{aligned}
\left(\frac{\sum_{i=1}^{\tau}\bar{y}_{i-1}^2}{\hat{\Omega}_{HAR}}\right)^{1/2}(\tilde{\rho}_\tau - 1) &= O_p\left(\frac{n^{1+\alpha+2d_1}\rho_n^{2(\tau_f-\tau_e)}}{n^{2d_1}\rho_n^{2(\tau-\tau_e)}}\right)^{1/2}\left(-\frac{c}{n^\alpha}\right) \\
&= -O_p\left(n^{\frac{1-\alpha}{2}}\right) \to -\infty.
\end{aligned}
$$

To show $\hat{r}_e^{HAR} \xrightarrow{p} r_e$ and $\hat{r}_f^{HAR} \xrightarrow{p} r_f$, note that if $\tau \in N_0$,

$$
\lim_{n\to\infty}\Pr(\widetilde{DF}_{\tau,HAR} > cv_{n,HAR}) = \Pr\left(F_{r,d} > \infty\right) = 0.
$$

If $\tau \in B$, $\lim_{n\to\infty}\Pr(\widetilde{DF}_{\tau,HAR} > cv_{n,HAR}) = 1$, given that $\frac{cv_{n,HAR}}{n^{(1-\alpha)/2}} \to 0$. If $\tau \in N_1$, $\lim_{n\to\infty}\Pr(\widetilde{DF}_{\tau,HAR} > cv_{n,HAR}) = 0$, as $\widetilde{DF}_{\tau,HAR} = -O_p\left(n^{\frac{1-\alpha}{2}}\right)$.

It follows that for any $\eta, \vartheta > 0$,

$$\Pr(\hat{r}_e^{HAR} > r_e + \eta) \to 0, \text{ and } \Pr(\hat{r}_f^{HAR} < r_f + \vartheta) \to 0,$$

due to the fact that $\Pr(\widetilde{DF}_{(\tau_e + \alpha_\eta / n), HAR} > r_e + \eta) \to 1$ for all $0 < \alpha_\eta < \eta$ and $\Pr(\widetilde{DF}_{(\tau_f - \alpha_\vartheta / n), HAR} > cv_{n,HAR}) \to 1$ for all $0 < \alpha_\vartheta < \vartheta$. As η and ϑ are arbitrary and $\Pr(\hat{r}_e^{HAR} < r_e) \to 0$ and $\Pr(\hat{r}_f^{HAR} > r_f) \to 0$, we can deduce $\Pr(\left|\hat{r}_e^{HAR} - r_e\right| > \eta) \to 0$ and $\Pr(\left|\hat{r}_f^{HAR} - r_f\right| > \vartheta) \to 0$ as $n \to \infty$, provided that

$$\frac{1}{cv_{n,HAR}} + \frac{cv_{n,HAR}}{n^{(1-\alpha)/2}} \to 0.$$