The Grid Bootstrap for Continuous Time Models *

Yiu Lim Lui
Dongbei University of Finance and Economics

Weilin Xiao
Zhejiang University

Jun Yu
Singapore Management University

July 1, 2020

Abstract

This paper 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.

JEL classification: C11, C12
Keywords: Grid bootstrap, In-fill asymptotics, Continuous time models, Long-span asymptotics, Double asymptotics

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:

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

*We would like to thank Peter Phillips for helpful discussions and comments. Yiu Lim Lui, The Institute for Advanced Economic Research, Dongbei University of Finance and Economics, Dalian, Liaoning 116025 China, Email: allenlui@dufe.edu.cn. Weilin Xiao, School of Management, Zhejiang University, Hangzhou, 310058, China. Email: wlxiao@zju.edu.cn. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Rd, Singapore 178903. Email for Jun Yu: yujun@smu.edu.sg. URL: http://www.mysmu.edu/faculty/yujun/.

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 $\kappa \geq 0$ (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 $\kappa \geq 0$. 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 $\kappa \geq 0$, 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 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 3, a more general continuous-time model is introduced. The LS estimator of κ and its in-fill asymptotic distribution are also discussed. Section 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 5. Section 6 reports CIs for κ based on US interest rate data and Chicago Board Options Exchange's volatility index data. Section 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.

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

[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 $J_c(r) = \int_0^r \exp(-c(r-s))dW(s)$.

(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 $\kappa > 0$, $\sqrt{T}(\hat{\kappa} - \kappa) \Rightarrow N\left(0, (\exp(2\kappa h) - 1)/h\right)$; 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.

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

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

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 $CI^A_{95\%} = \{\rho \in R : -1.96 \leq t_T(Y,\rho) \leq 1.96\}$.

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

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

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 4.1. 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.

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.

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 $(\Sigma, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, 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

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

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:

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

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

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

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

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

Remark 3.1 *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.*

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 3.1 *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 3.2 *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 3.3 *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 3.1. 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 3.1 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. $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2dr$) 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 $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2dr$ 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 $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2dr$.

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

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.

Remark 3.4 *If we define the t statistic for κ as $t(Y, \kappa, h) = h(\hat{\kappa}_h - \kappa)/\hat{\sigma}_{\widehat{\rho}_h}$, then as $h \to 0$,*

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

Remark 3.5 *By assuming $N \to \infty$ with h fixed , it can be shown that the long-span asymptotic distribution of $t(Y, \kappa, h)$ is $N(0, 1)$ when $\kappa > 0$. It becomes $-\int_0^1 \overline{W}(r)dW(r)/\sqrt{\int_0^1 \overline{W}(r)^2dr}$ with $\overline{W}(r) = W(r) - \int_0^1 W(s)ds$ being the de-meaned Brownian motion when $\kappa = 0$.*

Remark 3.6 *As shown in Phillips (1987b), when $c \to -\infty$, $\int_0^1 J_c(r)dW(r)/\sqrt{\int_0^1 J_c(r)^2dr} \Rightarrow N(0, 1)$. It implies that, when N is fixed but $\kappa \to \infty$, $t(Y, \kappa, h)$ converges to $N(0, 1)$, since all the terms that involve $\exp(c)$ and $1/c$ vanish, and so does the initial condition.*

4 Confidence Interval for κ

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 $z(Y^*, \kappa, h) = N(\hat{\kappa}_h^* - \kappa)$ 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 $CI_q^* = \{\kappa \in R : c_T^*(x_1|\kappa) \leq z(Y, \kappa, h) \leq c_T^*(x_2|\kappa)\}$, and $c_T^*(q|\kappa)$ is the quantile function of $z(Y^*, \kappa, h)$, $x_1 = (1-q)/2$ 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 4.1 *Under Model (7), as $h \to 0$,*

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

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 4.1 *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* $Y = \{y_{th}\}_{t=1}^T$,

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

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 $E(R^*(T,\kappa)) = J$. 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*

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

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

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

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

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 $\kappa \geq 0$. 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 4.1 *If we replace $z(Y,\kappa,h)$ and $z(Y^*,\kappa,h)$ in Theorem 4.1 by $t(Y,\kappa,h)$ and $t(Y^*,\kappa,h)$, Theorem 4.1 remains valid. This implies that we can use the t statistic to obtain BCIs which are also justifiable under the in-fill scheme.*

Remark 4.2 *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 4.3 *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. 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:*

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

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

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* $\kappa \geq 0$*, the grid bootstrap can be justified under the long-span scheme and the double scheme.*

Remark 4.4 *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.*

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 $\int_0^1 J_c(r)dW/\sqrt{\int_0^1 J_c(r)^2 dr}$ 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 4.2 *Assume that in Model (4), the assumptions in Theorem 4.1 hold, and additionally, the increment of the Lévy process* $L(t+h) - L(t)$ *has a bounded* r^{th} *moment for some* $r \geq 8$ *. 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 4.5 *When $\psi'(0) = 0$, $\psi''(0) = 1$, $y_0 = \mu$, $\kappa = 0$, $z^{y_0}(\kappa, \theta) = -\int_0^1 \overline{W}(r)dW(r)/\int_0^1 \overline{W}(r)^2dr$. 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 $\psi'(0) = 0$, $\psi''(0) = 1$, $y_0 = \mu$, $z^{y_0}(\kappa, \theta) = -\int_0^1 \overline{J}(r)dW(r)/\int_0^1 \overline{J}(r)^2dr$. 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 4.6 *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 4.7 *Comparing equation (13) with equation (14), the grid bootstrap provides a second-order improvement over the in-fill asymptotic distribution.*

Remark 4.8 *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.*

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 7.7 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 4.1 *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 4.1, 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 3.1. As a result, we do not have a uniform convergence result as in Theorem 3.1.

5 Simulation Studies

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 4.1. We generate the bootstrap data $\{y^{*b}_{th}\}_{t=1}^T$ based on $\{u^*_{th}\}_{t=1}^T$ 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 $\left\{\{y^{*b}_{th}\}_{t=1}^T\right\}_{b=1}^B$. For every set of bootstrap data, obtain the LS estimator of κ (denoted by $\hat{\kappa}^*_h$) and calculate the bootstrap coefficient-based statistic $z(Y^*, \kappa_G, h) = N(\hat{\kappa}^*_h - \kappa_G)$. Calculate the x^{th} quantile of the bootstrap statistic $z(Y^*, \kappa_G, h)$ 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 ($K(x) = \frac{3}{4}(1 - x^2)1(|x| \leq 1)$) and choose the bandwidth by LS cross-validation.

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

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

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., $\frac{1}{10000}\sum_{m=1}^{10000} 1\left(\kappa_L^{(m)} \leq \kappa_0 \leq \kappa_U^{(m)}\right)$, 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.

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)

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 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 4.2.

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 as the first observation.

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)

[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.

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)

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.

References

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.

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

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.

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.

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.

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.

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

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.

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.

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.

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

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

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

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

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.

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

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.

Appendix

7.1 Proof of Lemma 3.1 and Remark 3.4

Proof. Before proving Lemma 3.1, we define Υ_1, Υ_2 and Υ_3 in Lemma 3.1 by

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

Proof of Lemma 3.1 and Remark 3.4 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 $x_t = y_t/\left(\sigma_\psi\lambda_h\right)$, 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

$$\frac{1}{T^2}\sum_{t=1}^T x_{t-1}^2 \Rightarrow \Upsilon_1,\ \frac{1}{T^{3/2}}\sum_{t=1}^T x_t \Rightarrow \Upsilon_2,\ \frac{1}{T}\sum_{t=1}^T x_{t-1}u_t \Rightarrow \Upsilon_3. \tag{18}$$

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 $R(T,\kappa)=\frac{1}{\hat\sigma^2T^2}\sum_{t=1}^T y_{t-1}^2-\left(\frac{1}{\hat\sigma T^{\frac{3}{2}}}\sum_{t=1}^T y_{t-1}\right)^2$, where $\hat\sigma^2=\frac{1}{T}\sum_{t=1}^T\left(y_t-\hat g_h-\hat\rho_h(\kappa)y_{t-1}\right)^2$. By construction, it can be seen that

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

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

$$z(Y,\kappa,h) = h\varsigma_h'(\rho_h(\kappa))\left(1+\frac{\varsigma_h'(\widetilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right)z(Y,\rho,T). \tag{19}$$

Since $\hat{\kappa}_h = \frac{-\ln(\hat{\rho}_h(\kappa))}{h}$, applying the generalized Delta method and using the relationship in (19), $Th = N$, $\left(1+\frac{\varsigma_h'(\widetilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right) \to_p 1$, and $h\varsigma_h'(\widetilde{\rho}_h(\kappa)) \to_p -1$, we obtain the limiting result $z(Y,\kappa,h) \Rightarrow -\frac{\Upsilon_3-\Upsilon_2\int_0^1 dW(r)}{\Upsilon_1-\Upsilon_2^2}$.

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

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

By Lemma 4.1, $\frac{\sigma_\psi\lambda_h}{\hat{\sigma}_c\sqrt{h}} \to_p 1$. Applying results in (18), we can obtain the limit of $t(Y,\rho,T)$.

To show the limit of $t(Y,\kappa,h)$, similar to (19), we have

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

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, $t(Y,\kappa,h) = -t(Y,\rho,T) + o_p(1)$ under the in-fill scheme, giving the result in Remark 3.4. ■

7.2 Proof of Lemma 4.1

Before proving Lemma 4.1, 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 7.1 *For Model (7), let $\hat{\breve{g}}_h$ be the LS estimator of $\breve{g}_h$. Then under the in-fill scheme, for any $\kappa \geq 0$, 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 (18), we therefore establish the result in Lemma 7.1. ■

We now prove Lemma 4.1.

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

$$\begin{aligned}
\hat{\sigma}_c^2 &= \frac{1}{Th} \sum_{t=1}^T e_t^2 = \frac{1}{Th} \sum_{t=1}^T \left(\epsilon_t + (g_h - \hat{g}_h) + (\rho_h(\kappa) - \hat{\rho}_h) y_{t-1}\right)^2 \\
&= \frac{1}{Th} \sum_{t=1}^T \epsilon_t^2 + \frac{1}{Th} \sum_{t=1}^T (g_h - \hat{g}_h)^2 + (\rho_h(\kappa) - \hat{\rho}_h)^2 \frac{1}{Th} \sum_{t=1}^T y_{t-1}^2 \\
&\quad + 2(g_h - \hat{g}_h) \frac{1}{Th} \sum_{t=1}^T \epsilon_t + 2(g_h - \hat{g}_h)(\rho_h(\kappa) - \hat{\rho}_h) \frac{1}{Th} \sum_{t=1}^T y_{t-1} \\
&\quad + 2(\rho_h(\kappa) - \hat{\rho}_h) \frac{1}{Th} \sum_{t=1}^T y_{t-1} \epsilon_t. \qquad (20)
\end{aligned}$$

We now investigate the five terms on the right-hand side of (20) one-by-one.

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

$$\frac{1}{Th} \sum_{t=1}^T (g_h - \hat{g}_h)^2 = \frac{(g_h - \hat{g}_h)^2}{h} = \frac{\sigma_\psi^2 \lambda_h^2 (\hat{\breve{g}}_h - \breve{g}_h)^2}{h} = O_p(h) = o_p(1), \quad \text{(by Lemma 7.1)}$$

$$(\rho_h(\kappa) - \hat{\rho}_h)^2 \frac{1}{Th} \sum_{t=1}^T y_{t-1}^2$$

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

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

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

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. ■

7.3 Proof of Theorem 4.1 and Remark 4.1

Before proving Theorem 4.1 and Remark 4.1, 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 7.2 (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 $\sup_{0 \le t < 1} |w_T(t) - \varsigma_T(t)| = o_{as}(T^{-1/2+1/r+\varepsilon})$.*

Suppose that bootstrap error term $\{u_t^*\}_{t=1}^T$ drawn from our resampling method in Section 4.1, 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 $T \to \infty$, $\sup_{0 \le t < 1} |\eta_T(t) - \varsigma_T(t)| = o_p(T^{-\delta})$ for some $\delta > 0$. Thus, we have

$$\sup_{0 \le t < 1} |\eta_T(t) - w_T(t)| = o_p(T^{-\delta}) \text{ for some } \delta > 0. \tag{21}$$

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 7.3 *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. $\lim_{h \to 0} \sup_{\kappa \in K} \Pr\{|(\tilde{g}_h - g_h)/\sigma_\psi \lambda_h| > \varepsilon\} = 0$;
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 4.1 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,

$$\begin{aligned}
&\sup_{\kappa\in K}\sup_t\left|\frac{1}{\hat{\sigma}}\left(\frac{y_{t-1}}{\sqrt{T}} - \frac{y_{t-1}^*}{\sqrt{T}}\right)\right| \\
=\ &\sup_{\kappa\in K}\sup_t(1 + o_p(1))\left|\begin{array}{c}(\rho_h - 1)\sum_{i=1}^t \rho_h^{t-i-1}\left(\eta_T\left(\frac{i}{T}\right) - w_T\left(\frac{i}{T}\right)\right) \\ +\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right) + \frac{g_h - \tilde{g}_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^t \rho_h^i\end{array}\right| \\
\le\ &\sup_{\kappa\in K}\left[(1 + o_p(1))\sup_t\left(\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|\left|\frac{\rho_h - 1}{\rho_h}\sum_{i=1}^t \rho_h^{t-i} + 1\right|\right)\right] + \sup_{\kappa\in K}\sup_t\left|\frac{g_h - \tilde{g}_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^t \rho_h^i\right|
\end{aligned}$$

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

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

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

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

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

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

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

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

$$\begin{aligned}
A &= \frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\frac{x_T}{\sqrt{T}}\eta_T(1) - \frac{\lambda_h^2}{h}\frac{x^*_T}{\sqrt{T}}w_T(1),\ B = \frac{\lambda_h^2}{h}\frac{\check{\check{g}}_h}{\sqrt{T}}\sum_{t=1}^{T} w_T\left(\frac{t}{T}\right) - \frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\frac{\check{g}_h}{\sqrt{T}}\sum_{t=1}^{T}\eta_T\left(\frac{t}{T}\right),\\
C &= \frac{(\rho_h-1)\lambda_h^2}{h}\sum_{t=2}^{T}\frac{x^*_{t-1}}{\sqrt{T}}w_T\left(\frac{t}{T}\right) - \frac{(\rho_h-1)\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^{T}\frac{x_{t-1}}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right),\\
D &= \frac{\lambda_h^2}{h}\sum_{t=2}^{T}\frac{u^*_t}{\sqrt{T}}w_T\left(\frac{t}{T}\right) - \frac{\sigma^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^{T}\frac{u_t}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right),\ E = \frac{(\rho_h-1)\hat{\sigma}_c\lambda_h}{\hat{\sigma}^2 T}y_0 u^*_1 - \frac{(\rho_h-1)\sigma_\psi\lambda_h}{\hat{\sigma}^2 T}y_0 u_1,\\
F &= \frac{\lambda_h^2}{h}\left(\frac{z^*_1}{\sqrt{T}}\right)^2 - \frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\left(\frac{z_1}{\sqrt{T}}\right)^2,\ G = \frac{\tilde{g}_h\epsilon^*_1}{\hat{\sigma}^2 T} - \frac{g_h\epsilon_1}{\hat{\sigma}^2 T}.
\end{aligned}$$

We now examine these terms one-by-one.

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

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

$$\sup_{\kappa\in K}\sup_t |C| = \sup_{\kappa\in K}\sup_t \left|\frac{(\rho_h-1)\lambda_h^2}{h}\sum_{t=2}^{T}\frac{x^*_{t-1}}{\sqrt{T}}w_T\left(\frac{t}{T}\right) - \frac{(\rho_h-1)\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^{T}\frac{x_{t-1}}{\sqrt{T}}\eta_T\left(\frac{t}{T}\right)\right|$$

$$
\begin{aligned}
&= (1+o_p(1))\sup_{\kappa\in K}|\rho_h-1|\sup_t\sum_{t=2}^T\left|\begin{array}{c}\frac{x^*_{t-1}}{\sqrt{T}}\left(w_T\left(\frac{t}{T}\right)-\eta_T\left(\frac{t}{T}\right)\right)\\ +\eta_T\left(\frac{t}{T}\right)\left(\frac{x^*_{t-1}}{\sqrt{T}}-\frac{x_{t-1}}{\sqrt{T}}\right)\end{array}\right| \\
&\leq \sup_{\kappa\in K}|-\kappa h+o(h^2)|T\left(\begin{array}{c}\sup_t\left|\frac{x^*_{t-1}}{\sqrt{T}}\right|\sup_t\left|w_T\left(\frac{t}{T}\right)-\eta_T\left(\frac{t}{T}\right)\right|\\ +\sup_t\left|\eta_T\left(\frac{t}{T}\right)\right|\sup_t\left|\frac{x^*_{t-1}}{\sqrt{T}}-\frac{x_{t-1}}{\sqrt{T}}\right|\end{array}\right) \\
&\leq CN\left(\sup_t\left|\frac{x^*_{t-1}}{\sqrt{T}}\right|\sup_t\left|w_T\left(\frac{t}{T}\right)-\eta_T\left(\frac{t}{T}\right)\right|+\sup_t\left|\eta_T\left(\frac{t}{T}\right)\right|\sup_t\left|\frac{x^*_{t-1}}{\sqrt{T}}-\frac{x_{t-1}}{\sqrt{T}}\right|\right) \\
&= o_p(T^{-\delta}).
\end{aligned}
$$

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

$$
\begin{aligned}
\sup_{\kappa\in K}\sup_t|E| &= \sup_{\kappa\in K}\sup_t\left|\frac{(\rho_h-1)\hat{\sigma}_c\lambda_h}{\hat{\sigma}^2 T}y_0u^*_1-\frac{(\rho_h-1)\sigma_\psi\lambda_h}{\hat{\sigma}^2 T}y_0u_1\right| \\
&= \sup_{\kappa\in K}\sup_t\left|\kappa h\frac{1}{\sigma}\frac{\lambda_h}{\sqrt{h}}\frac{y_0}{\sqrt{hT}}\left[\frac{u_1}{\sqrt{T}}\right]-\kappa h\frac{1}{\sigma}\frac{\lambda_h}{\sqrt{h}}\frac{y_0}{\sqrt{hT}}\left[\frac{u^*_1}{\sqrt{T}}\right]\right|+o_p(h) \\
&\leq C_\kappa h\frac{1}{\sigma}\frac{|y_0|}{\sqrt{N}}\left[\sup_t\left|\eta_T\left(\frac{t}{T}\right)\right|+\sup_t\left|w_T\left(\frac{t}{T}\right)\right|\right]+o_p(1)=o_p(h),
\end{aligned}
$$

where C_κ depends on κ.

$$
\begin{aligned}
\sup_{\kappa\in K}\sup_t|F| &= \sup_{\kappa\in K}\sup_t\left|\frac{\lambda_h^2}{h}\left(\frac{u^*_1}{\sqrt{T}}\right)^2-\frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\left(\frac{u_1}{\sqrt{T}}\right)^2\right| \\
&= \sup_{\kappa\in K}\sup_t\left|(1+o_p(h))\left(\left(\frac{u^*_1}{\sqrt{T}}\right)^2-\left(\frac{u_1}{\sqrt{T}}\right)^2\right)\right| \\
&= \sup_{\kappa\in K}\sup_t\left|(1+o_p(h))\left[\left(\frac{u^*_1}{\sqrt{T}}\right)-\left(\frac{u_1}{\sqrt{T}}\right)\right]\left[\left(\frac{u^*_1}{\sqrt{T}}\right)+\left(\frac{u_1}{\sqrt{T}}\right)\right]\right| \\
&\leq (1+o_p(h))\sup_t\left|w_T\left(\frac{t}{T}\right)-\eta_T\left(\frac{t}{T}\right)\right|\left(\sup_t\left|w_T\left(\frac{t}{T}\right)\right|+\sup_t\left|\eta_T\left(\frac{t}{T}\right)\right|\right) \\
&= o_p(T^{-\delta}).
\end{aligned}
$$

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

Thus, we have established item 5.

For item 6,

$$\begin{aligned}
\sup_{\kappa\in K}\left|\frac{\sum_{t=1}^T y_{t-1}^2}{T\hat{\sigma}^2} - \frac{\sum_{t=1}^T y_{t-1}^{*2}}{T\hat{\sigma}^2}\right| &= \sup_{\kappa\in K}\left|\frac{1}{T\hat{\sigma}^2}\sum_{t=2}^T y_{t-1}^2 - \frac{1}{T\hat{\sigma}^2}\sum_{t=2}^T y_{t-1}^{*2} + \frac{1}{T\hat{\sigma}^2}y_0^2 - \frac{1}{T\hat{\sigma}^2}y_0^2\right| \\
&= \sup_{\kappa\in K}\left|\frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^T\left(\frac{x_{t-1}}{\sqrt{T}}\right)^2 - \frac{\lambda_h^2}{h}\sum_{t=2}^T\left(\frac{x_{t-1}^*}{\sqrt{T}}\right)^2\right| \\
&= \sup_{\kappa\in K}\left|(1+o_p(h))\left[\sum_{t=2}^T\left(\frac{x_{t-1}}{\sqrt{T}}\right)^2 - \sum_{t=2}^T\left(\frac{x_{t-1}^*}{\sqrt{T}}\right)^2\right]\right| \\
&\le (1+o_p(h))\sup_t\left|\frac{x_t}{\sqrt{T}} - \frac{x_t^*}{\sqrt{T}}\right|\left(\sup_t\left|\frac{x_t}{\sqrt{T}}\right| + \sup_t\left|\frac{x_t^*}{\sqrt{T}}\right|\right) \\
&= o_p(T^{-\delta}).
\end{aligned}$$

For item 7,

$$\begin{aligned}
&\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^T\epsilon_t\frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1} - \frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^T\epsilon_t^*\frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1}^*\right| \\
= &\sup_{\kappa\in K}\left|\frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c h}\frac{1}{T^{3/2}}\sum_{t=1}^T x_{t-1}\sum_{t=1}^T\frac{u_t}{\sqrt{T}} - \frac{\lambda_h^2}{h}\frac{1}{T^{3/2}}\sum_{t=1}^T x_{t-1}\sum_{t=1}^T\frac{u_t^*}{\sqrt{T}}\right. \\
&\left. + \frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=1}^T\frac{u_t^*}{\sqrt{T}}\left(\frac{1}{T^{3/2}}\sum_{t=1}^T x_{t-1} - \frac{1}{T^{3/2}}\sum_{t=1}^T x_{t-1}^*\right)\right| \\
\le &\sup_{\kappa\in K}\sup_t\left|\frac{1}{T}\sum_{t=1}^T\frac{x_{t-1}}{\sqrt{T}}\right|\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| \\
&+\frac{1}{T}\sup_t\left|w_T\left(\frac{t}{T}\right)\right|\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| + o_{a.s.}(1) = o_p(T^{-\delta}).
\end{aligned}$$

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,

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

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

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

Now we are in the position to show Theorem 4.1, that is,

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

$$\inf_{\kappa\in K}\Pr\{\kappa_0\in CI_q\}\to x_2-x_2=q. \tag{22}$$

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

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

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

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

The final claim (22) is a direct result of Lemma 1 in Mikusheva (2007). The result in Remark 4.1 is established based on the same argument and is omitted. ■

7.4 Proof of Theorem 4.2

Before we prove Theorem 4.2, 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 7.4 (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 $r \geq 8$. 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

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

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

Lemma 7.5 (Mikusheva (2015), Theorem 1) *Suppose $c \leq 0$ and z_j satisfies the assumption in Lemma 7.4. 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:*

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

$$+o_p(T^{-1/2}).$$

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

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

Lemma 7.6 *Under model (4), if $\kappa \geq 0$, 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

$$\begin{aligned}
x_t &= \sum_{j=1}^{t} e^{(t-j)c/T}z_j + \frac{b}{\sqrt{T}}\frac{e^{ct/T}-1}{e^{c/T}-1} + e^{ct/T}x_0 + o_p(T^{-1/2}) \\
&= \tilde{x}_t + \frac{b}{\sqrt{T}}\frac{e^{ct/T}-1}{e^{c/T}-1} + e^{ct/T}x_0 + o_p(T^{-1/2}). \qquad (23)
\end{aligned}$$

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 7.6.

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

The approximation to the first term is given in Lemma 7.5(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 7.4. 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 7.6, 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 7.5. 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

$$\begin{aligned}&= \frac{1}{T^2}\sum_{t=1}^{T}\frac{2b}{\sqrt{T}}\frac{e^{tc/T}-1}{e^{c/T}-1}\sum_{t=1}^{T}\sum_{j=0}^{t}e^{(t-j)c/T}z_j\\ &= \frac{2b}{T(e^{c/T}-1)}\frac{1}{T}\sum_{t=1}^{T}(e^{ct/T}-1)\frac{1}{\sqrt{T}}\sum_{j=1}^{t}e^{(t-j)c/T}z_j\end{aligned}$$

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

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 7.6. ■

Now we are in the position to prove Theorem 4.2.

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 7.6, we obtain

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

$$-\frac{1}{T^{3/4}}R_{2,T^{-1/2}}M(V)-\frac{1}{T}R_{2,T^{-1/2}}N(V)+o_p(T^{-1/2}),$$

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

$$\begin{aligned} z(Y,\rho,T)=&\frac{\Upsilon_3-\Upsilon_2W(1)}{\Upsilon_1-\Upsilon_2^2}+\frac{1}{T^{1/4}}\frac{R_{3,T^{-1/4}}-M(V)\Upsilon_2}{\Upsilon_1-\Upsilon_2^2}\\ &+\frac{1}{T^{1/2}}\left(\begin{array}{c}\frac{R_{3,T^{-1/2}}-N(V)\Upsilon_2-R_{2,T^{-1/2}}W(1)}{\Upsilon_1-\Upsilon_2^2}\\ -\frac{\Upsilon_3-\Upsilon_2W(1)}{(\Upsilon_1-\Upsilon_2^2)^2}\left(R_{1,T^{-1/2}}-2R_{2,T^{-1/2}}\right)\end{array}\right)+o_p\left(T^{-1/2}\right)\\ =&z^{y_0}(\rho,\theta)+T^{-1/4}\tilde{A}+T^{-1/2}\tilde{B}+o_p(T^{-1/2}),\end{aligned}$$

where

$$\begin{aligned}\tilde{A}&=\frac{R_{3,T^{-1/4}}-M(V)\Upsilon_2}{\Upsilon_1-\Upsilon_2^2},\\ \tilde{B}&=\frac{R_{3,T^{-1/2}}-N(V)\Upsilon_2-R_{2,T^{-1/2}}W(1)}{\Upsilon_1-\Upsilon_2^2}-\frac{\Upsilon_3-\Upsilon_2W(1)}{(\Upsilon_1-\Upsilon_2^2)^2}\left(R_{1,T^{-1/2}}-2R_{2,T^{-1/2}}\right).\end{aligned}$$

The expansion of $z(Y,\kappa,h)$ can be obtained from (19) and the Taylor series expansion of $h\varsigma_h'(\rho_h(\kappa))=-\exp(\kappa h)$.

Finally, for the last claim in Theorem 4.2, 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 4.1, we need the following lemma.

Lemma 7.7 (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

$$x_t=\rho_h(\kappa)x_{t-1}+\breve{g}_h+\sigma_tu_t, \tag{24}$$

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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor sT\rfloor}\sigma_tu_t\Rightarrow\psi''(0)\int_0^s\omega(s)dW(s)=\bar{\sigma}_\psi W_\omega(s), \tag{25}$$

where $W_\omega(s)=\bar{\sigma}_\psi^{-1}\psi''(0)\int_0^s\omega(s)dW(s)$, and $\bar{\sigma}_\psi^2=\psi''(0)^2\int_0^1\omega(r)^2dr$. Applying the continuous mapping theorem and following the proof of Lemma 3.1, we have

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_t\ \Rightarrow\ \frac{e^c-c-1}{c^2}\bar{b}+\bar{\sigma}_\psi\int_0^1J_c^\omega(r)dr+\frac{e^c-1}{c}\bar{\gamma}:=\bar{\Upsilon}_2;$$

$$
\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 \;&\Rightarrow\; \frac{e^{2c}+4e^{c}+2c+3}{2c^3}\bar{b}^2+\bar{\sigma}_{\psi}\int_0^1 \left(J_c^{\omega}(r)\right)^2 dr+\frac{2\bar{b}\bar{\sigma}_{\psi}}{c}\int_0^1 e^{rc-1}J_c^{\omega}(r)dr \\
&\quad +\frac{e^{2c}-2e^{c}+1}{c^2}\bar{b}\bar{\gamma}_0+2\bar{\gamma}_0\bar{\sigma}_{\psi}\int_0^1 e^{rc}J_c^{\omega}(r)dr+\bar{\gamma}_0^2\frac{e^{2c}-1}{2c}:=\bar{\Upsilon}_1; \qquad (26)\\
\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\sigma_t u_t \;&\Rightarrow\; \frac{2\bar{b}\bar{\sigma}_{\psi}}{c}\int_0^1 (e^{rc}-1)dW_{\omega}(r)+\bar{\sigma}_{\psi}^2\int_0^1 J_c^{\omega}(r)dW_{\omega}(r)+\bar{\gamma}\bar{\sigma}_{\psi}\int_0^1 dW_{\omega}(r)=\bar{\Upsilon}_3,
\end{aligned}
$$

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 4.1

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,

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor sT\rfloor}\frac{e_t}{\sqrt{h}}z_t^* \overset{p}{\Rightarrow} \bar{\sigma}_{\psi}W_{\omega}(s). \qquad (27)
$$

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 (25), analogous to (24), the bootstrap data generating process can be written as

$$
x_t^*=\rho_h(\kappa)x_{t-1}^*+\breve{g}_h+e_t h^{-1/2}z_t^*+o_p(1). \qquad (28)
$$

Applying the continuous mapping theorem, we obtain the analogous results to those in (26). Eventually we have

$$
z(Y^*,\kappa,h)\overset{p}{\Rightarrow}\bar{z}^{y_0}(\kappa,\theta). \qquad (29)
$$

The convergence in (29) 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

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

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