11-2018

The grid bootstrap for continuous time models

Yiu Lim LUI
Singapore Management University, yl.lui.2015@phdecons.smu.edu.sg

Weilin XIAO
Zhejiang University

Jun YU
Singapore Management University, yujun@smu.edu.sg

SMU ECONOMICS &
STATISTICS

The Grid Bootstrap for Continuous Time Models

Yiu Lim Lui, Weilin Xiao, Jun Yu

November 2018

Paper No. 20-2018

The Grid Bootstrap for Continuous Time Models*

Yiu Lim Lui
Singapore Management University

Weilin Xiao
Zhejiang University

Jun Yu
Singapore Management University

November 9, 2018

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 established by assuming the sampling interval (h) shrinks to zero. 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 gird bootstrap method performs better than the in-fill asymptotic theory and much better than the long-span theory. Empirical applications to U.S. interest rate data highlight 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.

1 Introduction

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

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

where κ, μ, and σ are all constant, 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

*Yiu Lim Lui, School of Economics, Singapore Management University, 90 Stamford Rd, Singapore 178903, Email: yl.lui.2015@smu.edu.sg. 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/.

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 least squares (LS) method. Denote the LS estimator by $\hat{\kappa}$. To make statistical inference about κ, one needs to obtain the exact finite sample distribution of $\hat{\kappa}$. Unfortunately, the exact finite sample distribution 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 non-zero; 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 asymptotic distribution cannot provide more accurate approximation than the other two asymptotic distributions due to an added restriction.

Which of the two asymptotic distributions, the in-fill asymptotic distribution or the long-span distribution, provides more accurate approximation to the finite sample distribution? Yu (2014), Zhou and Yu (2015) and Bao et al. (2017) provide the answer to this question. Yu (2014) and Zhou and Yu (2015) derived the in-fill asymptotic distribution of $\hat{\kappa}$ when μ is known and unknown, respectively, and approximated the exact finite sample distribution of $\hat{\kappa}$ by simulations. They showed when κ is reasonably close to zero, the in-fill asymptotic distribution substantially outperforms the long-span asymptotic distribution, even when h is not very small and N is moderately large. This is not surprising as the in-fill distribution depends on the initial condition and is asymmetric, two features that can be found in the finite sample distribution but not in the long-span asymptotic distribution. Moreover, Bao et al. (2017) approximated the exact finite sample distribution of $\hat{\kappa}$ by numerical integrations and, based on the exact finite sample distribution, found the superiority of the in-fill asymptotic distribution over the long-span asymptotic distribution.

This paper proposes to use the grid bootstrap method, which was initially introduced by Andrew (1993) and then by Hansen (1999), to construct confidence intervals (CIs) for κ. The grid bootstrap has been used extensively in the literature for constructing CIs for the autoregressive (AR) parameter in a discrete time AR(1) model. Mikusheva (2007) shows that it gives CIs that have correct coverage uniformly over the parameter space, including the unit root case, in long span samples. The asymptotic justification of bootstrap methods has traditionally been made by the long-span asymptotic scheme and asymptotic expansions have typically been made on the long-span asymptotic distribution. For example, in the stationary

AR(1) model, the standard bootstrap can be justified by Edgeworth-type expansions which uses the normal distribution as the leading term (Bose, 1988). In a unit root AR(1) model, a non-standard bootstrap method was justified by expansions which use the Dickey-Fuller-Phillips asymptotic distribution as the leading term; see Park (2003). In the local-to-unity AR(1) model, the grid bootstrap can be justified by expansions which use the local-to-unity asymptotic distribution as the leading term; see Mikusheva (2015).

The present paper justifies the grid bootstrap procedure under the in-fill asymptotic scheme by showing that CIs for κ obtained by the grid bootstrap have correct coverage uniformly over the parameter space, including the case where $\kappa = 0$. We use the in-fill scheme, instead of the long-span scheme, to justify the bootstrap procedure because the in-fill distribution provides a better finite sample approximation than the long-span distribution. Our expansion 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. Second, consistent estimation of κ and μ is not required for constructing a valid CI of κ under the in-fill scheme. Third, the grid bootstrap method, with a simple modification, is applicable in the presence of heteroskedasticity. Finally, 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.

We organize the paper as follows. Section 2 reviews some important 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 class of continuous time models is introduced. The LS estimator of κ and the 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. Also reported are probabilistic expansions which use the in-fill asymptotic distribution as the leading term and how to do the grid bootstrap when there is heteroskedasticity. Simulation studies which 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. 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, "$\to$" means convergence in real sequence, "$=^d$" means equivalence in distribution "$\to_p$" and "$\to_{a.s.}$" mean convergence in probability and almost surely, respectively.

2 A Literature Review

In this section, we review some important 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 $Y := \{y_{th}\}_{t=1}^{T}$ is data generated from the continuous time model given by (1). The exact discrete model corresponding to (1) is given by

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

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

$$x_{th} = e^{-\kappa h} x_{(t-1)h} + \frac{\mu \left(1 - e^{-\kappa h}\right)}{\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}} + \epsilon_t, x_0 = \frac{y_0}{\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}}, \quad (3)$$

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

Model (3) 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 $\widehat{\rho}_h(\kappa)$ and the LS estimator of κ be $\widehat{\kappa} = -\ln\left(\widehat{\rho}_h(\kappa)\right)/h$. 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 (3) 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 (3), it is expected that the in-fill asymptotic distribution of $\widehat{\rho}_h(\kappa)$ 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 $\widehat{\rho}_h(\kappa)$ when $y_0 = 0$ and μ is known ($= 0$). 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 ($= 0$) 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 $\widehat{\rho}_h(\kappa)$. For example, when $\kappa > 0$, $\sqrt{T}(\hat{\kappa} - \kappa) \Rightarrow N\left(0, \left(\exp(2\kappa h) - 1\right)/h\right)$; see Tang and Chen (2009). When $\kappa = 0$,

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

$N(\hat{\kappa}-\kappa) \Rightarrow -\int_0^1 \overline{W}(r)dW(r)/\int_0^1 \overline{W}(r)^2 dr$ with $\overline{W}(r) = W(r) - \int_0^1 W(s)ds$. The discontinuity in the long-span limit theory of κ (both in the rate and in the limiting distribution) echoes that of ρ in the discrete time AR(1) model.

When κ is positive but reasonably close to zero (such as $\kappa = 0.01, 0.1, 1, 10$), 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 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. Unfortunately, the in-fill asymptotic theory is infeasible as it involves κ. In practice, one can plug-in an estimated κ into the in-fill distribution. However, κ cannot be estimated by $\hat{\kappa}$ consistently under the in-fill scheme, replacing it with an inconsistent estimate of κ leads to CIs with incorrect coverage.

For the discrete time AR(1) model, the in-fill scheme is not available. When the autoregressive 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 autoregressive 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 moderate. This motivates Phillips (1977) and Tanaka (1983) to develop Edgeworth expansions to approximate the finite sample distribution of the LS estimator of the AR coefficient. While the leading term in Edgeworth expansions is a normal distribution, departure from normality manifests in higher order terms. Alternatively, the finite sample distribution can be approximated by the bootstrap method. Bose (1988) showed the linkage between Edgeworth expansions and the bootstrap method.

When the AR(1) model has a unit root, the long-span asymptotic distribution is nonstandard. 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 tests. Under the local-to-unity AR(1) model, Hansen (1999) introduced the grid bootstrap approach. Mikusheva (2015) obtained expansions of the t-statistic about the local-to-unity asymptotic distribution and showed 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 a much better approximation to the finite sample distribution than the normal asymptotic 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 used to estimate parameters in the model. Let $t_T(Y,\rho)$ be a test statistic whose exact sampling distribution is $F_T(x|\rho) = \Pr(t_T(Y,\rho) < x|\rho)$. 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

$$CI_q := \{\rho \in R : c_T(x_1|\rho) \leq t_T(Y,\rho) \leq c_T(x_2|\rho)\}, \tag{4}$$

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. If we replace $c_T(x_i|\rho)$ with $c(x_i|\rho)$ in Equation (4), we obtain an asymptotic CI, CI_q^A, which has the correct coverage probability asymptotically, i.e., $\lim_{T\to\infty} \Pr(\rho_0 \in CI_q^A) = q$. For example, if the asymptotic distribution is standard normal, then a 95% asymptotic CI is $C_{95\%}^A = \{\rho \in R : -1.96 \leq t_T(Y,\rho) \leq 1.96\}$.

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 (4) 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 (4) to obtain an asymptotic CI, CI_q^A. It is easy to show that $\lim_{T\to\infty} \Pr(\rho_0 \in CI_q^A) = q$.

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_q^B. For example, a BCI given by the standard parametric bootstrap procedure is given by

$$CI_q^B := \{\rho \in R : c_T^*(x_1|\widehat{\rho}) \leq t_T(Y,\rho) \leq c_T^*(x_2|\widehat{\rho})\}, \tag{5}$$

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

There are some advantages in 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 bootstrap distribution may "require less information" about asymptotic theory; 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 a better approximation to the finite sample distribution than asymptotic distributions; see

Hall (2013). Not surprisingly, bootstrap methods often lead to CIs that have a more accurate coverage probability than traditional asymptotic theory.

3 The Model and In-fill Theory

The present paper extends Model (1) by allowing for non-normality in the stochastic behavior. 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 show via simulations that in-fill asymptotic distribution provides much better approximations to the finite sample distribution than the long-span asymptotic distribution. We then propose the grid bootstrap to obtain BCIs for κ and obtain its coverage rate under the in-fill scheme. Asymptotic expansions for the coefficient-based statistic with in-fill asymptotic distribution as the leading term are developed. The expansions justify the bootstrap method and also shows that the bootstrap method offers a refinement of the in-fill asymptotic distribution.

3.1 The model

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

$$dy(t) = \kappa(\mu - y(t))dt + \sigma dL(t), y(0) = y_0 = O_p(1), \tag{6}$$

where σ and κ are strictly positive constants, $L(t)$ is a Lévy process defined on a probability space $(\Sigma, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, with $L(0) = 0$ $a.s.$, $\mathcal{F}_t = \sigma\left\{\{y(s)\}_{s=0}^{t}\right\}$, which satisfies the following properties

1. (Independent increment) For any increasing sequence of times, say $t_0 < t_1 < \ldots < t_n$, $L(t)$ has independent increments;
2. (Stationary increment) The distribution of $L\left(t+h\right) - L\left(t\right)$ is independent of t;
3. (Stochastic continuity) For any $\epsilon > 0, t \geq 0$, $\lim_{h \longrightarrow 0} P(|L\left(t+h\right) - L\left(t\right)| \geq \epsilon) = 0$;
4. The initial condition, $y(0) = y_0$, is assumed to be independent of $L(t)$.

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 all parameters in $L(t)$ are being treated as nuisance parameters.

The exact discrete time version of (6) is

$$y_{th} = e^{-\kappa h} y_{(t-1)h} + \mu(1 - \exp(-\kappa h)) + \sigma \int_{(t-1)h}^{th} \exp(-\kappa(th - s))dL(s), \tag{7}$$

where $t = 0, 1, \ldots, T := N/h$. Note that, the characterization of the Lévy process makes the errors $\left\{\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s)\right\}_{t=1}^{N/h}$ 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, Wang and Yu (2016) showed that the error term has the following moments:

$$E\left(\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s)\right) = \sigma i\psi'(0)\frac{1-\exp(-\kappa h)}{\kappa}, \tag{8}$$

$$Var\left(\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s)\right) = \sigma^2\psi''(0)\frac{1-\exp(-2\kappa h)}{2\kappa}. \tag{9}$$

To simplify notations, let

$$\begin{aligned}
\rho_h(\kappa) &:= \exp(-\kappa h), \\
\lambda_h &:= \sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}, \\
\sigma_\psi^2 &:= \sigma^2\psi''(0), \\
g_h &:= \left[\mu + \frac{\sigma i\psi'(0)}{\kappa}\right](1-\exp(-\kappa h)), \\
u_{th} &:= (\sigma_\psi\lambda_h)^{-1}\left(\sigma \int_{(t-1)h}^{th} \exp(-\kappa(th-s))dL(s) - \sigma i\psi'(0)\frac{1-\exp(-\kappa h)}{\kappa}\right).
\end{aligned} \tag{10}$$

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 (10), we can rewrite (7) as:

$$\begin{aligned}
y_t &= \rho_h(\kappa)y_{t-1} + g_h + \epsilon_t, y(0) = y_0 = O_p(1), \\
\epsilon_t &= \sigma_\psi\lambda_h u_t.
\end{aligned} \tag{11}$$

3.2 Estimation

In Model (7), we use the LS method to estimate $\rho_h(\kappa)$ and then obtain the estimator of κ by

$$\hat{\kappa}_h = -\ln(\widehat{\rho}_h(\kappa))/h, \tag{12}$$

where

$$\widehat{\rho}_h(\kappa) = \frac{T\sum_{t=1}^T y_{t-1}y_t - \sum_{t=1}^T y_t \sum_{t=1}^T y_{t-1}}{T\sum_{t=1}^T y_{t-1}^2 - \left(\sum_{t=1}^T y_{t-1}\right)^2}. \tag{13}$$

Define

$$\hat{g}_h = \frac{\sum_{t=1}^T y_t \sum_{t=1}^T y_{t-1}^2 - \sum_{t=1}^T y_{t-1}\sum_{t=1}^T y_{t-1}y_t}{T\sum_{t=1}^T y_{t-1}^2 - \left(\sum_{t=1}^T y_{t-1}\right)^2}. \tag{14}$$

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

$$z\left(Y,\rho,T\right)=T\left(\widehat{\rho}_h(\kappa)-\rho_h(\kappa)\right)\ \text{ and } t\left(Y,\rho,T\right)=\frac{\widehat{\rho}_h(\kappa)-\rho_h(\kappa)}{\hat{\sigma}_{\widehat{\rho}_h}},\tag{15}$$

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

The coefficient-based statistic for κ can be constructed similarly as,

$$z\left(Y,\kappa,h\right)=N\left(\hat{\kappa}_h-\kappa\right).\tag{16}$$

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

$$\hat{\kappa}_h-\kappa=\varsigma_h(\widehat{\rho}_h(\kappa))-\varsigma_h(\rho_h(\kappa))=\varsigma_h'(\widetilde{\rho}_h(\kappa))(\widehat{\rho}_h(\kappa)-\rho_h(\kappa)),\tag{17}$$

where $\widetilde{\rho}_h(\kappa)$ is a value between $\widehat{\rho}_h(\kappa)$ and $\rho_h(\kappa)$. We therefore can write:

$$\frac{T}{\varsigma_h'(\rho_h(\kappa))}(\hat{\kappa}_h-\kappa)=\left(1+\frac{\varsigma_h'(\tilde{\rho}_h(\kappa))-\varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right)T(\widehat{\rho}_h(\kappa)-\rho_h(\kappa)).\tag{18}$$

This implies

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

This functional relationship is used to show that a valid CI for κ can be constructed.

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 $t_T\left(Y,\kappa\right)=h(\hat{\kappa}_h-\kappa)/\hat{\sigma}_{\widehat{\rho}_h}$, 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

We now extend the in-fill asymptotic result of Zhou and Yu (2015) to Model (6).

Theorem 3.1 *For Model (6), define $z\left(Y,\kappa,h\right)$ by (16). Then, as $h\to 0$,*

$$z\left(Y,\kappa,h\right)\Rightarrow z^{y_0}(\kappa,\theta):=-\frac{\Upsilon_3-\Upsilon_2\int_0^1 dW(r)}{\Upsilon_1-\Upsilon_2^2},\tag{20}$$

where $\theta = (\mu, \sigma, \psi'(0), \psi''(0))$, *and*

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

This limiting distribution in (20) allows us to invert the coefficient-based statistic and construct (infeasible) CIs for κ. It can be seen that when we have an error term involving a Lévy process, the Lévy exponent enters the limiting distribution through σ_ψ and $\psi'(0)$. The approach is infeasible as there are a number of unknown parameters in the limiting distribution in (20), including κ, μ, σ, $\psi'(0)$, $\psi''(0)$.

Remark 3.2 *If Model (6) 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 (20) 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 (6) 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 Theorem 3.1. If* $y_0 = \mu$, *subtract* y_0 *both side in equation (7), 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, Theorem 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* $z^{y_0}(\kappa, \theta) = -\frac{\int_0^1 \overline{W}(r) dW(r)}{\int_0^1 \overline{W}(r)^2 dr}$ *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 Equation (3) 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 (20) is expected to perform better than $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2dr$ 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 denotes $-\int_0^1 \overline{J}_c(r)dW(r)/\int_0^1 \overline{J}_c(r)^2dr$.

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

		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 inference from the discrete-time local-to-unity model with intercept is similar to making inference in the continuous time model (6) by restricting $\mu = 0, y_0 = 0$ or $y_0 = \mu$. From the simulation results, it can be clearly seen that the distribution depends on the initial condition, and it is expected that the in-fill distribution $z^{y_0}(\kappa, \theta)$ should outperform the distribution 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 a fixed h, it can be shown that the long-span asymptotic distribution of $t(Y, \kappa, h)$ is $N(0, 1)$ when $\kappa > 0$, whereas it becomes $-\frac{\int_0^1 \overline{W}(r)dW(r)}{\sqrt{\int_0^1 \overline{W}(r)^2dr}}$ 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.*

3.4 Finite sample performance of in-fill distribution

We design several Monte Carlo experiments to compare the accuracy of the in-fill theory relative to the long-span theory. Discrete data with sampling interval h are generated from Model (6) where the Lévy process is set to a variance gamma process (also known as the Laplace motion) with $v = 0.5$; see Madan and Seneta (1990) and Madan et al. (1998) for definition of the variance gamma distribution and the variance gamma process. The following parameter settings are considered, $\kappa \in \{0.01, 0.1, 1\}$, $h \in \{1/12, 1/52\}$, $N = 5$, $\mu = 0.1$, $\sigma = 1$, $i\psi'(0) = 0.05$, $\psi''(0) = 1$, $y_0 = 0.1$. The 1%, 5%, 10%, 50%, 90%, 95%, and 99% percentiles of $N(\hat{\kappa}_h - \kappa)$ are obtained from 10,000 replications and reported in Tables 2-4. For the purpose of comparison, we also report the same set of percentiles of the in-fill asymptotic distribution and the long-span asymptotic distribution.

Table 2: Percentile of $N(\hat{\kappa}_h - \kappa)$ when $\kappa = 0.01$

$h = 1/12$	1%	5%	10%	50%	90%	95%	99%
Finite	-1.303	0.009	0.703	4.356	11.758	14.638	22.164
In-fill	-1.266	0	0.780	4.371	11.236	13.957	20.164
Long-span	-0.212	-0.150	-0.117	0	0.117	0.150	0.212
$h = 1/52$	1%	5%	10%	50%	90%	95%	99%
Finite	-1.177	0.040	0.780	4.336	11.507	14.306	21.371
In-fill	-1.070	0.094	0.818	4.349	11.315	14.021	21.002
Long-span	-0.102	-0.072	-0.056	0	0.056	0.072	0.102

Table 3: Percentile of $N(\hat{\kappa}_h - \kappa)$ when $\kappa = 0.1$

$h = 1/12$	1%	5%	10%	50%	90%	95%	99%
Finite	-1.520	-0.063	0.683	4.479	12.002	15.216	22.248
In-fill	-1.419	-0.041	0.717	4.449	11.503	14.091	20.057
Long-span	-0.674	-0.477	-0.371	0	0.371	0.477	0.674
$h = 1/52$	1%	5%	10%	50%	90%	95%	99%
Finite	-1.303	-0.041	0.771	4.424	11.633	14.616	21.504
In-fill	-1.197	0.019	0.784	4.419	11.448	14.325	21.256
Long-span	-0.323	-0.228	-0.178	0	0.178	0.228	0.323

Table 4: Percentile of $N(\hat{\kappa}_h - \kappa)$ when $\kappa = 1$

$h = 1/12$	1%	5%	10%	50%	90%	95%	99%
Finite	-3.452	-1.874	-0.911	4.013	13.151	17.048	26.093
In-fill	-3.376	-1.803	-0.953	3.531	11.276	14.136	19.934
Long-span	-2.215	-1.566	-1.220	0	1.220	1.566	2.215
$h = 1/52$	1%	5%	10%	50%	90%	95%	99%
Finite	-3.198	-1.750	-0.814	3.916	12.603	15.909	23.044
In-fill	-3.200	-1.863	-0.953	3.755	12.062	15.161	22.121
Long-span	-1.030	-0.728	-0.567	0	0.567	0.728	1.030

In terms of finite sample approximation, a sharp contrast can be seen in the performance of the two asymptotic distributions. The long-span asymptotic distributions are always symmetric, while skewness can be seen clearly in the finite sample distributions. Moreover, the percentiles in the finite sample distributions are very different from their counterparts in the long-span distributions. These simulation results suggest that the long-span theory performs poorly in finite samples. The in-fill distributions, on the other hand, perform much better to approximate the finite sample distributions. The superior performance of the in-fill asymptotic theory motivates us to justify the bootstrap method under the in-fill scheme and to do expansions about the in-fill distribution $z^{y_0}(\kappa, \theta)$. By developing expansions about the in-fill asymptotic distribution, we can obtain a finite sample refinement which allows us to outperform the in-fill theory. Moreover, although the in-fill distribution provides a good finite sample approximation, it is infeasible in practice. In our model (7), κ, μ, σ and $\psi'(0)$ and $\psi''(0)$ all enter the in-fill asymptotic distribution (20).

4 Confidence Interval for κ

In this section, we first show how to use the grid bootstrap to construct $c_T^*(q|\kappa)$ from which we obtain BCIs. Then we formally provide asymptotic justification to the grid bootstrap and show its finite sample refinement of the in-fill distribution by stochastic expansions.

4.1 Grid bootstrap confidence interval

We propose to use the grid bootstrap to obtain BCIs. The parametric grid bootstrap was first proposed by Andrew (1993) in the context of AR(1) model with a Gaussian error while the nonparametric grid bootstrap was first proposed by Hansen (1999) under the local-to-unity AR(1) model. To the best of our knowledge, the grid bootstrap has never been applied to continuous time models. Here we show how to use two grid bootstrap procedures, (a) to generate parametric bootstrap samples and (b) to generate non-parametric bootstrap samples.

Consider generating the following AR(1) pseudo time series $\{y_t^*\}_{t=0}^T$ with error u_t^*:

$$y_t^* = \rho_h(\kappa) y_{t-1}^* + \tilde{g}_h + \hat{\sigma}_c \lambda_h u_t^*, y^*(0) = y_0 = O_p(1), \tag{21}$$

where $\rho_h(\kappa) = \exp(-\kappa h)$. Let $\hat{\sigma}_c := \sqrt{\frac{1}{Th}\sum_{t=1}^T (y_t - \hat{g}_h - \widehat{\rho}_h(\kappa) y_{t-1})^2}$, $\lambda_h := \sqrt{\frac{1-\exp(-2\kappa h)}{2\kappa}}$, and $\tilde{g}_h$ is obtained from regressing $y_t - \rho_h(\kappa) y_{t-1}$ on a constant. Note that $\hat{\sigma}_c^2$ is a consistent estimator of σ_ψ^2 when h shrinks to 0. This result is presented in the following lemma.

Lemma 4.1 *Under Model (11), as* $h \to 0$*,*

$$\sup_{\sigma>0} \sup_{\kappa \in R} \Pr\left(\frac{\hat{\sigma}_c^2}{\sigma_\psi^2} - 1 > \epsilon \right) \to 0. \tag{22}$$

We obtain u_t^* 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} plus a constant by LS. Let $\{e_{x,t}\}_{t=1}^T$ be the LS residuals. In parametric bootstrap (a), we draw u_t^* from an i.i.d. $N\left(0, \hat{\sigma}_x^2/\hat{\sigma}_c^2\right)$, where $\hat{\sigma}_x^2 = \frac{1}{T}\sum_{t=1}^T e_{x,t}^2$. Following the strategy of proving the consistency of $\hat{\sigma}_c^2$, we can show $\hat{\sigma}_x^2 = \sigma_\psi^2 + O_p(T^{-1})$. This implies when h is small enough, u_t^* is an i.i.d. $N(0,1)$ sequence approximately. In nonparametric bootstrap (b), we first scale the residual $\{e_{x,t}\}_{t=1}^T$ by multiplying $1/\hat{\sigma}_c$, then we re-center the scaled residual. Finally, we draw u_t^* from the empirical distribution function of these re-centered and scaled residuals independently with replacement. Clearly, Equation (21) is a bootstrap version of Model (11) conditional on κ, with the same initial condition y_0.

We can then apply LS to bootstrap samples to obtain $\hat{\rho}^*$, $\hat{\kappa}^*(:= -\ln(\hat{\rho}^*)/h)$ and the bootstrap coefficient-based statistic $z\left(Y^*, \kappa, h\right) = N\left(\hat{\kappa}_h^* - \kappa\right)$ where $Y^* = \{y_{th}^*\}_{t=1}^T$ is a bootstrap sample. We define the BCI as in (4). Since κ is our parameter of interest, we express the BCI for κ as $CI_q^* = \{\kappa \in R : c_T^*(x_1|\kappa) \le z\left(Y, \kappa, h\right) \le c_T^*(x_2|\kappa)\}$, and $c_T^*(q|\kappa)$ is the quantile function of $z\left(Y^*, \kappa, h\right)$, $x_1 = (1-q)/2$ and $x_2 = 1-(1-q)/2$.

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 bootstrapped distribution with error term drawn from parametric (a) or non-parametric (b) method. Assume that*

1. *$\kappa_0 \in K$, where K is a compact set in the positive half line.*
2. *The increment of the Lévy process $L(t+h) - L(t)$ has a finite variance and bounded r^{th} moment with $r \in (2, 4]$.*
3. *$\mu, \sigma, i\psi'(0)$ and $\psi''(0)$ are all bounded by $C < \infty$.*

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

- $\sup\limits_{\kappa \in K} \sup\limits_{x} \left[\Pr\{z\left(Y, \kappa, h\right) < x|\kappa\} - \Pr^*\{z\left(Y^*, \kappa, h\right) < x|\kappa\}\right] \to 0;$
- $\inf\limits_{\kappa \in K} \Pr\{\kappa_0 \in CI_q^*|\kappa\} \to x_2 - x_1 = q.$

The first assumption requires the parameter space of κ to be compact in the nonnegative half line. In principle, obtaining in-fill asymptotic distribution does not require κ to be nonnegative. In most economic and financial models, however, focus has been placed on cases where $\kappa = 0$ and $\kappa > 0$. Therefore, we restrict our attention to the nonnegative region of κ. Assumption 2 and 3 effectively regulate the error term in the exact discrete time model, enabling us to apply the invariance principle to the sum of error terms.

Note that both results hold regardless of how the bootstrap sample is constructed, either parametrically via (a) or non-parametrically via (b). The first result shows that the distribution of the bootstrap statistic is closer 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^* is closer to q when the sampling interval is smaller. This theorem therefore justifies the grid bootstrap for being able to build a valid CI for κ asymptotically.

While we have made the asymptotic justification to the grid bootstrap under the in-fill asymptotic scheme, it is also possible to make the asymptotic justification of the grid bootstrap under the long-span scheme where h assumed to be fixed (therefore ρ_h and λ_h are also fixed) and $N \to \infty$. Hansen (1999) and later Mikusheva (2007) show that BCIs of ρ have correct coverage asymptotically when $N \to \infty$. It is easy to show that BCIs of κ have correct coverage asymptotically when $N \to \infty$. We choose not to justify the bootstrap by the long-span theory simply because the long-span distribution has a poor finite sample performance in the continuous time model that we consider.

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 (6), 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 $h \to 0$.*

4.3 Expansions and refinements

A very 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 follows Park (2003) and Mikusheva (2015) by developing the second order probabilistic expansions of the coefficient-based test statistic around the in-fill asymptotic distribution which is not only non-pivotal but also dependent on the initial condition. 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 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. 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 (6), 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 $z(Y, \kappa, h)$.*

$$z(Y, \kappa, h) = z^{y_0}(\kappa, \theta) + T^{-1/4} A + T^{-1/2} B + o_p(T^{-1/2}), \tag{23}$$

where the leading term $z^{y_0}(\kappa, \theta)$ is the in-fill asymptotic distribution given in (20), 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 the two bootstrap methods, we have the following results for distributional expansions. For the parametric grid bootstrap (a), we have

$$\Pr{}^*(z\left(Y^{*,P}, \kappa, h\right) < x) = \Pr(z^{y_0}(\kappa, \theta) < x) + o(T^{-1/2}), \tag{24}$$

where $Y^{*,P} = \{y_t^*\}_{t=0}^T$ is any parametric bootstrap sample.

For the nonparametric grid bootstrap (b), we have

$$\sup_x |\Pr{}^*(z\left(Y^{*,NP}, \kappa, h\right) < x) - \Pr(z\left(Y, \kappa, h\right) < x)| = o(T^{-1/2}), \tag{25}$$

where $Y^{*,NP} = \{y_t^*\}_{t=0}^T$ is any nonparametric bootstrap sample.

Remark 4.3 *When $\psi'(0) = 0$, $\psi''(0) = 1$, $y_0 = \mu$, $\kappa = 0$, $z^{y_0}(\kappa, \theta) = -\int_0^1 \overline{W}(r)dW(r)/\int_0^1 \overline{W}(r)^2 dr$. Equation (23) extends the result on G_n in Park (2003) from the unit root model without intercept to the unit root model with intercept. When $\psi'(0) = 0$, $\psi''(0) = 1$, $y_0 = \mu$, $z^{y_0}(\kappa, \theta) = -\int_0^1 \overline{J}(r)dW(r)/\int_0^1 \overline{J}(r)^2 dr$. Equation (23) 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 4.4 *According to (23), we have*

$$\Pr(z\left(Y, \kappa, h\right) < x) = \Pr(z^{y_0}(\kappa, \theta) < x) + O(T^{-1/2}), \tag{26}$$

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

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

Remark 4.5 *Comparing (25) with (26), the nonparametric grid bootstrap provides a second-order improvement compared with the in-fill asymptotic distribution.*

Remark 4.6 *According to (24) and (26), for the parametric grid bootstrap method, we have,*

$$\Pr(z\left(Y, \kappa, h\right) < x) = \Pr{}^*(z\left(Y^{*,P}, \kappa, h\right) < x) + O(T^{-1/2}). \tag{27}$$

Comparing (27) with (26), the nonparametric grid bootstrap provides a second-order improvement compared with the parametric grid bootstrap in the in-fill asymptotic approach.

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

$$dy_t = \kappa(\mu - y_t)dt + \sigma_t dL(t), \tag{28}$$

where $\sigma_t = g(t/T)$ and g is a measurable function on the interval $(0, 1]$ such that both the infimum and the supremum of g over $(0, 1]$ is bounded strictly above 0 and below infinity and g satisfies the Lipschitz condition except at a finite number of points of discontinuity. The discrete time model is given by

$$y_t = \rho_h(\kappa)y_{t-1} + \mu(1 - \exp(-\kappa h)) + \sigma_t \lambda_h u_t, \tag{29}$$

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

The in-fill asymptotic distribution for $N(\hat{\kappa}_h - \kappa)$ can be developed in this model. It turns out that one can apply the wild bootstrap principle with the grid bootstrap method to generate a bootstrap sample. Namely, we replace ϵ_t^* in the fourth step of the bootstrap method which will be discussed below with the product of the LS residual e_t and an i.i.d. random variable (so $\epsilon_t^* = e_t u_t^*$). Then, following a similar strategy in proving Theorem 4.1, one can show that this method has a correct probability coverage asymptotically. The in-fill asymptotic theory and the justification of the bootstrap method for this model can be obtained from authors upon request.

5 Simulation Studies

5.1 Implementation

Before we design experiments to check the performance of the grid bootstrap, we give the following 6 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 use $\{e_{th}\}_{t=1}^{T}$ to construct the consistent estimator for σ^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 an i.i.d. random variable (its distribution depends on whether bootstrap (a) or bootstrap (b) is adopted) for $t = 1, ..., T$. 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, the parametric grid bootstrap method and the nonparametric grid bootstrap method. To do so, we consider 3 parameter settings to generate data (called DGP1 to DGP3). In DGP1, we simulate discrete time observations with sampling interval h from Model (1). In this case, the feasible in-fill asymptotic distribution can be obtained by replacing the unknown κ, μ, and σ with their estimates. In DGP2 and DGP3, we simulate discrete time observations with sampling interval h from Model (6) where the Lévy motion is the variance gamma process with $v = 0.5$. In particular we set $\psi'(0) = 0$ and $\psi''(0) = 1$ in DGP2 and $i\psi'(0) = 0.05$ and $\psi''(0) = 1$ in DGP3. The following parameter settings are considered, $\kappa \in \{0.01, 0.1, 1\}$, $h \in \{1/12, 1/52\}$, $N = 5$, $\mu = 0.1$, $\sigma = 1$, $y_0 = 0.1$. The number of replications is always set at 10,000.

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

1. In-feasible 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. Clearly this approach is infeasible in practice. The CIs serve as a benchmark to evaluate the performance of other methods.

2. Feasible in-fill asymptotic distribution for DGP1. Since the in-fill distribution depends on κ, μ, σ, we replace them with their estimates.

3. Two grid bootstrap methods: (a) parametric and (b) non-parametric. To calculate BCIs we set the number of bootstrap iterations $B = 399$ with grid size $G = 50$.

4. 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. Tables 5-6 report the empirical coverage and the absolute difference between the nominal coverage and the empirical coverage for alternative methods when $h = 1/12$ and $h = 1/52$, respectively. 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 5: 95% and 90% Confidence Intervals ($h = 1/12$)

		$\kappa_0 = 0.01$		$\kappa_0 = 0.1$		$\kappa_0 = 1$	
DGP1	Long-span	0.018	(0.932)	0.059	(0.892)	0.265	(0.685)
	In-fill (infeasible)	0.938	(0.012)	0.937	(0.013)	0.916	(0.034)
	In-fill (feasible)	0.988	(0.038)	0.986	(0.036)	0.923	(0.027)
	Grid bootstrap (a)	0.954	(0.004)	0.954	(0.004)	**0.948**	(0.002)
	Grid bootstrap (b)	**0.952**	(0.002)	**0.952**	(0.002)	**0.948**	(0.002)
DGP2	Long-span	0.019	(0.931)	0.062	(0.889)	0.272	(0.679)
	In-fill (infeasible)	0.941	(0.009)	0.940	(0.060)	0.919	(0.031)
	Grid bootstrap (a)	0.956	(0.006)	0.955	(0.005)	**0.950**	(0)
	Grid bootstrap (b)	**0.952**	(0.002)	**0.953**	(0.003)	0.948	(0.002)
DGP3	Long-span	0.020	(0.93)	0.063	(0.887)	0.270	(0.68)
	In-fill (infeasible)	0.943	(0.007)	0.940	(0.060)	0.921	(0.029)
	Grid bootstrap (a)	0.954	(0.004)	0.953	(0.003)	**0.950**	(0)
	Grid bootstrap (b)	**0.953**	(0.003)	**0.951**	(0.001)	0.949	(0.001)

Table 6: Coverage of 95% Confidence Intervals ($h = 1/52$)

		$\kappa_0 = 0.01$		$\kappa_0 = 0.1$		$\kappa_0 = 1$	
DGP1	Long-span	0.009	(0.941)	0.027	(0.923)	0.135	(0.816)
	In-fill (infeasible)	**0.947**	(0.003)	**0.947**	(0.003)	0.945	(0.005)
	In-fill (feasible)	0.992	(0.042)	0.991	(0.041)	**0.948**	(0.002)
	Grid bootstrap (a)	0.959	(0.009)	0.957	(0.007)	0.947	(0.003)
	Grid bootstrap (b)	0.957	(0.007)	0.954	(0.004)	**0.948**	(0.002)
DGP2	Long-span	0.009	(0.941)	0.029	(0.921)	0.138	(0.812)
	In-fill (infeasible)	**0.950**	(0)	**0.950**	(0)	0.946	(0.004)
	Grid bootstrap (a)	0.960	(0.01)	0.959	(0.009)	**0.951**	(0.001)
	Grid bootstrap (b)	0.958	(0.008)	0.958	(0.008)	**0.951**	(0.001)
DGP3	Long-span	0.009	(0.941)	0.031	(0.919)	0.138	(0.812)
	In-fill (infeasible)	**0.949**	(0.001)	**0.950**	(0)	0.948	(0.002)
	Grid bootstrap (a)	0.959	(0.009)	0.959	(0.009)	0.951	(0.001)
	Grid bootstrap (b)	0.960	(0.01)	0.957	(0.007)	**0.950**	(0)

Several interesting conclusions can be found from Tables 5-6. 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 in Bao et al. (2017). Second, for the (infeasible) in-fill asymptotic theory, the empirical coverage is always close to the nominal one. However, the performance is worse when $h = 1/12$ than when $h = 1/52$, which is naturally expected. Again, this conclusion echoes that in Zhou and Yu (2015) and in Bao et al. (2017). Third, the (feasible) in-fill asymptotic theory tends to lead to worse empirical coverage than the (infeasible) in-fill asymptotic theory. Especially, when κ is closer to zero, the problem of over-coverage is serious. Finally, the two grid bootstrap methods always perform well, regardless of h and κ_0. In particular, when $h = 1/12$, they tend to outperform the (infeasible) in-fill asymptotic distribution. When $h = 1/52$, the performance of the bootstrap methods and the (infeasible) in-fill asymptotic distribution is comparable.

6 An Empirical Study

In this section, we use the proposed grid bootstrap methods to construct BCIs for κ in Model (1) and in Model (6) based on real monthly short Federal fund effective rate. The data are available from H-15 Federal Reserve Statistical Release and covers 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).

Assuming the initial condition y_0 is the same as the first observation, the LS estimator 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 $\hat{\kappa}_h = 0.1201$. Four

CIs of κ are constructed, based on the long-span asymptotic distribution, the feasible in-fill asymptotic distribution when the model is assumed to be (1), the proposed grid bootstrap methods under parametric and non-parametric settings, respectively. They are reported in Table 7. It can be seen that the CI constructed from the long-span distribution is very different from other CIs. It excludes 0, suggesting that we have to reject the null hypothesis of unit root under the long-span scheme. However, the other three CIs all contain 0, suggesting that we cannot reject the unit root hypothesis. The two BCIs are very close to each other, with similar width, left endpoint and right endpoint. If the model is assumed to be (1), we can obtain the CI by replacing the unknown κ, μ, and σ with their estimates in the in-fill asymptotic distribution. While the corresponding CI contains 0, it is much wider than the two BCIs. This finding is consistent with the over-coverage found in the simulation study.

Table 7: Coverage of 95% Confidence Intervals

	95% C.I.	90% C.I.
Long-span	(0.0852, 0.1551)	(0.0908, 0.1495)
In-fill (feasible)	(-0.2050, 0.2448)	(-0.1505, 0.2191)
Grid bootstrap (a)	(-0.0502, 0.2083)	(-0.0368, 0.1868)
Grid bootstrap (b)	(-0.0435, 0.2005)	(-0.0319, 0.1785)

7 Conclusion

In this paper, we have established the in-fill asymptotic distribution of the coefficient-based statistic for the persistence parameter in a Lévy-driven OU model. The in-fill asymptotic distribution is asymmetric and dependent on the initial condition and performs better than the long-span distribution under empirically realistic settings when they are used to approximate the finite sample approximation. It is not pivotal as it depends on unknown parameters. To make use of in-fill asymptotic distribution, these unknown parameters must be replaced with estimators. The feasible in-fill asymptotic distribution performs worse than the infeasible in-fill asymptotic distribution but still outperforms the long-span asymptotic distribution under empirically realistic settings.

Following Park (2003) and Mikusheva (2015), we then develop probabilistic expansions to the coefficient-based statistic around the in-fill distribution. The second-order expansions of the coefficient-based statistic provide refinement of the infeasible in-fill distribution up to order $o(T^{-1/2})$. We then show that the nonparametric bootstrap procedure of Hansen (1999) offers a second-order refinement of the infeasible in-fill distribution when $h \to 0$. The 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.

Monte Carlo studies reveal several important results. First, the CIs implied by the long-span asymptotic distribution are under coverage very seriously in all cases considered. Second, the CIs implied by the feasible in-fill asymptotic distribution are over coverage seriously unless

κ is large. In all cases the gird bootstrap method performs better than the feasible in-fill asymptotic theory and much better than the long-span theory.

Empirical applications to U.S. interest rate data show that the unit root hypothesis cannot be rejected by the bootstrap CIs and the CI obtained from the feasible in-fill asymptotic distribution, but has to be rejected by the CI obtained from the long-span asymptotic distribution. These differences can be well explained by the simulation results.

8 Appendix

8.1 Proof of Theorem 3.1 and Remark 3.4

Proof of Theorem 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 (11) by $\sigma_\psi \lambda_h$, and let $x_t = y_t / (\sigma_\psi \lambda_h)$, then we have $x_t = \rho_h x_{t-1} + \frac{g_h}{\sigma_\psi \lambda_h} + u_t$. Under the in-fill scheme, we have

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^{T} x_{t-1}^2 &\Rightarrow \Upsilon_1, \\ \frac{1}{T^{3/2}} \sum_{t=1}^{T} x_t &\Rightarrow \Upsilon_2, \\ \frac{1}{T} \sum_{t=1}^{T} x_{t-1} u_t &\Rightarrow \Upsilon_3. \end{aligned} \tag{30}$$

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

$$T(\hat{\rho}_h(\kappa) - \rho_h(\kappa)) = \frac{S(T,\kappa)}{R(T,\kappa)} \text{ and } t\left(Y, \rho, T\right) = \frac{S(T,\kappa)}{\sqrt{R(T,\kappa)}}. \tag{31}$$

Hence,

$$T(\hat{\rho}_h(\kappa) - \rho_h(\kappa)) = \frac{\frac{1}{T} \sum_{t=1}^T x_{t-1} u_t - \frac{1}{\sqrt{T}} \sum_{t=1}^T \epsilon_t \frac{1}{T^{3/2}} \sum_{t=1}^T x_t}{\frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 - \left(\frac{1}{T^{3/2}} \sum_{t=1}^T x_{t-1} \right)^2}. \tag{32}$$

Since $\hat{\kappa}_h = \frac{-\ln(\hat{\rho}_h(\kappa))}{h}$, applying the generalized Delta method (Theorem 1.12, Shao, 2003), and using the relationship in (19), $Th = N$, $\left(1 + \frac{\varsigma_h'(\tilde{\rho}_h(\kappa)) - \varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}\right) \to_p 1$, and $h \varsigma_h'(\widetilde{\rho}_h(\kappa)) \to_p -1$, we obtain the limiting result $z\left(Y, \kappa, h\right) \Rightarrow -\frac{\Upsilon_3 - \Upsilon_2 \int_0^1 dW(r)}{\Upsilon_1 - \Upsilon_2^2}$.

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

$$
\begin{aligned}
t(Y,\rho,T) &= \frac{\sum_{t=1}^{T} y_{t-1}\epsilon_t - \frac{1}{T}\sum_{t=1}^{T} y_{t-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t}{\sqrt{\hat{\sigma}^2\left(\sum_{t=1}^{T} y_{t-1}^2 - \frac{1}{T}\left(\sum_{t=1}^{T} y_{t-1}\right)^2\right)}} \\
&= \frac{\frac{1}{\hat{\sigma}^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t - \frac{1}{\hat{\sigma} T^{3/2}}\sum_{t=1}^{T} y_{t-1}\frac{1}{\hat{\sigma}\sqrt{T}}\sum_{t=1}^{T}\epsilon_t}{\sqrt{\frac{1}{\hat{\sigma}^2 T^2}\sum_{t=1}^{T} y_{t-1}^2 - \left(\frac{1}{\hat{\sigma} T^{3/2}}\sum_{t=1}^{T} y_{t-1}\right)^2}} \\
&= \frac{\sigma_\psi \lambda_h}{\hat{\sigma}_c\sqrt{h}}\left[\frac{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}u_t - \frac{1}{T^{3/2}}\sum_{t=1}^{T} x_{t-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t}{\sqrt{\frac{1}{T^2}\sum_{t=1}^{T} x_{t-1}^2 - \left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_{t-1}\right)^2}}\right].
\end{aligned} \tag{33}
$$

By Lemma 4.1, $\frac{\sigma_\psi\lambda_h}{\hat{\sigma}_c\sqrt{h}} \to_p 1$. Applying results in (30), we can obtain the limit of $t(Y,\rho,T)$.

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

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

We will show later $\frac{\varsigma_h'(\widetilde{\rho}_h(\kappa)) - \varsigma_h'(\rho_h(\kappa))}{\varsigma_h'(\rho_h(\kappa))}$ 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.

8.2 Proof of Lemma 4.1

Before we move on to prove Lemma 4.1, we need the following lemma to show that we can obtain a consistent estimator for g_h at the rate of $h^{-1/2}$.

Lemma 8.1 *For Model (11), let $\widehat{g}_h$ be the LS estimator. Then under the in-fill scheme, for any $\kappa \geq 0$, we have*

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

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

Proof. Using (11) and (14), we have

$$
\begin{aligned}
\hat{g}_h - g_h &= \frac{\sum_{t=1}^{T} y_{t-1}^2\sum_{t=1}^{T}\epsilon_t - \sum_{t=1}^{T} y_{t-1}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{T\sum_{t=1}^{T} y_{t-1}^2 - \left(\sum_{t=1}^{T} y_{t-1}\right)^2} \\
&= \sigma_\psi\lambda_h\left[\frac{\sum_{t=1}^{T} x_{t-1}^2\sum_{t=1}^{T} u_t - \sum_{t=1}^{T} x_{t-1}\sum_{t=1}^{T} x_{t-1}u_t}{T\sum_{t=1}^{T} x_{t-1}^2 - \left(\sum_{t=1}^{T} x_{t-1}\right)^2}\right].
\end{aligned}
$$

Therefore, we have

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

Note that $T = N/h$, and $\lambda_h/\sqrt{h} \to 1$. Using (30), we therefore establish the result in Lemma 8.1. ■

We now prove Lemma 4.1. Let the LS residual be $e_t = y_t - \hat{g}_h - \hat{\rho}_h(\kappa)y_{t-1}$ and

$$\begin{aligned}\hat{\sigma}_c^2 = \frac{1}{Th}\sum_{t=1}^{T} e_t^2 &= \frac{1}{Th}\sum_{t=1}^{T}\left(\epsilon_t + (g_h - \hat{g}_h) + (\rho_h(\kappa) - \hat{\rho}_h(\kappa))y_{t-1}\right)^2 \\ &= \frac{1}{Th}\sum_{t=1}^{T}\epsilon_t^2 + \frac{1}{Th}\sum_{t=1}^{T}(g_h - \hat{g}_h)^2 + (\rho_h(\kappa) - \hat{\rho}_h(\kappa))^2\frac{1}{Th}\sum_{t=1}^{T} y_{t-1}^2 \\ &\quad + 2(g_h - \hat{g}_h)\frac{1}{Th}\sum_{t=1}^{T}\epsilon_t + 2(g_h - \hat{g}_h)(\rho_h(\kappa) - \hat{\rho}_h(\kappa))\frac{1}{Th}\sum_{t=1}^{T} y_{t-1} \\ &\quad + 2(\rho_h(\kappa) - \hat{\rho}_h(\kappa))\frac{1}{Th}\sum_{t=1}^{T} y_{t-1}\epsilon_t.\end{aligned} \tag{34}$$

We now investigate the five terms on the right hand side of (34) one by one.

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

$$\frac{1}{Th}\sum_{t=1}^{T}(g_h - \hat{g}_h)^2 = \frac{(g_h - \hat{g}_h)^2}{h} = O_p(h) = o_p(1), \quad \text{(by Lemma 8.1)},$$

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

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

$$\begin{aligned}(g_h - \hat{g}_h)(\rho_h(\kappa) - \hat{\rho}_h(\kappa))\frac{1}{Th}\sum_{t=1}^{T} y_{t-1} &= \sqrt{Th^{-1}}(g_h - \hat{g}_h)(\rho_h(\kappa) - \hat{\rho}_h(\kappa))\sigma_\psi\frac{\lambda_h}{\sqrt{h}}\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_{t-1} \\ &= O_p(h^{1/2})O_p(T^{-1/2}) = o_p(1).\end{aligned}$$

And finally,

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

Thus,

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

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

8.3 Proof of Theorem 4.1 and Remark 4.1

Before proceeding to prove Theorem 4.1 and Remark 4.1, some notations are needed. Let us define $\epsilon_t^*=\hat{\sigma}_c\lambda_h u_t^*$ and a pair of statistics $(S^*(T,\kappa),R^*(T,\kappa))$ by

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

By construction, we have $z\left(Y^*,\rho,T\right)=S^*(T,\kappa)/R^*(T,\kappa)$ and $t\left(Y^*,\rho,T\right)=S^*(T,\kappa)/\sqrt{R^*(T,\kappa)}$. We first claim the following lemma.

Lemma 8.2 *Suppose $\kappa_0\in K$, where K is a compact set in the positive half line, then for every $\varepsilon>0$, we have:*

1. $\lim\limits_{h\to 0}\sup\limits_{\kappa\in K}P\left\{|\tilde{g}_h-g_h|>\varepsilon\right\}=0$;
2. $\sup\limits_{\kappa\in K}\sup\limits_{t}\left|\frac{1}{\hat{\sigma}}\left(\frac{y_t}{\sqrt{T}}-\frac{y_t^*}{\sqrt{T}}\right)\right|=o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon}),a.s.$;
3. $\sup\limits_{\kappa\in K}\sup\limits_{t}\left|\frac{y_t}{\hat{\sigma}\sqrt{T}}\right|=O_p(1)$;
4. $\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_T\left(\frac{t}{T}\right)u_t-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_T\left(\frac{t}{T}\right)u_t^*\right|=o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon}),a.s.$;
5. $\sup\limits_{\kappa\in K}\left|\frac{1}{\hat{\sigma}T}\sum_{t=1}^{T}y_{t-1}\epsilon_t-\frac{1}{\hat{\sigma}T}\sum_{t=1}^{T}y_{t-1}^*\epsilon_t^*\right|=o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon}),a.s.$;

6. $\sup_{\kappa\in K}\left|\frac{1}{T^2\hat{\sigma}^2}\sum_{t=1}^T y_{t-1}^2 - \frac{1}{T^2\hat{\sigma}^2}\sum_{t=1}^T y_{t-1}^{*2}\right| = o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon}), a.s.$;

7. $\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^T \epsilon_t \frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1} - \frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^T \epsilon_t^* \frac{1}{T^{3/2}}\sum_{t=1}^T y_{t-1}^*\right| = o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon}), a.s.$;

8. $\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^T y_{t-1} - \frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^T y_{t-1}^*\right| = o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon}), a.s.$;

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. Note that all the results in Lemma 8.2 can be applied to both the parametric grid bootstrap and the nonparametric grid bootstrap. To save space, we omit the proof of the nonparametric grid bootstrap. The only difference in the proof between two bootstrap procedures is in their convergence speeds in items 1-8 above. For the parametric bootstrap, the order is of $o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon})a.s.$, while for the nonparametric bootstrap, the order is of $o(T^{-\delta})a.s.$ with $\delta > 0$.[2]

Throughout the proof, we will utilize a strong approximation argument shown in the supplementary appendix in Lemma 2 of Mikusheva (2007). The lemma allows us to apply Skorohod embedding scheme and use a Brownian motion to approximate the normalized partial sum of the error term u_t. We restate the lemma for the sake of readability.

Let $S_t = \sum_{i=1}^t u_i$. Then we can construct a sequence of processes $\eta_T(t) = \frac{1}{\sqrt{T}}S_{\lfloor tT\rfloor}$ and a sequence of Brownian motions w_T on a common probability space. And we define $\frac{u_i^*}{\sqrt{T}} = w_T(i/T) - w_T(i-1/T)$. So for every $\varepsilon > 0, r > 2$ we have

$$\sup_{0\le t\le 1}|\eta_T(t) - w_T(t)| = o(T^{-\frac{1}{2}+\frac{1}{r}+\varepsilon}), a.s.$$

We now prove Lemma 8.2.

$$1.\ \ \tilde{g}_h - g_h = \frac{1}{T}\sum_{t=1}^T \epsilon_t \le \sup_{\kappa}|\sigma_\psi\lambda_h|\frac{1}{T}\sum_{t=1}^T u_t \to_{a.s.} 0.$$

Since $\sup_{\kappa}|\sigma_\psi\lambda_h|$ is bounded (by assumptions) and u_t is an i.i.d. random variable with

[2] The subtle difference indeed lies in the difference in Lemma 6 and Lemma 2 in Mikusheva (2007).

mean 0 and unit variance, the convergence is guaranteed by the ergodic theorem.

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

Also,

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

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,

$$\begin{aligned}
&\sup_\kappa\sup_t\left|\frac{1}{\hat{\sigma}}\left(\frac{y_{t-1}}{\sqrt{T}} - \frac{y_{t-1}^*}{\sqrt{T}}\right)\right| \\
&= \sup_\kappa\sup_t\left|(1+o_p(1))\left[\begin{array}{c}(\rho_h - 1)\sum_{i=1}^{t}\rho_h^{t-i-1}\left(\eta_T\left(\frac{i}{T}\right) - w_T\left(\frac{i}{T}\right)\right) + \eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right) \\ +\frac{g_h - \tilde{g}_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^{t}\rho_h^i\end{array}\right]\right| \\
&\leq \sup_\kappa(1+o_p(1))\left|\left[\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|\left(\frac{\rho_h - 1}{\rho_h}\sum_{i=1}^{t}\rho_h^{t-i} + 1\right)\right]\right| + \sup_\kappa\sup_t\left|\frac{g_h - \tilde{g}_h}{\hat{\sigma}\sqrt{T}}\sum_{i=1}^{t}\rho_h^i\right| \\
&\leq (1+o_p(1))\left|\left[\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|\sup_\kappa\left|\frac{\rho_h - 1}{\rho_h}\frac{1-\rho_h^t}{1-\rho_h} + 1\right|\right]\right| + \sup_\kappa\left|\frac{g_h - \tilde{g}_h}{\hat{\sigma}_c\sqrt{N}}\frac{\rho_h(1-\rho_h^T)}{1-\rho_h}\right| \\
&\leq (1+o_p(1))\left|\left[\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|\sup_\kappa\left|\frac{1}{\rho_h} + 1\right|\right]\right| + \sup_\kappa|g_h - \tilde{g}_h|\frac{C}{\hat{\sigma}_c\sqrt{N}} \\
&\leq (1+o_p(1))\left|\left[\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|(M+1)\right]\right| + \sup_\kappa|g_h - \tilde{g}_h|\frac{C}{\hat{\sigma}_c\sqrt{N}} \\
&= o(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}), a.s.
\end{aligned}$$

Since K is compact, and $\frac{1}{\rho_h} = e^{\kappa h}$, there exists a C such that $C > e^{\kappa h}$ for all $\kappa \in K$.

3. $$\sup_{\kappa} \sup_{t} \left| \frac{1}{\hat{\sigma}} \frac{y_t}{\sqrt{T}} \right|$$
$$= \sup_{\kappa} \sup_{t} \left| \frac{\sigma \lambda_h}{\hat{\sigma}} \left[(\rho_h - 1) \sum_{i=1}^{t} \rho_h^{t-i-1} \eta_T \left(\frac{i}{T} \right) + \eta_T \left(\frac{t}{T} \right) \right] + \frac{\rho_h^t y_0}{\hat{\sigma}\sqrt{T}} + \frac{g_h}{\hat{\sigma}\sqrt{T}} \sum_{i=1}^{t} \rho_h^i \right|$$
$$\leq \sup_{\kappa} \left[(1 + o_p(1)) \left(\frac{\rho_h - 1}{\rho_h} \sum_{i=1}^{t} \rho_h^{t-j} + 1 \right) \right] \sup_{t} |w_T(t)| + \sup_{\kappa} \left| \frac{y_0}{\hat{\sigma}_c \sqrt{N}} \right| + \frac{G_1 M}{\sigma \sqrt{N}}$$
$$= O_p(1).$$

4. See Lemma 4 c) in Mikusheva (2007).
5. Denote $\breve{g}_h = \frac{g_h}{\sigma_\psi \lambda_h}$ and we have

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

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

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

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

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

We now examine these terms one by one.

$$
\begin{aligned}
\sup_\kappa \sup_t |A| &= \sup_\kappa \sup_t \left| \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h} \frac{x_T}{\sqrt{T}} \eta_T(1) - \frac{\lambda_h^2}{h} \frac{x_T^*}{\sqrt{T}} w_T(1) \right| \\
&= \sup_\kappa \sup_t \left| (1 + o_p(1)) \left(\frac{x_T}{\sqrt{T}} \eta_T(1) - \frac{x_T^*}{\sqrt{T}} w_T(1) \right) \right| \\
&= \sup_\kappa \sup_t \left| (1 + o_p(1)) \left(\frac{x_T}{\sqrt{T}} (\eta_T(1) - w_T(1)) + \left(\frac{x_T}{\sqrt{T}} - \frac{x_T^*}{\sqrt{T}} \right) w_T(1) \right) \right| \\
&\leq (1 + o_p(1)) \left[\sup_\kappa \left| \frac{x_T}{\sqrt{T}} \right| \sup_t \left| \eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right) \right| + \sup_t \left| \frac{x_t}{\sqrt{T}} - \frac{x_t^*}{\sqrt{T}} \right| \sup_t \left| w_T(t) \right| \right] \\
&= o(T^{-\frac{1}{2} + \frac{1}{r} + \epsilon}), a.s.
\end{aligned}
$$

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

$$
\begin{aligned}
\sup_\kappa \sup_t |C| &= \sup_\kappa \sup_t \left| \frac{(\rho_h-1)\lambda_h^2}{h} \sum_{t=2}^{T} \frac{x_{t-1}^*}{\sqrt{T}} w_T\left(\frac{t}{T}\right) - \frac{(\rho_h-1)\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h} \sum_{t=2}^{T} \frac{x_{t-1}}{\sqrt{T}} \eta_T\left(\frac{t}{T}\right) \right| \\
&= (1+o_p(1)) \sup_\kappa |\rho_h - 1| \sup_t \sum_{t=2}^{T} \left[\begin{array}{c} \frac{x_{t-1}^*}{\sqrt{T}} \left(w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right) \\ + \eta_T\left(\frac{t}{T}\right)\left(\frac{x_{t-1}^*}{\sqrt{T}} - \frac{x_{t-1}}{\sqrt{T}}\right) \end{array} \right] \\
&\leq (1+o_p(1)) \sup_\kappa |-\kappa h + o(h^2)| T \left[\begin{array}{c} \sup_t \left|\frac{x_{t-1}^*}{\sqrt{T}}\right| \sup_t \left| w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| \\ + \sup_t \left|\eta_T\left(\frac{t}{T}\right)\right| \sup_t \left|\frac{x_{t-1}^*}{\sqrt{T}} - \frac{x_{t-1}}{\sqrt{T}}\right| \end{array} \right] \\
&\leq MN \left[\sup_t \left|\frac{x_{t-1}^*}{\sqrt{T}}\right| \sup_t \left| w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| + \sup_t \left|\eta_T\left(\frac{t}{T}\right)\right| \sup_t \left|\frac{x_{t-1}^*}{\sqrt{T}} - \frac{x_{t-1}}{\sqrt{T}}\right| \right] \\
&= o\left(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}\right), a.s.
\end{aligned}
$$

$$
\begin{aligned}
\sup_\kappa \sup_t |D| &= \sup_\kappa \sup_t \left| \frac{\lambda_h^2}{h} \sum_{t=2}^{T} \frac{u_t^*}{\sqrt{T}} w_T\left(\frac{t}{T}\right) - \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h} \sum_{t=2}^{T} \frac{u_t}{\sqrt{T}} \eta_T\left(\frac{t}{T}\right) \right| \\
&= \sup_t \left| (1+o_p(1)) \left[\sum_{t=2}^{T} \frac{u_t^*}{\sqrt{T}} w_T\left(\frac{t}{T}\right) - \sum_{t=2}^{T} \frac{u_t}{\sqrt{T}} \eta_T\left(\frac{t}{T}\right) \right] \right| \\
&= \sup_t \left| (1+o_p(1)) \left[\sum_{t=2}^{T} \frac{u_t^*}{\sqrt{T}} \left(w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right) + \sum_{t=2}^{T} \left(\frac{u_t^* - u_t}{\sqrt{T}}\right) \eta_T\left(\frac{t}{T}\right) \right] \right| \\
&\leq \sup_t \left| w_T\left(\frac{t}{T}\right)\right| \sup_t \left| w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| + \sup_t \left| w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| \sup_t \left|\eta_T\left(\frac{t}{T}\right)\right| \\
&= o(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}), a.s.
\end{aligned}
$$

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

$$
\begin{aligned}
\sup_\kappa \sup_t |F| &= \sup_\kappa \sup_t \left| \frac{\lambda_h^2}{h} \left(\frac{u_1^*}{\sqrt{T}}\right)^2 - \frac{\sigma_\psi^2 \lambda_h^2}{\hat{\sigma}_c^2 h} \left(\frac{u_1}{\sqrt{T}}\right)^2 \right| \\
&= \sup_\kappa \sup_t \left| (1+o_p(h)) \left[\left(\frac{u_1^*}{\sqrt{T}}\right)^2 - \left(\frac{u_1}{\sqrt{T}}\right)^2 \right] \right| \\
&= \sup_\kappa \sup_t \left| (1+o_p(h)) \left[\left(\frac{u_1^*}{\sqrt{T}}\right) - \left(\frac{u_1}{\sqrt{T}}\right) \right] \left[\left(\frac{u_1^*}{\sqrt{T}}\right) + \left(\frac{u_1}{\sqrt{T}}\right) \right] \right| \\
&\leq (1+o_p(h)) \sup_t \left| w_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t}{T}\right)\right| \left[\sup_t \left| w_T\left(\frac{t}{T}\right)\right| + \sup_t \left|\eta_T\left(\frac{t}{T}\right)\right| \right] \\
&= o(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}), a.s.,
\end{aligned}
$$

$$
\sup_\kappa \sup_t |G| \leq \sup_\kappa \left| (1+o_p(1)) \frac{g_h}{\hat{\sigma}^2 T} \sup_t \left| \eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| \right| = o(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}), a.s.
$$

Thus, we have established item 5.

6. $$\sup_{\kappa}\left|\frac{1}{T\hat{\sigma}^2}\sum_{t=1}^{T} y_{t-1}^2 - \frac{1}{T\hat{\sigma}^2}\sum_{t=1}^{T} y_{t-1}^{*2}\right| = \sup_{\kappa}\left|\frac{1}{T\hat{\sigma}^2}\sum_{t=2}^{T} y_{t-1}^2 - \frac{1}{T\hat{\sigma}^2}\sum_{t=2}^{T} y_{t-1}^{*2} + \frac{1}{T\hat{\sigma}^2}y_0^2 - \frac{1}{T\hat{\sigma}^2}y_0^2\right|$$
$$= \sup_{\kappa}\left|\frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=2}^{T}\left(\frac{x_{t-1}}{\sqrt{T}}\right)^2 - \frac{\lambda_h^2}{h}\sum_{t=2}^{T}\left(\frac{x_{t-1}^*}{\sqrt{T}}\right)^2\right|$$
$$= \sup_{\kappa}\left|(1+o_p(h))\left[\sum_{t=2}^{T}\left(\frac{x_{t-1}}{\sqrt{T}}\right)^2 - \sum_{t=2}^{T}\left(\frac{x_{t-1}^*}{\sqrt{T}}\right)^2\right]\right|$$
$$\leq (1+o_p(h))\sup_t\left|\frac{x_t}{\sqrt{T}} - \frac{x_t^*}{\sqrt{T}}\right|\left(\sup_t\left|\frac{x_t}{\sqrt{T}}\right| + \sup_t\left|\frac{x_t^*}{\sqrt{T}}\right|\right)$$
$$= o(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}), a.s.$$

7. $$\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\frac{1}{T^{3/2}}\sum_{t=1}^{T}y_{t-1} - \frac{1}{\hat{\sigma}^2\sqrt{T}}\sum_{t=1}^{T}\epsilon_t^*\frac{1}{T^{3/2}}\sum_{t=1}^{T}y_{t-1}^*\right|$$
$$= \sup_{\kappa\in K}\left|\frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c h}\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\sum_{t=1}^{T}\frac{u_t}{\sqrt{T}} - \frac{\lambda_h^2}{h}\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}\sum_{t=1}^{T}\frac{u_t^*}{\sqrt{T}}\right.$$
$$\left. + \frac{\sigma_\psi^2\lambda_h^2}{\hat{\sigma}_c^2 h}\sum_{t=1}^{T}\frac{u_t^*}{\sqrt{T}}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1} - \frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{t-1}^*\right)\right|$$
$$\leq \sup_{\kappa\in K}\sup_t\left|\frac{1}{T}\sum_{t=1}^{T}\frac{x_{t-1}}{\sqrt{T}}\right|\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right|$$
$$+ \frac{1}{T}\sup_t\left|w_T\left(\frac{t}{T}\right)\right|\sup_t\left|\eta_T\left(\frac{t}{T}\right) - w_T\left(\frac{t}{T}\right)\right| + o_{a.s.}(1)$$
$$= o(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}), a.s.$$

8. $$\sup_{\kappa\in K}\left|\frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^{T}y_{t-1} - \frac{1}{\hat{\sigma}T^{3/2}}\sum_{t=1}^{T}y_{t-1}^*\right| = \sup_{\kappa\in K}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{\hat{\sigma}}\left(\frac{y_{t-1}}{\sqrt{T}} - \frac{y_{t-1}^*}{\sqrt{T}}\right)\right]\right|$$
$$= o(T^{-\frac{1}{2}+\frac{1}{r}+\epsilon}), a.s.$$

9. $$\sup_{\kappa\in K}\Pr\{|z(Y,\rho,T) - z(Y^*,\rho,T)| > \epsilon\}$$
$$= \sup_{\kappa\in K}\Pr\left\{\left|\frac{S(T,\kappa)}{R(T,\kappa)} - \frac{S^*(T,\kappa)}{R^*(T,\kappa)}\right| > \epsilon\right\}$$
$$= \sup_{\kappa\in K}\Pr\left\{\left|\left(\frac{S(T,\kappa)}{R(T,\kappa)} - \frac{S^*(T,\kappa)}{R(T,\kappa)}\right) + \left(\frac{S^*(T,\kappa)}{R(T,\kappa)} - \frac{S^*(T,\kappa)}{R^*(T,\kappa)}\right)\right| > \epsilon\right\}$$
$$\leq \sup_{\kappa\in K}\Pr\left\{R(T,\kappa)^{-1}\left|S_n - S_n^*\right| + |S^*(T,\kappa)|\left|R(T,\kappa)^{-1} - R^*(T,\kappa)^{-1}\right| > \epsilon\right\}$$
$$\to 0.$$

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 $z(Y^*,\kappa,h)$.

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

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 us have 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:

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

■

Now we are in the position to show Theorem 4.1, i.e.,

$$\lim_{h\to0}\sup_{\kappa\in K}\sup_x\left|\Pr\{z(Y,\rho,T)<x\}-\Pr{}^*\{z(Y^*,\rho,T)<x\}\right|=0;$$

$$\lim_{h\to0}\sup_{\kappa\in K}\sup_x\left|\Pr\{z(Y,\kappa,h)<x\}-\Pr{}^*\{z(Y^*,\kappa,h)<x\}\right|=0;$$

$$\lim_{h\to0}\inf_{\kappa\in K}\Pr\{\kappa_0\in CI_q\}=x_2-x_2=q.$$

We only need to show that the distribution of $z(Y^*,\rho,T)$ is uniformly continuous. Note that we generate a bootstrap sample by based on the normal distribution

$$y_t^*=\sum_{i=1}^t\rho_h(\kappa)^{t-i}\epsilon_i^*+\rho_h(\kappa)^ty_0=\hat\sigma_c\lambda_h\sum_{i=1}^t\rho_h(\kappa)^{t-i}z_i^*+\rho_h(\kappa)^ty_0,$$

where y_t^* is constructed to be a linear sum of the standard normal distributed variables (plus the initial condition). This implies that both y_t^* and $z(Y^*,\rho,T)$ have continuous distribution functions uniformly. Therefore, we can establish:

$$\lim_{h\to0}\sup_{\kappa\in K}\sup_x\left|\Pr\{z(Y,\rho,T)<x\}-P_\kappa\{u_t^{\rho*}<x\}\right|=0.$$

Similarly, for $z(Y,\kappa,h)$, we have

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

From this, we have established that

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

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

8.4 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 8.3 (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*

$$\frac{1}{\sqrt{T}\sigma_z}\sum_{t=1}^{T}u_t=W(1)+\frac{1}{T^{1/4}}M(V)+\frac{1}{\sqrt{T}}N(V)+o_p(T^{-1/2}), \tag{35}$$

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

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

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

Lemma 8.4 (Mikusheva (2015), Theorem 1) *Suppose $c \leq 0$ and z_j satisfies the assumption in Lemma 8.3. 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:*

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{t-1}u_t =& \int_0^1 J_c(r)dW(r) + \frac{1}{T^{1/4}}J_c(1)M(V)\\ &+ \frac{1}{\sqrt{T}}\left(-c\int_0^1\int_0^r e^{c(r-s)}J_c(s)dV(s)dW(r) + J_c(1)N(V) + \frac{1}{2}M^2(V) - \frac{1}{2}U\right)\\ &+ o_p(T^{-1/2}).\end{aligned} \tag{36}$$

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{x}_t^2 =& \int_0^1 J_c^2(r)dr - \frac{2c}{\sqrt{T}}\int_0^1 J_c(r)\int_0^r e^{c(r-s)}J_c(s)dV(s)dr\\ &- \frac{1}{\sqrt{T}}\int_0^1 J_c^2(r)dV(r) + \frac{1}{\sqrt{T}}J_c^2(1)V - \frac{2\mu_3}{3\sqrt{T}}\int_0^1 J_c(r)dr + o_p(T^{-1/2}).\end{aligned} \tag{37}$$

$$\begin{aligned}\frac{1}{T^{3/2}}\sum_{t=1}^{T}\tilde{x}_t =& \int_0^1 J_c(r)dr - \frac{c}{\sqrt{T}}\int_0^1\int_0^r e^{c(r-s)}J_c(s)dV(s)dr - \frac{1}{\sqrt{T}}\int_0^1 J_c(r)dV(r)\\ &+ \frac{1}{\sqrt{T}}J_c(1)V - \frac{\mu_3}{3\sqrt{T}} + o_p(T^{-1/2}).\end{aligned} \tag{38}$$

Lemma 8.5 *Under Model 6, 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

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

Proof. By backward substitutions, we can write x_t as

$$\begin{aligned}x_t =& \sum_{j=1}^{t} e^{(t-j)c/T}z_j + \frac{b}{\sqrt{T}}\frac{e^{ct/T}-1}{e^{c/T}-1} + e^{ct/T}x_0 + o_p(T^{-1/2})\\ =& \tilde{x}_t + \frac{b}{\sqrt{T}}\frac{e^{ct/T}-1}{e^{c/T}-1} + e^{ct/T}x_0 + o_p(T^{-1/2}).\end{aligned} \tag{39}$$

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 $\frac{1}{T^{3/2}}\sum_{t=1}^{T} x_t$.

1. We now show the first claim in Lemma 8.5.

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

   The approximation to the first term is given in Lemma 8.4(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 8.3. For the third term, we have

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

2. To show the second claim of Lemma 8.5, 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 8.4. For the second term, as in Zhou and Yu (2015), we can write

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

For the third term, we have

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

Finally, for the last three terms, we have:

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

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

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

3. For the last claim, we have

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

By summing all three terms, we obtain the results in Lemma 8.5. ■

Now we are in the position to prove Theorem 4.2. 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 8.5, we obtain

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

and

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

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

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

where

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

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

For the second claim, we have

$$\Pr(z(Y^*,\kappa,T)<x)=\Pr(z^{y_0}\kappa,\theta<x)+o(T^{-1/2}).$$

This result follows from Park (2003) and Remark 2 to Remark 6 of Mikusheva (2015), since, under parametric bootstrap, $V(t)=0$, and U is independent of W. When comparing the distributional order of $z(Y,\kappa,T)$ and $z(Y^*,\kappa,T)$, the additional term from the expansion of $z(Y^*,\kappa,T)$ (under the normality in the parametric bootstrap setting) is of order of $o(T^{-1/2})$.

Finally, for the last claim in Theorem 4.2, following Theorem 3 in Mikusheva (2015), we can easily show the difference between the distribution of the coefficient-based statistic and that of the bootstrap statistic is of order $o(T^{-1/2})$.

References

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

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. Econometrica, 61(1), 139-165.

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.

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.

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

Shao, J. (2003). Mathematical statistics: Springer New York.

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, C. Y., & 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.