Journal of Econometrics 193 (2016) 35–53

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Double asymptotics for explosive continuous time models[☆]

Xiaohu Wang[a], Jun Yu[b,*]

[a] Department of Economics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

[b] School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Received 4 January 2012
Received in revised form
2 February 2016
Accepted 2 February 2016
Available online 15 February 2016

JEL classification:
C13
C22
G13

Keywords:
Explosive continuous time models
Lévy process
Moderate deviations from unity
Double asymptotics
Invariance principle
Initial condition

ABSTRACT

This paper establishes a double asymptotic theory for explosive continuous time Lévy-driven processes and the corresponding exact discrete time models. The double asymptotic theory assumes the sample size diverges because the sampling interval (h) shrinks to zero and the time span (N) diverges. Both the simultaneous and sequential double asymptotic distributions are derived. In contrast to the long-time-span asymptotics ($N \to \infty$ with fixed h) where no invariance principle applies, the double asymptotic distribution is derived without assuming Gaussian errors, so an invariance principle applies, as the asymptotic theory for the mildly explosive process developed by Phillips and Magdalinos (2007). Like the in-fill asymptotics ($h \to 0$ with fixed N) of Perron (1991), the double asymptotic distribution explicitly depends on the initial condition. The convergence rate of the double asymptotics partially bridges that of the long-time-span asymptotics and that of the in-fill asymptotics. Monte Carlo evidence shows that the double asymptotic distribution works well in practically realistic situations and better approximates the finite sample distribution than the asymptotic distribution that is independent of the initial condition. Empirical applications to real Nasdaq prices highlight the difference between the new theory and the theory without taking the initial condition into account.

1. Introduction

Following the recent global financial crisis, one of the worst financial crises in history, public policy makers and academic researchers alike have devoted much effort into finding the causes of this crisis. A widely believed cause is the birth and burst of the U.S. real estate bubble. Not surprisingly, the recent literature focuses on the econometric identification of bubbles; see, for example, Phillips et al. (2011), Phillips and Yu (2011), Homm and Breitung (2012), and Phillips et al. (2015a,b, 2014). A primary technique used in this literature relies on the asymptotic theory developed in Phillips and Magdalinos (2007, hereafter PM) for a mildly explosive discrete time model.[1] The asymptotic distribution of PM is empirically appealing as it does not rely on the assumption of Gaussian errors, unlike the conventional asymptotic theory for the standard explosive model developed in White (1958) and Anderson (1959). Explosive processes are used for bubble analysis because, according to the rational expectations theory, the presence of bubble implies the explosive sub-martingale property. In the discrete time autoregressive set-up, this property leads to an autoregressive root larger than unity; see Gurkaynak (2008) for a recent survey of the literature on bubbles. In an empirical study, based on a recursive method implemented in a discrete time model proposed in Phillips et al. (2011), Phillips and Yu (2011) analyzed the bubble episodes in various U.S. markets and documented the bubble migration mechanism during the financial crisis. It was found that the real estate bubble in the U.S. indeed predates the financial crisis.

However, it is well-known that the degree of deviations from unity is typically determined by data frequency in discrete time models. Consequently, the empirical results may be sensitive to the choice of data frequency. Another potential restriction in using the theory of PM is that the asymptotic distribution is independent

☆ We would like to thank two referees, an associate editor, Peter Robinson (the editor), Peter Phillips, and Yacine Aït-Sahalia for helpful comments. This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund Tier 2 grant with the MOE's official grant number MOE2011-T2-2-096.

* Corresponding author.
E-mail addresses: xiaohu.wang@cuhk.edu.hk (X. Wang), yujun@smu.edu.sg (J. Yu).

[1] The error term in PM is assumed to be a sequence of independent and identically distributed (i.i.d.) random variables. Magdalinos (2012) extended the asymptotic results of PM to the case where the errors are serially dependent. The asymptotics for the case in which the variance of the errors is infinite were established by Aue and Horvath (2007).

http://dx.doi.org/10.1016/j.jeconom.2016.02.014
0304-4076/

of the initial condition. The initial condition was assumed to be of a small order in PM, and therefore, the resulting asymptotic distribution may not provide an accurate approximation to the finite sample distribution when the initial value is large.

In this paper, we overcome the two aforementioned problems in the literature by developing a double asymptotic theory for an explosive continuous time Lévy-driven process. There are several important reasons leading us to focus on a continuous time Lévy-driven process. First, it is well-known that the persistence parameter in continuous time models does not depend on data frequency (Bergstrom, 1990). Therefore, inference about the explosive behavior in the continuous time framework is less sensitive to the choice of data frequency in empirical analysis. Second, the continuous time model provides a natural tool to accommodate an initial condition whose order is higher than that in PM. As a result, our asymptotic distribution explicitly depends on the initial condition. This feature is the same as in the in-fill asymptotics ($h \to 0$) developed in Phillips (1987a) and Perron (1991). Not surprisingly, we find that our asymptotic theory improves over that does not depend on the initial condition in the finite sample approximation. Third, the use of Lévy-driven process allows us to develop an invariance principle for the persistence parameter, thereby sharing the asymptotic property of PM. The invariance principle is desirable because in many empirical analyses of bubbles the assumption of Gaussian errors may not be realistic.

The results in our paper build upon and extend the important work of Perron (1991). Based on the in-fill asymptotic scheme, Perron established a connection between the continuous diffusion process and the local-to-unity process and derived an alternative approximation to the estimator of the autoregressive parameter. His asymptotic theory permits explicit consideration of the effects of different initial conditions.[2] In our paper, by letting $h \to 0$ and $N \to \infty$ simultaneously, we build a link between the continuous time model and the discrete time autoregressive model with root moderately deviated from unity. Like Perron (1991), our double asymptotic distribution for the explosive case explicitly depends on the initial condition. However, when the process is stationary, as expected, no role for the initial condition is found in the double asymptotics.

Instead of focusing on the Brownian-motion-driven diffusion process as in Perron (1991), we consider the continuous time models driven by the Lévy process. Not only is our model empirically more realistic, but it also allows for the establishment of an invariance principle. Moreover, we derive two types of sequential asymptotics ($N \to \infty$ followed by $h \to 0$, and $h \to 0$ followed by $N \to \infty$) to bridge the gap among the simultaneous double asymptotics, the long-time-span asymptotics, and the in-fill asymptotics.

The results in our paper also extend the seminal work of PM (2007). In the explosive case, our double asymptotic scheme leads to a mildly explosive autoregressive model that has an initial condition with a higher order of magnitude than that in PM. That is why our asymptotic distribution depends on the initial condition, unlike the asymptotic distribution in PM. Extensive simulations show that our asymptotic distribution provides better approximations to the finite sample distribution. With a larger initial condition, one may be able to extend PM's asymptotic theory so that the modified asymptotic distribution depends on the initial condition. However, such a new asymptotic distribution depends on some nuisance parameters which cannot be consistently estimated. In the double asymptotic theory developed in the present study, the nuisance parameters are either known by the setting of the data structure or consistently estimable, making pivotal limit theory possible in continuous time models.

Our study also closely relates to the continuous time literature developed in statistics. In this literature, the least squares (LS) estimator, which is based on the Euler approximation to continuous time models, is often used to estimate the persistence parameter, when only discrete observations are available. The discretization error introduced by the Euler approximation has some important implications. For example, in developing the simultaneous double asymptotics for the stationary case, an extra condition that governs the relative convergence rates of N and h is needed in order to control the size of the discretization error; see, for example, Shimizu (2009) and Hu and Long (2009). In the explosive case, Shimizu (2009) showed that no asymptotic distribution can be derived because the size of the discretization error cannot be well controlled any more. However, our estimator of the persistence parameter is constructed directly from the exact discretized model, and hence, not subject to the discretization error. Consequently, in the stationary case, there is no need to impose an extra condition to control the joint behavior of N and h. More importantly, in the explosive case, we can derive a double asymptotic distribution for our estimator.

The rest of the paper is organized as follows. Section 2 introduces the model, builds the connection between our model and the mildly explosive process of PM, and discusses the relationship between our results and those in the literature. Section 3 develops the simultaneous double asymptotics. Two types of sequential double asymptotics are established in Section 4. In Section 5, we use simulated data to check the performance of our asymptotic theory and use real data to highlight the implications of our theory for statistical inference. Section 6 concludes. All the proofs are collected in the Appendix.

2. The model and the estimator

2.1. The model

The model studied in the paper is an Ornstein–Uhlenbeck (OU) diffusion process of the form:

$$dy(t) = \kappa \left(\mu - y(t)\right) dt + \sigma dL(t), \qquad y(0) = y_0 = O_p(1), \quad (2.1)$$

where $L(t)$ is a Lévy process defined on a filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P\right)$ with $L(0) = 0$ *a.s.*, $\mathcal{F}_t = \sigma\left\{[y(s)]_{s=0}^{t}\right\}$, and satisfies the following three properties:

1. Independent increments: for every increasing sequence of times $t_0, \ldots, t_n$ the random variables $L(t_0), L(t_1) - L(t_0), \ldots, L(t_n) - L(t_{n-1})$ are independent;
2. Stationary increments: the law of $L(t+h) - L(t)$ is independent of t;
3. Stochastic continuity: for any $\varepsilon > 0$, $t \geq 0$, $\lim_{h\to 0} P(|L(t+h) - L(t)| \geq \varepsilon) = 0$.[3]

The initial value y_0 is assumed to be independent of $L(t)$.

[2] Phillips (1987a) established the in-fill asymptotics for the unit root case (i.e., setting the persistence parameter κ to be zero) to take into account the effect of the initial condition in the limiting distribution. In Phillips (1987b) the in-fill asymptotics were established for the case where $\kappa \neq 0$ and the initial condition is set to be zero.

[3] This property allows existence of jumps happening at random times in sample path. As an effective way to introduce discontinuity into sample path, various Lévy processes have been developed in the asset pricing literature; see, for example, Barndorff-Nielsen (1998), Madan et al. (1998), and Carr and Wu (2003).

The characteristic function of $L(t)$ takes the form of $E(\exp\{isL(t)\}) = \exp\{-t\psi(s)\}$, where i is the imaginary unit and the function $\psi(\cdot) : \mathbb{R} \to \mathbb{C}$ is called the Lévy exponent of $L(t)$. When a Lévy process is square-integrable, the first two moments can be calculated from the derivatives of the Lévy exponent as

$$i\psi'(0) = E[L(1)] = \frac{E[L(t)]}{t},$$

and

$$\psi''(0) = Var[L(1)] = \frac{Var[L(t)]}{t}.$$

Data in economics and finance are typically available at discrete points in time, say at T equally spaced points $\{th\}_{t=1}^{T}$ over a time interval $(0, N]$. So N is the time span and h represents the sampling interval. The sample size $T = N/h$ diverges when $N \to \infty$ or $h \to 0$ or both. The exact discrete time model corresponding to (2.1) is

$$y_{th} = e^{-\kappa h} y_{(t-1)h} + \mu\left[1 - e^{-\kappa h}\right] + \sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)} dL(s), \quad (2.2)$$

where the integral is defined as the L_2 limit of approximating Riemann–Stieltjes sums.

The first two moments of the errors in Eq. (2.2) are

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

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

Moreover, the characterization of the Lévy process makes the errors an i.i.d. sequence with the distribution depending on the specification of the Lévy measure.

For any fixed h, Eq. (2.2) gives a discrete time first-order autoregressive (AR(1)) model. The sign of κ determines whether the process is stationary or explosive, since $\kappa > 0$ and $\kappa < 0$ imply that $e^{-\kappa h} < 1$ and $e^{-\kappa h} > 1$ respectively. That is why κ captures the degree of persistence of $y(t)$. If $\mu = 0$, there is no intercept in (2.2). For this reason, we call μ the intercept parameter in this paper.

In the special case where $L(t)$ is a standard Brownian motion, denoted as $W(t)$, the stochastic process (2.1) is interpreted as an Itô equation with solution $\{y(t), t \geq 0\}$ satisfying

$$y(t) = e^{-\kappa h} y(0) + \mu\left[1 - e^{-\kappa h}\right] + \sigma \int_0^t e^{-\kappa(t-s)} dW(s). \quad (2.3)$$

The corresponding exact discrete time model is

$$y_{th} = e^{-\kappa h} y_{(t-1)h} + \mu\left[1 - e^{-\kappa h}\right] + \sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)} dW(s). \quad (2.4)$$

In Model (2.4) the errors follow the Gaussian distribution, $N\left(0, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa h})\right)$.

Although much of the literature focuses on the continuous time Brownian motion-driven diffusions, there are important reasons for us to study the Lévy-driven models. First, the Lévy measure can generate heavier tails than the Brownian motion. This generalization is relevant to the bubble analysis as many applications are based on asset prices.[4] Second, the use of the Lé vy process allows us to derive an invariance principle.

[4] However, our model specification does not allow for stochastic volatility, a stylized fact commonly found in asset prices.

2.2. The estimator

Our primary goal in this paper is to develop a double asymptotic theory for the LS estimator of κ in the continuous time model (2.1) and the LS estimator of $e^{-\kappa h}$ in the corresponding discrete time model (2.2) when the process (2.1) is explosive with $\kappa < 0$. We do so by assuming either $h \to 0$ and $N \to \infty$ simultaneously (termed the "simultaneous double asymptotics" in the paper) or $h \to 0$ and $N \to \infty$ sequentially with different orders (termed the "sequential double asymptotics").

To simplify the discrete time representation (2.2) and to facilitate the linking of our model to the model of PM, we first introduce a few notations. Let

$$a_h(\kappa) := \exp\{-\kappa h\} = \exp\{-\kappa/k_T\} \quad \text{with } k_T = 1/h,$$

$$g_h := \left[\mu + \frac{\sigma i\psi'(0)}{\kappa}\right]\left[1 - e^{-\kappa h}\right],$$

$$\lambda_h := \sigma\sqrt{\psi''(0)\left[1 - e^{-2\kappa h}\right]/2\kappa},$$

$$\varepsilon_{th} := \lambda_h^{-1}\left(\sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)} dL(s) - \sigma i\psi'(0)\,\frac{1 - e^{-\kappa h}}{\kappa}\right).$$

Then, the discrete time model (2.2) can be rewritten as

$$y_{th} = a_h(\kappa) y_{(t-1)h} + g_h + \lambda_h \varepsilon_{th}, \qquad y_{0h} = y_0 = O_p(1). \quad (2.5)$$

Whenever h is fixed, $\{\varepsilon_{th}\}_{t=1}^{T}$ is a sequence of i.i.d. random variables with zero mean and unit variance. The distribution of ε_{th} depends on the specification of $L(t)$. When h varies, the errors $\{\varepsilon_{th}\}$ form a martingale-difference array. This is because in general the distribution of ε_{th} may depend on the sampling interval, although the first two moments remain unaffected. The assumption on the independence between y_0 and $L(t)$ makes y_0 independent of $\sigma(\varepsilon_{1h}, \ldots, \varepsilon_{Th})$, the σ-field generated by $(\varepsilon_{1h}, \ldots, \varepsilon_{Th})$, for any h.

Let $\sum$ denote $\sum_{t=1}^{T}$. The LS estimators of the intercept and the AR coefficient are

$$\begin{bmatrix} \widehat{g}_h \\ \widehat{a}_h(\kappa) \end{bmatrix} = \begin{bmatrix} T & \sum y_{(t-1)h} \\ \sum y_{(t-1)h} & \sum y_{(t-1)h}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_{th} \\ \sum y_{(t-1)h} y_{th} \end{bmatrix}, \quad (2.6)$$

and hence

$$\begin{bmatrix} \widehat{g}_h - g_h \\ \widehat{a}_h(\kappa) - a_h(\kappa) \end{bmatrix} = \begin{bmatrix} T & \sum y_{(t-1)h} \\ \sum y_{(t-1)h} & \sum y_{(t-1)h}^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \lambda_h \sum \varepsilon_{th} \\ \lambda_h \sum y_{(t-1)h}\varepsilon_{th} \end{bmatrix}. \quad (2.7)$$

The LS estimator of κ in Model (2.1) considered in the paper is

$$\widehat{\kappa} = -\frac{1}{h}\ln\left(\widehat{a}_h(\kappa)\right). \quad (2.8)$$

The corresponding t-statistics for $a_h(\kappa)$ and g_h in Model (2.5) are

$$t_{a_h(\kappa)} = \frac{\widehat{a}_h(\kappa) - a_h(\kappa)}{\widehat{\sigma}_{\widehat{a}_h(\kappa)}} = \frac{\left[\widehat{a}_h(\kappa) - a_h(\kappa)\right]\left\{T\sum y_{(t-1)h}^2 - \left(\sum y_{(t-1)h}\right)^2\right\}^{1/2}}{\left\{T \cdot \widehat{\lambda}_h^2\right\}^{1/2}}, \quad (2.9)$$

and

$$t_{g_h} = \frac{\widehat{g}_h - g_h}{\widehat{\sigma}_{\widehat{g}_h}} = \frac{\left[\widehat{g}_h - g_h\right]\left\{T\sum y_{(t-1)h}^2 - \left(\sum y_{(t-1)h}\right)^2\right\}^{1/2}}{\left\{\widehat{\lambda}_h^2 \sum y_{(t-1)h}^2\right\}^{1/2}}, \quad (2.10)$$

respectively, where

$$\widehat{\lambda_h}^2 = \frac{1}{T}\sum \left[y_{th} - \widehat{a}_h(\kappa) y_{(t-1)h} - \widehat{g}_h\right]^2. \tag{2.11}$$

When $h \to 0$, the variance of the error term in Eq. (2.5) goes to zero as $\lambda_h \sim O_p\left(\sqrt{h}\right)$. To facilitate the comparison with PM's model where the variance of the error term is a positive constant, a standardization is needed. Letting $x_{th} = y_{th}/\lambda_h$ and $\tilde{g}_h = g_h/\lambda_h$, dividing both sides of Eq. (2.5) by λ_h, we have

$$x_{th} = a_h(\kappa) x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th} \quad \text{with } x_{0h} = y_{0h}/\lambda_h. \tag{2.12}$$

The error term in Model (2.12) has an unit variance for any h. When $h \to 0$ and $N \to \infty$ simultaneously, we have

$$k_T = \frac{1}{h} \to \infty, \qquad \frac{k_T}{T} = \frac{1}{N} \to 0,$$

$$a_h(\kappa) = \exp\{-\kappa/k_T\} = 1 - \kappa/k_T + O(k_T^{-2}) \to 1.$$

Clearly, the AR coefficient $a_h(\kappa)$ converges to unity at a rate slower than $1/T$. Using the terminology of PM, this model contains moderate deviations from unity because the AR roots belong to a larger neighborhood of one than the conventional local-to-unity case. The LS estimators and the corresponding t-statistics for the parameters in Model (2.12) are:

$$\begin{bmatrix} \widehat{\tilde{g}}_h - \tilde{g}_h \\ \widehat{a}_h(\kappa) - a_h(\kappa) \end{bmatrix} = \begin{bmatrix} T & \sum x_{(t-1)h} \\ \sum x_{(t-1)h} & \sum x_{(t-1)h}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum \varepsilon_{th} \\ \sum x_{(t-1)h}\varepsilon_{th} \end{bmatrix} = \begin{bmatrix} [\lambda(h)]^{-1}(\widehat{g}_h - g_h) \\ \widehat{a}_h(\kappa) - a_h(\kappa) \end{bmatrix},$$

$$t_{a_h(\kappa)} = \frac{\left(\widehat{a}_h(\kappa) - a_h(\kappa)\right)\left\{T\sum x_{(t-1)h}^2 - \left(\sum x_{(t-1)h}\right)^2\right\}^{1/2}}{\left\{T \cdot \widehat{\sigma}_\varepsilon^2\right\}^{1/2}},$$

$$t_{\tilde{g}_h} = \frac{\left(\widehat{\tilde{g}}_h - \tilde{g}_h\right)\left\{T\sum x_{(t-1)h}^2 - \left(\sum x_{(t-1)h}\right)^2\right\}^{1/2}}{\left\{\widehat{\sigma}_\varepsilon^2 \cdot \sum x_{(t-1)h}^2\right\}^{1/2}} = t_{g_h},$$

where

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum\left(x_{th} - \widehat{a}_h(\kappa) x_{(t-1)h} - \widehat{\tilde{g}}_h\right)^2.$$

2.3. Relationship of our results to the literature

First of all, our analysis is closely related to PM who studied an AR(1) model with a root moderately deviated from unity as in the following form[5]

$$x_t = \rho_T x_{t-1} + \varepsilon_t, \qquad \rho_T = \left(1 + \frac{-\kappa}{k_T}\right), \qquad k_T \to \infty,$$
$$\frac{k_T}{T} \to 0, \qquad x_0 = o_p\left(\sqrt{k_T}\right), \qquad \varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2), \tag{2.13}$$

with $t = 1, \ldots, T$. In this model the root ρ_T belongs to a larger neighborhood of one than the conventional local-to-unity case. PM showed that, when $\kappa < 0$ and $T \to \infty$,

$$\frac{k_T \rho_T^T}{-2\kappa}(\widehat{\rho}_T - \rho_T) \Rightarrow \text{standard Cauchy}, \tag{2.14}$$

where $\widehat{\rho}_T$ is the LS estimator of ρ_T. A very important and nice feature in PM's theory is that the limiting result does not rely on the assumption of Gaussian errors, and thus an invariance principle applies. This result is in sharp contrast to the conventional limit theory developed in White (1958) and Anderson (1959) for the standard explosive model of

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \rho > 1,\ x_0 = 0,\ \varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2), \tag{2.15}$$

in which $\rho^T(\widehat{\rho} - \rho)/(\rho^2 - 1) \Rightarrow$ Cauchy with $\widehat{\rho}$ being the LS estimator of ρ. No invariance principle applies in this conventional limit theory because the assumption that $\varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$ is required.

Table 1
Comparison between Model (2.12) in the present paper and Model (1) in PM.

PM's model	Our model
$\rho_T = 1 + (-\kappa)/k_T$	$a_h(\kappa) = e^{-\kappa h} = 1 + (-\kappa/k_T) + O(k_T^{-2})$
$k_T \to \infty$, $k_T/T \to 0$	$k_T = 1/h \to \infty$, $k_T/T = 1/N \to 0$
$\{\varepsilon_t\} \overset{i.i.d.}{\sim} (0, \sigma^2)$	$\{\varepsilon_{th}\}$ form a martingale-difference array
$x_0 \sim o_p\left(\sqrt{k_T}\right)$	$x_0 \sim O_p\left(\sqrt{k_T}\right)$
No intercept	Intercept with order $O\left(\sqrt{1/k_T}\right)$

Table 1 compares our model given by (2.12) and PM's model given by (2.13). By letting $k_T = 1/h$, under the double asymptotic scheme, the AR root in our model is $a_h(\kappa) = e^{-\kappa h} = 1 + (-\kappa/k_T) + O(k_T^{-2})$ which converges to 1 at a slower rate than $1/T$ because $k_T/T = 1/N \to 0$. Therefore, our model also implies moderate deviations from a unit root. However, there are four differences between the two models. First, the AR root in our model (2.12) is $e^{-\kappa h} = 1 + (-\kappa/k_T) + O(k_T^{-2})$ rather than $1 + (-\kappa/k_T)$ as in PM. This difference is quite small when $k_T \to \infty$. Hence, it is expected to have no impact on the limiting distribution. Second, the errors $\{\varepsilon_{th}\}$ in our model form a martingale difference array as $k_T \to \infty$, whereas they are assumed to be i.i.d. random variables in PM. This difference has a technical implication for choosing an appropriate central limit theorem (CLT) to develop our limit theory.

More importantly, the third and fourth differences lie on the initial condition and the intercept term, respectively, and lead to a material change in the limit theory. In our model $\lambda_h = O\left(\sqrt{1/k_T}\right)$. When $h \to 0$, we have

$$x_0 = y_0/\lambda_h = O_p\left(\sqrt{k_T}\right) \quad \text{with } x_0/\sqrt{k_T} \xrightarrow{a.s.} y_0/\left(\sigma\sqrt{\psi''(0)}\right),$$

and

$$\tilde{g}_h = g_h/\lambda_h = O\left(\sqrt{1/k_T}\right) \quad \text{with}$$
$$\sqrt{k_T}\tilde{g}_h \to \left(\kappa\mu + \sigma i\psi'(0)\right)/\left(\sigma\sqrt{\psi''(0)}\right).$$

In contrast, PM's model given by (2.13) is an AR(1) process without intercept and its initial value is assumed to be $o_p\left(\sqrt{k_T}\right)$. As a result, neither the intercept term nor the initial condition appears in the limiting distribution developed in PM. However, as it becomes clear later, the consideration of the intercept term and the larger initial condition lead to a different limit theory which allows for the explicit consideration of the effects of the intercept term and the initial condition.

To see the implications of the initial condition and the intercept term, note that Model (2.12) can be equivalently expressed as

$$x_{th} = \sum_{s=0}^{t-1}[a_h(\kappa)]^s \varepsilon_{(t-s)h} + [a_h(\kappa)]^t x_{0h} + \frac{1 - [a_h(\kappa)]^t}{1 - a_h(\kappa)}\tilde{g}_h. \tag{2.16}$$

When the process is explosive and the double asymptotic treatment is considered, the last two items on the right side of Eq. (2.16) have the same order of $O_p\left(\sqrt{k_T}\,[a_h(\kappa)]^t\right)$. It becomes

[5] Park (2003) considered an autoregressive process with a weak unit root of the form $\rho = 1 - m/n$ where $m, n \to \infty$ and $m/n \to 0$, and provided the asymptotics for the stationary side of moderate deviations from unity.

clear later that the first term has the order of $O_p\left(\sqrt{k_T}\,[a_h(\kappa)]^t\right)$ too. This is the reason why both the intercept term and the initial condition are important in the double asymptotics. On the contrary, if it is assumed that $x_0 \sim o_p\left(\sqrt{k_T}\right)$ and $\tilde{g}_h = 0$ as done in PM, the first term on the right side of Eq. (2.16) dominates. As a result, not surprisingly, the limit theory in PM does not depend on the initial condition. It is reasonable to believe that the finite sample distribution for the explosive process should depend on the initial condition and the intercept term. Hence, we expect our limit theory to better approximate the finite sample distribution.

The motivation for us to develop a limit theory that explicitly depends on the initial condition comes directly from Perron (1991). Perron used a continuous time model and the in-fill asymptotic scheme to derive a new asymptotic theory for the local-to-unity process that explicitly depends on the initial condition. To build a link between our continuous time model and the discrete time model of PM, we let not only $h \to 0$ as requested by the in-fill asymptotic scheme but also $N \to \infty$. As it is shown in Table 1, in the explosive case, our double asymptotic scheme leads to an initial condition that is of greater order of magnitude than that in PM. It is this increase in the order of magnitude in the initial condition that makes the corresponding asymptotic distribution to be explicitly dependent on the initial condition. However, as it becomes clear later, no role for the initial condition is found in the double asymptotics for the stationary process. This is expected because, when a process is stationary, the effect of the initial condition decays quickly as $N \to \infty$.

Our paper is also closely related to the continuous time literature developed in statistics. To explain the connection between our results and those in the statistics literature, consider the following diffusion process

$$dy(t) = -\kappa y(t)dt + \sigma dW(t). \tag{2.17}$$

When a continuous record ($h = 0$) is available, the LS estimator of κ takes the form of

$$\hat{\kappa}_{MLE} = -\int_0^N y(t)dy(t) \Big/ \int_0^N y^2(t)d(t).$$

When $y(t)$ is observed at discrete time points, the following approximate estimator is often used in the statistics literature,

$$\widehat{\kappa}_{Euler} = -\sum y_{(t-1)h}\left(y_{th} - y_{(t-1)h}\right) \Big/ \left(h\sum y^2_{(t-1)h}\right).$$

This estimator is from the LS method applied to the following Euler approximate discrete time model

$$y_{th} = (1-\kappa h)y_{(t-1)h} + \sigma(W_{th} - W_{(t-1)h}). \tag{2.18}$$

It is easy to see that

$$\widehat{\kappa}_{Euler} = -\frac{\widehat{a}_h(\kappa) - 1}{h} \quad \text{with } \widehat{a}_h(\kappa) = \frac{\sum y_{(t-1)h}y_{th}}{\sum y^2_{(t-1)h}},$$

where $\widehat{a}_h(\kappa)$ is the LS estimator of the AR coefficient in (2.18). Note that $\widehat{a}_h(\kappa)$ has the same form as the LS estimator in the exact discrete time model of (2.17) given by

$$y_{th} = a_h(\kappa)y_{(t-1)h} + \sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_{th} \quad \text{with } a_h(\kappa) = e^{-\kappa h} \text{ and}$$

$$\varepsilon_{th} \overset{i.i.d.}{\sim} N(0,1).$$

We then have

$$\begin{aligned}
-\frac{\widehat{a}_h(\kappa) - a_h(\kappa)}{h} &= -\frac{\widehat{a}_h(\kappa) - 1}{h} + \frac{a_h(\kappa) - 1}{h} \\
&= \widehat{\kappa}_{Euler} + \frac{a_h(\kappa) - 1}{h} \\
&= (\widehat{\kappa}_{Euler} - \kappa) + \frac{a_h(\kappa) - 1 + \kappa h}{h} \\
&= (\widehat{\kappa}_{Euler} - \kappa) + O(h).
\end{aligned}$$

If (2.17) is stationary with $\kappa > 0$, it is easy to show that $\sqrt{T/h}\,(\widehat{a}_h(\kappa) - a_h(\kappa)) \Rightarrow N(0, 2\kappa)$ when $N \to \infty$ and $h \to 0$ simultaneously. To develop the double asymptotics for $\widehat{\kappa}_{Euler} - \kappa$, an extra condition on the relative convergence rates of N and h is needed to control the size of the discretization error, $(a_h(\kappa) - 1 + \kappa h)/h$, introduced by the Euler approximation. In particular, Shimizu (2009) proved that, if $Nh^2 \to 0$,

$$\begin{aligned}
-\sqrt{N}\,(\widehat{\kappa}_{Euler} - \kappa) &= \sqrt{T/h}\,(\widehat{a}_h(\kappa) - a_h(\kappa)) + \sqrt{N}O(h) \\
&\Rightarrow N(0, 2\kappa).
\end{aligned}$$

The condition of $Nh^2 \to 0$ ensures that the discretization error, $(a_h(\kappa) - 1 + \kappa h)/h$, is asymptotically dominated by the estimation error, $(\widehat{a}_h(\kappa) - a_h(\kappa))/h$.

If (2.17) is explosive with $\kappa < 0$, Shimizu (2009) showed that no double asymptotic distribution for $\widehat{\kappa}_{Euler}$ is possible if $N \to \infty$, $h \to 0$, and $Nh^\delta \to \infty$ for a constant $\delta > 0$. This is because $(\widehat{a}_h(\kappa) - a_h(\kappa))/h$ is $O_p\left(e^{\kappa N}\right)$ while $(a_h(\kappa) - 1 + \kappa h)/h$ is $O(h)$. As a result, for any $q \geq 0$,

$$\begin{aligned}
-h^q e^{-\kappa N}\,(\widehat{\kappa}_{Euler} - \kappa) &= h^q e^{-\kappa N}(\widehat{a}_h(\kappa) - a_h(\kappa))/h + e^{-\kappa N}O(h^{q+1}) \\
&= O_p\left(h^q\right) + e^{-\kappa N}O(h^{q+1}) \to \infty,
\end{aligned}$$

where the last limit comes from the fact that $e^{-\kappa N}h^{q+1} \to \infty$ under the condition that $Nh^\delta \to \infty$. In this case, the discretization error asymptotically dominates the estimation error.

Compared to $\widehat{\kappa}_{Euler}$, our estimator $\widehat{\kappa} = -\ln(\widehat{a}_h(\kappa))/h$ is constructed directly from the exact discrete time model. It is not subject to the discretization error. As it becomes clear later, we have

$$-h\,(\widehat{\kappa} - \kappa) = \ln(\widehat{a}_h(\kappa)) - \ln(a_h(\kappa)) = \frac{1}{\beta_h(\kappa)}\,(\widehat{a}_h(\kappa) - a_h(\kappa)),$$

where $\beta_h(\kappa)$ takes values between $\widehat{a}_h(\kappa)$ and $a_h(\kappa)$, and $\beta_h(\kappa) \xrightarrow{p} 1$ when $h \to 0$ and $N \to \infty$. Therefore, the double asymptotic distribution of $\widehat{\kappa} - \kappa$ can be directly derived from that of $\widehat{a}_h(\kappa) - a_h(\kappa)$. Consequently, unlike Shimizu (2009), we do not need to impose any extra condition on N and h to control the size of the discretization error for developing the limiting distribution for $\widehat{\kappa}$ in the stationary case. More importantly, we can obtain the rate of convergence and the limiting distribution for $\widehat{\kappa}$ in the explosive case.

3. Simultaneous double asymptotics

3.1. Limit theory for the explosive case

In this subsection, we develop the simultaneous double asymptotic theory for the explosive process (2.1) with $\kappa < 0$.

We first denote

$$Z_{Th} = \frac{1}{\sqrt{T}}\sum \varepsilon_{th}, \tag{3.1}$$

whose distribution converges to $N(0,1)$ according to the CLT for martingale difference arrays (see e.g. Hall and Heyde, 1980, Corollary 3.1). Following PM, we define

$$\begin{aligned}
X_{Th} &= \frac{1}{\sqrt{k_T}}\sum (a_h(\kappa))^{-(T-t)-1}\,\varepsilon_{th} \quad \text{and} \\
Y_{Th} &= \frac{1}{\sqrt{k_T}}\sum (a_h(\kappa))^{-t}\,\varepsilon_{th}.
\end{aligned} \tag{3.2}$$

In Lemma 3.1, we give the limits of (X_{Th}, Y_{Th}, Z_{Th}) and show that they are independent of each other.

Lemma 3.1. *When $\kappa < 0$, $h \to 0$ and $N \to \infty$ simultaneously, we have:*

(a)

$$(a_h(\kappa))^{-T} = o\,(k_T/T) = o(1/N);$$

(b)

$$(X_{Th}, Y_{Th}, Z_{Th}) \Rightarrow (X, Y, Z);$$

where X and Y are independent $N\,(0, -1/2\kappa)$ random variables, and $Z \overset{d}{\equiv} N(0,1)$ is independent of (X, Y).

Note that $a_h(\kappa) \to 1$ when $h \to 0$, and $(a_h(\kappa))^{-T} \to 0$ when $N \to \infty$. Hence, Y_{Th} is dominated by the first few terms in the summation, such as $(a_h(\kappa))^{-1}\,\varepsilon_{1h}$, $(a_h(\kappa))^{-2}\,\varepsilon_{2h}$, etc. Similarly, X_{Th} is dominated by the last few terms in the summation, such as $(a_h(\kappa))^{-1}\,\varepsilon_{Th}$, $(a_h(\kappa))^{-2}\,\varepsilon_{(T-1)h}$, etc. For Z_{Th}, no single term dominates other terms as the weights are the same in the summation. Not surprisingly, X_{Th}, Y_{Th}, Z_{Th} are asymptotically independent.

Let

$$\xi = \sqrt{-2\kappa}X \quad \text{and} \quad \eta = \sqrt{-2\kappa}Y,$$

which are independent $N\,(0, 1)$ random variables. We report the simultaneous double asymptotics in Lemma 3.2 and Theorem 3.3.

Lemma 3.2. *For the explosive process (2.1) with $\kappa < 0$, when $h \to 0$ and $N \to \infty$ simultaneously:*

(a)

$$\frac{(a_h(\kappa))^{-T}}{k_T}\sum_{t=1}^{T}\sum_{j=t}^{T}(a_h(\kappa))^{t-j-1}\,\varepsilon_{jh}\varepsilon_{th} \xrightarrow{L_1} 0;$$

(b)

$$\frac{(a_h(\kappa))^{-T}\,[a_h(\kappa)-1]}{\sqrt{k_T}}\sum x_{(t-1)h} \Rightarrow \sqrt{\frac{1}{-2\kappa}}\,[\eta + D]\,;$$

(c)

$$\frac{(a_h(\kappa))^{-T}}{k_T}\sum x_{(t-1)h}\varepsilon_{th} \Rightarrow \frac{1}{-2\kappa}\xi\,[\eta + D]\,;$$

(d)

$$\frac{(a_h(\kappa))^{-2T}\left[(a_h(\kappa))^2-1\right]}{k_T}\sum x^2_{(t-1)h} \Rightarrow \frac{1}{-2\kappa}\,[\eta + D]^2\,;$$

where $D = \sqrt{2}\left(\kappa\mu + \sigma i\psi'(0) - \kappa y_0\right)/\left(\sigma\sqrt{-\kappa\psi''(0)}\right)$.

Theorem 3.3. *For the explosive process (2.1) with $\kappa < 0$, when $h \to 0$ and $N \to \infty$ simultaneously:*

(a)

$$\frac{(a_h(\kappa))^{T}}{(a_h(\kappa))^2-1}\left(\widehat{a}_h\,(\kappa) - a_h\,(\kappa)\right) \Rightarrow \frac{\xi}{\eta + D};$$

(b)

$$\sqrt{T}\left(\widehat{\tilde{g}}_h - \tilde{g}_h\right) \Rightarrow Z;$$

(c)

$$\sqrt{Tk_T}\,(\widehat{g}_h - g_h) \Rightarrow \sigma\sqrt{\psi''(0)}Z;$$

(d)

$$t_{a_h(\kappa)} \Rightarrow \frac{\xi}{\eta + D}\left\{(\eta + D)^2\right\}^{1/2} \equiv \begin{cases}\xi & \text{if } \eta + D > 0\\ -\xi & \text{if } \eta + D < 0;\end{cases}$$

(e)

$$t_{\tilde{g}_h} = t_{g_h} \Rightarrow Z;$$

(f)

$$\frac{e^{-\kappa N}}{2\kappa}\,(\widehat{\kappa} - \kappa) \Rightarrow \frac{\xi}{\eta + D};$$

where (ξ, η, Z) are independent $N\,(0, 1)$ random variables.

Remark 3.4. For the discrete time explosive AR(1) model without intercept, Anderson (1959) showed that the limiting distribution is dependent on the distribution of the errors and no invariance principle applies. Only under the assumption that the error distribution is Gaussian was he able to show that the limiting distribution is a standard Cauchy. However, the results in Lemma 3.2 and Theorem 3.3 suggest that although the invariance principle does not cover the discrete time explosive model, it covers the explosive continuous time model under the simultaneous double asymptotics.

Remark 3.5. It is known from Perron (1991) that the rate of convergence of $\widehat{a_h}\,(\kappa)$ under the in-fill asymptotics is T. From Anderson (1959) the rate of convergence of $\widehat{a_h}\,(\kappa)$ under the long-time-span asymptotics is known to be ρ^T. According to Theorem 3.3, the rate of convergence of $\widehat{a_h}\,(\kappa)$ under the double asymptotics is $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1} = \frac{(\exp\{-\kappa h\})^T}{\exp\{-2\kappa h\}-1} = \frac{\exp\{-\kappa N\}}{\exp\{-2\kappa h\}-1}$, which gives the rate of $(\exp\{-\kappa h\})^T\ \left(:= \rho^T\right)$ for fixed h, and the rate of T for fixed N. Hence, the rate of convergence under the double asymptotic scheme provides a link between the T and ρ^T rates.

Remark 3.6. The limiting distributions of $\widehat{a_h}\,(\kappa) - a_h(\kappa)$ and $\widehat{\kappa} - \kappa$ explicitly depend on the initial condition y_0 via the term D. Given the fact that $g_h/h \longrightarrow \kappa\mu + \sigma i\psi'(0)$ as $h \to 0$, we then can see that D has two components, one depending on the initial condition y_0 and the other depending on the intercept g_h in Model (2.5). Similar to the limit theory of Perron (1991), what matters in our double asymptotic distribution is not the initial condition and the intercept term per se, but their ratios to σ. This can be seen more clearly by considering the Brownian-motion-driven model (2.3), in which case we have $\sigma i\psi'(0) = 0$, $\psi''(0) = 1$, and

$$D = \sqrt{2}\frac{\kappa\mu - \kappa y_0}{\sigma\sqrt{-\kappa}} = \sqrt{-2\kappa}\left(\frac{y_0}{\sigma} - \frac{\mu}{\sigma}\right).$$

Moreover, if $\mu = y_0$, then $D = 0$, we have

$$\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}(\kappa) - a_h(\kappa)\right) \Rightarrow \text{Cauchy},$$

and

$$\frac{e^{-\kappa N}}{2\kappa}\,(\widehat{\kappa} - \kappa) \Rightarrow \text{Cauchy}.$$

So, even if $\mu \neq 0$ and hence the exact discrete time model has a nonzero intercept, choosing $y_0 = \mu$ will give rise to a standard Cauchy distribution.

Remark 3.7. To consistently estimate D, note that

$$\frac{1}{T}\sum\left[x_{th} - \widehat{a_h}\,(\kappa)\,x_{(t-1)h} - \widehat{\tilde{g}}_h\right]^2 \xrightarrow{p} 1 \quad \text{as } h \to 0 \text{ and } N \to \infty.$$

Given that $\sqrt{k_T}\lambda_h \to \sigma\sqrt{\psi''(0)}$ when $h \to 0$, we have

$$\frac{1}{N}\sum\left[y_{th} - \widehat{a_h}(\kappa)y_{(t-1)h} - \widehat{g}_h\right]^2 = \frac{k_T\lambda_h^2}{T}\sum\left[x_{th} - \widehat{a_h}\,(\kappa)\,x_{(t-1)h} - \widehat{\tilde{g}}_h\right]^2 \xrightarrow{p} \sigma^2\psi''(0)\,.$$

Remark 3.8. It is worth mentioning that the double asymptotic distributions of the AR coefficient and the intercept term are independent. So are the double asymptotic distributions of $t_{a_h(\kappa)}$ and t_{g_h}.

Consider a special case where $\mu = 0$ and the Lévy process $L(t)$ satisfies the conditions that $E(L(1)) = 0$ and $Var(L(1)) = 1$. Thus, Model (2.1) can be rewritten as

$$dy(t) = -\kappa y(t)dt + \sigma dL(t), \qquad y(0) = y_0 = O_p(1). \tag{3.3}$$

The exact discrete time model becomes

$$y_{th} = a_h(\kappa)y_{(t-1)h} + \sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}\varepsilon_{th}, \qquad y_{0h} = y_0. \tag{3.4}$$

The LS estimators of $a_h(\kappa)$ and κ are,

$$\widehat{a_h}(\kappa) = \frac{\sum y_{(t-1)h}y_{th}}{\sum y^2_{(t-1)h}} \quad \text{and} \quad \widehat{\kappa} = -\frac{1}{h}\ln\left(\widehat{a_h}(\kappa)\right). \tag{3.5}$$

Letting $x_{th} = y_{th}/\sigma\sqrt{\frac{1-e^{-2\kappa h}}{2\kappa}}$, we have

$$x_{th} = a_h(\kappa)x_{(t-1)h} + \varepsilon_{th}. \tag{3.6}$$

This equation is nearly the same as the model studied in PM but with one important difference. That is $x_{0h} = O_p\left(\sqrt{k_T}\right)$, but not $o_p\left(\sqrt{k_T}\right)$ as assumed in PM. Corollary 3.9 reports the double asymptotic theory for this special case.

Corollary 3.9. *For the explosive process (3.3) with $\kappa < 0$ and the estimators defined in (3.5), when $h \to 0$ and $N \to \infty$ simultaneously, we have:*

(a)

$$\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2 - 1}\left(\widehat{a_h}(\kappa) - a_h(\kappa)\right) \Rightarrow \frac{\xi}{\eta + d}; \tag{3.7}$$

(b)

$$\frac{e^{-\kappa N}}{2\kappa}\left(\widehat{\kappa} - \kappa\right) \Rightarrow \frac{\xi}{\eta + d}; \tag{3.8}$$

where ξ, η are independent $N(0,1)$ random variables, and $d = y_0\sqrt{-2\kappa}/\sigma$.

Remark 3.10. When κ is interpretable in a meaningful way only under a particular choice of h, such an interpretation will provide a practical guidance to the choice of h and hence N. A good example is in affine term structure modeling. In this case the interest rate data is typically available in the annualized term and κ determines the half-life in number of years. If the monthly (weekly, daily) data is available, one chooses $h = 1/12$ $(1/52, 1/252)$ to maintain this interpretation of κ. Only this choice of κ enters the bond pricing and options pricing formulae in a meaningful way; see Phillips and Yu (2005, 2009). If other values for h are used, the corresponding κ should not be used in the pricing equations. Similar examples may be found in equity options pricing and exchange rate options pricing.

Remark 3.11. When the data structure does not provide guidance for choosing h, an arbitrary choice of h (and hence N) can be made for a given sample size $T(=N/h)$. In this case h becomes a tuning parameter. It is clear that the double asymptotic distributions of normalized $\widehat{a_h}(\kappa)$ and $\widehat{\kappa}$ developed in Theorem 3.3 and Corollary 3.9 are independent of the choice of h and N. Moreover, once a dataset is given, $\widehat{a_h}(\kappa)$ is fixed no matter how h and N are chosen. It is easy to see that the value of the AR parameter $a_h(\kappa)$ in the discrete-time model is fixed too; so is the normalization $(a_h(\kappa))^T/((a_h(\kappa))^2 - 1)$. As a result, the finite sample distribution of normalized $\widehat{a_h}(\kappa)$ is not affected by the choice of h and N. A change in h causes κ to change so that $a_h(\kappa) = e^{-\kappa h}$ remains a constant. To show the robustness of the finite sample distribution of normalized $\widehat{\kappa}$ with respect to h and N, let $N_1/h_1 = N_2/h_2 = T$, $\kappa_1 = -\frac{1}{h_1}\ln a_h(\kappa)$, $\widehat{\kappa}_1 = -\frac{1}{h_1}\ln\widehat{a_h}(\kappa)$, $\kappa_2 = -\frac{1}{h_2}\ln a_h(\kappa)$, and $\widehat{\kappa}_2 = -\frac{1}{h_2}\ln\widehat{a_h}(\kappa)$. We then have

$$\begin{aligned}
&\frac{e^{-\kappa_2 N_2}}{2\kappa_2}\left(\widehat{\kappa}_2 - \kappa_2\right)\\
&= \exp\left(\frac{\ln a_h(\kappa)}{h_2}Th_2\right)\left(-\frac{\ln\widehat{a_h}(\kappa)}{h_2} + \frac{\ln a_h(\kappa)}{h_2}\right)\Big/\left(-2\frac{\ln a_h(\kappa)}{h_2}\right)\\
&= \exp\left(T\ln a_h(\kappa)\right)\left(-\ln\widehat{a_h}(\kappa) + \ln a_h(\kappa)\right)/\left(-2\ln a_h(\kappa)\right)\\
&= \exp\left(\frac{\ln a_h(\kappa)}{h_1}Th_1\right)\left(-\frac{\ln\widehat{a_h}(\kappa)}{h_1} + \frac{\ln a_h(\kappa)}{h_1}\right)\Big/\left(-2\frac{\ln a_h(\kappa)}{h_1}\right)\\
&= \frac{e^{-\kappa_1 N_1}}{2\kappa_1}\left(\widehat{\kappa}_1 - \kappa_1\right).
\end{aligned}$$

A change in h not only causes κ and $\widehat{\kappa}$ to change but also the normalization $e^{-\kappa N}/(2\kappa)$ to change, so that the finite sample distribution of the normalized $\widehat{\kappa}$ is not affected by the choice of h and N. Therefore, the choice of h and N does not affect the performance of the double asymptotic distribution in approximating the finite-sample distributions of normalized $\widehat{a_h}(\kappa)$ and $\widehat{\kappa}$.

Remark 3.12. Comparing the results in Corollary 3.9 with those in Theorem 3.3, it can be seen that the same double asymptotic distribution for the AR parameter is obtained, whether the intercept in the discrete time process (3.4) is estimated or not. This observation is in sharp contrast to the in-fill asymptotic theory which has different asymptotic distributions for different model specifications; see, for example, Perron (1991), Yu (2014) and Zhou and Yu (2015). This sensitivity naturally suggests that the in-fill asymptotic theory better reflects the finite-sample situation than the double asymptotic theory. The reason is that the in-fill asymptotic theory only requires $h \to 0$, while the double asymptotic theory requires not only $h \to 0$ but also $N \to \infty$. In general, the smaller the N, the bigger the difference between the double asymptotic distribution and the in-fill asymptotic distribution. For the stationary AR(1) model with a strong persistence, the simulation results in Zhou and Yu (2015) show that the in-fill asymptotic theory outperforms the double asymptotic theory in approximating finite-sample distributions. Interestingly, as we will show in Section 5, when the model is explosive, since the LS estimator of the AR parameter converges to the true value very fast, even for a small sample size, the double asymptotic distributions developed in Corollary 3.9 are very close to the finite sample distributions. Being robust to the model specification makes the double asymptotic theory convenient to use. Moreover, the double asymptotic distribution is easy to obtain whereas the in-fill asymptotic distributions have to be approximated using numerical methods. Furthermore, the in-fill asymptotic distributions, such as the one in (4.3) below, depend on κ, which is not consistently estimable with a finite N.

Remark 3.13. To facilitate a comparison of our results with those of PM, we may rewrite the limit theory in (3.7) as

$$\frac{(a_h(\kappa))^T k_T}{-2\kappa}\left(\widehat{a_h}(\kappa) - a_h(\kappa)\right) \Rightarrow \frac{X}{Y + y_0/\sigma}, \tag{3.9}$$

where X, Y are defined in Lemma 3.1. When $y_0 = 0$, the limiting distribution is a standard Cauchy and the same as in PM. Since the finite sample distribution depends on the initial value, we expect that the double asymptotic distribution in (3.9) provides a better approximation than the standard Cauchy distribution when

y_0 is different from 0. Monte Carlo evidence reported in Section 5 supports this argument.

Remark 3.14. There are some differences between our asymptotic theory given in (3.7) and (3.8) and that obtained by PM and given in (2.14). Our limiting distribution depends on the initial condition while theirs does not. Of course, a larger initial condition may be assumed in PM's model and hence one can accordingly extend PM's asymptotic theory. It is expected that the new asymptotic theory will depend on the localization parameter $c = -\kappa$ and $k_T (=T^\alpha)$ which in turn depends on α.[6] In general, c and α are not consistently estimable. Consequently, it is infeasible to implement PM's theory directly. Interestingly, the continuous time counterpart of T^α is $1/h$ and h is known for any given data. Hence, in our model there is no need to estimate α. It is well-known that κ can be consistently estimated as long as $N \to \infty$ (Tang and Chen, 2009). As a result, it is feasible to implement our theory directly due to this pivotal property.

Remark 3.15. By constructing the estimator directly from the exact discrete time model, we give the limiting distribution of $\widehat{\kappa}$ in (3.8) for the explosive model with $\kappa < 0$. This is very different from what Shimizu (2009) found when the estimator is constructed from the Euler discretized model. Clearly our result makes statistical inference possible. Moreover, our limit theory is derived for the Lévy process which allows us to develop an invariance principle.

3.2. Limit theory for the stationary case

Following the suggestion by a referee, in this subsection we develop the simultaneous double asymptotics for the stationary process (2.1) with $\kappa > 0$. Tang and Chen (2009) obtained the sequential double asymptotics ($N \to \infty$ followed by $h \to 0$) for κ when $\kappa > 0$ for the Brownian-motion-driven model (2.3). By letting $h \to 0$ followed by $N \to \infty$, Perron (1991) derived a sequential asymptotic distribution for a special case of Model (2.3) with $\mu = 0$. PM developed the asymptotic distribution for $\rho_T (:= 1 - \kappa/k_T)$ with $\kappa > 0$ in Model (2.13). In all these studies, the limiting distribution is $N(0, 2\kappa)$.

Lemma 3.16. *Let x_{th} follows Model (2.12) with $\kappa > 0$. Assuming that for some $\delta > 0$, $E|\varepsilon_{1h}|^{2+\delta} < \infty$, when $h \to 0$ and $N \to \infty$ simultaneously, we have:*

(a)

$$\frac{1}{T\sqrt{k_T}}\sum_{t=1}^{T} x_{(t-1)h} \xrightarrow{p} \frac{\kappa\mu + \sigma i\psi'(0)}{\kappa\sigma\sqrt{\psi''(0)}};$$

(b)

$$\frac{1}{\sqrt{Tk_T}}\sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th} \Rightarrow \frac{\kappa\mu + \sigma i\psi'(0)}{\kappa\sigma\sqrt{\psi''(0)}}Z + \Xi/\sqrt{2\kappa};$$

(c)

$$\frac{1}{Tk_T}\sum_{t=1}^{T}\left[x_{(t-1)h}\right]^2 \xrightarrow{p} \frac{1}{2\kappa} + \left(\frac{\kappa\mu + \sigma i\psi'(0)}{\kappa\sigma\sqrt{\psi''(0)}}\right)^2;$$

where Ξ and Z are independent $N(0, 1)$ random variables.

[6] In PM, an AR process with the root of $\rho_T = 1 + c/T^\alpha$ is used as an example of the mildly explosive process where c is a positive constant and $\alpha \in (0, 1)$.

Theorem 3.17. *Consider the continuous time model given by (2.1) and its exact discrete time representations (2.5) and (2.12) with $\kappa > 0$. Assume that for some $\delta > 0$, $E|\varepsilon_{1h}|^{2+\delta} < \infty$. When $h \to 0$ and $N \to \infty$ simultaneously, we have:*

(a)

$$\sqrt{Tk_T}\left(\widehat{a}_h(\kappa) - a_h(\kappa)\right) \Rightarrow \sqrt{2\kappa}\,\Xi;$$

(b)

$$\sqrt{T}\left(\widehat{\tilde{g}}_h - \tilde{g}_h\right) \Rightarrow Z - \sqrt{2}\frac{\kappa\mu + \sigma i\psi'(0)}{\sigma\sqrt{\kappa\psi''(0)}}\Xi;$$

(c)

$$\sqrt{Tk_T}\left(\widehat{g}_h - g_h\right) \Rightarrow \sigma\sqrt{\psi''(0)}Z - \sqrt{2}\frac{\kappa\mu + \sigma i\psi'(0)}{\sqrt{\kappa}}\Xi;$$

(d)

$$t_{a_h(\kappa)} \Rightarrow \Xi;$$

(e)

$$t_{\tilde{g}_h} = t_{g_h} \Rightarrow \frac{\left\{Z - \sqrt{2}\frac{\kappa\mu + \sigma i\psi'(0)}{\sigma\sqrt{\kappa\psi''(0)}}\Xi\right\}/\sqrt{2\kappa}}{\left\{\frac{1}{2\kappa} + \left(\frac{\kappa\mu + \sigma i\psi'(0)}{\sigma\kappa\sqrt{\psi''(0)}}\right)^2\right\}^{1/2}} \overset{d}{=} N(0, 1);$$

(f)

$$\sqrt{N}\left(\widehat{\kappa} - \kappa\right) \Rightarrow -\sqrt{2\kappa}\,\Xi;$$

where Z and Ξ are independent $N(0, 1)$ random variables.

Remark 3.18. Interestingly, the simultaneous double asymptotics for κ in the stationary case are the same as the sequential asymptotics derived in Perron (1991) and Tang and Chen (2009), and the same as the long-time-span asymptotics in PM (2007). Moreover, unlike the double asymptotics for κ in the explosive process, neither the initial condition nor the intercept term plays a role in the limiting distributions of $\widehat{a}_h(\kappa)$, $t_{a_h(\kappa)}$ and $\widehat{\kappa}$ in the stationary process. Furthermore, our double asymptotics for κ do not require $Nh^2 \to 0$, unlike Shimizu (2009). Once again, this advantage arises because our estimator is based on the exact discretized model.

Remark 3.19. The limiting distributions of $\widehat{g}_h$ and t_{g_h} depend on the intercept and are correlated with those of $\widehat{a}_h(\kappa)$ and $t_{a_h(\kappa)}$. When there is no drift in the discrete time model, i.e., $\mu = 0$ and $i\psi'(0) = E(L(1)) = 0$, we have

$$\sqrt{Tk_T}\left(\widehat{g}_h - g_h\right) = \sqrt{Tk_T}\widehat{g}_h \Rightarrow \sigma\sqrt{\psi''(0)}Z \quad \text{and} \quad t_{g_h} \Rightarrow Z,$$

which are independent of the limiting distributions of $\widehat{a}_h(\kappa)$ and $t_{a_h(\kappa)}$. This result contrasts with the double asymptotics for the explosive case in which, regardless of the value of the intercept, $\widehat{a}_h(\kappa)$ and $t_{a_h(\kappa)}$ are always asymptotically independent of $\widehat{g}_h$ and t_{g_h}.

Remark 3.20. Because of the non-trivial effect from the intercept, when the value of the intercept changes from 0 to $g_h \neq 0$, not only does the mean of the limiting distribution of $\widehat{g}_h$ change, but also the variance of the distribution increases. However, for the explosive case, the variance does not change. Therefore, the coefficient based test has a better local power for testing the hypothesis of zero intercept in the explosive case than in the stationary case. The same argument applies to the t-test for zero intercept. This property was also obtained in Wang and Yu (2015) in a discrete time AR(1) model based on the long-time-span asymptotic theory.

4. Sequential double asymptotics

To develop the sequential double asymptotics, without loss of generality, we confine our attention to a special case where

$\mu = 0$ and the Lévy process $L(t)$ satisfies the conditions that $E(L(1)) = 0$ and $Var(L(1)) = 1$. The continuous time model and its exact discrete time representation are given by (3.3) and (3.4), respectively. The LS estimators of $a_h(\kappa)$ and κ are given in (3.5).

The goal of this section is to obtain the two types of sequential double asymptotics given by (i) letting $N \to \infty$ followed by $h \to 0$; (ii) letting $h \to 0$ followed by $N \to \infty$. It can be seen that the two sequential asymptotics are the same as the simultaneous double asymptotics developed in Section 3. Therefore, we build the link between the long-time-span asymptotics and the simultaneous double asymptotics as well as the link between the in-fill asymptotics and the simultaneous double asymptotics. In this section, we will only focus our attention on the explosive case as the results for the stationary case can be obtained easily.

4.1. Sequential double asymptotics: $N \to \infty$ followed by $h \to 0$

For the explosive AR(1) model without drift defined in (3.4), Anderson (1959) proved that when h is fixed and $N \to \infty$,

$$\frac{(a_h(\kappa))^T [\widehat{a}_h(\kappa) - a_h(\kappa)]}{(a_h(\kappa))^2 - 1} \Rightarrow \frac{X_a}{Y_a + a_h(\kappa) x_0},$$

where

$$X_a \overset{d}{=} \lim_{T\to\infty} \sum_{t=1}^{T} (a_h(\kappa))^{-(T-t)} \varepsilon_{th} \quad \text{and}$$

$$Y_a \overset{d}{=} \lim_{T\to\infty} \sum_{t=1}^{T-1} (a_h(\kappa))^{-t+1} \varepsilon_{th}.$$

While X_a and Y_a are independent, their distributions are highly dependent on the distribution of the errors, therefore, no invariance principle applies. Anderson gave the proof under the condition that x_{0h} is a constant, but his result still holds when $x_{0h} \sim O_p(1)$.

In the Appendix, we show that when $h \to 0$ ($k_T = 1/h \to \infty$),

$$\left(\frac{X_a}{\sqrt{k_T}}, \frac{Y_a}{\sqrt{k_T}}\right) \Rightarrow (X, Y), \tag{4.1}$$

where X and Y are two independent $N(0, -1/2\kappa)$ random variables. Therefore, an invariance principle applies. With the fact that $x_0/\sqrt{k_T} \xrightarrow{p} y_0/\sigma$ as $h \to 0$, we now have

$$\lim_{h\to 0}\lim_{N\to\infty} \frac{(a_h(\kappa))^T [\widehat{a}_h(\kappa) - a_h(\kappa)]}{(a_h(\kappa))^2 - 1}$$
$$\overset{d}{=} \lim_{h\to 0} \frac{X_a/\sqrt{k_T}}{Y_a/\sqrt{k_T} + a_h(\kappa) x_0/\sqrt{k_T}} \overset{d}{=} \frac{X}{Y + y_0/\sigma} \overset{d}{=} \frac{\xi}{\eta + d}, \tag{4.2}$$

where $d = y_0\sqrt{-2\kappa}/\sigma$, ξ and η are two independent $N(0, 1)$ random variables defined as $\xi = \sqrt{-2\kappa}X$ and $\eta = \sqrt{-2\kappa}Y$. An immediate consequence of (4.2) is

$$\lim_{h\to 0}\lim_{N\to\infty} \frac{\exp\{-\kappa N\}}{2\kappa} (\widehat{\kappa} - \kappa) \overset{d}{=} \frac{\xi}{\eta + d}.$$

These sequential asymptotics are exactly the same as the simultaneous double asymptotics reported in Corollary 3.9. We collect these results in Theorem 4.1.

Theorem 4.1. *For the continuous time model defined in (3.3) with $\kappa < 0$, when $N \to \infty$ followed by $h \to 0$, we have,*

$$\frac{(a_h(\kappa))^T [\widehat{a}_h(\kappa) - a_h(\kappa)]}{(a_h(\kappa))^2 - 1} \Rightarrow \frac{\xi}{\eta + d} \quad \text{and}$$
$$\frac{\exp\{-\kappa N\}}{2\kappa} (\widehat{\kappa} - \kappa) \Rightarrow \frac{\xi}{\eta + d}.$$

4.2. Sequential double asymptotics: $h \to 0$ followed by $N \to \infty$

Assuming $L(t) = W(t)$ in (3.3), we obtain the exact discrete time model (3.4) where the error term is normally distributed. Perron (1991) developed the in-fill asymptotic distribution of $\widehat{a}_h(\kappa)$ as

$$T(\widehat{a}_h(\kappa) - a_h(\kappa)) \Rightarrow \frac{A(\gamma, c)}{B(\gamma, c)}, \tag{4.3}$$

where

$$A(\gamma, c) = \gamma \int_0^1 \exp\{cr\} dW(r) + \int_0^1 J_c(r) dW(r), \tag{4.4}$$

$$B(\gamma, c) = \gamma^2 \frac{\exp\{2c\} - 1}{2c} + 2\gamma \int_0^1 \exp\{cr\} J_c(r) dr + \int_0^1 J_c(r)^2 dr, \tag{4.5}$$

$c = -\kappa N$, $\gamma = y_0/\left(\sigma\sqrt{N}\right)$, the initial value y_0 is a fixed constant, and $J_c(r) = \int_0^r \exp\{c(r - s)\} dW(s)$ is generated by the stochastic differential equation

$dJ_c(r) = cJ_c(r) dr + dW(r)$,

with the initial condition of $J_c(0) = 0$. Perron (1991) also derived the joint moment generating function (MGF) of $A(\gamma, c)$ and $B(\gamma, c)$. Based on Perron's results, in the following Theorem, we first derive the limit of the joint MGF of $(2c)e^{-c}A(\gamma, c)$ and $(2c)^2 e^{-2c}B(\gamma, c)$ when $c = -\kappa N \to +\infty$ and then obtain the sequential limiting distribution of $\widehat{a}_h(\kappa)$ and $\widehat{\kappa}$.

Theorem 4.2. *Letting $d = y_0\sqrt{-2\kappa}/\sigma$, when $N \to +\infty$ (therefore, $c = -\kappa N \to +\infty$), we have:*

(a) *the joint MGF of $(2c)e^{-c}A(\gamma, c)$ and $(2c)^2 e^{-2c}B(\gamma, c)$ has the limit as*

$$\lim_{c\to+\infty} M(\widetilde{v}, \widetilde{u}) = \lim_{c\to+\infty} E\left[\exp\left(\widetilde{v}(2c)e^{-c}A(\gamma, c) + \widetilde{u}(2c)^2 e^{-2c}B(\gamma, c)\right)\right] = \frac{1}{\{1 - 2\widetilde{u} - \widetilde{v}^2\}^{1/2}} \exp\left\{\frac{d^2\left[2\widetilde{u} + \widetilde{v}^2\right]}{2\left(1 - 2\widetilde{u} - \widetilde{v}^2\right)}\right\};$$

(b) *letting ξ and η be two independent $N(0, 1)$ random variables, then*

$$\left((2c)e^{-c}A(\gamma, c), (2c)^2 e^{-2c}B(\gamma, c)\right) \Rightarrow \left(\xi[d + \eta], [d + \eta]^2\right);$$

(c)

$$\lim_{N\to\infty}\lim_{h\to 0} \frac{e^{-\kappa N}(\widehat{a}_h(\kappa) - a_h(\kappa))}{-2\kappa h} = \lim_{c\to\infty} \frac{(2c)e^{-c}A(\gamma, c)}{(2c)^2 e^{-2c}B(\gamma, c)} = \frac{\xi[d + \eta]}{[d + \eta]^2} = \frac{\xi}{d + \eta},$$

$$\lim_{N\to\infty}\lim_{h\to 0} \frac{e^{-\kappa N}(\widehat{\kappa} - \kappa)}{2\kappa} = \frac{\xi}{d + \eta}.$$

Remark 4.3. Based on the in-fill asymptotic distribution of $\widehat{a}_h(\kappa)$, Perron (1991) also derived the sequential asymptotics in his Corollary 1, which are

$$\lim_{N\to\infty}\lim_{h\to 0} \frac{e^{-\kappa N}(\widehat{a}_h(\kappa) - a_h(\kappa))}{-2\kappa h} = \lim_{c\to\infty} \frac{(2c)e^{-c}A(\gamma, c)}{(2c)^2 e^{-2c}B(\gamma, c)} = \frac{d\eta + \xi\eta}{[d + \eta]^2},$$

Table 2
The finite sample distributions of $e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)/(2\kappa)$ and $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$ are compared with the asymptotic distribution of PM and the double asymptotic distribution developed in the present paper when the initial condition $y_0=0$. Reported are six percentiles (1%, 2.5%, 10%, 90%, 97.5%, 99%) of each distribution.

	Daily ($h=1/252$)	Weekly ($h=1/52$)	Monthly ($h=1/12$)
PM ($N=5$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
New($N=5$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
Finite sample$^{\kappa}$	−36.0, −13.9, −3.15, 3.03, 12.0, 28.5	−37.6, −14.1, −3.22, 3.09, 12.3, 29.7	−42.6, −15.8, −3.72, 3.55, 13.8, 33.3
Finite sample^{a_h}	−35.8, −13.8, −3.13, 3.01, 12.0, 28.3	−36.2, −13.7, −3.10, 2.98, 11.9, 28.6	−35.9, −13.4, −3.14, 2.99, 11.7, 28.1
PM ($N=10$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
New ($N=10$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
Finite sample$^{\kappa}$	−31.5, −12.6, −3.10, 3.11, 12.6, 32.4	−31.6, −13.0, −3.17, 3.21, 13.3, 33.9	−37.4, −14.6, −3.62, 3.64, 15.1, 37.9
Finite sample^{a_h}	−31.3, −12.6, −3.07, 3.09, 12.6, 32.2	−30.5, −12.6, −3.06, 3.09, 12.8, 32.7	−31.5, −12.3, −3.05, 3.07, 12.8, 32.0

and

$$\lim_{N\to\infty}\lim_{h\to 0}\frac{e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)}{2\kappa}=\frac{d\eta+\xi\eta}{[d+\eta]^2}.$$

These results are different from the sequential asymptotics obtained in Theorem 4.2. The reason for the discrepancy is that Perron's results were obtained under the assertion that

$$\left((2c)^{3/2}e^{-2c}\int_0^1\exp\{cr\}J_c(r)\,dr,\,(2c)^{1/2}e^{-c}\int_0^1\exp\{cr\}\,dW(r)\right)\Rightarrow(\eta,\eta)\,,\tag{4.6}$$

whereas, as we prove in the Appendix, the limit of the joint distribution is

$$\left((2c)^{3/2}e^{-2c}\int_0^1\exp\{cr\}J_c(r)\,dr,\,(2c)^{1/2}e^{-c}\int_0^1\exp\{cr\}\,dW(r)\right)\Rightarrow(\eta,\xi)\,,\tag{4.7}$$

where ξ and η are two independent $N(0,1)$ random variables.

Remark 4.4. The sequential asymptotics reported in Theorem 4.2 (c) are derived under the condition of $L(t)=W(t)$ and a constant initial value y_0. It is easy to see that the same sequential asymptotics still hold when $y_0\sim O_p(1)$. Also, all the results in Theorem 4.2 continue to hold true when $L(t)\neq W(t)$. This is because an invariance principle is applicable to the functional central limit theorem used in Perron (1991).

5. Simulation and empirical results

In this section, we first conduct Monte Carlo simulations to (i) examine the sensitivity of our double asymptotic distribution and the finite sample distribution with respect to the initial condition; (ii) check the finite sample performance of our double asymptotic theory; and (iii) compare the performance of our double asymptotic theory with that of PM. To do so, we simulate 100,000 sample paths from Model (2.3) with $\kappa=-2,\mu=0,\sigma=1$. However, we allow both h and N to take different values. In particular, we choose $h=1/252,1/52,1/12$, corresponding to the daily, weekly and monthly data.[7] Moreover, we choose $N=5,10$. For each simulated path, we estimate κ and $a_h(\kappa)$ and calculate $e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)/(2\kappa)$ and $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$.

In Table 2, we choose the initial value $y_0=0$ (which implies $d=0$) and report six percentiles (1%, 2.5%, 10%, 90%, 97.5%, 99%) of the finite sample distributions of $e^{-\kappa N}(\widehat{\kappa}-\kappa)/(2\kappa)$ and $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$, PM's asymptotic distribution, and the new asymptotic distribution. Since $d=0$, our double asymptotic distribution is the same as that of PM which is a standard Cauchy. Not surprisingly, both sets of percentiles are identical. Moreover, in all cases, these two sets of percentiles are close to those of the corresponding finite sample distributions. This finding indicates that the two asymptotic distributions work well when $y_0=0$. When N is as small as 5 and h is smaller than 1/12, our double asymptotic distribution is very close to the finite sample distributions. This finding is encouraging as it is typically found in stationary but near unit root continuous time models that a much larger N is needed for the double asymptotics to work well; see, for example, Jeong and Park (2011). Compared to the finite sample distribution of $e^{-\kappa N}(\widehat{\kappa}-\kappa)/(2\kappa)$, the finite sample distribution of $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$ is slightly closer to the developed double asymptotic distribution.

In Table 3, we choose the initial value $y_0=3.5$ (which implies $d=7$, a value that is close to the one in the empirical study below) and report six percentiles (1%, 2.5%, 10%, 90%, 97.5%, 99%) of the finite sample distributions of $e^{-\kappa N}(\widehat{\kappa}-\kappa)/(2\kappa)$ and $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$, PM's asymptotic distribution, and the new asymptotic distribution. Since d is different from 0, our double asymptotic distribution is different from that of PM. An immediate finding from Table 3 is that the two sets of percentiles are very different. For example, the 1, 2.5, 10 percentiles under PM's distribution are 90 times, 43 times and 16 times of those under our asymptotic distribution. This indicates strongly that the new double asymptotic distribution is very sensitive to the initial condition. Moreover, in all cases the percentiles under the new asymptotics are close to those of the corresponding finite sample distributions. This finding indicates that our double asymptotic distribution works well when $d=7$. On the other hand, PM's distribution has a very large spread and is very far away from the finite sample distributions. As before, when N is as small as 5 and h is smaller than 1/12, our double asymptotic distribution is close to the finite sample distributions. The big difference between our asymptotic theory and PM's theory has important practical implications for statistical inference. Our theory not only provides a much better approximation to the finite sample distributions, but also results in asymptotic distributions that second-order-stochastically dominate PM's asymptotic distributions. Consequently, we expect that the test based on our theory should have higher power and shorter confidence intervals at the same level of confidence.[8] Compared to the finite sample distribution of $e^{-\kappa N}(\widehat{\kappa}-\kappa)/(2\kappa)$, the finite

[7] This choice of h comes from the fact that many financial variables are measured in the annualized term; see, for example, Aït-Sahalia (1999) and Phillips and Yu (2005).

[8] We are grateful to a referee for making this point for us. However, it should be emphasized that the model of PM is different from ours as PM assumes a smaller initial condition.

Table 3
The finite sample distributions of $e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)/(2\kappa)$ and $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$ are compared with the asymptotic distribution of PM and the double asymptotic distribution developed in the present paper when the initial condition $y_0 = 3.5$. Reported are six percentiles (1%, 2.5%, 10%, 90%, 97.5%, 99%) of each distribution.

	Daily ($h = 1/252$)	Weekly ($h = 1/52$)	Monthly ($h = 1/12$)
PM ($N = 5$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
New ($N = 5$)	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35
Finite sample$^{\kappa}$	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35	−0.36, −0.30, −0.19, 0.19, 0.30, 0.36	−0.41, −0.34, −0.22, 0.22, 0.34, 0.41
Finite sample^{a_h}	−0.35, −0.29, −0.19, 0.18, 0.29, 0.35	−0.35, −0.29, −0.19, 0.19, 0.29, 0.35	−0.35, −0.29, −0.19, 0.19, 0.29, 0.35
PM ($N = 10$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
New ($N = 10$)	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35
Finite sample$^{\kappa}$	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35	−0.37, −0.30, −0.19, 0.19, 0.30, 0.36	−0.35, −0.29, −0.18, 0.18, 0.29, 0.35
Finite sample^{a_h}	−0.35, −0.29, −0.19, 0.19, 0.29, 0.35	−0.36, −0.30, −0.19, 0.19, 0.29, 0.35	−0.35, −0.29, −0.19, 0.19, 0.29, 0.35

Table 4
The finite sample distributions of $e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)/(2\kappa)$ and $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$ are compared with the asymptotic distribution of PM and the double asymptotic distribution developed in the present paper when the initial condition $y_0 = 10$. Reported are six percentiles (1%, 2.5%, 10%, 90%, 97.5%, 99%) of each distribution.

	Daily ($h = 1/252$)	Weekly ($h = 1/52$)	Monthly ($h = 1/12$)
PM ($N = 5$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
New ($N = 5$)	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11
Finite sample$^{\kappa}$	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11	−0.12, −0.10, −0.06, 0.06, 0.10, 0.12	−0.13, −0.11, −0.07, 0.07, 0.11, 0.13
Finite sample^{a_h}	−0.12, −0.10, −0.06, 0.06, 0.10, 0.12	−0.12, −0.10, −0.06, 0.06, 0.10, 0.12	−0.12, −0.11, −0.06, 0.06, 0.11, 0.12
PM ($N = 10$)	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8	−31.8, −12.7, −3.08, 3.08, 12.7, 31.8
New ($N = 10$)	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11
Finite sample$^{\kappa}$	−0.11, −0.09, −0.06, 0.06, 0.09, 0.11	−0.12, −0.10, −0.06, 0.06, 0.10, 0.12	−0.14, −0.11, −0.07, 0.07, 0.11, 0.14
Finite sample^{a_h}	−0.12, −0.10, −0.06, 0.06, 0.10, 0.12	−0.12, −0.10, −0.06, 0.06, 0.10, 0.12	−0.12, −0.10, −0.06, 0.06, 0.10, 0.12

sample distribution of $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$ continues to be slightly closer to the developed double asymptotic distribution.

In Table 4, we choose the initial value $y_0 = 10$ (which implies $d = 20$) and report six percentiles (1%, 2.5%, 10%, 90%, 97.5%, 99%) of the finite sample distributions of $e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)/(2\kappa)$ and $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$, PM's asymptotic distribution, and the new asymptotic distribution. Similarly, the two sets of percentiles under the two asymptotic distributions are very different. For example, the 1, 2.5, 10 percentiles under PM's distribution are 270 times, 129 times and 48 times of those under our asymptotic distribution. This suggests a larger discrepancy between the two distributions associated with a larger initial condition. However, in all cases the percentiles under the new double asymptotic distribution are close to those of the corresponding finite sample distributions. This finding indicates that the good performance of the new double asymptotic distribution is not lost with an increase in the initial condition. As before, when N is as small as 5 and h is smaller than 1/12, our double asymptotic distribution is close to the finite sample distributions. In this case the finite sample distribution of $\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a_h}\left(\kappa\right)-a_h(\kappa)\right)$ is very close to the finite sample distribution of $e^{-\kappa N}\left(\widehat{\kappa}-\kappa\right)/(2\kappa)$.

To further appreciate the difference between the two limiting distributions, we apply them to a real data set—the monthly log Nasdaq real price between January 1990 to December 2000. The same data were used in Phillips et al. (2011). Model (2.4) without intercept is fitted to the data with $h = 1/12, N = 10, T = 120$, and $y_0 = 5.0628$ which is the log Nasdaq real price in December 1989. The estimated value of σ is 0.2014, implying $d = 6.7784$. The results, including the estimated κ, the corresponding 95% confidence interval for κ using PM's theory, the corresponding quantiles and the 95% confidence interval for κ using our theory, the estimated $\widehat{a_h}$, the corresponding 95% confidence interval for a_h using PM's theory, the corresponding quantiles and the 95% confidence interval for a_h using our theory, are reported in Table 5. While PM's confidence intervals contain $\kappa = 0$ or $a_h = 1$, the confidence intervals based on the new theory do not contain $\kappa = 0$ or $a_h = 1$, suggesting strong evidence of explosiveness (i.e., $\kappa < 0$ and $a_h > 1$). The empirical conclusion is different because d is very different from zero. The 2.5 and 97.5 percentiles of PM's distribution are −12.7 and 12.7. In our distribution, these two percentiles are −0.3016 and 0.3016, which lead to much tighter confidence intervals that exclude $\kappa = 0$ and $a_h = 1$, respectively.

6. Conclusion

This paper develops the double asymptotic theory for the explosive Lévy-driven diffusion process and the corresponding discrete time model under the scheme of a large number of time span (N) and a small number of sampling interval (h). A link between the continuous time model and the discrete time AR(1) model with root moderately deviated from unity is established. The double asymptotic theory contributes to the literature in three aspects. First, our theory permits explicit consideration of the effects from the initial condition and the intercept term. Monte Carlo evidence suggests that the new asymptotic theory provides a better approximation to the finite sample distribution than the limit theory that is independent of the initial condition. Second, the theory is developed for a continuous time Lévy-driven process to facilitate the derivation of an invariance principle. Third, the double asymptotic theory bridges the gap between the in-fill asymptotic theory and the long-time-span asymptotic theory, in the sense that the convergence rate is between the rate of in-fill and the rate of long-time-span asymptotics.

For the unit root case where $\kappa = 0$, Model (2.1) becomes

$$dy(t) = \sigma dL(t), \qquad y(0) = y_0 = O_p(1).$$

Its exact discrete time model is

$$y_{th} = y_{(t-1)h} + \sigma\sqrt{\Psi''(0)\,\frac{1-e^{-2\kappa h}}{2\kappa}}\,\varepsilon_{th}, \qquad y_{0h} = y_0,$$

or

$$x_{th} = x_{(t-1)h} + \varepsilon_{th}, \quad \text{and} \quad x_{th} = y_{th}/\sigma\sqrt{\Psi''(0)\,\frac{1-e^{-2\kappa h}}{2\kappa}},$$

$$x_{0h} = x_0 = O_p\left(\sqrt{k_T}\right).$$

It is well known that in order for the initial condition to have an impact on the limit theory of unit root model, the order of the initial

Table 5
Empirical results using the monthly log Nasdaq real price between January 1990 and December 2000.

$\widehat{\kappa}$	PM's 95% CI	2.5 percentile of $\frac{\xi}{\eta+d}$	97.5 percentile of $\frac{\xi}{\eta+d}$	New 95% CI
−0.0363	(−0.6776, 0.6050)	−0.3016	0.3016	(−0.0211, −0.0515)
$\widehat{a}_h(\kappa)$	PM's 95% CI	2.5 percentile of $\frac{\xi}{\eta+d}$	97.5 percentile of $\frac{\xi}{\eta+d}$	New 95% CI
1.003	(0.9497, 1.0563)	−0.3016	0.3016	(1.0017, 1.0043)

condition should be $O_p\left(\sqrt{T}\right)$ (see, for example, Phillips (1987a) and Phillips and Magdalinos (2009)). Under the double asymptotic scheme, $x_{0h} = O_p\left(\sqrt{k_T}\right)$ with $k_T/T \to 0$. We therefore expect the double asymptotic theory for the unit root process to be the same as the conventional limit theory for the discrete time unit root model with zero initial condition.

Appendix

Proof of Lemma 3.1. (a) As $k_T = 1/h$ and $T = N/h$, when $\kappa < 0$, $N \to \infty$, we have

$$\frac{(a_h(\kappa))^{-T}}{k_T/T} = \frac{\exp\{\kappa hT\}}{1/(Th)} = \frac{\exp\{\kappa N\}}{1/N} = \frac{N}{\exp\{-\kappa N\}} \to 0.$$

(b) Denote $a_h(\kappa) = a_h$ when there is no confusion. By the Cramér–Wold device (e.g. Kallenberg, 2002, Corollary 5.5), it is sufficient to show that

$$aX_{Th} + bY_{Th} + cZ_{Th} \Rightarrow aX + bY + cZ \quad \text{for all } a, b, c \in \mathbb{R}, \tag{A.1}$$

where X and Y are independent $N(0, 1/(-2\kappa))$ random variables, Z is a standard normal distribution and independent of (X, Y). If $\Upsilon \sim N\left(0, \frac{a^2+b^2}{-2\kappa} + c^2\right)$, $aX + bY + cZ \overset{d}{=} \Upsilon$, so $aX_{Th} + bY_{Th} + cZ_{Th} \Rightarrow \Upsilon$ is sufficient for (A.1).

We can write $aX_{Th} + bY_{Th} + cZ_{Th} = \sum_{t=1}^{T} \zeta_{Tt}$, where

$$\{\zeta_{Tt}\}_{t=1}^{T} = \left\{\left(a[a_h]^{-(T-t)-1} + b[a_h]^{-t} + \frac{c\sqrt{k_T}}{\sqrt{T}}\right)\frac{\varepsilon_{th}}{\sqrt{k_T}}\right\}_{t=1}^{T},$$

is a martingale difference array as h and N varies, as $\{\varepsilon_{th}\}_{t=1}^{T} \overset{i.i.d.}{\sim} (0, 1)$ for any fixed h. Let $\mathcal{F}_{T,t} = \sigma(x_0, \varepsilon_{1h}, \ldots, \varepsilon_{th})$ be the information set. Note that $k_T/T = 1/N$, $k_T(a_h - 1) \to -\kappa$ and $k_T\left([a_h]^2 - 1\right) \to -2\kappa$ when $h \to 0$, and $(a_h(\kappa))^{-T} = o(k_T/T) \to 0$ as $N \to \infty$. We then derive the conditional variance as,

$$\begin{aligned}
V_{TT} &= \sum_{t=1}^{T} E\left((\zeta_{Tt})^2 \middle| \mathcal{F}_{T,t-1}\right) \\
&= \frac{1}{k_T}\sum_{t=1}^{T}\left(a[a_h]^{-(T-t)-1} + b[a_h]^{-t} + \frac{c\sqrt{k_T}}{\sqrt{T}}\right)^2 \\
&= \frac{1}{k_T}\sum_{t=1}^{T}\left(a[a_h]^{-(T-t)-1} + b[a_h]^{-t}\right)^2 + c^2 \\
&\quad + \frac{2}{\sqrt{Tk_T}}\sum_{t=1}^{T}\left(a[a_h]^{-(T-t)-1} + b[a_h]^{-t}\right)c \\
&= \frac{a^2+b^2}{-2\kappa} + c^2 + \frac{2ac}{\sqrt{Tk_T}}\sum_{t=1}^{T}[a_h]^{-(T-t)-1} \\
&\quad + \frac{2bc}{\sqrt{Tk_T}}\sum_{t=1}^{T}[a_h]^{-t} + o(1) \\
&= \frac{a^2+b^2}{-2\kappa} + c^2 + \frac{2ac\sqrt{k_T}}{\sqrt{T}}\frac{1-[a_h]^{-T}}{k_T(a_h-1)} \\
&\quad + \frac{2bc\sqrt{k_T}}{\sqrt{T}}\frac{1-[a_h]^{-T}}{k_T(a_h-1)} + o(1) \\
&\to \frac{a^2+b^2}{-2\kappa} + c^2 \quad \text{when } h \to 0 \text{ and } N \to \infty,
\end{aligned}$$

where the third equation comes from the fact that

$$\begin{aligned}
&\frac{1}{k_T}\sum_{t=1}^{T}\left(a[a_h]^{-(T-t)-1} + b[a_h]^{-t}\right)^2 \\
&= \frac{1}{k_T}\sum_{t=1}^{T}a^2[a_h]^{-2(T-t)-2} + \frac{1}{k_T}\sum_{t=1}^{T}b^2[a_h]^{-2t} \\
&\quad + \frac{2ab}{k_T}\sum_{t=1}^{T}[a_h]^{-T-1} \\
&= \frac{a^2\left(1-[a_h]^{-2T}\right)}{k_T\left([a_h]^2-1\right)} + \frac{b^2\left(1-[a_h]^{-2T}\right)}{k_T\left([a_h]^2-1\right)} + 2ab\frac{T[a_h]^{-T-1}}{k_T} \\
&\to \frac{a^2+b^2}{-2\kappa} \quad \text{as } h \to 0,\ N \to \infty.
\end{aligned}$$

The conditional Lindeberg condition holds because for any $\varepsilon > 0$

$$\begin{aligned}
&\sum_{t=1}^{T} E\left((\zeta_{Tt})^2 1\{|\zeta_{Tt}| > \varepsilon\} \middle| \mathcal{F}_{T,t-1}\right) \\
&= \frac{1}{k_T}\sum_{t=1}^{T}\left[\left(a(a_h)^{-(T-t)-1} + b(a_h)^{-t} + \frac{c\sqrt{k_T}}{\sqrt{T}}\right)^2\right. \\
&\quad \times E\left((\varepsilon_{th})^2 1\left\{\left(a(a_h)^{-(T-t)-1} + b(a_h)^{-t} + \frac{c\sqrt{k_T}}{\sqrt{T}}\right)^2\right.\right. \\
&\quad \left.\left.\left.(\varepsilon_{th})^2 > k_T\varepsilon^2\right\}\right)\right] \\
&\le V_{TT}\max_{1\le t\le T} E\left((\varepsilon_{th})^2 1\left\{\left(a(a_h)^{-(T-t)-1} + b(a_h)^{-t}\right.\right.\right. \\
&\quad \left.\left.\left. + \frac{c\sqrt{k_T}}{\sqrt{T}}\right)^2(\varepsilon_{th})^2 > k_T\varepsilon^2\right\}\right) \\
&\le V_{TT}\max_{1\le t\le T} E\left((\varepsilon_{th})^2 1\left\{3\left(a^2(a_h)^{-2(T-t)-2} + b^2(a_h)^{-2t}\right.\right.\right. \\
&\quad \left.\left.\left. + \frac{c^2k_T}{T}\right)(\varepsilon_{th})^2 > k_T\varepsilon^2\right\}\right) \\
&\le V_{TT}\max_{1\le t\le T} E\left((\varepsilon_{th})^2 1\left\{3\left(a^2+b^2+c^2\right)(\varepsilon_{th})^2 > k_T\varepsilon^2\right\}\right) \\
&= V_{TT}\, E\left((\varepsilon_{1h})^2 1\left\{3\left(a^2+b^2+c^2\right)(\varepsilon_{1h})^2 > k_T\varepsilon^2\right\}\right) \\
&\to 0 \quad \text{when } h \to 0,\ N \to \infty,
\end{aligned}$$

where the third inequality comes from the fact that $(a_h)^{-2(T-t)-2} < 1$ and $(a_h)^{-2t} < 1$ for any $1 \le t \le T$ when $\kappa < 0$, and $c^2k_T/T = c^2/N < c^2$ when $N > 1$, the last equation is because ε_{th} are identically distributed, and the limit result is from the integrability of ε_{1h}^2 and the condition $k_T = 1/h \to \infty$.

Then, from the CLT for martingale difference arrays (see, e.g. Hall and Heyde, 1980, Corollary 3.1), we can get $aX_{Th} + bY_{Th} + cZ_{Th} \Rightarrow \Upsilon$, for all $a, b, c \in \mathbb{R}$.

Proof of Lemma 3.2. The proof of (a) is similar to PM and is omitted here.

(b) Let $a_h(\kappa) = a_h$ when there is no confusion. From model (2.12) we get

$$x_{th} = a_h x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th} = \tilde{g}_h \frac{1-a_h^t}{1-a_h} + \sum_{j=1}^{t} a_h^{t-j}\varepsilon_{jh} + a_h^t x_0.$$

Hence, when $h \to 0$ and $N \to \infty$,

$$\begin{aligned}
\frac{a_h^{-T}}{\sqrt{k_T}} x_{Th} &= \frac{a_h^{-T}}{\sqrt{k_T}}\left(\tilde{g}_h \frac{1-a_h^T}{1-a_h} + \sum_{j=1}^{T} a_h^{T-j}\varepsilon_{jh} + a_h^T x_0\right)\\
&= \sqrt{k_T}\tilde{g}_h \frac{1-a_h^{-T}}{k_T(a_h-1)} + \frac{1}{\sqrt{k_T}}\sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh} + \frac{x}{\sqrt{k_T}}\\
&\Rightarrow \frac{\kappa\mu + \sigma i\psi'(0)}{-\kappa\sigma\sqrt{\psi''(0)}} + Y + \frac{y_0}{\sigma\sqrt{\psi''(0)}}\\
&= \sqrt{\frac{1}{-2\kappa}}\left(\sqrt{-2\kappa}Y + \sqrt{-2\kappa}\frac{\kappa\mu + \sigma i\psi'(0) - \kappa y_0}{-\kappa\sigma\sqrt{\psi''(0)}}\right)\\
&= \sqrt{\frac{1}{-2\kappa}}[\eta + D]
\end{aligned}$$

where Y is defined as in the Lemma 3.1 and η is defined as $\eta = \sqrt{-2\kappa}Y \overset{d}{=} N(0,1)$.

The relation that $x_{th} - x_{(t-1)h} = (a_h - 1)x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th}$ leads to

$$x_{Th} - x_{0h} = (a_h - 1)\sum_{t=1}^{T} x_{(t-1)h} + T\tilde{g}_h + \sum_{t=1}^{T}\varepsilon_{th}.$$

Therefore,

$$\begin{aligned}
&\frac{a_h^{-T}(a_h-1)}{\sqrt{k_T}}\sum_{t=1}^{T} x_{(t-1)h}\\
&= \frac{a_h^{-T}}{\sqrt{k_T}}x_{Th} - a_h^{-T}\frac{x_{0h}}{\sqrt{k_T}} - \frac{a_h^{-T}T}{k_T}\sqrt{k_T}\tilde{g}_h - \frac{a_h^{-T}\sqrt{T}}{\sqrt{k_T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th}\\
&= \frac{a_h^{-T}}{\sqrt{k_T}}x_{Th} + o_p(1)\\
&\Rightarrow \sqrt{\frac{1}{-2\kappa}}[\eta + D] \quad \text{as } h \to 0 \text{ and } N \to \infty,
\end{aligned}$$

where the second equality comes from the facts that $x_{0h}/\sqrt{k_T} = O_p(1)$, $\sqrt{k_T}\tilde{g}_h = O(1)$ and $a_h^{-T} = o(k_T/T)$.

(c) Since

$$x_{th} = a_h x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th} = \tilde{g}_h \frac{1-a_h^t}{1-a_h} + \sum_{j=1}^{t} a_h^{t-j}\varepsilon_{jh} + a_h^t x_{0h},$$

we have

$$\begin{aligned}
&\sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th}\\
&= \tilde{g}_h\sum_{t=1}^{T}\frac{1-a_h^{t-1}}{1-a_h}\varepsilon_{th} + \sum_{t=2}^{T}\left[\sum_{j=1}^{t-1} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th}\\
&\quad + x_{0h}\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\\
&= \tilde{g}_h\sum_{t=1}^{T}\frac{1-a_h^{t-1}}{1-a_h}\varepsilon_{th} + \sum_{t=2}^{T}\left[\sum_{j=1}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th}\\
&\quad - \sum_{t=2}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} + x_{0h}\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\\
&= \tilde{g}_h\sum_{t=1}^{T}\frac{1-a_h^{t-1}}{1-a_h}\varepsilon_{th} + \left[\sum_{t=2}^{T} a_h^{t-1}\varepsilon_{th}\right]\left[\sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh}\right]\\
&\quad - \sum_{t=2}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} + x_{0h}\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\\
&= \tilde{g}_h\sum_{t=1}^{T}\frac{1-a_h^{t-1}}{1-a_h}\varepsilon_{th} + \left[\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\right]\left[\sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh}\right]\\
&\quad - \sum_{t=1}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} + x_{0h}\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\\
&= \frac{\tilde{g}_h}{1-a_h}\sum_{t=1}^{T}\varepsilon_{th} + \left[\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\right]\left[\frac{\tilde{g}_h}{a_h-1} + \sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh} + x_{0h}\right]\\
&\quad - \sum_{t=1}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th}.
\end{aligned}$$

When $h \to 0$, $N \to \infty$, we have $\frac{a_h^{-T}}{k_T}\sum_{t=1}^{T}\left[\sum_{j=t}^{T} a_h^{t-j-1}\varepsilon_{jh}\right]\varepsilon_{th} = o_p(1)$ from (a), and it is easy to see that

$$\begin{aligned}
\frac{a_h^{-T}}{k_T}\frac{\tilde{g}_h}{(1-a_h)}\sum_{t=1}^{T}\varepsilon_{th} &= \frac{a_h^{-T}\sqrt{T}}{\sqrt{k_T}}\frac{\tilde{g}_h\sqrt{k_T}}{(1-a_h)k_T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th}\right)\\
&= o_p(1),
\end{aligned}$$

since $a_h^{-T} = o(k_T/T)$. We then have

$$\begin{aligned}
&\frac{a_h^{-T}}{k_T}\sum_{t=1}^{T} x_{(t-1)h}\varepsilon_{th}\\
&= \frac{a_h^{-T}}{k_T}\left[\sum_{t=1}^{T} a_h^{t-1}\varepsilon_{th}\right]\left[\frac{\tilde{g}_h}{a_h-1} + \sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh} + x_0\right] + o_p(1)\\
&= \left[\frac{1}{\sqrt{k_T}}\sum_{t=1}^{T} a_h^{-(T-t)-1}\varepsilon_{th}\right]\\
&\quad \times\left[\frac{\tilde{g}_h\sqrt{k_T}}{(a_h-1)k_T} + \frac{1}{\sqrt{k_T}}\sum_{j=1}^{T} a_h^{-j}\varepsilon_{jh} + \frac{x_0}{\sqrt{k_T}}\right] + o_p(1)\\
&= X_{Th}\left[\frac{\tilde{g}_h\sqrt{k_T}}{(a_h-1)k_T} + Y_{Th} + \frac{x_0}{\sqrt{k_T}}\right] + o_p(1)\\
&\Rightarrow X\left[\frac{\kappa\mu + \sigma i\psi'(0)}{-\kappa\sigma\sqrt{\psi''(0)}} + Y + \frac{y_0}{\sigma\sqrt{\psi''(0)}}\right]\\
&= \frac{1}{-2\kappa}\left[\sqrt{-2\kappa}X\right]\left[\sqrt{-2\kappa}Y + \sqrt{-2\kappa}\frac{\kappa\mu + \sigma i\psi'(0) - \kappa y_{0h}}{-\kappa\sigma\sqrt{\psi''(0)}}\right]\\
&= \frac{1}{-2\kappa}\xi[\eta + D],
\end{aligned}$$

where the third equation is obtained from Eq. (3.2), $\xi = \sqrt{-2\kappa}X \overset{d}{=} N(0,1)$ and $\eta = \sqrt{-2\kappa}Y \overset{d}{=} N(0,1)$.

(d) The function $x_{th} = a_h x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th}$ leads to

$$x_{th}^2 = a_h^2 x_{(t-1)h}^2 + 2\tilde{g}_h a_h x_{(t-1)h} + 2a_h x_{(t-1)h}\varepsilon_{th} + \tilde{g}_h^2 + \varepsilon_{th}^2 + 2\tilde{g}_h\varepsilon_{th},$$

and

$$x_{th}^2 - x_{(t-1)h}^2 = \left(a_h^2 - 1\right)x_{(t-1)h}^2 + 2\tilde{g}_h a_h x_{(t-1)h} + 2a_h x_{(t-1)h}\varepsilon_{th}$$

$$+\tilde{g}_h^2+\varepsilon_{th}^2+2\tilde{g}_h\varepsilon_{th}.$$

Hence,

$$\left(a_h^2-1\right)\sum_{t=1}^{T}x_{(t-1)h}^2$$
$$=\left(x_{Th}^2-x_{0h}^2\right)-2\tilde{g}_h a_h\sum_{t=1}^{T}x_{(t-1)h}-2a_h\sum_{t=1}^{T}x_{(t-1)h}\varepsilon_{th}-T\tilde{g}_h^2$$
$$-\sum_{t=1}^{T}\varepsilon_{th}^2-2\tilde{g}_h\sum_{t=1}^{T}\varepsilon_{th}.$$

When $h\to 0$, $N\to\infty$, the results in parts (b) and (c) provide

$$\frac{a_h^{-2T}}{k_T}\left(2\tilde{g}_h a_h\sum_{t=1}^{T}x_{(t-1)h}\right)$$
$$=2\tilde{g}_h\sqrt{k_T}\frac{a_h^{-T+1}}{[a_h-1]\,k_T}\left(\frac{a_h^{-T}\,[a_h-1]}{\sqrt{k_T}}\sum_{t=1}^{T}x_{(t-1)h}\right)=o_p\left(1\right),$$
$$\frac{a_h^{-2T}}{k_T}\left(2a_h\sum_{t=1}^{T}x_{(t-1)h}\varepsilon_{th}\right)$$
$$=2a_h^{-T+1}\left(\frac{a_h^{-T}}{k_T}\sum_{t=1}^{T}x_{(t-1)h}\varepsilon_{th}\right)=o_p\left(1\right).$$

And, since $E\left|\frac{a_h^{-2T}}{k_T}\sum_{t=1}^{T}\varepsilon_{th}^2\right|=\frac{a_h^{-2T}}{k_T}\sum_{t=1}^{T}E\left(\varepsilon_{th}^2\right)=\frac{a_h^{-2T}T}{k_T}\to 0$, we get $\frac{a_h^{-2T}}{k_T}\sum_{t=1}^{T}\varepsilon_{th}^2\xrightarrow{L_1}0$. Together with the results that $a_h^{-2T}\frac{x_{0h}^2}{k_T}\to 0$, $\frac{a_h^{-2T}}{k_T}T\tilde{g}_h^2\to 0$ and $\frac{a_h^{-2T}}{k_T}2\tilde{g}_h\sum_{t=1}^{T}\varepsilon_{th}=\frac{a_h^{-2T}\sqrt{T}}{k_T}2\tilde{g}_h\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th}\to 0$, when $h\to 0$, $N\to\infty$, we have

$$\left(a_h^2-1\right)\frac{a_h^{-2T}}{k_T}\sum_{t=1}^{T}x_{(t-1)h}^2=\left(\frac{a_h^{-T}}{\sqrt{k_T}}x_{Th}\right)^2+o_p(1)$$
$$\Rightarrow\frac{1}{-2\kappa}\left(\eta+D\right)^2,$$

where the final limit result is proved in part (a).

Proof of Theorem 3.3. Parts (a) and (b) are immediate consequences of Lemmas 3.1 and 3.2. Based on the facts that $\widehat{\tilde{g}}_h-\tilde{g}_h=[\lambda_h]^{-1}\left(\widehat{g}_h-g_h\right)$ and

$$\sqrt{k_T}\lambda_h=\sqrt{k_T}\sigma\sqrt{\psi''\left(0\right)\left[1-e^{-2\kappa h}\right]/2\kappa}$$
$$\to\sigma\sqrt{\psi''\left(0\right)}\quad\text{as } h\to 0,$$

part (c) can be obtained straightforwardly from part (b).

Together with the results in (a), (b) and Lemma 3.2, the limit

$$\frac{1}{T}\sum_{t=1}^{T}\left(x_{th}-\widehat{a}_h\left(\kappa\right)x_{t-1h}-\widehat{\tilde{g}}_h\right)^2\to 1\quad\text{as } h\to 0\text{ and } N\to\infty$$

leads to the results in parts (d) and (e) straightforwardly. The above limit holds because, when $h\to 0$ and $N\to\infty$,

$$\frac{1}{T}\sum_{t=1}^{T}\left(x_{th}-\widehat{a}_h\left(\kappa\right)x_{t-1h}-\widehat{\tilde{g}}_h\right)^2$$
$$=\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{th}-\left[\widehat{a}_h\left(\kappa\right)-a_h\left(\kappa\right)\right]x_{t-1h}-\left[\widehat{\tilde{g}}_h-\tilde{g}_h\right]\right)^2$$
$$=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{th}^2+o_p\left(1\right)\to 1,$$

where the limit follows the law of large number for martingale difference arrays (see e.g. Hall and Heyde, 1980, Theorem 2.23).

(f) Since $\kappa=-(1/h)\ln\left(a_h\left(\kappa\right)\right)$ and $\widehat{\kappa}=-(1/h)\ln\left(\widehat{a}_h\left(\kappa\right)\right)$, by the mean value theorem,

$$-h\left(\widehat{\kappa}-\kappa\right)=\ln\left(\widehat{a}_h\left(\kappa\right)\right)-\ln\left(a_h(\kappa)\right)=\frac{1}{\beta_h(\kappa)}\left(\widehat{a}_h\left(\kappa\right)-a_h(\kappa)\right)$$

for some $\beta_h(\kappa)$ whose value is between $\widehat{a}_h(\kappa)$ and $a_h(\kappa)$. The *Delta* method is not directly applicable since $a_h(\kappa)$ is not a constant but a real sequence that goes to 1 as $h\to 0$. However, if we can show $\beta_h(\kappa)\xrightarrow{p}1$, we can obtain the limiting distribution of $\widehat{\kappa}$.

For any $\varepsilon>0$, when h is small enough, $|a_h(\kappa)-1|<\varepsilon/2$. Then,

$$\begin{aligned}\Pr\left\{|\beta_h(\kappa)-1|>\varepsilon\right\}&=\Pr\left\{|\beta_h(\kappa)-a_h(\kappa)+a_h(\kappa)-1|>\varepsilon\right\}\\&\leq\Pr\left\{|\beta_h(\kappa)-a_h(\kappa)|+|a_h(\kappa)-1|>\varepsilon\right\}\\&\leq\Pr\left\{|\widehat{a}_h\left(\kappa\right)-a_h(\kappa)|+|a_h(\kappa)-1|>\varepsilon\right\}\\&\leq\Pr\left\{|\widehat{a}_h\left(\kappa\right)-a_h(\kappa)|>\varepsilon/2\right\}\\&\to 0,\quad\text{as } h\to 0,\ \text{and } N\to\infty,\end{aligned}$$

where the first inequality is the triangular inequality, the second comes from the fact that $\beta_h(\kappa)$ is between $\widehat{a}_h\left(\kappa\right)$ and $a_h(\kappa)$, and the final result is based on the fact that $\widehat{a}_h\left(\kappa\right)-a_h(\kappa)\xrightarrow{p}0$. Hence, $\beta_h(\kappa)\xrightarrow{p}1$ and

$$\frac{e^{-\kappa N}}{2\kappa}\left(\widehat{\kappa}-\kappa\right)$$
$$=\frac{1-(a_h(\kappa))^2}{2\kappa h}\frac{1}{\beta_h(\kappa)}\left[\frac{(a_h(\kappa))^T}{(a_h(\kappa))^2-1}\left(\widehat{a}_h\left(\kappa\right)-a_h(\kappa)\right)\right]$$
$$\Rightarrow\frac{\xi}{\eta+D}.$$

Proof of Corollary 3.9. The results are straightforward consequences of Lemma 3.2, so proofs are omitted.

Proof of Lemma 3.16. Let us first define $x_{th}^0=x_{th}-\frac{1-[a_h(\kappa)]^t}{1-a_h(\kappa)}\tilde{g}_h$ and hence $x_{th}^0=a_h\left(\kappa\right)x_{(t-1)h}^0+\varepsilon_{th}$ is an AR process with the initial value $x_{0h}^0=x_{0h}=O_p\left(\sqrt{k_T}\right)$. Before we obtain the results in Lemma 3.16, let us first show that

$$\frac{1}{\sqrt{T}k_T}\sum_{t=1}^{T}x_{(t-1)h}^0\Rightarrow Z/\kappa\overset{d}{=}N\left(0,1/\kappa^2\right),$$
$$\frac{1}{Tk_T}\sum_{t=1}^{\lfloor Ts\rfloor}\left[x_{(t-1)h}^0\right]^2\xrightarrow{p}\frac{s}{2\kappa},\qquad\text{for any } s\in[0,1],$$
$$\frac{1}{\sqrt{Tk_T}}\sum_{t=1}^{T}x_{(t-1)h}^0\varepsilon_{th}\Rightarrow\Xi/\sqrt{2\kappa}\overset{d}{=}N\left(0,1/2\kappa\right),$$

where Ξ and Z are independent $N\left(0,1\right)$ random variables and Z is the limit of Z_{Th} defined in (3.1).

Let $a_h(\kappa)=a_h$ when there is no confusion. From the definition of x_{th}^0, we have, for any $s\in[0,1]$

$$x_{\lfloor Ts\rfloor h}^0=\left(a_h\right)^{\lfloor Ts\rfloor}x_{0h}+\sum_{j=0}^{\lfloor Ts\rfloor-1}\left(a_h\right)^j\varepsilon_{(\lfloor Ts\rfloor-j)h}.$$

Note that when $h\to 0$ and $N\to\infty$,

$$\ln\left(a_h\right)^{\lfloor Ts\rfloor}=-\kappa h\left\lfloor Ts\right\rfloor=-\kappa N\left\lfloor Ts\right\rfloor/T\to-\infty.$$

Hence $(a_h)^{\lfloor Ts\rfloor} \to 0$. Together with the facts that $x_{0h} = O_p\left(\sqrt{k_T}\right)$ and that

$$E\left(\frac{1}{\sqrt{T}}\sum_{j=0}^{\lfloor Ts\rfloor-1}(a_h)^j\,\varepsilon_{(\lfloor Ts\rfloor-j)h}\right)^2 = \frac{1}{T}\sum_{j=0}^{\lfloor Ts\rfloor-1}(a_h)^{2j} = \frac{1}{T}\frac{1-(a_h)^{2\lfloor Ts\rfloor}}{1-(a_h)^2} = \frac{k_T}{T}\frac{1-(a_h)^{2\lfloor Ts\rfloor}}{k_T\left[1-(a_h)^2\right]} \to 0,$$

we get

$$\frac{x^0_{\lfloor Ts\rfloor h}}{\sqrt{T}} = \frac{(a_h)^{\lfloor Ts\rfloor}\sqrt{k_T}}{\sqrt{T}}\frac{x_{0h}}{\sqrt{k_T}} + \frac{1}{\sqrt{T}}\sum_{j=0}^{\lfloor Ts\rfloor-1}(a_h)^j\,\varepsilon_{(\lfloor Ts\rfloor-j)h} \xrightarrow{p} 0.$$

As $x^0_{th} - x^0_{(t-1)h} = (a_h-1)\,x^0_{(t-1)h} + \varepsilon_{th}$, we then have

$$x^0_{Th} - x_{0h} = (a_h-1)\sum_{t=1}^{T}x^0_{(t-1)h} + \sum_{t=1}^{T}\varepsilon_{th},$$

and

$$\frac{a_h-1}{\sqrt{T}}\sum_{t=1}^{T}x^0_{(t-1)h} = \frac{x^0_{Th}-x_{0h}}{\sqrt{T}} - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th} + o_p(1).$$

Therefore, when $h \to 0$ and $N \to \infty$,

$$\frac{1}{\sqrt{T}k_T}\sum_{t=1}^{T}x^0_{(t-1)h} = \frac{1}{[a_h-1]\,k_T}\frac{a_h-1}{\sqrt{T}}\sum_{t=1}^{T}x^0_{(t-1)h} = \frac{1}{-\kappa+O(h)}\left(-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th} + o_p(1)\right) = \frac{1}{\kappa}Z_{Th} + o_p(1) \Rightarrow \frac{1}{\kappa}Z \overset{d}{=} N\left(0, 1/\kappa^2\right).$$

Moreover,

$$\frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor}x^0_{(t-1)h}\varepsilon_{th} = \frac{1}{T}\left\{\sum_{t=2}^{\lfloor Ts\rfloor}\left((a_h)^{(t-1)}x_{0h} + \sum_{j=1}^{t-1}(a_h)^{t-1-j}\varepsilon_{jh}\right)\varepsilon_{th} + x_{0h}\varepsilon_{1h}\right\}$$
$$= \frac{x_{0h}}{\sqrt{k_T}}\frac{\sqrt{k_T}}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Ts\rfloor}(a_h)^{(t-1)}\varepsilon_{th} + \frac{1}{T}\sum_{t=2}^{\lfloor Ts\rfloor}\sum_{j=1}^{t-1}(a_h)^{t-1-j}\varepsilon_{jh}\varepsilon_{th} \xrightarrow{p} 0.$$

This is right because when $h \to 0$ and $N \to \infty$,

$$E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Ts\rfloor}(a_h)^{(t-1)}\varepsilon_{th}\right)^2 = \frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor}(a_h)^{2(t-1)} = \frac{k_T}{T}\frac{1-(a_h)^{2\lfloor Ts\rfloor}}{k_T\left[1-(a_h)^2\right]} \to 0,$$

and

$$E\left(\frac{1}{T}\sum_{t=2}^{\lfloor Ts\rfloor}\sum_{j=1}^{t-1}(a_h)^{t-1-j}\varepsilon_{jh}\varepsilon_{th}\right)^2 = \frac{1}{T^2}\sum_{t=2}^{\lfloor Ts\rfloor}\sum_{j=1}^{t-1}(a_h)^{2(t-1-j)} = \frac{1}{T^2}\sum_{t=2}^{\lfloor Ts\rfloor}\frac{1-(a_h)^{2(t-1)}}{1-(a_h)^2}$$
$$= \frac{\lfloor Ts\rfloor}{T^2}\frac{1}{1-(a_h)^2} - \frac{1}{T^2\left(1-(a_h)^2\right)}\frac{(a_h)^2\left(1-(a_h)^{2\lfloor Ts\rfloor-2}\right)}{1-(a_h)^2}$$
$$= O\left(k_T/T\right) + O\left((k_T/T)^2\right) \to 0.$$

From the definition of x^0_{th}, we have

$$\left(x^0_{th}\right)^2 - \left(x^0_{(t-1)h}\right)^2 = \left[(a_h)^2-1\right]\left(x^0_{(t-1)h}\right)^2 + \varepsilon^2_{th} + 2a_hx^0_{(t-1)h}\varepsilon_{th},$$

and

$$\left[(a_h)^2-1\right]\sum_{t=1}^{\lfloor Ts\rfloor}\left(x^0_{(t-1)h}\right)^2 = \left(x^0_{\lfloor Ts\rfloor h}\right)^2 - x^2_{0h} - \sum_{t=1}^{\lfloor Ts\rfloor}\varepsilon^2_{th} - 2a_h\sum_{t=1}^{\lfloor Ts\rfloor}x^0_{(t-1)h}\varepsilon_{th}.$$

Therefore, when $h \to 0$ and $N \to \infty$,

$$\frac{1}{Tk_T}\sum_{t=1}^{\lfloor Ts\rfloor}\left(x^0_{(t-1)h}\right)^2 = \frac{1}{k_T\left[(a_h)^2-1\right]}\frac{\left[(a_h)^2-1\right]}{T}\sum_{t=1}^{\lfloor Ts\rfloor}\left(x^0_{(t-1)h}\right)^2 = \frac{1}{-2\kappa+O(h)}\left\{-\frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor}\varepsilon^2_{th} + o_p(1)\right\} \xrightarrow{p} \frac{s}{2\kappa}.$$

Denote

$$Z_{Th} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th} \quad\text{and}\quad \Xi_{Th} = \frac{\sqrt{2\kappa}}{\sqrt{Tk_T}}\sum_{t=1}^{T}x^0_{(t-1)h}\varepsilon_{th}.$$

We now prove that Z_{Th} and Ξ_{Th} are asymptotically independent with standard normal limiting distributions. By the Cramér–Wold device, it is sufficient to show that

$$aZ_{Th} + b\Xi_{Th} \Rightarrow aZ + b\Xi, \quad \text{for all } a, b \in \mathbb{R},$$

where Z and Ξ are independent $N(0,1)$ random variables. If Υ is an $N\left(0, a^2+b^2\right)$ random variable, $aZ+b\Xi \overset{d}{=} \Upsilon$, then the sufficient condition becomes $aZ_{Th} + b\Xi_{Th} \Rightarrow \Upsilon$, for all $a, b \in \mathbb{R}$.

We can write $aZ_{Th} + b\Xi_{Th} = \sum_{t=1}^{T}\xi_{Tt}$ where

$$\{\xi_{Tt}\}_{t=1}^{T} = \left\{a\frac{\varepsilon_{th}}{\sqrt{T}} + b\frac{\sqrt{2\kappa}x^0_{(t-1)h}\varepsilon_{th}}{\sqrt{Tk_T}}\right\}_{t=1}^{T},$$

is a martingale difference array as h and N varies, because $\varepsilon_{th} \overset{iid}{\sim} (0,1)$ whenever h is fixed and x_{0h} is independent with $\sigma(\varepsilon_{1h}, \ldots, \varepsilon_{Th})$. Thus, the weak convergence to a Gaussian random variable can be derived as a consequence of the CLT for martingale difference arrays.

The conditional variance is obtained as

$$V^2_{TT} = \sum_{t=1}^{T}E\left(\xi^2_{Tt}\,\middle|\,\mathcal{F}_{T,t-1}\right) = \sum_{t=1}^{T}E\left(\frac{(a\varepsilon_{th})^2}{T} + \frac{2\kappa b^2\left(x^0_{(t-1)h}\varepsilon_{th}\right)^2}{Tk_T}\right.$$

$$
\begin{aligned}
&\quad + \left. \frac{2\sqrt{2\kappa}abx^0_{(t-1)h}\varepsilon^2_{th}}{T\sqrt{k_T}} \right| \mathcal{F}_{T,t-1} \Bigg) \\
&= a^2 + \sum_{t=1}^{T} \frac{2\kappa b^2 \left(x^0_{(t-1)h}\right)^2}{Tk_T} + 2\sqrt{2\kappa}ab \sum_{t=1}^{T} \frac{x^0_{(t-1)h}}{T\sqrt{k_T}} \\
&= a^2 + 2\kappa b^2 \frac{1}{Tk_T} \sum_{t=1}^{T} \left(x^0_{(t-1)h}\right)^2 \\
&\quad + 2\sqrt{2\kappa}ab \frac{\sqrt{k_T}}{\sqrt{T}} \frac{1}{k_T\sqrt{T}} \sum_{t=1}^{T} x^0_{(t-1)h} \\
&= a^2 + 2\kappa b^2 \left(\frac{1}{2\kappa} + o_p(1)\right) + O_p\left(\frac{\sqrt{k_T}}{\sqrt{T}}\right) \xrightarrow{p} a^2 + b^2.
\end{aligned}
$$

To prove the conditional Lindeberg condition, we first get, for any $\varepsilon > 0$,

$$
\begin{aligned}
&\sum_{t=1}^{T} E\left(\xi^2_{Tt} 1\{|\xi_{Tt}| > \varepsilon\} \middle| \mathcal{F}_{T,t-1}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\frac{a}{\sqrt{T}} + \frac{\sqrt{2\kappa}bx^0_{(t-1)h}}{\sqrt{Tk_T}}\right)^2 \\
&\quad \times E\left(\varepsilon^2_{th} 1\left\{\left|\frac{a\varepsilon_{th}}{\sqrt{T}} + \frac{\sqrt{2\kappa}bx^0_{(t-1)h}\varepsilon_{th}}{\sqrt{Tk_T}}\right| > \varepsilon\right\} \middle| \mathcal{F}_{T,t-1}\right) \\
&\le V^2_{TT} \cdot \max_{1\le t\le T} E\left(\varepsilon^2_{th} 1\left\{\left|\frac{a\varepsilon_{th}}{\sqrt{T}} + \frac{\sqrt{2\kappa}bx^0_{(t-1)h}\varepsilon_{th}}{\sqrt{Tk_T}}\right| > \varepsilon\right\} \middle| \mathcal{F}_{T,t-1}\right) \\
&\le V^2_{TT} \cdot \max_{1\le t\le T} E\left(\varepsilon^2_{th} 1\left\{\left|\frac{a\varepsilon_{th}}{\sqrt{T}}\right| + \left|\frac{\sqrt{2\kappa}bx^0_{(t-1)h}\varepsilon_{th}}{\sqrt{Tk_T}}\right| > \varepsilon\right\} \middle| \mathcal{F}_{T,t-1}\right) \\
&\le V^2_{TT} \cdot \max_{1\le t\le T} E\left(\varepsilon^2_{th} 1\left\{\left|\frac{\sqrt{2\kappa}bx^0_{(t-1)h}\varepsilon_{th}}{\sqrt{Tk_T}}\right| > \frac{\varepsilon}{2}\right\} \middle| \mathcal{F}_{T,t-1}\right) \\
&\quad \text{when } T \text{ is large enough,}
\end{aligned}
$$

where the last inequality is based on the fact that $\varepsilon_{th}/\sqrt{T} \to 0$ almost everywhere as $T \to \infty$. Hence, the conditional Lindeberg condition will be satisfied if

$$
\max_{1\le t\le T} E\left(\varepsilon^2_{th} 1\left\{\left|x^0_{(t-1)h}\varepsilon_{th}\right| > \frac{\sqrt{Tk_T}\varepsilon}{2\sqrt{2\kappa}\,|b|}\right\} \middle| \mathcal{F}_{T,t-1}\right) \xrightarrow{p} 0
$$
as $h \to 0$ and $N \to \infty$.

Applying the Hölder and Chebyshev inequalities, we obtain, for some $\delta > 0$

$$
\begin{aligned}
&E\left(\varepsilon^2_{th} 1\left\{\left|x^0_{(t-1)h}\varepsilon_{th}\right| > \frac{\sqrt{Tk_T}\varepsilon}{2\sqrt{2\kappa}\,|b|}\right\} \middle| \mathcal{F}_{T,t-1}\right) \\
&\le \left\{E\left(\left|\varepsilon^2_{th}\right|^{(2+\delta)/2} \middle| \mathcal{F}_{T,t-1}\right)\right\}^{2/(2+\delta)} \\
&\quad \times \left[P\left\{\left|x^0_{(t-1)h}\varepsilon_{th}\right| > \frac{\sqrt{Tk_T}\varepsilon}{2\sqrt{2\kappa}\,|b|} \middle| \mathcal{F}_{T,t-1}\right\}\right]^{\delta/(2+\delta)} \\
&\le \left\{E\left(\left|\varepsilon_{th}\right|^{(2+\delta)}\right)\right\}^{2/(2+\delta)} \left(\frac{E\left\{\left|x^0_{(t-1)h}\varepsilon_{th}\right|^2 \middle| \mathcal{F}_{T,t-1}\right\}}{Tk_T\varepsilon^2/8\kappa b^2}\right)^{\delta/(2+\delta)} \\
&= \left\{E\left(\left|\varepsilon_{th}\right|^{(2+\delta)}\right)\right\}^{2/(2+\delta)} \left(\frac{\left(x^0_{(t-1)h}\right)^2}{Tk_T}\right)^{\delta/(2+\delta)} \left(\frac{8\kappa b^2}{\varepsilon^2}\right)^{\delta/(2+\delta)},
\end{aligned}
$$
for each $t \in \{1, \ldots, T\}$.

Since ε_{th} are identically distributed once h is fixed, the sufficient condition for conditional Lindeberg condition now changes to be

$$
\max_{1\le t\le T} \frac{\left(x^0_{(t-1)h}\right)^2}{Tk_T} \xrightarrow{p} 0 \quad \text{as } h \to 0 \text{ and } N \to \infty.
$$

For $m \in \{1, \ldots, T\}$, define the sets

$$
B_{T,m} := \bigcap_{j=1}^{m} \left\{\omega : \left|\frac{1}{Tk_T} \sum_{t=1}^{\lfloor T(j/m)\rfloor} \left[x^0_{(t-1)h}(\omega)\right]^2 - \frac{j}{m}\frac{1}{2\kappa}\right| \le \frac{1}{m}\right\}.
$$

As $(Tk_T)^{-1}\sum_{t=1}^{\lfloor Ts\rfloor}\left[x^0_{(t-1)h}\right]^2 \xrightarrow{p} s/2\kappa$ for any $s \in [0, 1]$, we have $P\left(B_{T,m}\right) \to 1$ for each m when $h \to 0$ and $N \to \infty$. Next, note that[9]

$$
\max_{1\le t\le T} \frac{\left(x^0_{(t-1)h}\right)^2}{Tk_T} \le \frac{1}{Tk_T} \sup_{s\in[0,1]} \sum_{t=\lfloor Ts\rfloor+1}^{\lfloor T(s+1/m)\rfloor} \left(x^0_{(t-1)h}\right)^2.
$$

For any given $s \in [0, 1]$, choose $j \in \{1, \ldots, m\}$ so that $s \in [(j-1)/m, j/m]$. Then, for any $s \in [0, 1]$, $\omega \in B_{T,m}$ implies

$$
\begin{aligned}
&\frac{1}{Tk_T} \sum_{t=\lfloor Ts\rfloor+1}^{\lfloor T(s+1/m)\rfloor} \left(x^0_{(t-1)h}\right)^2 \\
&\le \frac{1}{Tk_T} \sum_{t=\lfloor T(j-1)/m\rfloor+1}^{\lfloor T(j+1)/m\rfloor} \left(x^0_{(t-1)h}\right)^2 \\
&= \left(\frac{1}{Tk_T} \sum_{t=1}^{\lfloor T(j+1)/m\rfloor} \left(x^0_{(t-1)h}\right)^2 - \frac{j+1}{m}\frac{1}{2\kappa}\right) \\
&\quad - \left(\frac{1}{Tk_T} \sum_{t=1}^{\lfloor T(j-1)/m\rfloor} \left(x^0_{(t-1)h}\right)^2 - \frac{j-1}{m}\frac{1}{2\kappa}\right) + \frac{1}{m\kappa} \\
&\le \frac{1}{m} + \frac{1}{m} + \frac{1}{m\kappa} = \frac{2\kappa+1}{m\kappa}.
\end{aligned}
$$

Therefore, for any $m \in \{1, \ldots, T\}$,

$$
\begin{aligned}
1 &= \lim_{h\to 0, N\to\infty} P\left(B_{T,m}\right) \\
&\le \lim_{h\to 0, N\to\infty} P\left\{\max_{1\le t\le T} \frac{\left(x^0_{(t-1)h}\right)^2}{Tk_T} \le \frac{2\kappa+1}{m\kappa}\right\}.
\end{aligned}
$$

As $h \to 0$ and $N \to \infty$, we have $T \to \infty$ and m can take an arbitrarily large number. Therefore, $\max_{1\le t\le T}\frac{\left(x^0_{(t-1)h}\right)^2}{Tk_T} \xrightarrow{p} 0$, and the conditional Lindeberg condition is proved.

Now, we can prove (a)–(c).

(a) From the relation $x_{th} = \frac{1-[a_h]^t}{1-a_h}\tilde{g}_h + x^0_{th}$, we have

$$
\frac{1}{T\sqrt{k_T}} \sum_{t=1}^{T} x_{(t-1)h}
$$

9 As we only have observations up to x^0_{Th}, for the values of s and m such that $\lfloor T(s+1/m)\rfloor > T+1$, we terminate the summation in the following display at $T+1$. For example, when $s = 1$, we define it as $\sum_{t=T+1}^{\lfloor T+T/m\rfloor}\left(x^0_{(t-1)h}\right)^2 = \left(x^0_{Th}\right)^2$ for any value of $m \in \{1, 2, \ldots, T\}$.

$$= \frac{\sqrt{k_T}\tilde{g}_h}{k_T(1-a_h)} - \frac{k_T}{T}\frac{\sqrt{k_T}\tilde{g}_h\left(1-a_h^T\right)}{k_T^2(1-a_h)^2} + \frac{\sqrt{k_T}}{\sqrt{T}}\frac{1}{\sqrt{T}k_T}\sum_{t=1}^{T}x_{(t-1)h}^0$$

$$= \frac{\sqrt{k_T}\tilde{g}_h}{k_T(1-a_h)} + O(k_T/T) + O_p\left(\sqrt{k_T}/\sqrt{T}\right) \to \frac{\kappa\mu+\sigma i\psi'(0)}{\kappa\sigma\sqrt{\psi''(0)}}$$

as $h \to 0$ and $N \to \infty$.

(b) Also starting from the relation $x_{th} = \frac{1-[a_h]^t}{1-a_h}\tilde{g}_h + x_{th}^0$, we have

$$\frac{1}{\sqrt{Tk_T}}\sum_{t=1}^{T}x_{(t-1)h}\varepsilon_{th}$$
$$= \frac{\sqrt{k_T}\tilde{g}_h}{k_T(1-a_h)}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th} - \frac{\sqrt{k_T}\tilde{g}_h}{k_T(1-a_h)}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[a_h]^{t-1}\varepsilon_{th} + \frac{1}{\sqrt{Tk_T}}\sum_{t=1}^{T}x_{(t-1)h}^0\varepsilon_{th}$$
$$= \frac{\sqrt{k_T}\tilde{g}_h}{k_T(1-a_h)}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{th} + \frac{1}{\sqrt{Tk_T}}\sum_{t=1}^{T}x_{(t-1)h}^0\varepsilon_{th} + O_p\left(\frac{\sqrt{k_T}}{\sqrt{T}}\right)$$
$$\Rightarrow \frac{\kappa\mu+\sigma i\psi'(0)}{\kappa\sigma\sqrt{\psi''(0)}}Z + \Xi/\sqrt{2\kappa}.$$

(c) The function $x_{th} = a_h x_{(t-1)h} + \tilde{g}_h + \varepsilon_{th}$ leads to

$$x_{th}^2 - x_{(t-1)h}^2 = \left(a_h^2-1\right)x_{(t-1)h}^2 + 2\tilde{g}_h a_h x_{(t-1)h} + 2a_h x_{(t-1)h}\varepsilon_{th} + \tilde{g}_h^2 + \varepsilon_{th}^2 + 2\tilde{g}_h\varepsilon_{th}.$$

Hence,

$$\left(a_h^2-1\right)\sum_{t=1}^{T}x_{(t-1)h}^2$$
$$= \left(x_{Th}^2 - x_{0h}^2\right) - 2\tilde{g}_h a_h\sum_{t=1}^{T}x_{(t-1)h} - 2a_h\sum_{t=1}^{T}x_{(t-1)h}\varepsilon_{th} - T\tilde{g}_h^2 - \sum_{t=1}^{T}\varepsilon_{th}^2 - 2\tilde{g}_h\sum_{t=1}^{T}\varepsilon_{th}.$$

Together with the results in parts (a) and (b), the fact that

$$\frac{x_{Th}}{\sqrt{k_T}} = \left(\sqrt{k_T}\tilde{g}_h\right)\frac{1-[a_h]^T}{(1-a_h)k_T} + \frac{x_{Th}^0}{\sqrt{k_T}} = O_p(1)$$

as $h \to 0$, $N \to \infty$,

leads to

$$\frac{a_h^2-1}{T}\sum_{t=1}^{T}x_{(t-1)h}^2$$
$$= -\frac{2\sqrt{k_T}\tilde{g}_h a_h}{T\sqrt{k_T}}\sum_{t=1}^{T}x_{(t-1)h} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{th}^2 + o_p(1)$$
$$\to -\frac{2}{\kappa}\left(\frac{\kappa\mu+\sigma i\psi'(0)}{\sigma\sqrt{\psi''(0)}}\right)^2 - 1.$$

Therefore, when $h \to 0$, $N \to \infty$, we have

$$\frac{1}{Tk_T}\sum_{t=1}^{T}x_{(t-1)h}^2 = \frac{1}{k_T\left(a_h^2-1\right)}\frac{a_h^2-1}{T}\sum_{t=1}^{T}x_{(t-1)h}^2 \to \frac{1}{2\kappa} + \left(\frac{\kappa\mu+\sigma i\psi'(0)}{\sigma\kappa\sqrt{\psi''(0)}}\right)^2.$$

Proof of Eq. (4.1). Let $\{N_i\}$ be any integer sequence diverging to ∞. Then for any fixed k_T and N_i, we have

$$\frac{Y_a}{\sqrt{k_T}} \overset{d}{=} \frac{1}{\sqrt{k_T}}\sum_{t=1}^{\infty}(a_h(\kappa))^{-t+1}\varepsilon_{th} = \frac{1}{\sqrt{k_T}}\left(\sum_{t=1}^{N_ik_T}(a_h(\kappa))^{-t+1}\varepsilon_{th} + \sum_{t=N_ik_T+1}^{\infty}(a_h(\kappa))^{-t+1}\varepsilon_{th}\right).$$

Therefore,

$$\lim_{k_T\to\infty}\frac{Y_a}{\sqrt{k_T}} \overset{d}{=} \lim_{k_T\to\infty,\ N_i\to\infty}\frac{1}{\sqrt{k_T}}\left(\sum_{t=1}^{N_ik_T}(a_h(\kappa))^{-t+1}\varepsilon_{th} + \sum_{t=N_ik_T+1}^{\infty}(a_h(\kappa))^{-t+1}\varepsilon_{th}\right).$$

With the same spirit of the proof of Lemma 3.1, when $N_i \to \infty$ and $k_T = h \to \infty$ simultaneously, the CLT for martingale difference arrays leads to

$$\frac{1}{\sqrt{k_T}}\sum_{t=1}^{N_ik_T}(a_h(\kappa))^{-t+1}\varepsilon_{th} \Rightarrow Y \overset{d}{=} N(0,-1/2\kappa).$$

Together with the fact that

$$\lim_{k_T\to\infty,N_i\to\infty}\frac{1}{k_T}E\left(\sum_{t=N_ik_T+1}^{\infty}(a_h(\kappa))^{-t+1}\varepsilon_{th}\right)^2 = \lim_{k_T\to\infty,N_i\to\infty}\frac{1}{k_T}\sum_{t=N_ik_T+1}^{\infty}(a_h(\kappa))^{-2(t-1)} = \lim_{k_T\to\infty,N_i\to\infty}\frac{(a_h(\kappa))^{-2N_ik_T}}{k_T\left[1-(a_h(\kappa))^{-2}\right]} = 0,$$

when $k_T = h \to \infty$, we have

$$\frac{Y_a}{\sqrt{k_T}} \Rightarrow Y \overset{d}{=} N(0,-1/2\kappa).$$

As $\varepsilon_{th} \overset{iid}{\sim} (0,1)$, it is easy to see that $\lim_{k_T\to\infty}X_a/\sqrt{k_T}$ is a $N(0,-1/2\kappa)$ random variable. We can denote it as X. Note that

$$\lim_{k_T\to\infty}\lim_{T\to\infty}\frac{1}{k_T}E\left(\sum_{t=\lfloor T/2\rfloor+1}^{T-1}(a_h(\kappa))^{-t+1}\varepsilon_{th}\right)^2 = 0,$$

and that

$$\lim_{k_T\to\infty}\lim_{T\to\infty}\frac{1}{k_T}E\left(\sum_{t=1}^{\lfloor T/2\rfloor}(a_h(\kappa))^{-(T-t)}\varepsilon_{th}\right)^2 = 0.$$

Therefore, the sequential asymptotics of

$$\lim_{k_T\to\infty}\frac{Y_a}{\sqrt{k_T}} \overset{d}{=} \lim_{k_T\to\infty}\lim_{T\to\infty}\frac{1}{\sqrt{k_T}}\left(\sum_{t=1}^{\lfloor T/2\rfloor}(a_h(\kappa))^{-t+1}\varepsilon_{th} + \sum_{t=\lfloor T/2\rfloor+1}^{T-1}(a_h(\kappa))^{-t+1}\varepsilon_{th}\right)$$
$$\overset{d}{=} \lim_{k_T\to\infty}\lim_{T\to\infty}\frac{1}{\sqrt{k_T}}\sum_{t=1}^{\lfloor T/2\rfloor}(a_h(\kappa))^{-t+1}\varepsilon_{th} + o_p(1),$$

and

$$\lim_{k_T\to\infty}\frac{X_a}{\sqrt{k_T}} \overset{d}{=} \lim_{k_T\to\infty}\lim_{T\to\infty}\frac{1}{\sqrt{k_T}}\left(\sum_{t=1}^{\lfloor T/2\rfloor}(a_h(\kappa))^{-(T-t)}\varepsilon_{th}\right.$$

$$+ \sum_{t=\lfloor T/2\rfloor+1}^{T} (a_h(\kappa))^{-(T-t)} \varepsilon_{th} \Bigg)$$
$$\overset{d}{=} \lim_{k_T\to\infty} \lim_{T\to\infty} \frac{1}{\sqrt{k_T}} \sum_{t=\lfloor T/2\rfloor+1}^{T} (a_h(\kappa))^{-(T-t)} \varepsilon_{th} + o_p(1),$$

are independent.

Proof of Theorem 4.2. (a) With $\gamma = y_0/\left(\sigma\sqrt{N}\right)$, Perron (1991) derived the joint MGF of $A(\gamma, c)$ and $B(\gamma, c)$ as

$$\begin{aligned} &MGF(v,u)\\ &= E\left[\exp\left(vA(\gamma,c)+uB(\gamma,c)\right)\right]\\ &= \Psi_c(v,u)\exp\left\{-\left(\frac{\gamma^2}{2}\right)(v+c-\lambda)\right.\\ &\quad\times\left.\left[1-\exp(v+c+\lambda)\Psi_c^2(v,u)\right]\right\}\\ &= \underbrace{\Psi_c(v,u)}_{I}\underbrace{\exp\left\{-\left(\frac{\gamma^2}{2}\right)(v+c-\lambda)\right\}}_{II}\\ &\quad\times\underbrace{\exp\left\{\left(\frac{\gamma^2}{2}\right)(v+c-\lambda)\exp(v+c+\lambda)\Psi_c^2(v,u)\right\}}_{III}, \end{aligned}$$

where

$$\lambda = \left(c^2+2cv-2u\right)^{1/2},$$
$$\Psi_c(v,u) = \left[\frac{2\lambda\exp\{-(v+c)\}}{(\lambda+(v+c))\exp\{-\lambda\}+(\lambda-(v+c))\exp\{\lambda\}}\right]^{1/2}.$$

Let $v = \widetilde{v}(2c)e^{-c}$ and $u = \widetilde{u}(2c)^2e^{-2c}$. The joint MGF of $(2c)e^{-c}A(\gamma,c)$ and $(2c)^2e^{-2c}B(\gamma,c)$ is

$$M(\widetilde{v},\widetilde{u}) = E\left[\exp\left(\widetilde{v}(2c)e^{-c}A(\gamma,c)+\widetilde{u}(2c)^2e^{-2c}B(\gamma,c)\right)\right].$$

We get

$$\begin{aligned} \lambda &= \left\{c^2+(2c)^2e^{-c}\widetilde{v}-2(2c)^2e^{-2c}\widetilde{u}\right\}^{1/2}\\ &= \left\{\left[c+(2c)e^{-c}\widetilde{v}-2(2c)e^{-2c}\widetilde{u}-(2c)e^{-2c}\widetilde{v}^2\right]^2 + O\left(e^{-3c}\right)\right\}^{1/2}\\ &= c+(2c)e^{-c}\widetilde{v}-2(2c)e^{-2c}\widetilde{u}-(2c)e^{-2c}\widetilde{v}^2+O\left(e^{-3c}\right),\\ \lambda+(v+c) &= 2c+2(2c)e^{-c}\widetilde{v}-2(2c)e^{-2c}\widetilde{u}-(2c)e^{-2c}\widetilde{v}^2+O\left(e^{-3c}\right), \end{aligned}$$
$$\lambda-(v+c) = -2(2c)e^{-2c}\widetilde{u}-(2c)e^{-2c}\widetilde{v}^2+O\left(e^{-3c}\right),$$
$$e^{-\lambda} = e^{-c}-(2c)e^{-2c}\widetilde{v}+O\left(e^{-3c}\right),$$

and

$$(\lambda-(v+c))e^{\lambda} = -(2c)e^{-c}\left[2\widetilde{u}+\widetilde{v}^2\right]+O\left(e^{-2c}\right).$$

The denominator of $\Psi_c^2(v,u)$ is

$$\begin{aligned} &(\lambda+(v+c))e^{-\lambda}+(\lambda-(v+c))e^{\lambda}\\ &= (2c)e^{-c}\left[1-2\widetilde{u}-\widetilde{v}^2\right]+O\left(e^{-2c}\right). \end{aligned}$$

The numerator of $\Psi_c^2(v,u)$ is

$$\begin{aligned} 2\lambda\exp\{-(v+c)\} &= 2\lambda\exp\left\{-(2c)e^{-c}\widetilde{v}-c\right\}\\ &= (2c)e^{-c}+O\left(c^2e^{-2c}\right). \end{aligned}$$

Hence,

$$\begin{aligned} I &= \Psi_c(v,u) = \left\{\frac{(2c)e^{-c}+O\left(c^2e^{-2c}\right)}{(2c)e^{-c}\left[1-2\widetilde{u}-\widetilde{v}^2\right]+O\left(e^{-2c}\right)}\right\}^{1/2}\\ &\to \left\{\frac{1}{1-2\widetilde{u}-\widetilde{v}^2}\right\}^{1/2}. \end{aligned}$$

It is easy to show that $II \to 1$ because

$$\begin{aligned} &-\left(\frac{\gamma^2}{2}\right)(v+c-\lambda)\\ &= \left(\frac{-y_0^2\kappa}{2\sigma^2c}\right)\left[-2(2c)e^{-2c}\widetilde{u}-(2c)e^{-2c}\widetilde{v}^2+O\left(e^{-3c}\right)\right]\to 0. \end{aligned}$$

Since

$$\begin{aligned} &\exp\{\lambda+v+c\}\\ &= e^{2c}\exp\left\{2(2c)e^{-c}\widetilde{v}-2(2c)e^{-2c}\widetilde{u}-(2c)e^{-2c}\widetilde{v}^2+O\left(e^{-3c}\right)\right\}, \end{aligned}$$

together with the definition $d = y_0\sqrt{-2\kappa}/\sigma$, we get $(2c)\gamma^2 = d^2$ and

$$\begin{aligned} &\left(\frac{\gamma^2}{2}\right)(v+c-\lambda)\exp\{\lambda+v+c\}\\ &= \frac{d^2}{2}\left[2\widetilde{u}+\widetilde{v}^2+O\left(e^{-c}\right)\right]\exp\left\{2(2c)e^{-c}\widetilde{v}-2(2c)e^{-2c}\widetilde{u}\right.\\ &\quad\left.-(2c)e^{-2c}\widetilde{v}^2+O\left(e^{-3c}\right)\right\}\\ &\to \frac{d^2}{2}\left[2\widetilde{u}+\widetilde{v}^2\right]. \end{aligned}$$

Therefore,

$$III \to \exp\left\{\frac{d^2\left[2\widetilde{u}+\widetilde{v}^2\right]}{2\left[1-2\widetilde{u}-\widetilde{v}^2\right]}\right\}.$$

The limit behavior of I, II and III gives rise to the limit joint MGF of $(2c)e^{-c}A(\gamma,c)$ and $(2c)^2e^{-2c}B(\gamma,c)$.

(b) Since ξ and η are independent $N(0,1)$ random variables and d is a constant, we have

$$\begin{aligned} M(\widetilde{v},\widetilde{u}) &= E\left\{\exp\left(\xi[d+\eta]\widetilde{v}+[d+\eta]^2\widetilde{u}\right)\right\}\\ &= E\left\{E\left[\exp\left(\xi[d+\eta]\widetilde{v}+[d+\eta]^2\widetilde{u}\right)\middle|\mathcal{F}_\xi\right]\right\}\\ &= E\left\{\exp\left([d+\eta]^2\widetilde{u}\right)\exp\left(\frac{[d+\eta]^2\widetilde{v}^2}{2}\right)\right\}\\ &= \frac{1}{\left\{1-2\widetilde{u}-\widetilde{v}^2\right\}^{1/2}}\exp\left\{\frac{d^2\left[2\widetilde{u}+\widetilde{v}^2\right]}{2\left[1-2\widetilde{u}-\widetilde{v}^2\right]}\right\}, \end{aligned}$$

where $\mathcal{F}_\xi$ is the σ-field generated by ξ. This is the joint MGF of $\xi[d+\eta]$ and $[d+\eta]^2$ and is equivalent to the result in (a).

(c) This is an immediate consequence of (b).

Derivation of the joint limiting distribution reported in (4.7): Phillips (1987b) showed that, when $c = -\kappa N \to +\infty$,

$$(2c)^2e^{-2c}\int_0^1 J_c(r)^2dr \Rightarrow \eta^2 \quad\text{and}$$
$$(2c)e^{-c}\int_0^1 J_c(r)\,dW(r) \Rightarrow \xi\eta,$$

where ξ and η are two independent $N(0,1)$ random variables. To derive the joint limiting distribution reported in (4.7), we only need to show separately that $\left\{(2c)^{3/2}e^{-2c}\int_0^1\exp\{cr\}J_c(r)\,dr\right\}^2 \Rightarrow \eta^2$ and $(2c)^{1/2}e^{-c}\int_0^1\exp\{cr\}\,dW(r) \Rightarrow \xi$.

$$\frac{(2c)\, e^{-c}A\,(\gamma, c)}{(2c)^2\, e^{-2c}B\,(\gamma, c)} = \frac{(2c)^{1/2}\,\gamma \left[(2c)^{1/2}\, e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right] + (2c)\, e^{-c} \int_0^1 J_c(r)\, dW(r)}{\gamma^2 (2c)\left[1 - e^{-2c}\right] + 2\gamma\,(2c)^{1/2}\left[(2c)^{3/2}\, e^{-2c} \int_0^1 \exp\{cr\} J_c(r)\, dr\right] + (2c)^2\, e^{-2c} \int_0^1 J_c(r)^2\, dr}$$
$$= \frac{d\left[(2c)^{1/2}\, e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right] + \left[\xi\eta + o_p(1)\right]}{\left[d^2 + o(1)\right] + 2d\left[\eta^2 + o_p(1)\right]^{1/2} + \left[\eta^2 + o_p(1)\right]}$$
$$= \frac{d\left[(2c)^{1/2}\, e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right] + \left[\xi\eta + o_p(1)\right]}{[d+\eta]^2 + o_p(1)},$$

where the first equation comes from the definition of $A(\gamma, c)$ and $B(\gamma, c)$ in equations of (4.4) and (4.5). Theorem 4.2 shows that

$$\frac{(2c)\, e^{-c}A\,(\gamma, c)}{(2c)^2\, e^{-2c}B\,(\gamma, c)} \Rightarrow \frac{\xi\,[d+\eta]}{[d+\eta]^2} \quad \text{as } c = -\kappa N \to +\infty.$$

Box I.

The fact that $2c \int_0^1 \exp\{cr\} J_c(r)\, dr = e^c J_c(1) - \int_0^1 \exp\{cr\}\, dW(r)$ leads to

$$\left\{(2c)^{3/2}\, e^{-2c} \int_0^1 \exp\{cr\} J_c(r)\, dr\right\}^2$$
$$= (2c)\, e^{-4c}\left\{e^c J_c(1) - \int_0^1 \exp\{cr\}\, dW(r)\right\}^2$$
$$= (2c)\, e^{-2c} J_c(1)^2 + e^{-2c}\left[(2c)^{1/2}\, e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right]^2$$
$$-2e^{-c}\left[(2c)^{1/2}\, e^{-c} J_c(1)\right]\left[(2c)^{1/2}\, e^{-c} \int_0^1 \exp\{cr\}\, dW(r)\right].$$

From the stochastic differentiation of $\left\{\int_0^r \exp\{-cs\}\, dW(s)\right\}^2$, following Phillips (1987b), we have,

$$\{J_c(1)\}^2 = 1 + 2c \int_0^1 J_c(r)^2\, dr + 2\int_0^1 J_c(r)\, dW(r).$$

We, therefore, have the limit as

$$\left\{(2c)^{1/2}\, e^{-c} J_c(1)\right\}^2$$
$$= (2c)\, e^{-2c} + (2c)^2\, e^{-2c} \int_0^1 J_c(r)^2\, dr + 2\,(2c)\, e^{-2c}$$
$$\times \int_0^1 J_c(r)\, dW(r) \Rightarrow \eta^2.$$

Since $\int_0^1 \exp\{cr\}\, dW(r) \sim N\left(0, \frac{\exp\{2c\}-1}{2c}\right)$, we then have $(2c)^{1/2}\, e^{-c} \int_0^1 \exp\{cr\}\, dW(r)$ is $O_p(1)$, and hence,

$$\left\{(2c)^{3/2}\, e^{-2c} \int_0^1 \exp\{cr\} J_c(r)\, dr\right\}^2$$
$$= \left\{(2c)^{1/2}\, e^{-c} J_c(1)\right\}^2 + o_p(1) \Rightarrow \eta^2.$$

Note that $2c\gamma^2 = -2\kappa y_0^2/\sigma^2 = d^2$. Then, based on the limiting results above, we have the equation given in Box I.
We, therefore, have

$$(2c)^{1/2}\, e^{-c} \int_0^1 \exp\{cr\}\, dW(r) \Rightarrow \xi.$$

References

Anderson, T.W., 1959. On asymptotic distribution of estimates of parameters of stochastic difference equations. Ann. Math. Stat. 30, 676–687.
Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. J. Finance 54, 1361–1395.
Aue, A., Horvath, L., 2007. A limit theorem for mildly explosive autoregression with stable errors. Econometric Theory 23, 201–220.
Barndorff-Nielsen, O.E., 1998. Processes of normal inverse gaussian type. Finance Stoch. 2, 41–68.
Bergstrom, A.R., 1990. Continuous Time Econometric Modeling. Oxford University Press, Oxford.
Carr, P., Wu, L., 2003. The finite moment log stable process and option pricing. J. Finance 58, 753–778.
Gurkaynak, R.S., 2008. Econometric tests of asset price bubbles: Taking stock. J. Econ. Surv. 22, 166–186.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press.
Homm, U., Breitung, J., 2012. Testing for speculative bubbles in stock markets: a comparison of alternative methods. J. Financ. Econom. 10, 198–231.
Hu, Y., Long, H., 2009. Least squares estimator for ornstein–uhlenbeck processes driven by α-stable motions. Stochastic Process. Appl. 119, 2465–2480.
Jeong, M., Park, J., 2011. Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model. Department of Economics, University of Indiana, Working Paper.
Kallenberg, O., 2002. Foundations of Modern Probability. Springer, Berlin.
Madan, D.B., Carr, P., Chang, E., 1998. The variance gamma processes and option pricing. Eur. Finance Rev. 2, 79–105.
Magdalinos, T., 2012. Mildly explosive autoregression under weak and strong dependence. J. Econometrics 169, 179–187.
Park, J.Y., 2003. Weak unit roots. In: Mimeographed. Department of Economics, Rice University.
Perron, P., 1991. A continuous time approximation to the unstable first order autoregressive processes: The case without an intercept. Econometrica 59, 211–236.
Phillips, P.C.B., 1987a. Time series regression with a unit root. Econometrica 55, 277–301.
Phillips, P.C.B., 1987b. Toward a unified asymptotic theory for autoregression. Biometrika 74, 533–547.
Phillips, P.C.B., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. J. Econometrics 136, 115–130.
Phillips, P.C.B., Magdalinos, T., 2009. Unit root and cointegrating limit theory when the initialization is in the infinite past. Econometric Theory 25, 1682–1715.
Phillips, P.C.B., Shi, S.P., Yu, J., 2014. Specification sensitivity in right-tailed unit root testing for explosive behavior. Oxford Bull. Econ. Stat. 76, 315–333.
Phillips, P.C.B., Shi, S.P., Yu, J., 2015a. Testing for multiple bubbles: historical episodes of exuberance and collapse in the s&p500. Internat. Econom. Rev. 56 (4), 1043–1078.
Phillips, P.C.B., Shi, S.P., Yu, J., 2015b. Testing for multiple bubbles: Limit theory of real time detector. Internat. Econom. Rev. 56 (4), 1079–1134.
Phillips, P.C.B., Wu, Y., Yu, J., 2011. Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? Internat. Econom. Rev. 52, 201–226.
Phillips, P.C.B., Yu, J., 2005. Jackknifing bond option prices. Rev. Financ. Stud. 18, 707–742.
Phillips, P.C.B., Yu, J., 2009. Simulation-based estimation of contingent-claims prices. Rev. Financ. Stud. 22, 3669–3705.
Phillips, P.C.B., Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. Quant. Econ. 2, 455–491.
Tang, C.Y., Chen, S.X., 2009. Parameter estimation and bias correction for diffusion processes. J. Econometrics 149, 65–81.
Shimizu, Y., 2009. Notes on drift estimation for certain non-recurrent diffusion processes from sampled data. Statist. Probab. Lett. 79, 2200–2207.
Wang, X., Yu, J., 2015. Limit theory for an explosive autoregressive process. Econ. Lett. 126, 176–180.
White, J.S., 1958. The limiting distribution of the serial correlation coefficient in the explosive case. Ann. Math. Stat. 29, 1188–1197.
Yu, J., 2014. Econometric analysis of continuous time models: a survey of peter phillips' work and some new results. Econometric Theory 30, 737–774.
Zhou, Q., Yu, J., 2015. Asymptotic theory for linear diffusions under alternative sampling schemes. Economic Letters 128, 1–5.