3-2013

Bias in the Mean Reversion Estimator in Continuous-Time Gaussian and Lévy Processes

Yong BAO
Purdue University

Aman ULLAH
University of California, Riverside

Yun WANG
University of International Business and Economics

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

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Bias in the Mean Reversion Estimator in Continuous-Time Gaussian and Lévy Processes

Yong Bao, Aman Ullah, Yun Wang & Jun Yu

March 2013

Paper No. 02-2013

Bias in the Mean Reversion Estimator in Continuous-Time Gaussian and Lévy Processes

Yong Bao*
Purdue University

Aman Ullah†
University of California, Riverside

Yun Wang‡
University of International Business and Economics

Jun Yu§
Singapore Management University

February 28, 2013

ABSTRACT

This paper develops the approximate finite-sample bias of the ordinary least squares or quasi maximum likelihood estimator of the mean reversion parameter in continuous-time Lévy processes. For the special case of Gaussian processes, our results reduce to those of Tang and Chen (2009) (when the long-run mean is unknown) and Yu (2012) (when the long-run mean is known). Simulations show that in general the approximate bias works well in capturing the true bias of the mean reversion estimator under difference scenarios. However, when the time span is small and the mean reversion parameter is approaching its lower bound, we find it more difficult to approximate well the finite-sample bias.

JEL Classification: C10, C22

*Corresponding Author: Department of Economics, Purdue University, 403 W. State Street, West Lafayette, IN 47907. Email: ybao@purdue.edu.

†Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu.

‡School of International Trade and Economics, University of International Business and Economics, Beijing, China. E-mail: wyuncolor@gmail.com.

§Sim Kee Boon Institute for Financial Economics, School of Economics and Lee Kong Chian School of Business, Singapore Management University, Singapore 178903. E-mail: yujun@smu.edu.sg.

1 Introduction

There is an extensive literature for using diffusion processes to model the dynamic behaviour of financial asset prices. For example, Vasicek (1977) used the following Ornstein-Uhlenbeck (OU) process to model the spot interest rate,

$$\mathrm{d}x(t) = \kappa(\mu - x(t))\mathrm{d}t + \sigma \mathrm{d}B(t), \tag{1.1}$$

where $B(t)$ is a standard Brownian motion. This is a Gaussian Markov process and possesses a stationary distribution when $\kappa > 0$. In this case, κ captures the rate of convergence towards its long-run mean μ. Tang and Chen (2009) considered a more general form of Brownian-motion-based continuous-time model, namely, a diffusion process,

$$\mathrm{d}x(t) = \kappa(\mu - x(t))\mathrm{d}t + \sigma(x(t);\theta)\mathrm{d}B(t), \tag{1.2}$$

where $\sigma(x(t);\theta)$ is the diffusion function of $x(t)$ at time t. If $\sigma(x(t);\theta) = \sigma\sqrt{x(t)}$, the diffusion process becomes the CIR model (Cox, Ingersoll, and Ross, 1985). A even more general diffusion process is given by,

$$\mathrm{d}x(t) = \mu(x(t);\theta)dt + \sigma(x(t);\theta)\mathrm{d}B(t), \tag{1.3}$$

with a general drift function $\mu(x(t);\theta)$. An important special case is when $\mu(x(t);\theta) = \mu x(t)$ and $\sigma(x(t);\theta) = \sigma x(t)$. Black and Scholes (1973) used it to model the spot price of a stock.

All these processes are based on the Brownian motion. Under some smoothness conditions on the drift and the diffusion functions, the sample path generated from $x(t)$ is continuous everywhere. In recent years, however, strong evidence of infinite activity jumps in financial variables has been reported. To capture the infinite activity jumps, continuous-time Lévy processes have become increasingly popular and various Lévy models have been developed in the asset pricing literature, see, among others, Barndorff-Nielsen (1998), Madan, Carr and Chang (1998), and Carr and Wu (2003).

In practice, one can only obtain the observations at discrete points from a finite time span. Based on discrete-time observations, different methods have been used to estimate continuous-time models. Phillips and Yu (2009) provided an overview of some widely used estimation methods. When the drift function is linear and slowly mean reverting, it is found that there is serious estimation bias in the mean reversion parameter κ by almost all the methods. Because this parameter is of important implications for asset pricing, risk management and forecasting, accurate estimation of it has received considerable attentions in the literature. For example, Yu (2012) approximated the bias of the maximum likelihood estimator (MLE) of κ when the long-run mean is known and the start-up value is random for the Gaussian OU process. Tang and Chen (2009) approximated the bias of MLE of κ when the long-run mean is unknown for the Gaussian OU process and the CIR model. To reduce the estimation bias of κ, Phillips and Yu (2005) proposed the jackknife method. While the jackknife increases the variance, a carefully designed jackknifing procedure can offer substantial reduction of the bias, leading to a decrease in the root mean square errors (RMSE). To further reduce RMSE, Phillips and Yu (2009) proposed the indirect inference method, whereas Tang and Chen (2009) proposed a parametric bootstrapping method. These two methods are simulation-based and hence numerically more demanding.

The difficulty in the estimation of κ is related to the finite-sample bias problem well documented for the discrete autoregressive model, see, for example, Kendall (1954). However, in contrast to the finite-sample bias of the estimated autoregressive parameter, which is inversely proportional to the sample size, the bias in the estimated κ can be severe when the time span is small, regardless of the sample size. In practically relevant cases, this estimation bias can be very large, and thus a thorough understanding of the bias becomes very important. For example, Phillips and Yu (2005) demonstrated that the bias of MLE of κ in the CIR model can be over 200% even with 25 years of data used (regardless of the sampling frequency). They further reported evidence that the estimation bias in the drift term has more serious implications for asset pricing than the bias caused by discretization and sometimes by misspecification of the diffusion function. The simulation results of Phillips and Yu (2005) and Tang and Chen (2009) indicated that the biases of the estimated long-run mean and parameters in the diffusion function are virtually zero. In the stationary Vasicek model, Tang and Chen (2009) further showed that the bias of the estimated κ is up to $O(T^{-1})$, while estimation biases for σ^2 and μ are $O(n^{-1})$ and $O(n^{-2})$, respectively, as $T \to \infty$ with h fixed. (Throughout, T, h, and $n(= T/h)$ denote the time span, sampling frequency, and number of observations/sample size, respectively.)

While the bias in estimating κ has been well studied in continuous-time diffusion processes, to the best of our knowledge, nothing has been reported on the analytical bias issue in continuous-time Lévy processes. The objective of this paper is to develop the approximate bias of the quasi maximum likelihood (QML) estimator of κ under the Lévy measure, and then study the effects of nonnormality and initial condition on the estimation bias.

The structure of this paper is as follows. Section 2 develops the main results of the paper. In Section 3, we report Monte Carlo evidence to check the quality of our approximation. Section 4 concludes. The proof of the main results is collected in Appendix.

2 Main Results

A Lévy-driven OU process is

$$\mathrm{d}x(t) = \kappa(\mu - x(t))\mathrm{d}t + \sigma \mathrm{d}L(t), \;\; x(0) = x_0,$$

where $L(t)$, $t \geq 0$, is a Lévy process with $L(0) = 0$ a.s. In the special case when $L(t)$ is a Brownian motion, the process is Ornstein-Uhlenbeck (OU) Gaussian process used by Vasicek (1977) to model the dynamics of short term interest rates.

It is well known that the QML estimator of κ is

$$\hat{\kappa} = -\frac{\ln(\hat{\phi})}{h}, \tag{2.1}$$

where $\hat{\phi}$ is the least-squares (LS) estimator of the autoregression coefficient ϕ from the discretized AR(1) model

$$x_{th} = \alpha + \phi x_{(t-1)h} + \varepsilon_{th}, \tag{2.2}$$

in which $\alpha = \mu(1 - e^{-\kappa h})$, $\phi = e^{-\kappa h}$, $\varepsilon_{th} = \sigma \int_{(t-1)h}^{th} e^{-\kappa(th-s)}\mathrm{d}L(s)$, h is the sampling interval, $t =$

$1, \cdots, n$ such that the observed data are discretely recorded at $(0, h, 2h, \cdots, nh)$ in the time interval $[0, T]$ and $nh = T$. By the properties of Lévy process, the sequence of $\{\varepsilon_{th}\}_{t=1}^{T}$ consists of iid random variables. We assume that the moments of ε_{th} exist, up to order 4, with variance σ_ε^2, and skewness and excess kurtosis coefficients γ_1 and γ_2, respectively.[1]

We are interested in studying the properties of $\hat{\kappa}$ estimated from the discrete sample via $\hat{\phi}$. As can be expected, the properties of $\hat{\kappa}$ depend on how we spell out the initial observation $x(0) = x_0 : x_0$ can be fixed at a constant, possibly zero, or x_0 can be a random draw, independent of $(\varepsilon_1, \cdots, \varepsilon_n)$, such that the time series $(x_0, x_1, \cdots, x_n)$ is stationary.

For notational convenience, we drop the subscript h, and throughout, $\boldsymbol{x} = (x_1, \cdots, x_n)'$, $\boldsymbol{x}_{-1} = (x_0, \cdots, x_{n-1})'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$. For a given ϕ, $\boldsymbol{f}_1$ is an $n \times 1$ vector with $f_{1,i} = \phi^i$, $\boldsymbol{f}_2 = \boldsymbol{f}_1/\phi$, $\boldsymbol{C}_1$ is a lower-triangular matrix with $c_{1,ij} = \phi^{i-j}$, $i \geq j$, $\boldsymbol{C}_2$ is a strict lower-triangular matrix with $c_{2,ij} = \phi^{i-j-1}$, $i > j$. Note that by definition, $\boldsymbol{C}_2 = \phi^{-1}(\boldsymbol{C}_1 - \boldsymbol{I})$. The dimensions of vectors/matrices are to be read from the context, and thus we suppress the dimension subscripts in our notation.

For a class of $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\theta}}$ by the condition $\psi(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$, Bao (2013) presented the second-order bias as

$$\mathrm{B}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}^{-1}\mathbb{E}(\boldsymbol{H}_1 \otimes \boldsymbol{\psi}')\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbb{E}\left(\boldsymbol{H}_2\right)(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathrm{vec}\left[\mathbb{E}\left(\boldsymbol{\psi}\boldsymbol{\psi}'\right)\right], \tag{2.3}$$

where $\boldsymbol{\psi} = \psi(\boldsymbol{\theta})$, $\boldsymbol{H}_l = \nabla^l \boldsymbol{\psi}$, $l = 1, 2$, ∇ denotes the derivative with respect to $\boldsymbol{\theta}$, and $\boldsymbol{\Sigma}^{-1} = -[\mathbb{E}(\boldsymbol{H}_1)]^{-1}$. This expression is equivalent to that in Bao and Ullah (2007), but might be easier to work with. For the scalar case, the bias result can be written as

$$\mathrm{B}(\hat{\theta}) = \frac{1}{[\mathbb{E}(H_1)]^2}\mathbb{E}(H_1\psi) - \frac{1}{2[\mathbb{E}(H_1)]^3}\mathbb{E}(H_2)\mathbb{E}(\psi^2). \tag{2.4}$$

[1] This might rule out some Lévy processes. Also, in general, the moments of ε_{th} depend on the parameters κ and σ and sampling frequency h.

2.1 $\mu = 0$ and Known

When $\mu = 0$ and is known *a priori*, we can write $\boldsymbol{x} = x_0 \boldsymbol{f}_1 + \boldsymbol{C}_1 \boldsymbol{\varepsilon}$, $\boldsymbol{x}_{-1} = x_0 \boldsymbol{f}_2 + \boldsymbol{C}_2 \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} = \boldsymbol{x} - \exp(-\kappa h)\boldsymbol{x}_{-1}$.[2] The moment condition, up to some scaling constant, for estimating κ is

$$\psi(\kappa) = \frac{1}{n}\boldsymbol{x}'_{-1}\boldsymbol{\varepsilon}, \tag{2.5}$$

Upon taking derivatives, we have

$$H_l = \frac{-(-h)^l \phi}{n}\boldsymbol{x}'_{-1}\boldsymbol{x}_{-1}, \quad l = 1, 2, \tag{2.6}$$

Appendix A derives the approximate bias of $\hat{\kappa}$ from (2.4) as follows when x_0 is fixed,

$$\begin{aligned} \mathrm{B}(\hat{\kappa}) \quad = \quad & \frac{1 + 3e^{-2\kappa h} + 4e^{-2n\kappa h}}{2Te^{-2\kappa h}} - \frac{\left(1 - e^{-2n\kappa h}\right)\left(1 + 7e^{-2\kappa h}\right)}{2Tne^{-2\kappa h}\left(1 - e^{-2\kappa h}\right)} \\ & - \frac{4e^{-2n\kappa h}\left(1 - e^{-2\kappa h}\right)x_0^2}{2T\sigma_\varepsilon^2 e^{-2\kappa h}} + \frac{\left(1 + 3e^{-2\kappa h}\right)\left(1 - e^{-2n\kappa h}\right)x_0^2}{2Tn\sigma_\varepsilon^2 e^{-2\kappa h}} \\ & + \frac{2(1 + e^{-\kappa h})\left(1 - e^{-n\kappa h}\right)\left(e^{-\kappa h} - e^{-n\kappa h}\right)x_0\gamma_1}{2Tn\sigma_\varepsilon e^{-2\kappa h}}, \end{aligned} \tag{2.7}$$

and when x_0 is random,

$$\mathrm{B}(\hat{\kappa}) = \frac{1}{2T}(3 + e^{2\kappa h}) - \frac{2(1 - e^{-2n\kappa h})}{Tn(1 - e^{-2\kappa h})}. \tag{2.8}$$

Remark 1: We can see that the skewness parameter γ_1 matters for the bias of $\hat{\kappa}$. Its effect, however, disappears for the special case of $x_0 = 0$, where the bias expression simplifies to

$$\mathrm{B}(\hat{\kappa}) = \frac{1 + 3e^{-2\kappa h} + 4e^{-2n\kappa h}}{2Te^{-2\kappa h}} - \frac{\left(1 - e^{-2n\kappa h}\right)\left(1 + 7e^{-2\kappa h}\right)}{2Tne^{-2\kappa h}\left(1 - e^{-2\kappa h}\right)}.$$

Remark 2: (2.8) suggests that the result in Yu (2012) is in fact robust to nonnormality.

[2] When μ is known but may not be 0, one just needs to define $y_t = x_t - \mu$ and work with y_t.

2.2 μ is Unknown

When μ is unknown and has to be estimated, $\boldsymbol{x} = x_0\boldsymbol{f}_1 + \alpha\boldsymbol{C}_1\boldsymbol{\iota} + \boldsymbol{C}_1\boldsymbol{\varepsilon},\ \boldsymbol{x}_{-1} = x_0\boldsymbol{f}_2 + \alpha\boldsymbol{C}_2\boldsymbol{\iota} + \boldsymbol{C}_2\boldsymbol{\varepsilon}$, $\alpha = \mu\left(1 - \exp(-\kappa h)\right), \boldsymbol{\varepsilon} = \boldsymbol{x} - \alpha\boldsymbol{\iota} - \exp(-\kappa h)\boldsymbol{x}_{-1}$. Since the pairs (α, ϕ), (α, κ), and (μ, κ) have one-to-one mapping into each other, and we focus on deriving the finite-sample bias of $\hat{\kappa}$, the reparametrized model $x_t = \alpha + \exp\left(-\kappa h\right) x_{t-1} + \varepsilon_t$ with parameter vector $\boldsymbol{\theta} = (\alpha, \kappa)$ gives exactly the same $\hat{\kappa}$ as estimated from the original $x_t = \mu(1 - \exp\left(-\kappa h\right)) + \exp\left(-\kappa h\right) x_{t-1} + \varepsilon_t$ with parameter vector (μ, κ). Thus, we define the moment condition, up to some scaling constant, as

$$\psi(\boldsymbol{\theta}) = \frac{1}{n}\begin{pmatrix} \boldsymbol{\iota}'\boldsymbol{\varepsilon} \\ -h\phi\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon} \end{pmatrix}. \tag{2.9}$$

By taking derivatives, we have

$$\begin{aligned} \boldsymbol{H}_1 &= \frac{1}{n}\begin{pmatrix} -n & h\phi\boldsymbol{\iota}'\boldsymbol{x}_{-1} \\ h\phi\boldsymbol{\iota}'\boldsymbol{x}_{-1} & h^2\phi\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon} - h^2\phi^2\boldsymbol{x}_{-1}'\boldsymbol{x}_{-1} \end{pmatrix}, \\ \boldsymbol{H}_2 &= \frac{1}{n}\begin{pmatrix} 0 & 0 & 0 & -h^2\phi\boldsymbol{\iota}'\boldsymbol{x}_{-1} \\ 0 & -h^2\phi\boldsymbol{\iota}'\boldsymbol{x}_{-1} & -h^2\phi\boldsymbol{\iota}'\boldsymbol{x}_{-1} & -h^3\phi\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon} + 3h^3\phi^2\boldsymbol{x}_{-1}'\boldsymbol{x}_{-1} \end{pmatrix}. \end{aligned} \tag{2.10}$$

Appendix B derives the approximate bias of $\hat{\kappa}$ when x_0 is fixed,

$$\begin{aligned} \mathrm{B}(\hat{\kappa}) = {} & \frac{5 + 2e^{h\kappa} + e^{2h\kappa} + 4e^{-2(n-1)h\kappa}}{2T} + \frac{2[e^{-2nh\kappa} - e^{-2(n-1)h\kappa}](x_0-\mu)^2}{T\sigma_\varepsilon^2} \\ & + \frac{\left(1 - e^{-nh\kappa}\right)[2e^{h\kappa} + 13e^{2h\kappa} + 4e^{3h\kappa} + e^{4h\kappa} + e^{-(n-4)h\kappa} + 2e^{-(n-3)h\kappa} + 9e^{-(n-2)h\kappa}]}{2\left(1 - e^{2h\kappa}\right)Tn} \\ & + \frac{\left(1 - e^{-nh\kappa}\right)[e^{h\kappa} + 5e^{-(n-1)h\kappa}]\left(x_0^2 + \mu^2\right)}{Tn\sigma_\varepsilon^2} \\ & + \frac{\left(1 - e^{-nh\kappa}\right)[5 + e^{2h\kappa} + 5e^{-(n-2)h\kappa} + 9e^{-nh\kappa}](x_0-\mu)^2}{2Tn\sigma_\varepsilon^2} \\ & - \frac{2\left(1 - e^{-nh\kappa}\right)[e^{-h\kappa} - e^{h\kappa} + e^{3h\kappa} + 5e^{-(n-1)h\kappa}]x_0\mu}{Tn\sigma_\varepsilon^2} \\ & - \frac{\gamma_1\left(1 - e^{-nh\kappa}\right)[e^{-(n-1)h\kappa} + e^{-(n-2)h\kappa}](x_0-\mu)}{Tn\sigma_\varepsilon}, \end{aligned} \tag{2.11}$$

and when x_0 is random,

$$
\begin{aligned}
\mathrm{B}(\hat{\kappa}) \;=\; & \frac{5+2e^{h\kappa}+e^{2h\kappa}}{2T}-\frac{2e^{-h\kappa}\left(1-e^{-nh\kappa}\right)\left(1-e^{2h\kappa}\right)^{2}\mu^{2}}{Tn\sigma_{\varepsilon}^{2}} \\
& +\frac{\left(1-e^{-nh\kappa}\right)\left[e^{h\kappa}+\ 4e^{2h\kappa}+e^{3h\kappa}+2e^{-(n-2)h\kappa}\right]}{\left(1-e^{2h\kappa}\right)Tn}. \qquad (2.12)
\end{aligned}
$$

Remark 3: The leading term (of order $O(T^{-1})$) in (2.12) is the same as that derived from Tang and Chen (2009). Moreover, (2.12) suggests that the approximate bias of $\hat{\kappa}$ under the case of random x_0 is robust to nonnormality.

Remark 4: Similar as before, the skewness matters for the approximate bias. In contrast, for the special case when x_0 is fixed at 0, its effect does not disappear:

$$
\begin{aligned}
\mathrm{B}(\hat{\kappa}) \;=\; & \frac{5+2e^{h\kappa}+e^{2h\kappa}+4e^{-2(n-1)h\kappa}}{2T}+\frac{2[e^{-2nh\kappa}-e^{-2(n-1)h\kappa}]\mu^{2}}{T\sigma_{\varepsilon}^{2}} \\
& +\frac{\left(1-e^{-nh\kappa}\right)\left[2e^{h\kappa}+13e^{2h\kappa}+4e^{3h\kappa}+e^{4h\kappa}+e^{-(n-4)h\kappa}+2e^{-(n-3)h\kappa}+9e^{-(n-2)h\kappa}\right]}{2\left(1-e^{2h\kappa}\right)Tn} \\
& +\frac{\left(1-e^{-nh\kappa}\right)\left[5+2e^{h\kappa}+e^{2h\kappa}+10e^{-(n-1)h\kappa}+5e^{-(n-2)h\kappa}+9e^{-nh\kappa}\right]\mu^{2}}{2Tn\sigma_{\varepsilon}^{2}} \\
& +\frac{\gamma_1\left(1-e^{-nh\kappa}\right)\left[e^{-(n-1)h\kappa}+e^{-(n-2)h\kappa}\right]\mu}{Tn\sigma_{\varepsilon}}.
\end{aligned}
$$

Remark 5: When x_0 is fixed at μ, however, the effect of skewness disappears on the approximate bias:

$$
\begin{aligned}
\mathrm{B}(\hat{\kappa}) \;=\; & \frac{5+2e^{h\kappa}+e^{2h\kappa}+4e^{-2(n-1)h\kappa}}{2T}-\frac{2\left(1-e^{-nh\kappa}\right)\left(e^{-h\kappa}-2e^{h\kappa}+e^{3h\kappa}\right)\mu^{2}}{Tn\sigma_{\varepsilon}^{2}} \\
& +\frac{\left(1-e^{-nh\kappa}\right)\left[2e^{h\kappa}+13e^{2h\kappa}+4e^{3h\kappa}+e^{4h\kappa}+e^{-(n-4)h\kappa}+2e^{-(n-3)h\kappa}+9e^{-(n-2)h\kappa}\right]}{2\left(1-e^{2h\kappa}\right)Tn}.
\end{aligned}
$$

Remark 6: For the random case, if further $\mu = 0$ (i.e., the true model has no drift term but we still estimate the discrete AR model with an intercept), the result reduces to

$$
\mathrm{B}(\hat{\kappa})=\frac{5+2e^{h\kappa}+e^{2h\kappa}}{2T}+\frac{\left(1-e^{-nh\kappa}\right)\left[e^{h\kappa}+\ 4e^{2h\kappa}+e^{3h\kappa}+2e^{-(n-2)h\kappa}\right]}{\left(1-e^{2h\kappa}\right)Tn}.
$$

3 Numerical Results

In this section, we conduct Monte Carlo simulations to demonstrate the performance of our bias formulae in finite samples. In practice we observe only the discrete sample $\{x_0, \cdots, x_n\}$ and we can always estimate σ_ε^2 and γ_1 from the sample residuals. So we simulate discrete AR(1) processes with some nonnormal ε_t without specifying which Lévy process will generate the corresponding nonnormal distribution.

We simulate $\varepsilon_t = \sigma_\varepsilon \epsilon_t$, where ϵ_t follows a standardized noncentral t distribution with noncentrality parameter 1 and degrees of freedom 10. (This gives $\gamma_1 = 0.3999$). We set $\sigma_\varepsilon^2 = 0.1$, $\mu = 0, 0.1$, $x_0 = \mu$ or $x_0 \sim \mathrm{N}(\mu, \sigma^2/(2\kappa))$, $h = 1/12, 1/52, 1/252$ (corresponding to monthly, weekly, and daily data, respectively). Figures 1–4 plot the true and feasible biases of $\hat{\kappa}$ with $T = 10, 50$ when κ goes from 0.1 to 4. The true bias is the averaged actual bias of $\hat{\kappa}$ and the feasible bias is the averaged $\mathrm{B}(\hat{\kappa})$ with all the unknown parameters replaced with their sample consistent estimates, both from 100,000 replications. We observe first that our bias formulae (2.7), (2.8), (2.11), and (2.12) generally do a good job in capturing the true bias of $\hat{\kappa}$ under various scenarios. $\hat{\kappa}$ always over estimates and in some cases the degree of overestimation can be severe. The feasible bias captures most of the overestimation. Second, we notice that when κ is small, our bias formulae provide less satisfying results compared with when κ is big. Recall that our second-order bias results are developed under the assumption that $\kappa > 0$. We have truncated $o((Tn)^{-1})$ terms involving $\exp(-n\kappa h)$ in (2.7), (2.8), (2.11), and (2.12). When κ is close to zero, however, these terms can become quite significant in finite samples and thus make our bias formulae less satisfying. Third, the bias results are more sensitive to the data span T than the data frequency h. This can be seen clearly from (2.7), (2.8), (2.11), and (2.12), where the leading terms are $O(T^{-1})$ and the remaining terms are $O((Tn)^{-1})$. Increasing data frequency alone does little help in reducing the finite-sample bias of $\hat{\kappa}$, but expanding the data span can lead to significant lower bias of $\hat{\kappa}$. Fourth, when the drift term is not zero, $\hat{\kappa}$ tends to be more biased compared with the case when μ is known. Also, when x_0 is fixed, $\hat{\kappa}$ tends to be more biased compared with the case when x_0 is random.

Given that our bias formulae (2.7), (2.8), (2.11), and (2.12) are less satisfying when κ is small and that the true bias decreases with data span T, but not sensitive to the sampling frequency h, in Figures

Figure 1: True and Feasible Biases of $\hat{\kappa}$, $T = 10$, x_0 Fixed

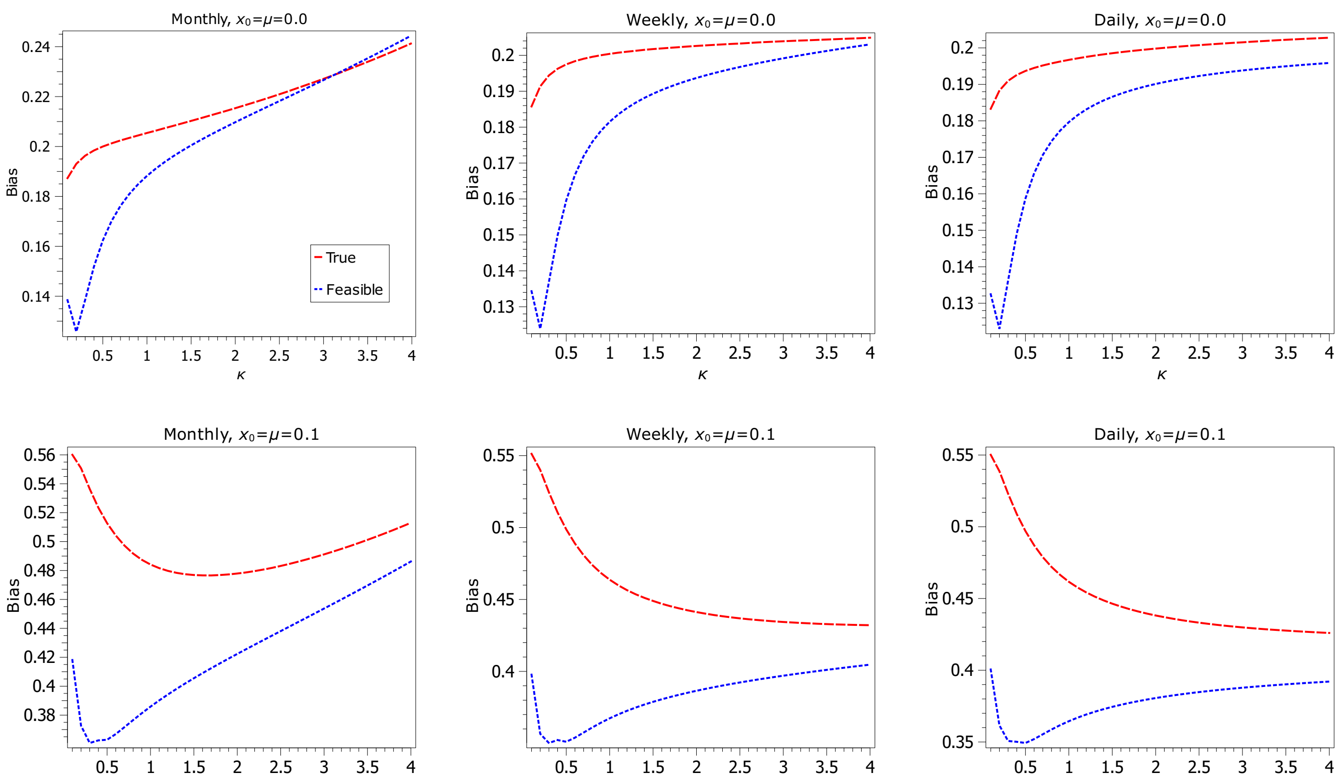


Figure 2: True and Feasible Biases of $\hat{\kappa}$, $T = 10$, x_0 Random

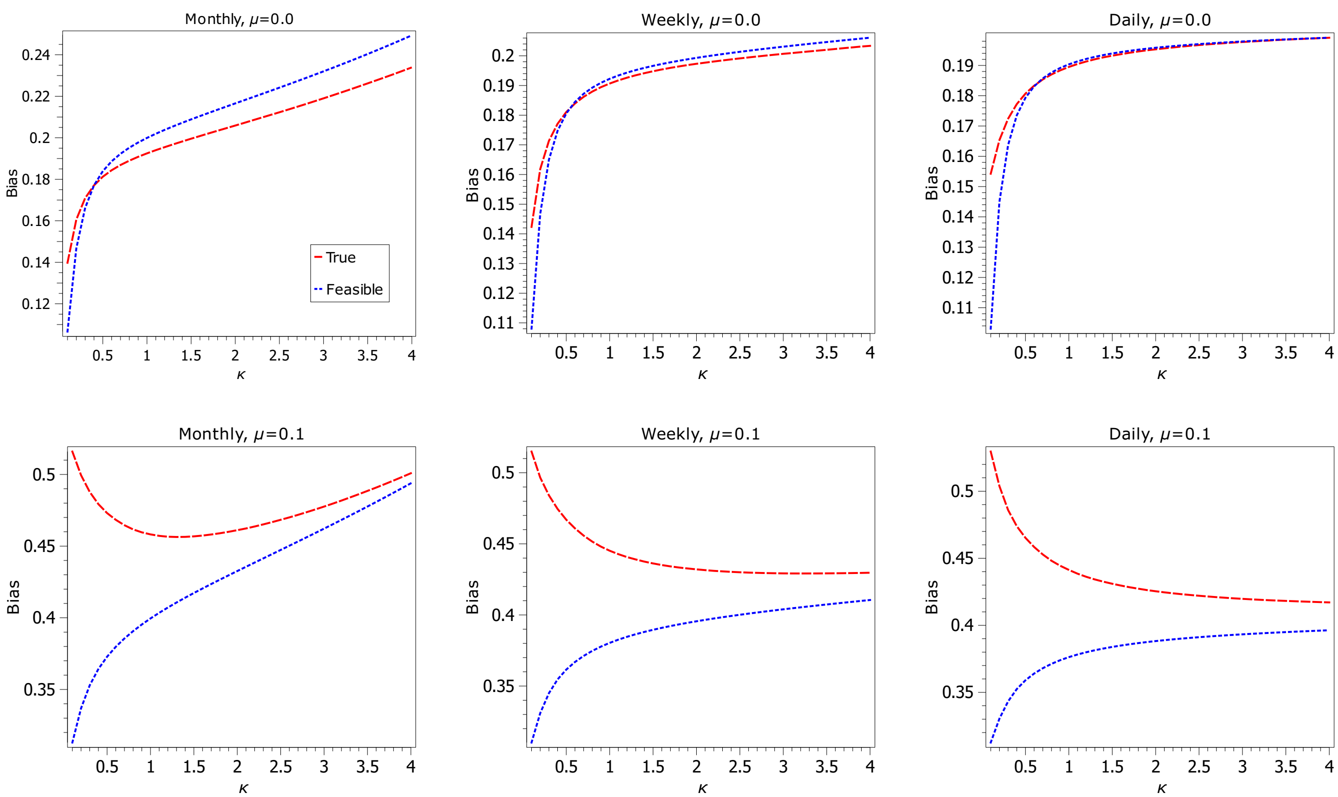


Figure 3: True and Feasible Biases of $\hat{\kappa}$, $T = 50$, x_0 Fixed

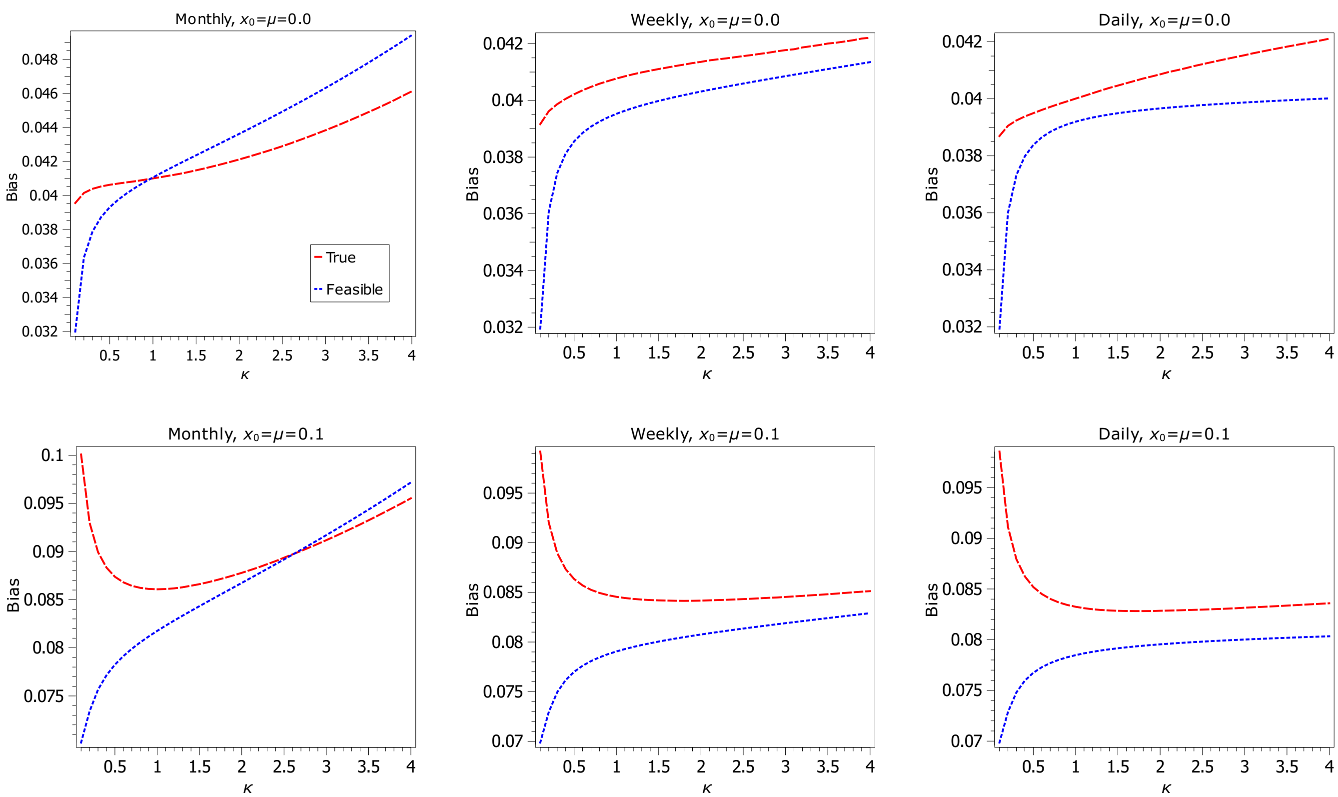
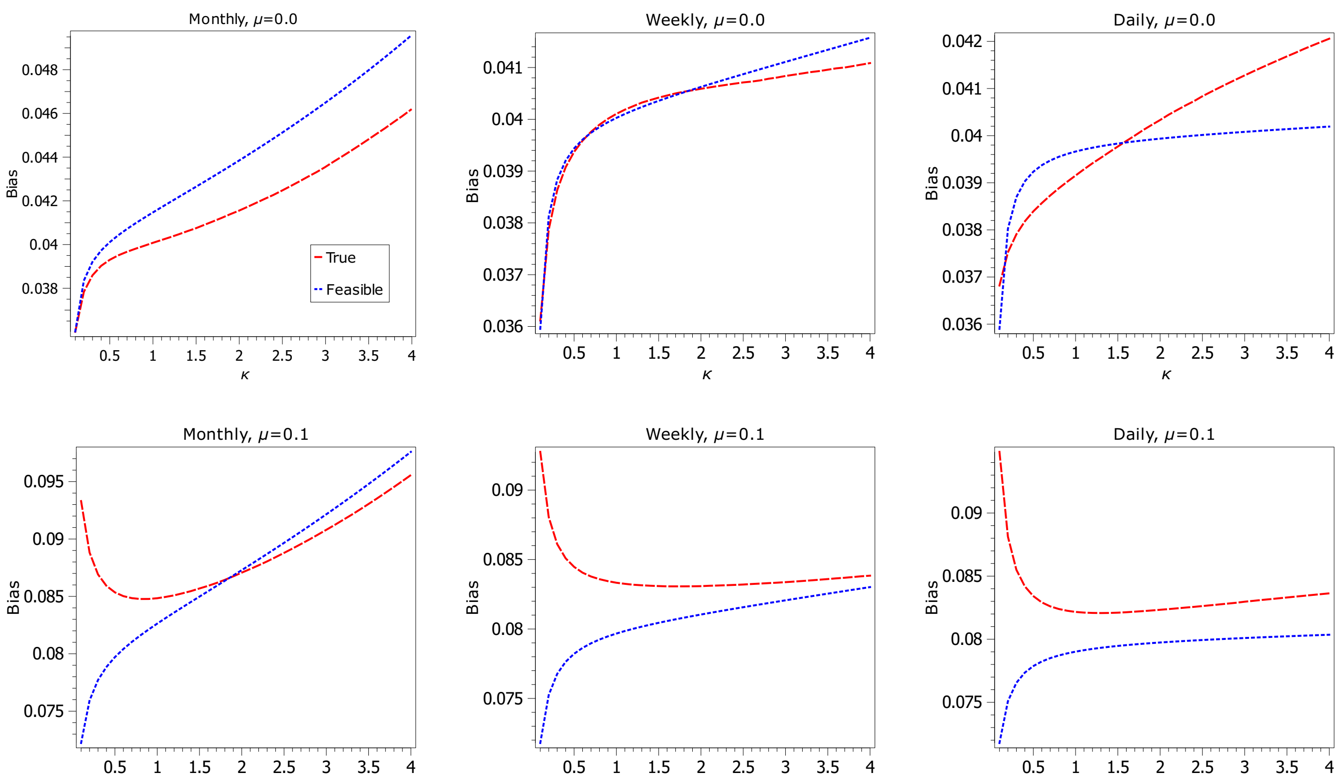


Figure 4: True and Feasible Biases of $\hat{\kappa}$, $T = 50$, x_0 Random



5–6 we display the biases of $\hat{\kappa}$ when T goes from 10 to 100, with $h = 1/12$ and small values of κ (0.10, 0.20, 0.50). We see that for the case of random x_0, the feasible bias results are relatively better to capture the true biases compared with the case of fixed x_0. As the data span increases, the gap between the true and feasible biases does not necessarily get smaller initially. But eventually when the data span is relatively big (around $T = 100$), there is really no much difference between the true and feasible biases.

Figure 5: True and Feasible Biases of $\hat{\kappa}$, $h = 1/12$, x_0 Fixed

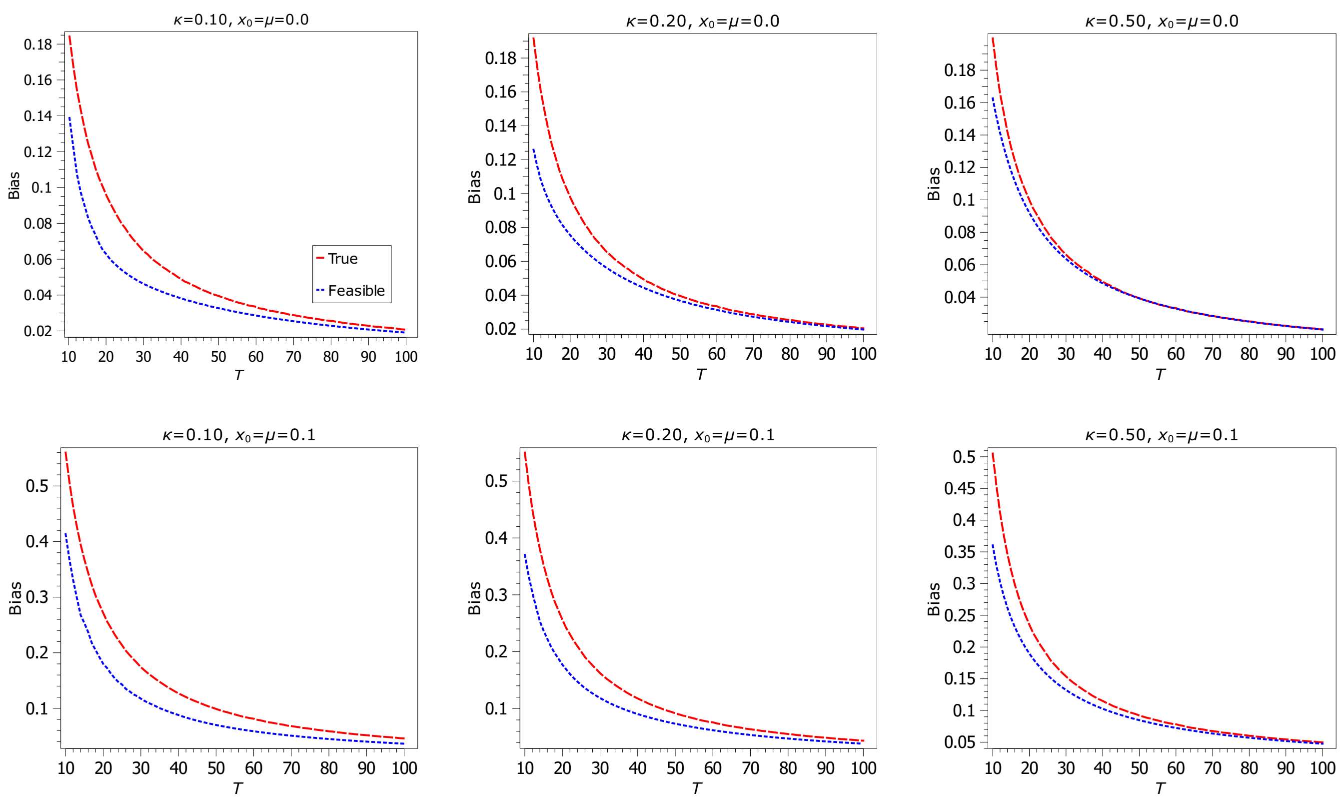
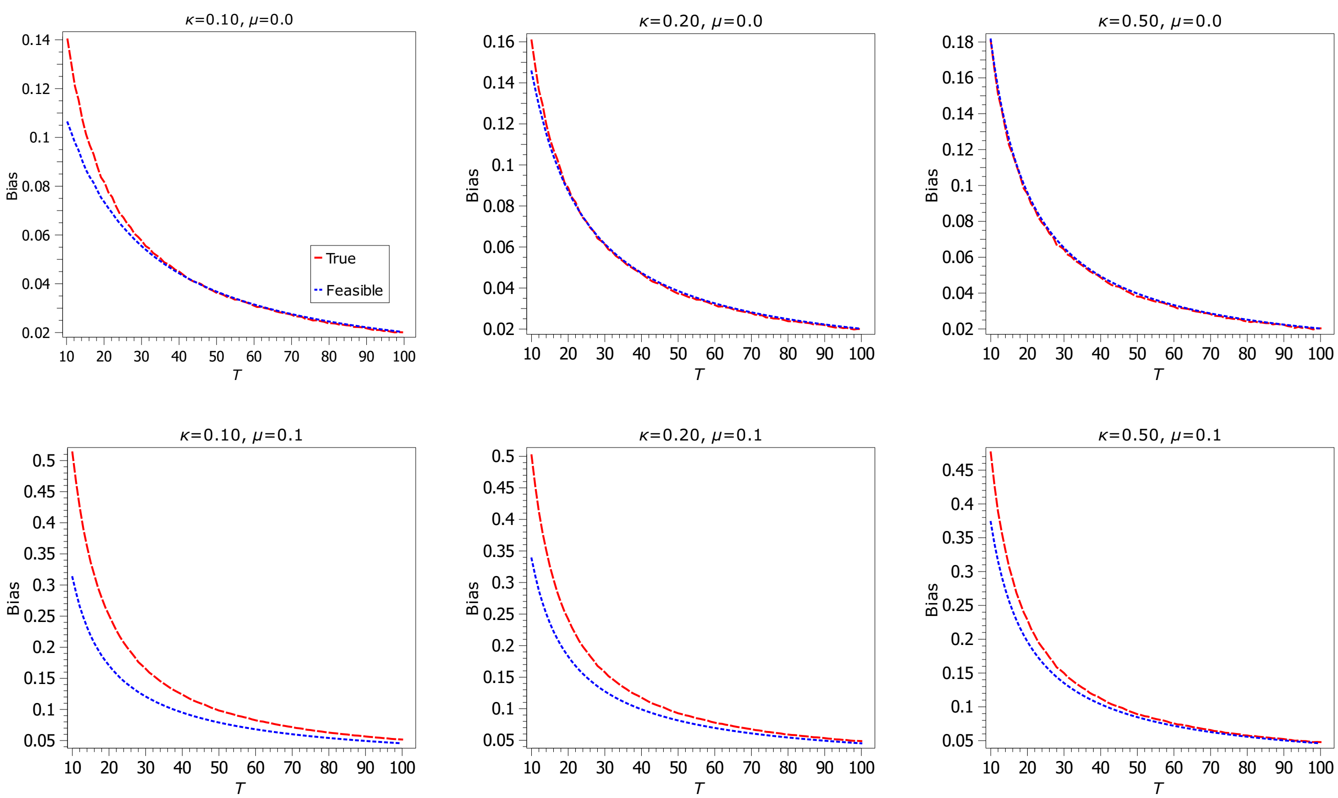


Figure 6: True and Feasible Biases of $\hat{\kappa}$, $h = 1/12$, x_0 Random



4 Conclusions

Lévy processes have found increasing applications in economics and finance. It has been documented, however, that the typical quasi maximum likelihood estimation procedure tends to over estimate the mean reversion parameter in continuous-time Lévy processes. Based on the technique of Bao (2013), we have derived several analytical formulae to approximate the finite-sample bias of the estimated mean reversion parameter under different cases: known or unknown long-run mean, fixed or random initial condition. Our simulation results indicate in general good performance of the approximate bias formulae in capturing the true bias behaviours of the mean reversion estimator. When the time span is small and the mean reversion parameter is close to its lower bound, we find that it is more difficult to approximate well the true finite-sample bias.

Appendix A

Given (2.5) and (2.6), we take expectations and use the results from Ullah (2004, P. 187):

$$\begin{aligned}
\mathbb{E}(H_l) &= \frac{-(-h)^l\phi}{n}[x_0^2\boldsymbol{f}_2'\boldsymbol{f}_2 + \sigma_\varepsilon^2\mathrm{tr}(\boldsymbol{C}_2'\boldsymbol{C}_2)], \quad l = 1, 2, \\
\mathbb{E}(\psi^2) &= \frac{1}{n^2}\{x_0^2\sigma_\varepsilon^2\boldsymbol{f}_2'\boldsymbol{f}_2 + \mathbb{E}[(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon})^2]\}, \\
\mathbb{E}(H_1\psi) &= \frac{h\phi}{n^2}[x_0\sigma_\varepsilon^3\gamma_1\boldsymbol{f}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2) + 2x_0^2\sigma_\varepsilon^2\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{f}_2 + \mathbb{E}(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\varepsilon})].
\end{aligned}$$

When x_0 is random and the process is strictly stationary, x_0 should be replaced with $\mathbb{E}(x_0) = 0$ and x_0 be replaced with $\sigma_\varepsilon^2/(1-\phi^2)$ in the above expectations.

With the special structures of $\boldsymbol{f}_2$ and $\boldsymbol{C}_2$, the expressions for $\boldsymbol{f}_2'\boldsymbol{f}_2$, $\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{f}_2$, $\boldsymbol{f}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2)$, $\mathrm{tr}(\boldsymbol{C}_2'\boldsymbol{C}_2)$, $\mathbb{E}[(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon})^2]$, and $\mathbb{E}(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\varepsilon})]$ can be derived as follow:

$$\begin{aligned}
\boldsymbol{f}_2'\boldsymbol{f}_2 &= \frac{1-\phi^{2n}}{1-\phi^2}, \\
\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{f}_2 &= \frac{\phi - \phi^{2n+1} - n\phi^{2n-1}(1-\phi^2)}{(1-\phi^2)^2},
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{f}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2) &= \frac{(1-\phi^n)\left(1-\phi^{n-1}\right)}{(1-\phi)(1-\phi^2)}, \\
\mathrm{tr}(\boldsymbol{C}_2'\boldsymbol{C}_2) &= \frac{n(1-\phi^2)-(1-\phi^{2n})}{\left(1-\phi^2\right)^2}, \\
\mathbb{E}[(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon})^2] &= \frac{\sigma_\varepsilon^4[n(1-\phi^2)-(1-\phi^{2n})]}{\left(1-\phi^2\right)^2}, \\
\mathbb{E}(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\varepsilon})] &= \frac{2\sigma_\varepsilon^4[n\left(1-\phi^2\right)\left(\phi^2+\phi^{2n}\right)-2\phi^2(1-\phi^{2n})]}{\phi\left(1-\phi^2\right)^3}.
\end{aligned}
$$

Upon substitution and replacing ϕ with $e^{-\kappa h}$, the bias results (2.7) and (2.8) follow.

Appendix B

Let $a_1 = n^{-1}\mathbb{E}(\boldsymbol{\iota}'\boldsymbol{x}_{-1})$, $a_2 = n^{-1}\mathbb{E}(\boldsymbol{x}_{-1}'\boldsymbol{x}_{-1})$, $a_3 = n^{-2}\mathbb{E}(\boldsymbol{\iota}'\boldsymbol{x}_{-1}\boldsymbol{\iota}'\boldsymbol{\varepsilon})$, $a_4 = n^{-2}\mathbb{E}(\boldsymbol{\iota}'\boldsymbol{x}_{-1}\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon})$, $a_5 = n^{-2}\mathbb{E}(\boldsymbol{x}_{-1}'\boldsymbol{x}_{-1}\boldsymbol{\iota}'\boldsymbol{\varepsilon})$, $a_6 = n^{-2}\mathbb{E}(\boldsymbol{x}_{-1}'\boldsymbol{x}_{-1}\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon})$, $a_7 = n^{-2}\mathbb{E}(\boldsymbol{\iota}'\boldsymbol{\varepsilon}\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon})$, $a_8 = n^{-2}\mathbb{E}(\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon}\boldsymbol{x}_{-1}'\boldsymbol{\varepsilon})$. Then we can write

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{H}_1) &= \begin{pmatrix} -1 & h\phi a_1 \\ h\phi a_1 & -h^2\phi^2 a_2 \end{pmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{h^2\phi^2 a_2 - h^2\phi^2 a_1^2}\begin{pmatrix} h^2\phi^2 a_2 & h\phi a_1 \\ h\phi a_1 & 1 \end{pmatrix}, \\
\mathbb{E}\left(\boldsymbol{H}_2\right) &= \begin{pmatrix} 0 & 0 & 0 & -h^2\phi a_1 \\ 0 & -h^2\phi a_1 & -h^2\phi a_1 & 3h^3\phi^2 a_2 \end{pmatrix}, \quad \mathbb{E}\left(\boldsymbol{\psi}\boldsymbol{\psi}'\right) = \begin{pmatrix} n^{-1}\sigma^2 & -h\phi a_7 \\ -h\phi a_7 & h^2\phi^2 a_8 \end{pmatrix}, \\
\mathbb{E}(\boldsymbol{H}_1\otimes\boldsymbol{\psi}') &= \begin{pmatrix} 0 & 0 & h\phi a_3 & -h^2\phi^2 a_4 \\ h\phi a_3 & -h^2\phi^2 a_4 & h^2\phi a_7 - h^2\phi^2 a_5 & -h^3\phi^2 a_8 + h^3\phi^3 a_6 \end{pmatrix}.
\end{aligned}
$$

From (2.3), the second-order bias of $\hat{\kappa}$ simplifies to

$$
\mathrm{B}(\hat{\kappa}) = \frac{a_1^2\sigma^2}{2nh\phi^2\left(a_1^2-a_2\right)^2} + \frac{a_8 - 2a_1a_7 + 2(a_1^2a_3 + a_2a_3 - 2a_1a_4 - a_1a_5 + a_6)\phi}{2h\phi^2\left(a_1^2-a_2\right)^2}.
$$

By substituting $\boldsymbol{x}_{-1} = x_0\boldsymbol{f}_2 + \alpha\boldsymbol{C}_2\boldsymbol{\iota} + \boldsymbol{C}_2\boldsymbol{\varepsilon}$ and use the results from Ullah (2004, P. 187), we have

$$
\begin{aligned}
a_1 &= \frac{1}{n}\left(x_0\boldsymbol{\iota}'\boldsymbol{f}_2 + \alpha\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{\iota}\right), \\
a_2 &= \frac{1}{n}[x_0^2\boldsymbol{f}_2'\boldsymbol{f}_2 + \alpha^2\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\iota} + 2\alpha x_0\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{\iota} + \sigma_\varepsilon^2\mathrm{tr}(\boldsymbol{C}_2'\boldsymbol{C}_2)], \\
a_3 &= \frac{\sigma_\varepsilon^2}{n^2}\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{\iota}, \\
a_4 &= \frac{\sigma_\varepsilon^2}{n^2}(x_0\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{f}_2 + \alpha\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota}), \\
a_5 &= \frac{1}{n^2}[\sigma_\varepsilon^3\gamma_1\boldsymbol{\iota}'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2) + 2x_0\sigma_\varepsilon^2\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{\iota} + 2\alpha\sigma_\varepsilon^2\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\iota}], \\
a_6 &= \frac{1}{n^2}[\sigma_\varepsilon^3\gamma_1 x_0\boldsymbol{f}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2) + \sigma_\varepsilon^3\gamma_1\alpha\boldsymbol{\iota}'\boldsymbol{C}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2) + \mathbb{E}(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\varepsilon}) \\
&\quad +2x_0^2\sigma_\varepsilon^2\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{f}_2 + 2x_0\alpha\sigma_\varepsilon^2\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota} + 2x_0\alpha\sigma_\varepsilon^2\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{f}_2 + 2\alpha^2\sigma_\varepsilon^2\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota}], \\
a_7 &= \frac{\sigma_\varepsilon^2}{n^2}(x_0\boldsymbol{\iota}'\boldsymbol{f}_2 + \alpha\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{\iota}), \\
a_8 &= \frac{1}{n^2}\{x_0^2\sigma_\varepsilon^2\boldsymbol{f}_2'\boldsymbol{f}_2 + \alpha^2\sigma_\varepsilon^2\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\iota} + \mathbb{E}[(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon})^2] + 2x_0\alpha\sigma_\varepsilon^2\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{\iota}\}.
\end{aligned}
$$

When x_0 is random and the process is strictly stationary, x_0 should be replaced with $\mathbb{E}(x_0) = \alpha/(1-\phi)$ and x_0^2 with $\mathbb{E}(x_0^2) = \sigma_\varepsilon^2/(1-\phi^2) + \alpha^2/(1-\phi)^2$ in the expressions for a_i's.

Note that $\boldsymbol{f}_2'\boldsymbol{f}_2$, $\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{f}_2$, $\boldsymbol{f}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2)$, $\mathrm{tr}(\boldsymbol{C}_2'\boldsymbol{C}_2)$, $\mathbb{E}[(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon})^2]$, and $\mathbb{E}(\boldsymbol{\varepsilon}'\boldsymbol{C}_2\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\varepsilon})]$ are already derived in Appendix A. In addition, $\boldsymbol{\iota}'\boldsymbol{f}_2$, $\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{\iota}$, $\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{\iota}$, $\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{f}_2$, $\boldsymbol{\iota}'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2)$, $\boldsymbol{\iota}'\boldsymbol{C}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2)$, $\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{f}_2$, $\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\iota}$, $\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota}$, $\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota}$, $\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota}$ are needed in evaluating a_i, $i = 1, \cdots, 8$. Upon some algebra, we can verify

$$
\begin{aligned}
\boldsymbol{\iota}'\boldsymbol{f}_2 &= \frac{1-\phi^n}{1-\phi}, \\
\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{\iota} &= \frac{n(1-\phi)-(1-\phi^n)}{(1-\phi)^2}, \\
\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{\iota} &= \frac{(1-\phi^n)(\phi-\phi^n)}{(1+\phi)(1-\phi^2)}, \\
\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{f}_2 &= \frac{1-n(1-\phi)\phi^{n-1}-\phi^n}{(1-\phi)^2}, \\
\boldsymbol{\iota}'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2) &= \frac{n(1-\phi^2)-(1-\phi^{2n})}{\left(1-\phi^2\right)^2},
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\iota}'\boldsymbol{C}_2'\mathrm{diag}(\boldsymbol{C}_2'\boldsymbol{C}_2) &= \frac{n\left(1-\phi^2\right)-2-\phi-\phi^{2n-1}+\phi^{n-1}(1+\phi)^2}{(1-\phi)(1-\phi^2)^2},\\
\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{f}_2 &= \frac{1+\phi+\phi^2+n\phi^{2n-1}\left(1-\phi^2\right)+\phi^{2n+1}-[n(1-\phi)+\phi]\phi^{n-1}(1+\phi)^2}{(1-\phi)\left(1-\phi^2\right)^2},\\
\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{\iota} &= \frac{n\left(1-\phi^2\right)-\left(1-\phi^n\right)\left(1+2\phi-\phi^n\right)}{(1-\phi)^2(1-\phi^2)},\\
\boldsymbol{\iota}'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota} &= \frac{n(1-\phi)+n(1-\phi)\phi^{n-1}+2\phi^n-2}{(1-\phi)^3},\\
\boldsymbol{f}_2'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota} &= \frac{\phi^n(1+\phi)^2(1-2\phi)+\phi^{1+2n}\left[n(1-\phi^2)+\phi+2\phi^2\right]-(1-2\phi^2)}{(1-\phi)(1-\phi^2)^2\phi^2},\\
\boldsymbol{\iota}'\boldsymbol{C}_2'\boldsymbol{C}_2\boldsymbol{C}_2\boldsymbol{\iota} &= \frac{\left(1-\phi^n\right)\left[\phi^n(1+2\phi)-\phi(4+3\phi)-2\right]}{(1-\phi)^2\left(1-\phi^2\right)^2}\\
&\quad+\frac{n\left(1-\phi^2\right)\left[(1+\phi^{n-1})(1+\phi)-\phi^{2n-1}\right]}{(1-\phi)^2\left(1-\phi^2\right)^2}.
\end{aligned}
$$

One can see that a_1 and a_2 are $O(1)$, while $a_i, i=3,\cdots,8$ are $O(n^{-1})$. By substituting the a_i's, ignoring terms of smaller orders, and replacing ϕ with $e^{-\kappa h}$, the bias results (2.11) and (2.12) follow.

References

Bao, Y., 2013, Finite sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* 29, 68–88. (Erratum at *Econometric Theory*, 2013, 29, 89.)

Bao, Y. and A. Ullah, 2007, The second-order bias and mean squared error of estimators in time series models. *Journal of Econometrics* 140, 650–669.

Barndorff-Nielsen, O. E., 1998, Processes of normal inverse Gaussian type. *Finance and Stochastics* 2, 41–68.

Black, F. and M. Scholes, 1973, The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.

Carr, P. and L. Wu, 2003, The finite moment log stable process and option pricing. *Journal of Finance* 58, 753–778.

Cox, J., J. Ingersoll, and S. Ross, 1985, A theory of the term structure of interest rates. *Econometrica* 53, 385–407.

Kendall, M.G., 1954, Note on bias in the estimation of autocorrelation. *Biometrika* 41, 403–404.

Madan, D., P. Carr, and E. Chang, 1999, The variance gamma processes and option pricing. *European Finance Review* 2, 79–105.

Phillips, P.C.B. and J. Yu, 2005, Jackknifing bond option prices. *Review of Financial Studies* 18, 707–742.

Phillips, P.C.B. and J. Yu, 2009, Maximum likelihood and Gaussian estimation of continuous time models in finance, in: T. G. Andersen, R.A. Davis, J.-P. Kreiß, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, Springer-Verlag, New York, pp. 497–530.

Tang, C.Y. and S.X. Chen, 2009, Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149, 65–81.

Ullah, A., 2004, *Finite Sample Econometrics.* Oxford University Press, New York.

Vasicek, O., 1977, An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

Yu, J., 2012, Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics* 169, 114–122.