11-2009

Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models

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

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models

Jun YU

November 2009

Paper No. 16-2009

Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models[1]

Jun Yu[2]

First version: April 2007; *This version*: September 2009

[1]I gratefully acknowledge financial support from the Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS. I would like to thank Peter Phillips for extensive discussions on the subject and Yacine Aït-Sahalia, Aman Ullah, Qiankun Zhou, the participants of the conference in honor of Peter Phillipcs, and especially two referees for comments.

[2]School of Economics and Sim Keen Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; email: yujun@smu.edu.sg.

Abstract

It is well known that for continuous time models with a linear drift standard estimation methods yield biased estimators for the mean reversion parameter both in finite discrete samples and in large in-fill samples. In this paper, we obtain two expressions to approximate the bias of the least squares/maximum likelihood estimator of the mean reversion parameter in the Ornstein-Uhlenbeck process with a known long run mean when discretely sampled data are available. The first expression mimics the bias formula of Marriott and Pope (1954) for the discrete time model. Simulations show that this expression does not work satisfactorily when the speed of mean reversion is slow. Slow mean reversion corresponds to the near unit root situation and is empirically realistic for financial time series. An improvement is made in the second expression where a nonlinear correction term is included into the bias formula. It is shown that the nonlinear term is important in the near unit root situation. Simulations indicate that the second expression captures the magnitude, the curvature and the non-monotonicity of the actual bias better than the first expression.

Keywords: Least squares, Maximum likelihood, Discrete sampling, Continuous record, Near unit root.

JEL Classifications: C22, C32

1 Introduction

There is an extensive literature on using continuous time models in economic theory (e.g., Merton, 1990). Motivated by this success, econometricians have developed methods for estimating continuous time models, aiming to provide a basis from which these models may be used in empirical applications. While Ito's lemma facilitates the mathematical treatment of continuous time models in economic applications, continuous time models are more difficult to deal with econometrically than their discrete time counterparts. In recent years, however, several exciting developments have been made on estimating and testing continuous time models based on discrete time observations. In terms of parameter estimation, important contributions include Lo (1986), Bergstrom (1990), Duffie and Singleton (1993), Pedersen (1995), Aït-Sahalia (1996a, 1999, 2002), Stanton (1997), Elerian, Chib and Shephard (2001), Bandi and Phillips (2002, 2007), and Aït-Sahalia and Yu (2006). In terms of specification analysis, important contributions include Chan, Karolyi, Longstaff, and Sanders (1992), Aït-Sahalia (1996a, 1996b), Dai and Singleton (2000), Bandi (2002), and Hong and Li (2005). While there are abundant continuous time specifications available, much of the focus in the asset pricing literature has been on the continuous time diffusion equations with an affine structure (see Duffie and Kan, 1996). This is the main motivation why we choose to focus our attention on continuous time diffusion models with a linear drift function. However, the methodology employed here is general and is applicable to non-affine models including those with nonlinear drift.

One problem with utilizing continuous time models is estimation bias.[1] Standard estimation methods, such as least squares (LS), maximum likelihood (ML) or generalized method of moments (GMM), produce biased estimators for the mean reversion parameter. The bias is essentially of the Hurwicz type that Hurwicz (1950) developed in the context of dynamic discrete time models. However, as it will be clear later, the percentage bias is much more pronounced in continuous time models than their discrete time counterpart. On the other hand, estimation is fundamentally important for many practical applications. For example, it provides parameter estimators which are used directly for estimating prices of financial assets and derivatives. For another example, parameter estimation serves as an important stage for the empirical analysis

[1]The bias in this article refers to bias arising from estimation. This is different from the bias induced by discretizing continuous time models.

of specification and comparative diagnostics. Not surprisingly, it has been found in the literature that the bias in the mean reversion estimator has important implications for the specification analysis of continuous time models (Prisker, 1998) and for pricing financial assets (Phillips and Yu, 2005).

Several methods have been proposed to reduce the bias in the mean reversion estimator.[2] Ball and Torous (1996) suggested utilizing more cross-sectional information for estimating continuous time term structure models. Obviously this approach is subject to data availability. In Phillips and Yu (2005) the jackknife method of Quenouille (1956) was suggested to reduce the bias. While the jackknife method cannot completely remove the bias, it can be very useful in practice as it is computationally simple and is applicable to a very broad range of models, including the models for which it is impossible or difficult to develop the explicit form of an asymptotic expansion of the bias. Another method whose performance was examined in Phillips and Yu (2005) is the median unbiased estimator of Andrews (1993). This estimator is closely related to the indirect inference method and the bootstrap method. The indirect inference method was originally proposed by Smith (1993) and Gouriéroux, Monfort and Renault (1993) and subsequently applied to reduce the bias in the mean reversion estimator by Phillips and Yu (2009a). The bootstrap method was recently proposed to reduce the bias in the mean reversion estimator by Tang and Chen (2009). All three methods are simulation-based, and hence computationally demanding.

In an independent and concurrent study, Tang and Chen (2009) derived an analytical formula for approximating the bias of certain estimators for the Ornstein-Uhlenbeck (OU) process and the square root process, both with an unknown long run mean. The bias formula corresponds to that of Marriott and Pope (1954) and Kendall (1954) for the discrete time autoregressive (AR) model with an intercept. It was shown that the bias of the mean reversion estimator is of order T^{-1} but not of order n^{-1}, where T is the data span and n is the number of observations. As a result, increasing the sample size, by the way of increasing the sampling frequency, cannot yield a consistent LS estimator. This result confirms what has been known in the literature; see, for example, Merton (1980). However, the performance of their bias formula is unsatisfactory in the near unit root situations.

[2]Bias has been under extensive study in the context of discrete time models. Some recent studies include Abdir (1993), Rilstone, Srivastava, and Ullah (1996), Vinod and Shenton (1996), MacKinnon and Smith (1998), and Bao and Ullah (2007).

In this paper we derive an analytical formula for approximating the bias of ML/LS estimators for the OU process with a known long run mean. Thus, our results complement those of Tang and Chen (2009). We make several contributions to the literature. First, we point out that the true bias of the mean reversion estimate has an interesting curvature and goes to zero when the mean reversion parameter is closer to zero. This result echoes the conjecture of Hurwicz (1950) about the bias in the autoregressive (AR) estimate in the discrete time AR(1) model. Second, we show that the bias formula, which mimics that of Marriott and Pope (1954) and Kendall (1954) for the discrete time model and that of Tang and Chen (2009) for continuous time models, is essentially linear in coefficient. Consequently, the bias predicted by the formula does not disappear in the unit root case. One reason why this bias formula does not work well is that the Cesaro sums are badly approximated in the unit root and the near unit root situations. Since many financial time series have roots extremely near unity, there is considerable interest in improving the bias formula.

As a third contribution, we derive an alternative bias formula which includes an extra term. The extra term arises from the exact evaluation of the Cesaro sums. It is of smaller order and hence can be ignored when the mean reversion parameter is far away from zero. Interestingly, it does not have a smaller order effect when the mean reversion parameter is close to zero. Monte Carlo studies show that the alternative bias formula is more accurate. It reproduces the nonlinear feature in the true bias function and goes to zero when the mean reversion parameter goes to zero. Finally, we approximate the bias and the mean square errors (MSE) up to a higher order term.

The paper is organized as follows. Section 2 derives the formulae for approximating the bias and the mean square error. In Section 3 we assess the accuracy of the analytical expressions using Monte Carlo experiments. Section 4 obtains the bias and the MSE in a higher order term. Section 5 concludes the paper. The Appendix collects proofs of the main results.

2 OU Process with a Known Mean

The model considered here is the Ornstein-Uhlenbeck (OU) process:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\, dB(t), X(0) \sim N(\mu, \sigma^2/2\kappa) \tag{1}$$

with μ being known, where $B(t)$ is a standard Brownian motion. This model has been previously used to explain the dynamics of short-term interest rates (Vasicek, 1977) and log-volatilities (Taylor, 1982). Since we assume the long run mean, μ, is known apriori, without loss of generality, it is set to zero. The parameter of interest is the speed of mean reversion, κ, which is assumed to be positive.[3] Phillips (1972) showed that the exact discrete time model corresponding to (1), is given by the following AR(1) structure

$$X_{ih} = \phi X_{(i-1)h} + \sigma \sqrt{\frac{1 - e^{-2\kappa h}}{2\kappa}} \epsilon_i, \tag{2}$$

where $\phi = e^{-\kappa h}$, $\epsilon_i \sim$ i.i.d. $N(0, 1)$ and h is the sampling interval. Obviously the covariance structure of any discrete sample in Model (1) is the same as that in Model (2) and there is a one-to-one correspondence between κ and ϕ. Also, it is easy to see that $\kappa > 0$ implies $\phi < 1$ and hence stationarity; $\kappa \to 0$ or $h \to 0$ implies $\phi \to 1$ and the model converges to a unit root model. For a small value of κ or a small value of h (high frequency), both being empirically relevant, the model has a root near unity. This situation is the primary interest of the present study. Moreover, since the distribution of the LS estimator of ϕ is invariant to σ^2, the same property holds for κ. The observed data are assumed to be recorded discretely at $(0, h, 2h, \cdots, nh(= T))$ in the time interval $[0, T]$. So $n + 1$ is the total number of observations and T is the data span. With a finite value of T, $n \to \infty$ when $h \to 0$ and vice versa. In the limit as $h \to 0$, a continuous sample path from the interval is observed. This in-fill asymptotics has become very popular in recent years in financial econometrics following the work on realized volatility; see, for example, Andersen, Bollerslev, Diebold, Labys (2001) and Barndorff-Nielson and Shephard (2002). For financial time series, $X(t)$ is often recorded monthly, weekly, or daily and hence $h = 1/12$, $1/52$ or $1/252$. However, higher frequencies are possible in the setup with an even smaller value for h. When there is no confusion, we simply write X_{ih} as X_i. Unless specified, the summation sign $\sum$ is always referred to summation from $i = 1$ to $i = n$.

The LS estimator of κ (denoted by $\hat{\kappa}$) can be obtained by

$$\min_{\kappa} \sum \left(X_i - e^{-\kappa h} X_{i-1} \right)^2. \tag{3}$$

[3] It is known, from the simulations conducted in Phillips and Yu (2005) and the theoretical work in Tang and Chen (2009), that the ML estimators of the long mean parameter and the diffusion parameter have little bias. For this reason, we focus our attention to the mean reversion parameter in the present paper.

It can be shown that the LS estimator is equivalent to the ML estimator which maximizes the following log-likelihood function (conditional on $X_0 = X(0)$),

$$\sum \ln pdf(X_i|X_{i-1}). \tag{4}$$

where *pdf* represents the conditional density. For Model (1) with $\mu = 0$, the conditional distribution is given by

$$X_i|X_{i-1} \sim N\left(e^{-\kappa h}X_{i-1}, \sigma^2(1 - e^{-2\kappa h})/(2\kappa)\right). \tag{5}$$

The ML estimator has been widely used in the literature (see, for example, Aït-Sahalia, 1999). The equivalence is the main reason why we focus on LS.

It is well known from the discrete time dynamic literature that the LS estimator can be downward biased. For example, in the AR(1) model without intercept

$$X_i = \phi X_{i-1} + \sigma\epsilon_i, \ \epsilon_i \sim N(0,1) \tag{6}$$

Marriott and Pope (1954) derived the following expression to approximate the bias of the LS estimator

$$E(\hat{\phi}) - \phi = -\frac{2\phi}{n} + o(n^{-1}). \tag{7}$$

Bartlett (1946) derived the following expression to approximate the variance of $\hat{\phi}$

$$Var(\hat{\phi}) = \frac{1-\phi^2}{n} + o(n^{-1}). \tag{8}$$

Equations (7) and (8) are obtained by replacing the Cesaro sum

$$\sum_{j=-n}^{n}\left(1 - \frac{|j|}{n}\right)\phi^{|j|}$$

with

$$\sum_{j=-\infty}^{\infty}\phi^{|j|}.$$

Obviously the quality of the approximation deteriorates when $\phi \to 1$. When $|\phi| < 1$, the model is stationary and the limiting theory of $\hat{\phi}$ is given by

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, 1 - \phi^2). \tag{9}$$

Since $\phi = e^{-\kappa h}$, it is reasonable to believe that the bias in $\hat{\phi}$ translates into $\widehat{\kappa}$. In fact, Phillips and Yu (2005, 2009a) provided extensive Monte Carlo evidence of severe finite sample bias in $\widehat{\kappa}$ and many other estimators of κ.

When κ is not close to zero, for $\widehat{\kappa}$, we take a Taylor expansion up to the second order term,

$$
\begin{aligned}
\widehat{\kappa} &= -\ln(\widehat{\phi})/h \\
&= -\frac{1}{h}\left(\ln\phi + \frac{1}{\phi}(\widehat{\phi}-\phi) - \frac{1}{2\phi^2}(\widehat{\phi}-\phi)^2 + o_p(n^{-1})\right) \\
&= -\frac{1}{h}\left(\ln\phi + \frac{1}{\phi}(\widehat{\phi}-\phi) - \frac{1}{2\phi^2}(Var(\widehat{\phi}) + (E(\widehat{\phi})-\phi)^2) + o_p(n^{-1})\right) \\
&= \kappa - \frac{1}{h\phi}(\widehat{\phi}-\phi) + \frac{1}{2h\phi^2}Var(\widehat{\phi}) + o_p(T^{-1}). \qquad (10)
\end{aligned}
$$

From Equations (7), (8) and (10), it is straightforward to show that

$$
\begin{aligned}
E(\widehat{\kappa}) - \kappa &= \frac{2}{hn} + \frac{1}{2h\phi^2}\left(\frac{1-\phi^2}{n}\right) + o(T^{-1}) \qquad (11) \\
&= \frac{1}{2T}\left(3 + e^{2\kappa h}\right) + o(T^{-1}). \qquad (12)
\end{aligned}
$$

Bias formula (12) is analogous to that of Marriott and Pope for the AR(1) model and corresponds to that of Tang and Chen (2009) for the OU process with an unknown mean. The first term in (11) arises from the bias in $\hat{\phi}$ while the second term arises from the variance of $\widehat{\kappa}$ and the nonlinear dependence of κ in ϕ. By including only the first two terms in Taylor expansion, the bias due to the skewness and the kurtosis in $\hat{\phi}$ is obviously omitted. This omission trades off the quality of the approximation against algebraic tractability. The bias formula (12) has several implications for the behavior of the bias. First, according to (12), the size of the bias is mainly determined by the data span T but not by the sample size n. Second, the bias converges to $2/T$ when $h \to 0$. According to this in-fill asymptotics, the bias does not go away unless T goes to infinity. Third, when κ is reasonably small, $e^{2\kappa h} \approx 1 + 2\kappa h \approx 1$. Hence, (12) implies that the bias is essentially linear in κ and that the bias is about $2/T$ and hence insensitive to κ. According to the second and the third implications, the approximate bias is $2/T$ when either $h \to 0$ or $\kappa \to 0$. Fourth, the predicted bias will not disappear when $\kappa \to 0$. The first implication seems to be consistent with what have been found in literature (Phillips and Yu, 2005). The second and the third implications are rather surprising because (7) suggests that the bias in $\hat{\phi}$ is sensitive to the true value. The last implication seems at odds with the conjecture made by Hurwicz (1950) that the bias in $\hat{\phi}$ is zero in the discrete time unit root case (i.e. $\phi = 1$).

To understand the behavior of the actual bias in Model (1) and the performance of (12), we simulate 756 daily observations (i.e. $T = 3$) from the model with κ taking

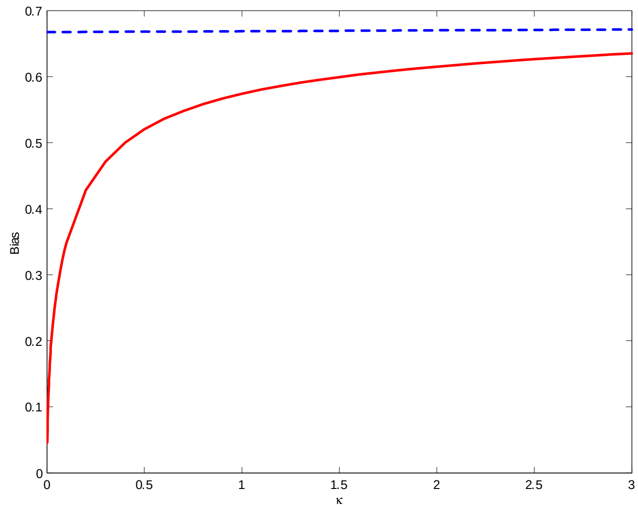

Figure 1: The bias as a function of κ for daily frequency (ie $h = 1/252$) when $T = 3$ (i.e. $n = 756$). The solid line is from the simulations. The dashed line is from formula (12).

various values from the region of (0,3] and estimate κ using the LS estimator (3). The experiment is replicated 10,000 times to get the actual bias. Fig. 1 plots the true bias and the expression (12), both as a function of κ. Obviously there is a great deal of discrepancy between them. The smaller κ, the bigger the difference. The actual bias goes to zero when $\kappa \to 0$, echoing the conjecture made by Hurwicz (1950) in the discrete time model, whereas according to (12) the expected value of $\hat{\kappa}$ is about $2/3 \approx 0.67$ when κ or h is close to zero. The discrepancy is due to the error arising from approximating the Cesaro sum since the derivation of (12) makes use of (7) and (8). Moreover, there is a strong nonlinearity in the actual bias function while the expression (12) is nearly linear. Therefore, there are good reasons to find a better bias formula than (12).

To derive the bias, we adopt the approach of Bao and Ullah (2007) which is briefly reviewed here. Suppose $\widehat{\beta}$ is an estimator of β, based on a sample of n observations, which satisfies the following estimation equation:

$$\psi_n(\widehat{\beta}) = \frac{1}{n} \sum q_i(\widehat{\beta}) = 0. \tag{13}$$

The identification condition is given by $E(\psi_n(\beta)) = 0$. Under a set of regular conditions, Bao and Ullah (2007) obtained the stochastic expansion of $\widehat{\beta}$ as[4]

$$\widehat{\beta} - \beta = a_{-1/2} + a_{-1} + a_{-3/2} + o_p(n^{-3/2}), \tag{14}$$

where $a_{-1/2} = -Q\psi_n$, $a_{-1} = -QVa_{-1/2} - \frac{1}{2}Q\overline{H_2}a_{-1/2}^2$, $a_{-3/2} = -QVa_{-1} - \frac{1}{2}QWa_{-1/2}^2 - Q\overline{H_2}a_{-1/2}a_{-1} - \frac{1}{6}Q\overline{H_3}a_{-1/2}^3$, with $\psi_n = \psi_n(\kappa)$, $\bar{\cdot} = E(\cdot)$, $H_i = \partial^i\psi_n/\partial\kappa^i$, $Q = \left(\overline{H_1}\right)^{-1}$, $V = H_1 - \overline{H_1}$, $W = H_2 - \overline{H_2}$. By the identification condition, $E(a_{-1/2}) = 0$.[5] The second order and the third order bias of $\widehat{\beta}$ is, respectively,

$$E(a_{-1}), E(a_{-1} + a_{-3/2}) \tag{15}$$

and the first order and the second order MSE of $\widehat{\beta}$ is, respectively,

$$E(a_{-1/2}^2), E(a_{-1/2}^2 + 2a_{-1/2}a_{-1}). \tag{16}$$

When the parameter of interest is ϕ in the AR(1) model, it is easy to see that $H_2 = H_3 = V = W = 0$, greatly simplifying the analysis. The parameter of interest in the present study is κ for which the estimation equation is a nonlinear function in κ, Consequently, none of these quantities is zero and hence the derivation of the bias is more complex in continuous time models. Working with $E(a_{-1})$ without approximating the Cesaro sums in the OU model, we get a new second order bias for $\hat{\kappa}$.

THEOREM 2.1 (New Approximation to the Bias of $\hat{\kappa}$): *Under Model (1) with a known μ, when κ is close to 0, we have the following second order bias of $\hat{\kappa}$,*

$$\frac{1}{2T}\left(3 + e^{2\kappa h}\right) - \frac{2(1 - e^{-2n\kappa h})}{Tn(1 - e^{-2\kappa h})}. \tag{17}$$

Remark 2.1 Compared with (12), the bias formula (17) has an extra term, which arises from the exact calculation of the Cesaro sums, as shown in the Appendix. This term is of order $(Tn)^{-1}$ and hence smaller than $1/T$, when κ is far away from zero. In this case it is negligible and (17) becomes (12). However, if κ is close to zero, the extra term is negligible, even for a large n. To see this, applying L'Hospital's rule to the second term, we have

$$\lim_{\kappa\to 0} \frac{1 - e^{-2n\kappa h}}{n(1 - e^{-2\kappa h})} = 1,$$

[4]The expansion was first derived in the i.i.d. framework by Rilstone, Srivastava, and Ullah (1996).

[5]The asymptotic normaylity theory, such as (9), follows from the fact that $\sqrt{n}a_{-1/2}$ converges to a normal distribution.

and

$$\lim_{\kappa \to 0} E(a_{-1}) = 0. \tag{18}$$

As a result, the extra term is of order T^{-1} but not of order $(Tn)^{-1}$. Indeed, (18) suggests that the bias is close to zero when κ is close to 0, which is consistent with what is found in Fig. 1. The bias, however, decreases when the span of data (T) becomes larger. This observation is consistent with the Monte Carlo results reported in Phillips and Yu (2005). Compared with (12), the bias formula (17) is much more nonlinear due to the inclusion of the extra term.

Remark 2.2 When h converges to 0, $n(1-e^{-2\kappa h}) \to 2T\kappa$ and $1-e^{-2n\kappa h} = 1-e^{-2T\kappa}$. Thus,

$$\lim_{h \to 0} \frac{2(1-e^{-2n\kappa h})}{n(1-e^{-2\kappa h})} = \frac{1-e^{-2T\kappa}}{T\kappa} \tag{19}$$

and

$$\lim_{h \to 0} E(a_{-1}) = \frac{1}{T}\left\{2 - \frac{1-e^{-2T\kappa}}{T\kappa}\right\}. \tag{20}$$

The implication for $h \to 0$ is very different from that for $\kappa \to 0$ although both cases lead to a unit root in the exact discrete time representation. The difference arises because as $\kappa \to 0$ the initial condition becomes dominant whereas as $h \to 0$ the error variance goes to 0. The bias formula (20) is also remarkably different from the limit case of (12) when $h \to 0$. It is easy to see that the bias formula (20) works well for practically relevant values for h. For example, if $T = 3$ and $\kappa = 3$, (20) suggests that bias is about 0.63 as $h \to 0$; if $T = 3$ and $\kappa = 1$, (20) suggests that bias is about 0.46 as $h \to 0$. These values appear to match very well with what we have found in Fig. 1 when $h = 1/252$.

Remark 2.3 Formulae (17) and (20) suggest feasible ways for bias correction. If κ is reasonably close to zero, we can estimate κ by

$$\widehat{\kappa} - \frac{1}{2T}\left(3 + e^{2\widehat{\kappa}h}\right) + \frac{2(1-e^{-2n\widehat{\kappa}h})}{Tn(1-e^{-2\widehat{\kappa}h})}.$$

If in addition, h is small, we can then estimate κ by

$$\widehat{\kappa} - \frac{1}{T}\left\{2 - \frac{1-e^{-2T\widehat{\kappa}}}{T\widehat{\kappa}}\right\}.$$

To obtain the limiting theory for $\hat{\kappa}$ when $\kappa > 0$, we apply the delta method to (9)

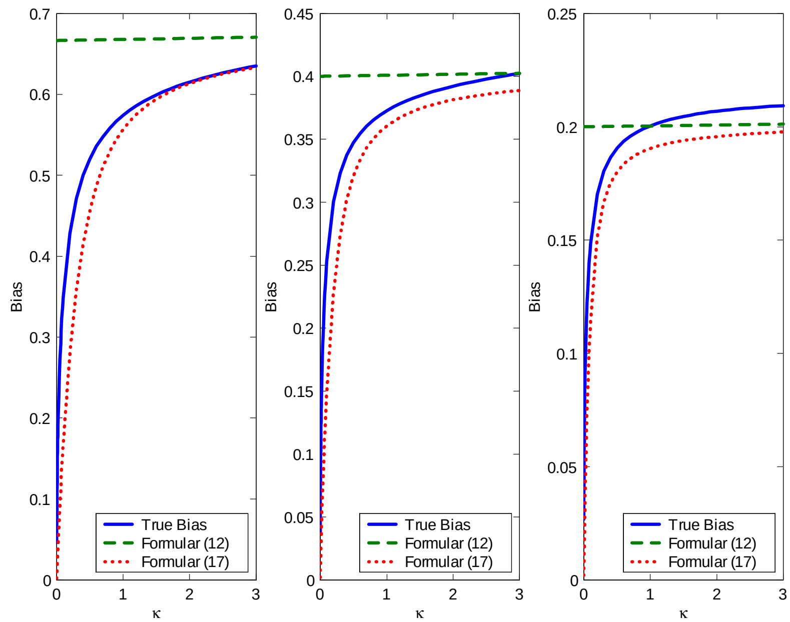

Figure 2: The bias as a function of κ for daily frequency (ie $h = 1/252$). The three graphs correspond to $T = 3, 5, 10$ (ie, $n = 756, 1260, 2520$), respectively. The solid line is from the simulations. The dashed line is from formula (12). The dotted line is from formula (17).

$$\sqrt{T}(\hat{\kappa} - \kappa) \xrightarrow{d} N(0, (e^{2\kappa h} - 1)/h). \tag{21}$$

The variance in the limiting distribution is identical to what was found in Tang and Chen (2009).

Working with $E(a_{-1}^2)$ without approximating the Cesaro sums, we get the expression for the first order MSE.

THEOREM 2.2 (The first order MSE of $\hat{\kappa}$): *Under Model (1) with a known μ, we have the following the first order MSE,*

$$MSE(\hat{\kappa}) \approx \frac{e^{2kh} - 1}{Th}. \tag{22}$$

Remark 2.4 Interestingly, the exact calculation of the Cesaro sums does not make any difference for the first order MSE as it is the same as the asymptotic variance $\hat{\kappa}$

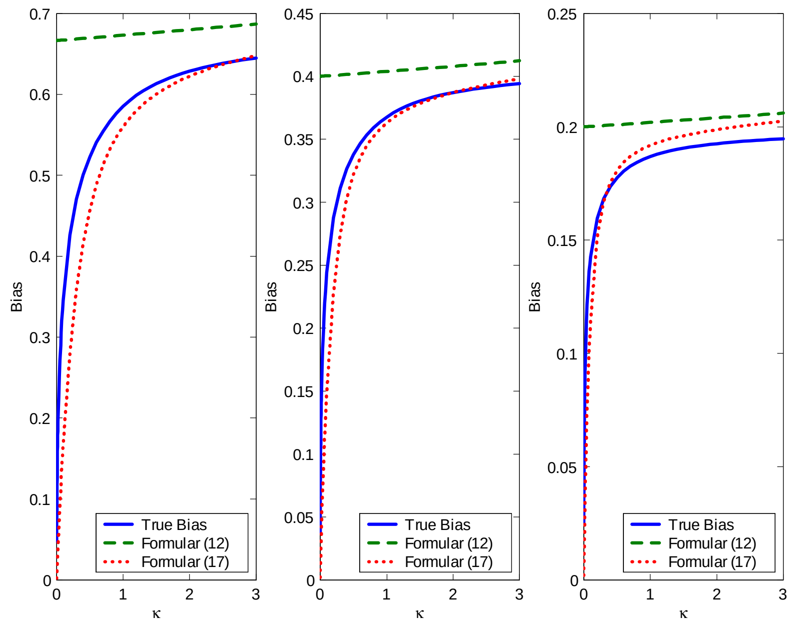

Figure 3: The bias as a function of κ for weekly frequency (ie $h = 1/52$). The three graphs correspond to $T = 3, 5, 10$ (ie, $n = 156, 260, 520$), respectively. The solid line is from the simulations. The dashed line is from formula (12). The dotted line is from formula (17).

given by (21). Furthermore, when $h \to 0$, $MSE \approx 2\kappa/T$ and $\hat{\kappa} \overset{a}{\sim} N(\kappa, 2\kappa/T)$, the latter of which is well known in the statistics literature – see, for example, Brown and Hewitt (1975).

3 Monte Carlo Results

To examine the performance of the two alternative bias formulae, we estimate κ in Model (1) using the LS estimators (3), assuming κ takes various values from the region of (0,3]. This range covers empirically reasonable values of κ for real data on interest rates and volatilities. The mean reversion parameter is estimated with 3, 5 10 years of daily, weekly and monthly data. The experiment is replicated 10,000 times to get the bias. Since the number of simulated paths is large, the bias can be regarded as the actual bias.

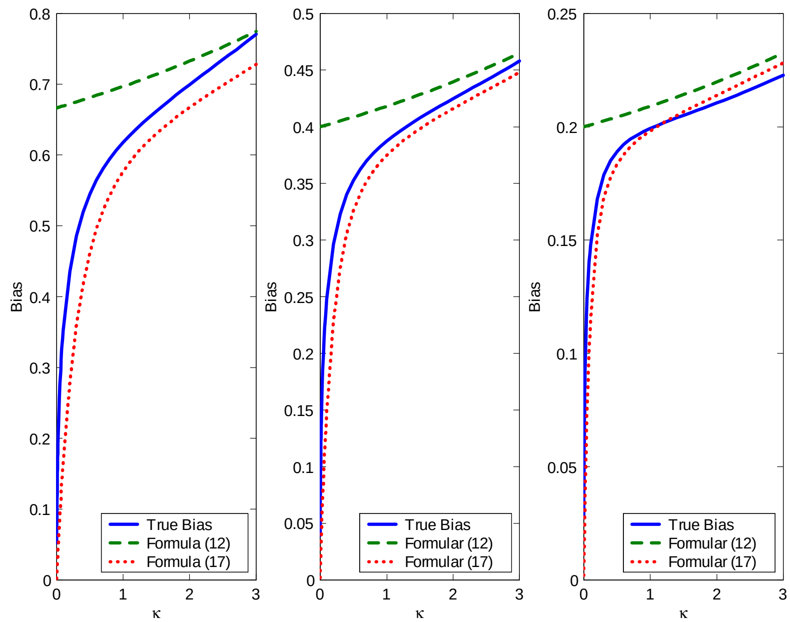

Figure 4: The bias as a function of κ for monthly frequency (ie $h = 1/12$). The three graphs correspond to $T = 3, 5, 10$ (ie, $n = 36, 60, 120$), respectively. The solid line is from the simulations. The dashed line is from formula (12). The dotted line is from formula (17).

Figures 2-4 report the simulation results for the daily, weekly and monthly frequency, respectively. In the figures, we plot the actual bias, the bias expression (12) and the bias expression (17) as a function of κ.

Several features are apparent in the figures. First, the actual bias can be substantial. The bias is especially large for small T both in percentage and absolute terms. For example, if data from a three-year time interval are used to estimate κ when $\kappa = 0.1$, regardless of the frequency at which the data are collected, the percentage bias is about 250% and the absolute bias is about 0.25. This bias is very big and has important economic implications for asset pricing. When κ is small, the bias formula (12) does not perform well and the bias formula (17) offers substantial improvement to (12). The bad performance of (12) is not surprising since it is known to be difficult to correct the bias when ϕ is close to 1 (Hurwicz, 1950). Because a small value for κ is empirically reasonable, the improvement in the bias formula (17) is practically useful.

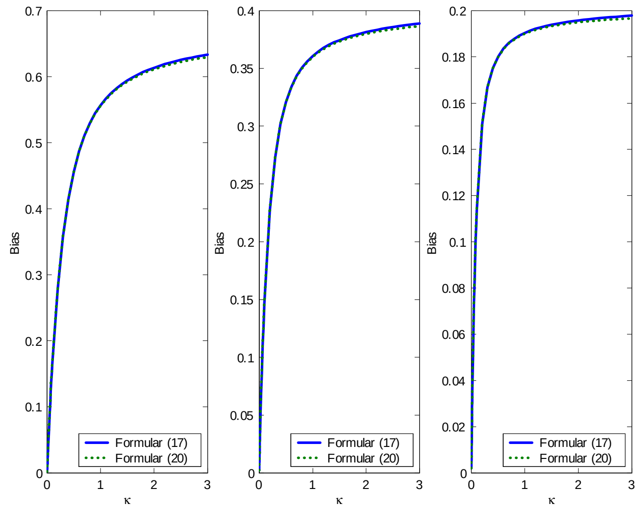

Figure 5: Approximate bias from (17) and (20) for the daily frequency with $T = 3, 5, 10$. The solid line is from formula (17). The dotted line is from formula (20).

Second, the actual bias is always a highly nonlinear function of κ, especially when κ is small. The bias formula (12) is virtually linear in κ whereas the bias formula (17) reproduces the curvature in the actual bias function quite well.

Third, as κ gets close to zero, the true bias seems to decrease toward zero. Interestingly, the bias formula (17) but not the bias formula (12) has the same feature. Fourth, the actual bias seems to be dependent upon the data span but not the sampling frequency, consistent with the two bias formulae.

To examine the performance of (20) relative to (17) (i.e. the effect of small h), we adopt the same simulation design as before but now plot the bias formulae (17) and (20). Fig. 3-6 are for the daily, weekly and monthly frequency, respectively. Obviously, the difference between (17) and (20) is the largest for monthly data and the least for daily data, consistent with the prediction of (20). Similarly to (17), (20) also suggests the bias converges to 0 as $\kappa \to 0$. Finally, when the true value of κ is closer to 0, the difference between (17) and (20) is very small, suggesting that we can replace (17) with (20) to approximate the bias in practice.

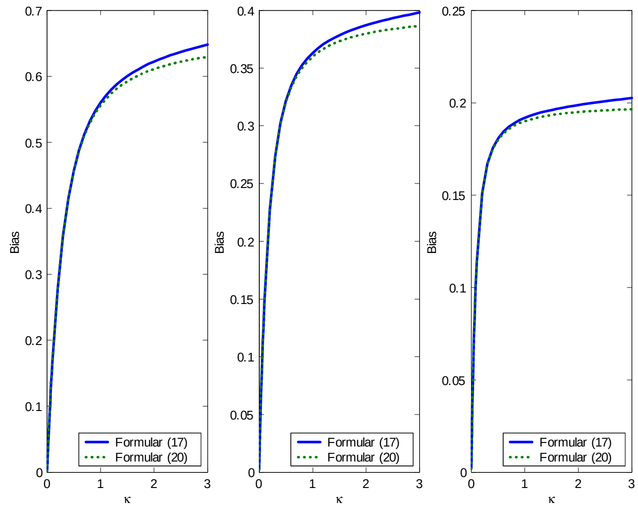

Figure 6: Approximate bias from (17) and (20) for the weekly frequency with $T = 3, 5, 10$. The dashed line is from formula (17). The dotted line is from formula (20).

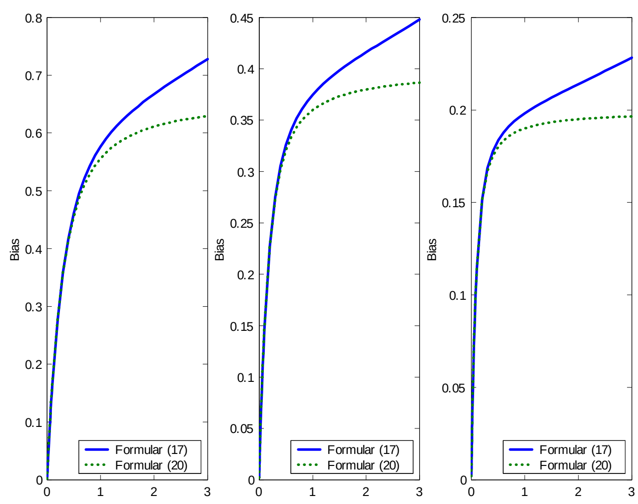

Figure 7: Approximate bias from (17) and (20) for the monthly frequency with $T = 3, 5, 10$. The dashed line is from formula (17). The dotted line is from formula (20).

4 Higher Order Bias and MSE

Under the framework of Bao and Ullah (2007), the third order bias of $\hat{\kappa}$ and the the second-order MSE of $\hat{\kappa}$, namely, $E(a_{-1}+a_{-3/2})$ and $E(a_{-1/2}^2+2a_{-1/2}a_{-1})$, can be worked out in the same manner.

THEOREM 3.1 (The Third Order Bias and The Second Order MSE): *Under Model (1) with a known μ, when κ is close to 0, the third order bias of $\hat{\kappa}$ is*

$$\frac{1}{2T}\left(3+e^{2\kappa h}\right)-\frac{2(3e^{2\kappa h}+9+e^{-2\kappa h}+4e^{-2(n-1)\kappa h}+7e^{-2n\kappa h})}{Tn(1-e^{-2\kappa h})}+\frac{4(1-e^{-2n\kappa h})(e^{2\kappa h}+7+4e^{-2\kappa h})}{Tn^2(1-e^{-2\kappa h})^2}. \tag{23}$$

and the second order MSE of $\hat{\kappa}$

$$\frac{e^{2\kappa h}-1}{Th}-\frac{2(5+7e^{2\kappa h})+16e^{-2(n-1)\kappa h}}{T^2}+\frac{10(3+e^{2\kappa h})(1-e^{-2n\kappa h})}{T^2n(1-e^{-2\kappa h})}. \tag{24}$$

Remark 3.1 By L'Hospital's rule, it can be shown that as $\kappa\to 0$, $E(a_{-1}+a_{-3/2})\to 0$. If $h\to 0$,

$$E(a_{-1}+a_{-3/2})\to\frac{2}{T}-\frac{13+11e^{-2T\kappa}}{T^2\kappa}+\frac{12(1-e^{-2T\kappa})}{T^3\kappa^2}, \tag{25}$$

and

$$E(a_{-1/2}^2+2a_{-1/2}a_{-1})\to\frac{2\kappa}{T}-\frac{24+16e^{-2T\kappa}}{T^2}+\frac{20(1-e^{-2T\kappa})}{T^3\kappa}. \tag{26}$$

5 Conclusions

We have presented two alternative expressions for approximating the bias of the mean reversion estimator in a continuous time diffusion model, based on the method proposed by Bao and Ullah (2007). The simpler expression mimics the bias formula derived by Marriott and Pope (1954) for the discrete time AR model and corresponds to the bias formula derived independently by Tang and Chen (2009) for the same model but with unknown mean. The complicated one includes an additional term from the exact evaluation of the Cesaro sums. We show that the additional term is important for improving the quality of bias approximation, especially when the mean reversion parameter is close to zero. This near unit case is practically realistic for financial time series.

One drawback is that our formulae only work for the simple univariate OU process. However, Bao and Ullah's method is general and is applicable to more complicated models, so long as the estimator can be obtained from an estimation equation which is available analytically. Interesting generalizations include the square-root model, non-affine models, models with a nonlinear drift, and multivariate models. The results for more general models will be reported in future work. Finally, the initial condition is assumed to be the stationary distribution in our treatment. This initial condition is known to have important implications for the finite sample theory (White, 1961) and even for asymptotic theory in the unit root case (Phillips and Magdalinos, 2009). It is useful to derive the bias formula for alternative initial conditions for the mean reversion parameter.

6 Appendix

Before proving Theorem 2.1, we first introduce a lemma.

Lemma 1

1. If $X \sim N(0, \Sigma)$, A, A_1, A_2 and A_3 are all symmetric matrices, then

$$E(X'AX) = tr(A\Sigma), \tag{27}$$

$$E(X'AX)^2 = (tr(A\Sigma))^2 + 2tr(A\Sigma A\Sigma), \tag{28}$$

$$E(X'A_1XX'A_2X) = tr(A_1\Sigma)tr(A_2\Sigma) + 2tr(A_1\Sigma A_2\Sigma), \tag{29}$$

and

$$\begin{aligned} E(X'A_1XX'A_2XX'A_3X) &= tr(A_1\Sigma)tr(A_2\Sigma)tr(A_3\Sigma) + 2\{tr(A_1\Sigma)tr(A_2\Sigma A_3\Sigma) \\ &\quad + tr(A_2\Sigma)tr(A_1\Sigma A_3\Sigma) + tr(A_3\Sigma)tr(A_1\Sigma A_2\Sigma)\} \\ &\quad + 8tr(A_1\Sigma A_2\Sigma A_3\Sigma), \end{aligned} \tag{30}$$

where tr denotes the trace of a matrix.

2. $\sum i\phi^{-i} = \frac{\phi - \phi^{1-n}(1+n) + n\phi^{-n}}{(1-\phi)^2}$.

3. $\sum\sum \phi^{|t-s|} = n\frac{1+\phi}{1-\phi} - \frac{2\phi(1-\phi^n)}{(1-\phi)^2}$.

4. $\sum\sum \phi^{|t-s|+|t-s-1|} = n\frac{2\phi}{1-\phi^2} - \frac{\phi(1+\phi^2)(1-\phi^{2n})}{(1-\phi^2)^2}$.

5. $\sum\sum(\phi^{2|t-s|}+\phi^{|t-s+1|+|t-s-1|}) = n\frac{1+4\phi^2-\phi^4}{1-\phi^2} - \frac{2\phi^2(1-\phi^{2n})}{(1-\phi^2)^2}$.

Proof of Lemma 1:

1. Equations (27) and (28) are straightforward consequences of Exercise 3 in Ullah (2004, Page 12). To get Equations (29) and (30), we need to define $y = X\Sigma^{-1/2}$ and assume $\mu = 0$ in Exercise 4 in Ullah (2004, Page 12).

2. Working from the derivatives, we have

$$\begin{aligned}\sum i\phi^{-i} &= -\phi\frac{\partial(\sum\phi^{-i})}{\partial\phi} = -\phi\frac{\partial((1-\phi^{-n})/(\phi-1))}{\partial\phi} \\ &= \frac{\phi-\phi^{1-n}(1+n)+n\phi^{-n}}{(1-\phi)^2}\end{aligned}$$

3. Following from the last equation, we have

$$\begin{aligned}\sum\sum\phi^{|t-s|} &= n+2\phi^n\sum_{i=1}^{n-1}i\phi^{-i} = n+2\frac{\phi^{n+1}-n\phi^2+(n-1)\phi}{(1-\phi)^2} \\ &= n\frac{1+\phi}{1-\phi}+\frac{2\phi^{n+1}-2\phi}{(1-\phi)^2},\end{aligned}$$

$$\begin{aligned}\sum\sum\phi^{|t-s|+|t-s-1|} &= n\phi+\sum_{i=1}^{n-1}(n-i)\phi^{2i-1}+\sum_{i=1}^{n-1}(n-i)\phi^{2i+1} \\ &= n\phi+(1+\phi^2)\frac{\phi^2(\phi^{2n-1}-\phi)-(n-1)(\phi^2-1)\phi}{(1-\phi)^2} \\ &= n\frac{2\phi}{1-\phi^2}+\frac{(1+\phi^2)(\phi^{2n+1}-\phi)}{(1-\phi^2)^2},\end{aligned}$$

and

$$\begin{aligned}\sum\sum(\phi^{2|t-s|}+\phi^{|t-s+1|+|t-s-1|}) &= \sum\sum\phi^{2|t-s|}+\sum\sum\phi^{|t-s+1|+|t-s-1|} \\ &= n\frac{1+4\phi^2-\phi^4}{1-\phi^2}-\frac{2\phi^2(1-\phi^{2n})}{(1-\phi^2)^2}\end{aligned}$$

Proof of Theorem 2.1: Denote $X = (X_0, \ldots, X_n)'$,

$$C_1 = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \text{ and } C_2 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{31}$$

Note that the LS estimator of κ is obtained from the following estimation equation,

$$\frac{1}{n}\sum X_{i-1}\left(X_i - e^{-\widehat{\kappa}h}X_{i-1}\right) = \frac{1}{n}X'C_1X - e^{-\widehat{\kappa}h}\frac{1}{n}X'C_2X := U_n - e^{-\widehat{\kappa}h}V_n. \tag{32}$$

with $U_n = \frac{1}{n}X'C_1X$ and $V_n = \frac{1}{n}X'C_2X$.

Since the property of $\widehat{\kappa}$ is independent of σ^2, without loss of generality, we assume $\sigma^2 = 2\kappa$. As a result, $X_t \sim N(0,1)$ and $X \sim N(0,\Sigma)$ where Σ is an $(n+1)\times(n+1)$ matrix with ij-th element $\phi^{|i-j|}$. By Lemma 1, $E(U_n) = \phi$ and $E(V_n) = 1$. Moreover,

$$\begin{aligned} E(U_nV_n) &= \frac{1}{n^2}E\left(X'C_1XX'C_2X\right) \\ &= \frac{1}{n^2}\left\{tr(C_1\Sigma)tr(C_2\Sigma) + tr(C_1\Sigma C_2\Sigma)\right\} \\ &= \phi + \frac{4\phi}{n(1-\phi^2)} - \frac{2\phi(1+\phi^2)(1-\phi^{2n})}{n^2(1-\phi^2)^2}, \end{aligned} \tag{33}$$

where the second and third equalities follow from Lemma 1. Similarly

$$\begin{aligned} E(V_n^2) &= \frac{1}{n^2}E\left(X'C_2X\right)^2 \\ &= 1 + \frac{2(1+\phi^2)}{n(1-\phi^2)} - \frac{4\phi^2(1-\phi^{2n})}{n^2(1-\phi^2)^2}, \end{aligned} \tag{34}$$

and

$$\begin{aligned} E(U_n^2) &= \frac{1}{n^2}E\left(X'C_1X\right)^2 \\ &= \phi^2 + \frac{1+4\phi^2-\phi^4}{n(1-\phi^2)} - \frac{4\phi^2(1-\phi^{2n})}{n^2(1-\phi^2)^2}, \end{aligned} \tag{35}$$

From the estimation equation (32), using the same notations as in Bao and Ullah (2004), we have $H_1 = \phi h V_n$, $Q = 1/(\phi h)$, $\overline{H_1} = \phi h$, $V = \phi h(V_n - 1)$, $H_2 = -\phi h^2 V_n$, $\overline{H_2} = -\phi h^2$, $H_3 = \phi h^3 V_n$, $W = \phi h^2(1 - V_n)$, and $\overline{H_3} = \phi h^3$. Substituting all these expressions to the individual terms in the stochastic expansion of $\hat{\kappa}$ given by Equation (14), we obtain

$$a_{-1/2} = -\frac{U_n - \phi V_n}{\phi h}, \tag{36}$$

and

$$a_{-1} = \frac{U_n^2 - \phi^2 V_n^2}{2\phi^2 h} - \frac{U_n - \phi V_n}{\phi h}. \tag{37}$$

Substituting (33), (34) and (35) into (36) and (37), taking expectation, and collecting terms, we have

$$E(a_{-1/2}) = 0, \tag{38}$$

$$\begin{aligned}
E(a_{-1}) &= \frac{E(U_n^2) - \phi^2 E(V_n^2)}{2\phi^2 h} \\
&= \frac{1}{2\phi^2 h}\left\{\phi^2 + \frac{1 + 4\phi^2 - \phi^4}{n(1-\phi^2)} - \frac{4\phi^2(1-\phi^{2n})}{n^2(1-\phi^2)^2}\right\} \\
&\quad - \frac{\phi^2}{2\phi^2 h}\left\{1 + \frac{2(1+\phi^2)}{n(1-\phi^2)} - \frac{4\phi^2(1-\phi^{2n})}{n^2(1-\phi^2)^2}\right\} \\
&= \frac{1}{2T}\left(3 + \phi^{-2}\right) - \frac{2(1-\phi^{2n})}{Tn(1-\phi^2)} \\
&= \frac{1}{2T}\left(3 + e^{2\kappa h}\right) - \frac{2(1-e^{-2n\kappa h})}{Tn(1-e^{-2\kappa h})}
\end{aligned}$$

This proves Equation (18).

Proof of Theorem 2.2: For the OU model, the first order MSE is of the form

$$a_{-1/2}^2 = \frac{(U_n - \phi V_n)^2}{\phi^2 h^2}. \tag{39}$$

Substituting (33), (34) and (35) into (39), taking expectation, and collecting terms, we have

$$\begin{aligned} E(a^2_{-1/2}) &= \frac{1}{\phi^2 h^2}\left[\phi^2 + \frac{1+4\phi^2-\phi^4}{n(1-\phi^2)} - \frac{4\phi^2(1-\phi^{2n})}{n^2(1-\phi^2)^2}\right] \\ &\quad + \frac{1}{\phi^2 h^2}\left[\phi^2 + \frac{2\phi^2(1+\phi^2)}{n(1-\phi^2)} - \frac{4\phi^4(1-\phi^{2n})}{n^2(1-\phi^2)^2}\right] \\ &\quad - \frac{1}{\phi^2 h^2}\left[2\phi^2 + \frac{8\phi^2}{n(1-\phi^2)} - \frac{2\phi^2(1+\phi^2)(1-\phi^{2n})}{n^2(1-\phi^2)^2}\right] \\ &= \frac{\phi^{-2}-1}{Th} = \frac{e^{2\kappa h}-1}{Th}. \end{aligned}$$

Interestingly, the terms that involve $1/n^2$ are cancelled out. Hence, the exact calculation of the Cesaro sums does not make a difference to the first order MSE. This proves Theorem 2.2.

Before proving Theorem 3.1, we introduce another lemma.

Lemma 2 Suppose Σ is an $(n+1)\times(n+1)$ matrix with ij-th element $\phi^{|i-j|}$, and C_1 and C_2 are defined in Equation (31), then

$$8tr(C_1\Sigma C_1\Sigma C_1\Sigma) = 2n\phi\left[\phi^2 - 3 + \frac{12(1+\phi^2)(1+\phi^{2n})}{(1-\phi^2)^2}\right] - \frac{12\phi(1+6\phi^2+\phi^4)(1-\phi^{2n})}{(1-\phi^2)^3}, \tag{40}$$

$$tr(C_2\Sigma C_2\Sigma C_2\Sigma) = n + \frac{6n\phi^2(1+\phi^{2n})}{(1-\phi^2)^2} - \frac{6\phi^2(1+\phi^2)(1-\phi^{2n})}{(1-\phi^2)^3}, \tag{41}$$

$$tr(C_1\Sigma C_2\Sigma C_2\Sigma) = \frac{3n\phi(1+\phi^2)(1+\phi^{2n})}{(1-\phi^2)^2} - \frac{2\phi(1+4\phi^2+\phi^4)(1-\phi^{2n})}{(1-\phi^2)^3}, \tag{42}$$

and

$$\begin{aligned} 8tr(C_1\Sigma C_1\Sigma C_2\Sigma) &= 4n\left[1 + 2\phi^{2n} + \frac{12\phi^2(1+\phi^{2n})}{(1-\phi^2)^2}\right] \qquad (43) \\ &\quad -2\left[1+\phi^{2n} + \frac{2(13\phi^2+10\phi^4+\phi^6-\phi^{2n}-10\phi^{2n+2}-13\phi^{2n+4})}{(1-\phi^2)^3}\right], \end{aligned}$$

Proof of Lemma 2: To prove Equation (40), first note that

$$8tr(C_1\Sigma C_1\Sigma C_1\Sigma) = 2n\phi(\phi^2-3) + 12\sum(2t-1)(2n-2t+1)\phi^{2n-2t+1}, \tag{44}$$

Applying Lemma 1 and the induction method to the right hand side of (44), one derives (40). Similarly, Equations (41)-(43) can be derived.

Proof of Theorem 3.1: First, we obtain expressions for $E(U_n^3)$, $E(V_n^3)$, $E(U_n^2V_n)$, and $E(U_nV_n^2)$. By Lemma 1.1, (33), (34), (35), (41), (42) and (43), we have

$$
\begin{aligned}
E(U_n^3) &= \frac{1}{n^3} E\left(X'C_1X\right)^3 \\
&= \frac{1}{n^3}\left\{(tr(C_1\Sigma))^3 + 3tr(C_1\Sigma)2tr(C_1\Sigma C_1\Sigma) + 8tr(C_1\Sigma C_1\Sigma C_1\Sigma)\right\} \\
&= \phi^3 + \frac{3\phi}{n^2}\left\{\frac{n(1+4\phi^2-\phi^4)}{1-\phi^2} - \frac{4\phi^2(1-\phi^{2n})}{(1-\phi^2)^2}\right\} \qquad (45) \\
&\quad + \frac{2\phi}{n^2}\left\{\phi^2 - 3 + \frac{12(1+\phi^2)(1+\phi^{2n})}{(1-\phi^2)^2}\right\} \\
&\quad - \frac{12\phi(1-\phi^{2n})(\phi^4+6\phi^2+1)}{n^3(1-\phi^2)^3},
\end{aligned}
$$

$$
\begin{aligned}
E(V_n^3) &= \frac{1}{n^3} E\left(X'C_2X\right)^3 \\
&= \frac{1}{n^3}\left\{(tr(C_2\Sigma))^3 + 3tr(C_2\Sigma)2tr(C_2\Sigma C_2\Sigma) + 8tr(C_2\Sigma C_2\Sigma C_2\Sigma)\right\} \\
&= 1 + \frac{3}{n^2}\left\{\frac{2n(1+\phi^2)}{1-\phi^2} - \frac{4\phi^2(1-\phi^{2n})}{(1-\phi^2)^2}\right\} \qquad (46) \\
&\quad + \frac{1}{n^2}\left\{8 + \frac{48\phi^2(1+\phi^{2n})}{(1-\phi^2)^2}\right\} - \frac{48\phi^2(1+\phi^2)(1-\phi^{2n})}{n^3(1-\phi^2)^3},
\end{aligned}
$$

$$
\begin{aligned}
E(U_nV_n^2) &= \frac{1}{n^3} E\left[X'C_1X\left(X'C_2X\right)^2\right] \\
&= \frac{1}{n^3}\{tr(C_1\Sigma)(tr(C_2\Sigma))^2 + 2tr(C_1\Sigma)tr(C_2\Sigma C_2\Sigma) \\
&\quad + 2tr(C_2\Sigma)2tr(C_1\Sigma C_2\Sigma) + 8tr(C_1\Sigma C_2\Sigma C_2\Sigma)\} \\
&= \phi + \frac{\phi}{n^2}\left\{\frac{2n(1+\phi^2)}{1-\phi^2} - \frac{4\phi^2(1-\phi^{2n})}{(1-\phi^2)^2}\right\} \qquad (47) \\
&\quad + \frac{4}{n^2}\left\{\frac{2n\phi}{1-\phi^2} - \frac{\phi(1+\phi^2)(1-\phi^{2n})}{(1-\phi^2)^2}\right\} \\
&\quad + \frac{24\phi(1+\phi^2)(1+\phi^{2n})}{n^2(1-\phi^2)^2} - \frac{16\phi(1+4\phi^2+\phi^4)(1-\phi^{2n})}{n^3(1-\phi^2)^3},
\end{aligned}
$$

$$
\begin{aligned}
E(U_n^2 V_n) &= \frac{1}{n^3} E\left[(X'C_1X)^2 X'C_2X\right] \\
&= \frac{1}{n^3}\{(tr(C_1\Sigma))^2 tr(C_2\Sigma) + 2tr(C_1\Sigma)2tr(C_1\Sigma C_2\Sigma) \\
&\quad +2tr(C_2\Sigma)tr(C_1\Sigma C_1\Sigma) + 8tr(C_1\Sigma C_1\Sigma C_2\Sigma)\} \\
&= \phi^2 + \frac{4\phi}{n^2}\left\{\frac{2n\phi}{1-\phi^2} - \frac{\phi(1+\phi^2)(1-\phi^{2n})}{(1-\phi^2)^2}\right\} \qquad (48)\\
&\quad +\frac{1}{n^2}\left\{\frac{n(1+4\phi^2-\phi^4)}{1-\phi^2} + \frac{4\phi^2(1-\phi^{2n})}{(1-\phi^2)^2}\right\} \\
&\quad +\frac{4}{n^2}\left\{1+2\phi^{2n} + \frac{12\phi^2(1+\phi^{2n})}{(1-\phi^2)^2}\right\} \\
&\quad -\frac{2}{n^3}\left\{1+\phi^{2n} + \frac{2(13\phi^2+10\phi^4+\phi^6-\phi^{2n}-10\phi^{2n+2}-13\phi^{2n+4})}{(1-\phi^2)^3}\right\}.
\end{aligned}
$$

In the second step of the proof, as in the proof of Theorem 2.1, we obtain expressions of the high order terms in the stochastic expansion of $\hat{\kappa}$ given by Equation (14), i.e.,

$$
\begin{aligned}
a_{-3/2} &= -QVa_{-1} - \frac{1}{2}QWa_{-1/2}^2 - Q\overline{H_2}a_{-1/2}a_{-1} - \frac{1}{6}Q\overline{H_3}a_{-1/2}^3 \\
&= -\frac{U_n^3-\phi^3V_n^3}{3\phi^3 h} + \frac{U_n^2-\phi^2V_n^2}{\phi^2 h} - \frac{U_n-\phi V_n}{\phi h}. \qquad (49)
\end{aligned}
$$

Substituting (40), (41), (45), and (46) into (49), taking expectation, and collecting terms, we have

$$
\begin{aligned}
E(a_{-3/2}) &= E\left[-\frac{U_n^3-\phi^3V_n^3}{3\phi^3 h} + \frac{U_n^2-\phi^2V_n^2}{\phi^2 h} - \frac{U_n-\phi V_n}{\phi h}\right] \\
&= -\frac{E(U_n^3)-\phi^3(V_n^3)}{3\phi^3 h} + E(a_{-1}) \\
&= -\frac{2(3e^{2\kappa h}+8+e^{-2\kappa h}+4e^{-2(n-1)\kappa h}+8e^{-2n\kappa h})}{Tn(1-e^{-2\kappa h})} \\
&\quad +\frac{4(1-e^{-2n\kappa h})(e^{2\kappa h}+7+4e^{-2\kappa h})}{Tn^2(1-e^{-2\kappa h})^2}.
\end{aligned}
$$

and hence

$$
\begin{aligned}
E(a_{-1}+a_{-3/2}) &= \frac{1}{2T}\left(3+e^{2\kappa h}\right) - \frac{2(3e^{2\kappa h}+9+e^{-2\kappa h}+4e^{-2(n-1)\kappa h}+7e^{-2n\kappa h})}{Tn(1-e^{-2\kappa h})} \\
&\quad +\frac{4(1-e^{-2n\kappa h})(e^{2\kappa h}+7+4e^{-2\kappa h})}{Tn^2(1-e^{-2\kappa h})^2}.
\end{aligned}
$$

This gives rise to (23).

Similarly, the second order MSE to $\hat{\kappa}$ is

$$\begin{aligned} E(a_{-1/2}^2 + 2a_{-1/2}a_{-1}) &= \frac{-1}{\phi^3 h^2}\left\{E(U_n^3) + \phi^3 E(V_n^3) - \phi^2 E(U_n V_n^2) - \phi E(U_n^2 V_n)\right\} + 3E(a_{-1/2}^2) \\ &= \frac{e^{2kh} - 1}{Th} - \frac{2(5 + 7e^{2\kappa h}) + 16e^{-2(n-1)\kappa h}}{T^2} + \frac{10(3 + e^{2\kappa h})(1 - e^{-2n\kappa h})}{T^2 n(1 - e^{-2\kappa h})}. \end{aligned}$$

This proves the first part of Theorem 3.1. The proof of the second part is straightforward and hence omitted.

References

[1] Abadir, K.M., 1993, OLS Bias in a Nonstationary Autoregression, *Econometric Theory*, 9, 81-93.

[2] Aït-Sahalia, Y., 1996a, Nonparametric Pricing of Interest Rate Derivative Securities, *Econometrica*, 64, 527–560.

[3] Aït-Sahalia, Y., 1996b, Testing Continuous-time Models of Spot Interest Rate Derivative Securities, *Review of Financial Studies*, 9, 385–426.

[4] Aït-Sahalia, Y., 1999, Transition Densities for Interest Rate and Other Nonlinear Diffusions, *Journal of Finance*, 54, 1361–1395.

[5] Aït-Sahalia, Y., 2002, Maximum Likelihood Estimation of Discretely Sampled Diffusion: A Closed-form Approximation Approach. *Econometrica*, 70, 223–262.

[6] Aït-Sahalia, Y. and J. Yu, 2006, Saddlepoint Approximation for Continuous-time Markov Processes. *Journal of Econometrics*, 134, 507-551.

[7] Andersen, T., T. Bollerslev, F.X. Diebold, and P. Labys 2001. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96, 42–55.

[8] Andrews, D. W. K., 1993, Exactly Median-unbiased Estimation of First Order Autoregressive/unit Root Models, *Econometrica*, 61, 139–166.

[9] Ball, C. A., and W. N. Torous, 1996, Unit Roots and the Estimation of Interest Rate Dynamics, *Journal of Empirical Finance*, 3, 215–238.

[10] Bandi, F., 2002, Short Term Interest Rate Dynamics: A Spatial Approach, *Journal of Financial Economics*, 65, 73–110.

[11] Bandi, F. and P.C.B. Phillips, 2003, Fully Nonparametric Estimation of Scalar Diffusion Models, *Econometrica*, 71, 241–283.

[12] Bandi, F. & P.C.B. Phillips, 2007, A simple approach to the parametric estimation of potentially nonstationary diffusions. *Journal of Econometrics*, 137, 354-395.

[13] Bartlett, M.S. 1946, On the Theoretical Specification and Sampling Properties of Autocorrelated Time-series. *Journal of Royal Statistical Society*, 8, 27-41.

[14] Barndorff-Nielsen, O. and N. Shephard, 2002, Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society, Series B*, 64, 253-280.

[15] 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.

[16] Bergstrom, Albert R. 1990, *Continuous Time Econometric Modelling.* Oxford University Press, Oxford.

[17] Brown, B.M. and J.I. Hewitt, 1975. Asymptotic Likelihood Theory for Diffusion Processes, *Journal Applied Probability*, 12, 228-238.

[18] Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders 1992, An Empirical Comparison of Alternative Models of Short Term Interest Rates, *Journal of Finance*, 47, 1209–1227.

[19] Dai, Q., and K. J. Singleton, 2000, Specification Analysis of Affine Term Structure Models, *Journal of Finance*, 55, 1943–78.

[20] Duffie, D., and R. Kan, 1996, A Yield-factor Model of Interest Rate, *Mathematical Finance*, 6, 379–406.

[21] Duffie, D., and K. J. Singleton, 1993, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, 61, 929–952.

[22] Elerian, O., Chib, S. and N. Shephard. 2001. Likelihood Inference for Discretely Observed Non-linear Diffusions. *Econometrica* 69, 959–993.

[23] Gouriéroux, C., A. Monfort, and E. Renault, 1993, Indirect Inference, *Journal of Applied Econometrics*, 8, S85–S118.

[24] Hong, Y., and H. Li, 2005, Nonparametric Specification Testing for Continuous Time Model with Application to Spot Interest Rates, *Review of Financial Studies.* 18, 37 - 84.

[25] Hurwicz, L., 1950, Least Square Bias in Time Series, in T. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, New York, Wiley, 365-383.

[26] Lo, A. W., 1988, Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data, *Econometric Theory*, 4, 231–247.

[27] MacKinnon, J. G., and A. A. Smith 1998, "Approximate Bias Correction in Econometrics," *Journal of Econometrics*, 85, 205-230.

[28] Marriott, F. and J. Pope, 1954, Bias in the Estimation of Autocorrelations. *Biometrika* 41, 390-402.

[29] Merton, R. C., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics*, 8, 323–361.

[30] Merton, R. C. 1990, Continuous-time Finance. Blackwell, Massachusetts.

[31] Pedersen, A., 1995, A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observation. *Scandinavian Journal of Statistics*, 22, 55-71.

[32] Phillips, P.C.B., 1972, The Structural Estimation of a Stochastic Differential Equation System, *Econometrica*, 40, pp. 1021-1041.

[33] Phillips, P.C.B. and T. Magdolinos, 2009, Unit root and cointegrating limit theory when initialization is in the infinite past. *Econometric Theory*, forthcoming.

[34] Phillips, P.C.B. and J. Yu, 2005, Jackknifing Bond Option Prices. *Review of Financial Studies*, 18, 707-742.

[35] Phillips, P.C.B. and J. Yu, 2009a, Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Handbook of Financial Time Series*, 497-530.

[36] Phillips, P.C.B. and J. Yu, 2009b, Simulation-based Estimation of Contingent-claims Prices. *Review of Financial Studies*, 22, 3669-3705..

[37] Pritsker, M., 1998, Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models. *Review of Financial Studies*, 11, 449-487.

[38] Quenouille, M. H., 1956, Notes on Bias in Estimation, *Biometrika*, 43, 353-360.

[39] Rilstone, P, Srivastava, V. K., and A. Ullah, 1996, The second-order bias and mean squared error of nonlinear estimators, *Journal of Econometrics*, 75, 369-395.

[40] Smith, A.A., 1993, Estimating nonlinear time-series models using simulated vector autoregressions, *Journal of Applied Econometrics*, 8, S63–S84.

[41] Stanton, R., 1997, A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk, *Journal of Finance*, 52, 1973–2002.

[42] Tang, C.Y. and S.X. Chen, 2009, Parameter Estimation and Bias Correction for Diffusion Processes. *Journal of Econometrics*, forthcoming.

[43] Ullah, A., 2004, *Finite Sample Econometrics*. Oxford University Press, Oxford.

[44] Vasicek, O., 1977, An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177–186.

[45] Vinod. H.D. and L. R. Shenton, 1996, Exact Moments for Autoregressive and Random Walk Models for a Zero or Stationary Initial Value, *Econometric Theory*, 12, 481-499.

[46] White, J. 1961, Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika* 48, 85-94.