BIAS IN ESTIMATING MULTIVARIATE AND UNIVARIATE DIFFUSIONS

By

Xiaohu Wang, Peter C. B. Phillips and Jun Yu

January 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1778

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

Bias in Estimating Multivariate and Univariate Diffusions*

Xiaohu Wang
Singapore Management University

Peter C. B. Phillips
Yale University, University of Auckland,
University of Southampton & Singapore Management University

Jun Yu
Singapore Management University

November 26, 2010

Abstract

Multivariate continuous time models are now widely used in economics and finance. Empirical applications typically rely on some process of discretization so that the system may be estimated with discrete data. This paper introduces a framework for discretizing linear multivariate continuous time systems that includes the commonly used Euler and trapezoidal approximations as special cases and leads to a general class of estimators for the mean reversion matrix. Asymptotic distributions and bias formulae are obtained for estimates of the mean reversion parameter. Explicit expressions are given for the discretization bias and its relationship to estimation bias in both multivariate and in univariate settings. In the univariate context, we compare the performance of the two approximation methods relative to exact maximum likelihood (ML) in terms of bias and variance for the Vasicek process. The bias and the variance of the Euler method are found to be smaller than the trapezoidal method, which are in turn smaller than those of exact ML. Simulations suggest that when the mean reversion is slow the approximation methods work better than ML, the bias formulae are accurate, and for scalar models the estimates obtained from the two approximate methods have smaller bias and variance than exact ML. For the square root process, the Euler method outperforms the Nowman method in terms of both bias and variance. Simulation evidence indicates that the Euler method has smaller bias and variance than exact ML, Nowman's method and the Milstein method.

Keywords: Bias; Diffusion, Euler approximation; Trapezoidal approximation; Milstein approximation

JEL classification: C15, G12

1 Introduction

Continuous time models, which are specified in terms of stochastic differential equations, have found wide applications in economics and finance. Empirical interest in systems of this type has grown particularly rapidly in recent years with the availability of high frequency

*Xiaohu Wang and Jun Yu (author for correspondence), School of Economics and Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: xiaohu.wang.2008@phdecons.smu.edu.sg and yujun@smu.edu.sg. Peter Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, Yale Station, New Haven, Connecticut 06520-8281. Email: peter.phillips@yale.edu.

financial data. Correspondingly, growing attention has been given to the development of econometric methods of inference. In order to capture causal linkages among variables and allow for multiple determining factors, many continuous systems are specified in multivariate form. The literature is now wide-ranging. Bergstrom (1990) motivated the use of multivariate continuous time models in macroeconomics; Sundaresan (2000) provided a list of multivariate continuous time models, particularly multivariate diffusions, in finance; and Piazzesi (2009) discusses how to use multivariate continuous time models to address various macro-finance issues.

Data in economics and finance are typically available at discrete points in time or over discrete time intervals and many continuous systems are formulated as Markov processes. These two features suggest that the log likelihood function can be expressed as the product of the log transition probability densities (TPD). Consequently, the implementation of maximum likelihood (ML) requires evaluation of TPD. But since the TPD is unavailable in closed form for many continuous systems and several methods have been proposed as approximations.

The simplest approach is to approximate the model using some discrete time system. Both the Euler approximation and the trapezoidal rule have been suggested in the literature. Sargan (1974) and Bergstrom (1984) showed that the ML estimators (MLEs) based on these two approximations converge to the true MLE as the sampling interval $h \to 0$, at least under a linear specification. This property also holds for more general diffusions (Florens-Zmirou, 1989). Of course, the quality of the approximation depends on the size of h. However, the advantage of the approximation approach is that it is computationally simple and often works well when h is small, for example at the daily frequency.

More accurate approximations have been proposed in recent years. The two that have received the most attention are in-fill simulations and closed-form approximations. Studies of in-fill simulations include Pedersen (1995) and Durham and Gallant (2002). For closed-form approximations, seminal contributions include Aït-Sahalia (1999, 2002, 2008), Aït-Sahalia and Kimmel (2007), and Aït-Sahalia and Yu (2006). These approximations have the advantage that they can control the size of the approximation errors even when h is not small. Aït-Sahalia (2008) provides evidence that the closed-form approximation is highly accurate and allows for fast repeated evaluations. Since the approximate TPD takes a complicated form in both these approaches, no closed form expression is available for the MLE. Consequently, numerical optimizations are needed to obtain the MLE.

No matter which of the above methods is used, when the system variable is persistent, the resulting estimator of the speed of mean reversion can suffer from severe bias in finite samples. This problem is well known in scalar diffusions (Phillips and Yu, 2005a, 2005b, 2009a, 2009b) but has also been reported in multivariate models (Phillips and Yu, 2005a and Tang and Chen, 2009). In the scalar case, Tang and Chen (2009) and Yu (2009) give explicit expressions to approximate the bias. To obtain these explicit expressions, the corresponding estimators must have a closed-form expression. That is why explicit bias results are presently available only for the scalar Vasicek model (Vasicek, 1977) and the Cox-Ingersoll-Ross (CIR,

1985) model.

The present paper focuses on extending existing bias formulae to the multivariate continuous system case. We partly confine our attention to linear systems so that explicit formulae are possible for approximating the estimation bias of the mean reversion matrix. It is known from previous work that bias in the mean reversion parameter has some robustness to specification changes in the diffusion function (Tang and Chen, 2009), which gives this approach a wider relevance. Understanding the source of the mean reversion bias in linear systems can also be helpful in more general situations where there are nonlinearities.

The paper develops a framework for studying estimation in multivariate continuous time models with discrete data. In particular, we show how the estimator that is based on the Euler approximation and the estimator based on the trapezoidal approximation can be obtained by taking Taylor expansions to the first and second orders. Moreover, the uniform framework simplifies the derivation of the asymptotic bias order of the ordinary least squares estimator and the two stage least squares estimator of Bergstrom (1984). Asymptotic theory is provided under long time span asymptotics and explicit formulae for the matrix bias approximations are obtained. The bias formulae are decomposed into the discretization bias and the estimation bias. Simulations reveal that the bias formulae work well in practice. The results are specialized to the scalar case, giving two approximate estimators of the mean reversion parameter which are shown to work well relative to the exact MLE when the mean reversion is slow.

The results confirm that bias can be severe in multivariate continuous time models for parameter values that are empirically realistic, just as it is in scalar models. Specializing our formulae to the univariate case yields some useful alternative bias expressions. Simulations are reported that detail the performance of the bias formulae in the multivariate setting and in the univariate setting.

The rest of the paper is organized as follows. Section 2 introduces the model and the setup and reviews four existing estimation methods. Section 3 outlines our unified framework for estimation, establishes the asymptotic theory, and provides explicit expressions for approximating the bias in finite samples. Section 4 discusses the relationship between the new estimators and two existing estimators in the literature, and derives a new bias formula in the univariate setting. Section 5 compares the performance of the estimator based on the Euler scheme relative to that the method proposed by Nowman (1997) in the context of the square root process and a diffusion process with a linear drift but a more general diffusion. Simulations are reported in Section 6. Section 7 concludes and the Appendix collects together proofs of the main results.

2 The Model and Existing Methods

We consider an M-dimensional multivariate diffusion process of the form (cf. Phillips, 1972):

$$dX(t) = (A(\theta)X(t) + B(\theta))dt + \zeta(dt), \quad X(0) = X_0, \tag{2.1}$$

where $X(t) = (X_1(t), \cdots, X_M(t))'$ is an M-dimensional continuous time process, $A(\theta)$ and $B(\theta)$ are $M \times M$ and $M \times 1$ matrices, whose elements depend on unknown parameters $\theta = (\theta_1, \cdots, \theta_K)$ that need to be estimated, $\zeta(dt)$ ($:= (\zeta_1(dt), \cdots, \zeta_M(dt))$) is a vector random process with uncorrelated increments and covariance matrix Σdt. The particular model receiving most attention in finance is when $\zeta(dt)$ is a vector of Brownian increments (denoted by $dW(t)$) with covariance Σdt, viz.,

$$dX(t) = (A(\theta)X(t) + B(\theta))dt + dW(t), \quad X(0) = X_0, \tag{2.2}$$

corresponding to a multivariate version of the Vasicek model (Vasicek, 1977).

Although the process follows a continuous time stochastic differential equation system, observations are available only at discrete time points, say at n equally spaced points $\{th\}_{t=0}^{n}$, where h is the sampling interval and is taken to be fixed. In practice, h might be very small, corresponding to high-frequency data. In this paper, we use $X(t)$ to represent a continuous time process and X_t to represent a discrete time process. When there is no confusion, we simply write X_{th} as X_t.

Bergstrom (1990) provided arguments why it is useful for macroeconomists and policy makers like central bankers to formulate models in continuous time even when discrete observations only are available. In finance, early fundamental work by Black and Scholes (1973) and much of the ensuing literature such as Duffie and Kan (1996) successfully demonstrated the usefulness of both scalar and multivariate diffusion models in the development of financial asset pricing theory.

Phillips (1972) showed that the exact discrete time model corresponding to (2.1) is given by

$$X_t = \exp\{A(\theta)h\}X_{t-1} - A^{-1}(\theta)[\exp\{A(\theta)h\} - I]B(\theta) + \varepsilon_t. \tag{2.3}$$

where $\varepsilon_t = (\varepsilon_1, \cdots, \varepsilon_M)'$ is a martingale difference sequence (MDS) with respect to the natural filtration and

$$E(\varepsilon_t \varepsilon_t') = \int_0^h \exp\{A(\theta)s\}\Sigma \exp\{A(\theta)'s\}ds := G.$$

Letting $F(\theta) := \exp\{A(\theta)h\}$ and $g(\theta) := -A^{-1}(\theta)[\exp\{A(\theta)h\} - I]B(\theta)$, we have the system

$$X_t = F(\theta)X_{t-1} + g(\theta) + \varepsilon_t, \tag{2.4}$$

which is a vector autoregression (VAR) model of order 1 with MDS$(0, G)$ innovations.

In general, identification of θ from the implied discrete model (2.3) generating discrete observations $\{X_{th}\}$ is not automatically satisfied. The necessary and sufficient condition for identifiability of θ in model (2.3) is that the correspondence between θ and $[F(\theta), g(\theta)]$ be one-to-one, since (2.3) is effectively a reduced form for the discrete observations. Phillips (1973) studied the identifiability of $(A(\theta), \Sigma)$ in (2.3) in terms of the identifiability of the matrix $A(\theta)$ in the matrix exponential $F = \exp(A(\theta)h)$ under possible restrictions implied by the structural functional dependence $A = A(\theta)$ in (2.1). In general, a one-to–one correspondence between $A(\theta)$ and F, requires the structural matrix $A(\theta)$ to be restricted. This is because if $A(\theta)$ satisfies $\exp\{A(\theta)h\} = F$ and some of its eigenvalues are complex, $A(\theta)$ is not uniquely identified. In fact, adding to each pair of conjugate complex eigenvalues the imaginary numbers $2ik\pi$ and $-2ik\pi$ for any integer k, leads to another matrix satisfying $\exp\{Ah\} = F$. This phenomenon is well known as aliasing in the signal processing literature. When restrictions are placed on the structural matrix $A(\theta)$ identification is possible. Phillips (1973) gave a rank condition for the case of linear homogeneous relations between the elements of a row of A. A special case is when $A(\theta)$ is triangular. Hansen and Sargent (1983) extended this result by showing that the reduced form covariance structure $G > 0$ provides extra identifying information about A, reducing the number of potential aliases.

To deal with the estimation of (2.1) using discrete data and indirectly (because it was not mentioned) the problem of identification, two approximate discrete time models were proposed in earlier studies. The first is based on the Euler approximation given by

$$\int_{(t-1)h}^{th} A(\theta)X(r)dr \approx A(\theta)hX_{t-1},$$

which leads to the approximate discrete time model

$$X_t - X_{t-1} = A(\theta)hX_{t-1} + B(\theta)h + u_t. \tag{2.5}$$

The second, proposed by Bergstrom (1966), is based on the trapezoidal approximation

$$\int_{(t-1)h}^{th} A(\theta)X(r)dr \approx \frac{1}{2}A(\theta)h(X_t + X_{t-1}),$$

which gives rise to the approximate nonrecursive discrete time model

$$X_t - X_{t-1} = \frac{1}{2}A(\theta)h(X_t + X_{t-1}) + B(\theta)h + \nu_t. \tag{2.6}$$

The discrete time models are then estimated by standard statistical methods, namely OLS for the Euler approximation and systems estimation methods such as two-stage or three-stage least squares for the trapezoidal approximation. As explained by Lo (1988) in the univariate context, such estimation strategies inevitably suffer from discretization bias. The size of the discretization bias depends on the sampling interval, h, and does not disappear even if $n \to \infty$. The bigger is h, the larger is the discretization bias. Sargan (1974) showed that the asymptotic discretization bias of the two-stage and three-stage least squares estimators for the trapezoidal approximation is $O(h^2)$ as $h \to 0$. Bergstrom (1984) showed that the asymptotic discretization bias of the OLS estimator for the Euler approximation is $O(h)$.

For the more general multivariate diffusion

$$dX(t) = \kappa(\mu - X(t))dt + \Sigma(X(t);\psi)dW(t), \quad X(0) = X_0, \tag{2.7}$$

where W is standard Brownian motion, two other approaches have been used to approximate the continuous time model (2.7). The first, proposed by Nowman (1997), approximates the diffusion function within each unit interval, $[(i-1)h, ih)$ by its left end point value leading to the approximate model

$$dX(t) = \kappa(\mu - X(t))dt + \Sigma(X_{(i-1)h};\psi)dW(t) \quad \text{for } t \in [(i-1)h, ih). \tag{2.8}$$

Since (2.8) is a multivariate Vasicek model within each unit interval, there is a corresponding exact discrete model as in (2.3). This discrete time model, being an approximation to the exact discrete time model of (2.7), facilitates direct Gaussian estimation.

To reduce the approximation error introduced by the Euler scheme, Milstein (1978) suggested taking the second order term in a stochastic Taylor series expansion when approximating the drift function and the diffusion function. Integrating (2.7) gives

$$\int_{(i-1)h}^{ih} dX(t) = \int_{(i-1)h}^{ih} \kappa(\mu - X(t))dt + \int_{(i-1)h}^{ih} \Sigma(X(t);\psi)dW(t). \tag{2.9}$$

By Itô's lemma, the linearity of the drift function in (2.7), and using tensor summation notation for repeated indices (p, q), we obtain

$$d\mu(X(t);\theta) = \frac{\partial \mu(X(t);\theta)}{\partial X_p} dX_p(t),$$

and

$$d\Sigma(X(t);\psi) = \frac{\partial \Sigma(X(t);\psi)}{\partial X_p} dX_p(t) + \frac{1}{2}\frac{\partial^2 \Sigma(X(t);\psi)}{\partial X_p \partial X_q'} dX_p(t)dX_q(t), \tag{2.10}$$

where $\mu_j(X(t);\theta)$ is the j^{th} element of the (linear) drift function $\kappa(\mu - X(t))$, Σ_{pq} is the $(p,q)^{th}$ element of Σ and X_p is the p^{th} element of X. These expressions lead to the approximations

$$\mu(X(t);\theta) \simeq \mu(X_{(i-1)h};\theta),$$

and

$$\Sigma(X(t);\psi) \simeq \Sigma(X_{(i-1)h};\theta) + \frac{\partial \Sigma(X_{(i-1)h};\psi)}{\partial X_p} \Sigma_{pq}(X_{(i-1)h};\psi) \int_{(i-1)h}^{t} dW_q(\tau).$$

Using these approximations in (2.9) we find

$$\begin{aligned} X_{ih} - X_{(i-1)h} &= \int_{(i-1)h}^{ih} \kappa(\mu - X(t))dt + \int_{(i-1)h}^{ih} \Sigma(X(t);\psi)dW(t) \\ &\simeq \mu(X_{(i-1)h};\theta)h + \Sigma(X_{(i-1)h};\psi) \int_{(i-1)h}^{ih} dW\,(t) \\ &+ \frac{\partial \Sigma(X_{(i-1)h};\psi)}{\partial X_p} \Sigma_{pq}(X_{(i-1)h};\psi) \int_{(i-1)h}^{ih} \int_{(i-1)h}^{t} dW_q(\tau)dW\,(t)\,. \end{aligned} \tag{2.11}$$

The multiple (vector) stochastic integral in (2.11) reduces as follows:

$$\begin{aligned} \int_{(i-1)h}^{ih} \int_{(i-1)h}^{t} dW_q(\tau)dW_p\,(t) &= \int_{(i-1)h}^{ih} \left(W_q(t) - W_{q(i-1)h}\right) dW_p\,(t) \\ &= \begin{cases} \frac{1}{2}\left\{\left(W_{qih} - W_{q(i-1)h}\right)^2 - h\right\} & p = q \\ \int_{(i-1)h}^{ih} \left(W_q(t) - W_{q(i-1)h}\right) dW_p\,(t) & p \neq q \end{cases}. \end{aligned} \tag{2.12}$$

The approximate model under a Milstein second order discretization is then

$$\begin{aligned} X_{ih} - X_{(i-1)h} &\simeq \mu(X_{(i-1)h};\theta)h + \Sigma(X_{(i-1)h};\psi)\left(W_{ih} - W_{(i-1)h}\right) \\ &+ \frac{\partial \Sigma(X_{(i-1)h};\psi)}{\partial X_p} \Sigma_{pq}(X_{(i-1)h};\psi) \int_{(i-1)h}^{ih} \int_{(i-1)h}^{t} dW_q(\tau)dW_p\,(t)\,. \end{aligned} \tag{2.13}$$

In view of the calculation (2.12), when the model is scalar the discrete approximation has the simple form (c.f., Phillips and Yu, 2009)

$$\begin{aligned} X_{ih} - X_{(i-1)h} \simeq & \left[\mu(X_{(i-1)h};\theta) - \frac{1}{2}\sigma'(X_{(i-1)h};\psi)\sigma(X_{(i-1)h};\psi) \right] h \\ & + \sigma(X_{(i-1)h};\psi) \left(W_{ih} - W_{(i-1)h} \right) \\ & + \sigma'(X_{(i-1)h};\psi)\sigma(X_{(i-1)h};\psi)\frac{1}{2} \left(W_{ih} - W_{(i-1)h} \right)^2 . \end{aligned} \tag{2.14}$$

Since $\frac{1}{2}\left\{ \left(W_{qih} - W_{q(i-1)h} \right)^2 - h \right\}$ has mean zero, the net contribution to the drift from the second order term is zero.

In the multivariate Vasicek model, $\Sigma(X(t);\psi) = \Sigma$, and the Milstein approximation (2.13) reduces to

$$X_{ih} - X_{(i-1)h} \simeq \mu(X_{(i-1)h};\theta)h + \Sigma(X_{(i-1)h};\psi) \left(W_{ih} - W_{(i-1)h} \right).$$

Thus, for the multivariate Vasicek model, the Milstein and Euler schemes are equivalent.

3 Estimation Methods, Asymptotic Theory and Bias

In this paper, following the approach of Phillips (1972), we estimate θ directly from the exact discrete time model (2.3). In particular, we first estimate $F(\theta)$ and θ from (2.3), assuming throughout that $A(\theta)$ and θ are identifiable and that all the eigenvalues in $A(\theta)$ have negative real parts. The latter condition ensures that X_t is stationary and is therefore mean reverting. The exact discrete time model (2.3) in this case is a simple VAR(1) model which has been widely studied in the discrete time series literature. We first review some relevant results from this literature.

Let $Z_t = [X_t', 1]'$. The OLS estimator of $H = [F, g]$ is

$$\hat{H} = [\hat{F}, \hat{g}] = \left[n^{-1} \sum_{t=1}^{n} X_t Z_{t-1}' \right] \cdot \left[n^{-1} \sum_{t=1}^{n} Z_{t-1} Z_{t-1}' \right]^{-1}. \tag{3.1}$$

If we have prior knowledge that $B(\theta) = 0$ and hence $g = 0$, the OLS estimator of F is:

$$\hat{F} = \left[n^{-1} \sum_{t=1}^{n} X_t X_{t-1}' \right] \cdot \left[n^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' \right]^{-1}, \tag{3.2}$$

for which the standard theory first order limit theory (e.g., Fuller (1976, p.340) and Hannan (1970, p.329)) is well known.

Lemma 3.1 *For the stationary VAR(1) model (2.4), if h is fixed and $n \to \infty$, we have*
(a) $\hat{F} \xrightarrow{p} F$;
(b) $\sqrt{n}\{Vec(\hat{F}) - Vec(F)\} \xrightarrow{d} N(0, (\Gamma(0))^{-1} \otimes G)$,
where $\Gamma(0) = Var(X_t) = \sum_{i=0}^{\infty} F^i \cdot G \cdot F'^i$ *and* $G = E(\varepsilon_t \varepsilon_t')$

Under different but related conditions, Yamamoto and Kunitomo (1984) and Nicholls and Pope (1988) derived explicit bias expressions for the OLS estimator $\hat{F}$. The proof of the following lemma is given in Yamamoto and Kunitomo (1984, theorem 1).

Lemma 3.2 *(Yamamoto and Kunitomo (1984)) Assume:*
(A1) X_t *is a stationary VAR(1) process whose error term is iid* $(0, G)$ *with* G *nonsingular;*
(A2) For some $s_0 \geq 16$, $E|\varepsilon_{ti}|^{s_0} < \infty$, *for all* $i = 1, \cdots, M$;
(A3) $E \left\| \left[n^{-1} \sum_{t=1}^{n} Z_{t-1} Z'_{t-1} \right]^{-1} \right\|^2$ *is bounded, where the operator* $\| \cdot \|$ *is defined by*

$$\| Q \| = \sup_{\beta} (\beta' Q' Q \beta)^{1/2} (\beta' \beta \leq 1),$$

for any vector β;
Under (A1)-(A3) if $n \to \infty$, *the bias of OLS estimator of* F *in the VAR(1) model with an unknown intercept is*

$$BIAS(\hat{F}) = -n^{-1} G \sum_{k=0}^{\infty} \{ F'^k + F'^k tr(F^{k+1}) + F'^{2k+1} \} D^{-1} + O(n^{-\frac{3}{2}}), \tag{3.3}$$

where

$$D = \sum_{i=0}^{\infty} F^i G F'^i,$$

and the bias of the OLS estimator of F *for the VAR(1) model with a known intercept is*

$$BIAS(\hat{F}) = -\frac{1}{n} G \sum_{k=0}^{\infty} \{ F'^k tr(F^{k+1}) + F'^{2k+1} \} D^{-1} + O(n^{-\frac{3}{2}}). \tag{3.4}$$

We now derive a simplified bias formulae in the two models which facilitates the calculation of the bias formulae in continuous time models.

Lemma 3.3 *Assume (A1)-(A3) hold,* h *is fixed and* $n \to \infty$. *The bias of the least squares estimator for* F *in the VAR(1) is given by*

$$B_n = E(\hat{F}) - F = -\frac{b}{n} + O(n^{-\frac{3}{2}}). \tag{3.5}$$

When the model has a unknown intercept,

$$b = G[(I - C)^{-1} + C(I - C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda (I - \lambda C)^{-1}] \Gamma(0)^{-1}, \tag{3.6}$$

where $C = F'$, $\Gamma(0) = Var(X_t) = \sum_{t=0}^{\infty} F^i \cdot G \cdot F'^i$, $G = E(\varepsilon_t \varepsilon'_t)$, *and* $Spec(C)$ *denotes the set of eigenvalues of* C. *When the model has a known intercept,*

$$b = G[C(I - C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda (I - \lambda C)^{-1}] \Gamma(0)^{-1}. \tag{3.7}$$

Remark 3.1 *The alternative bias formula (3.5) is exactly the same as that given by Nicholls and Pope (1988) for the Gaussian case, although here the expression is obtained without Gaussianity and in a simpler way. If the bias is calculated to a higher order, Bao and Ullah (2009) showed that skewness and excess kurtosis of the error distribution figure in the formulae. In a related contribution, Ullah et al (2010) obtain the second order bias in the mean reversion parameter for a (scalar) continuous time Lévy process.*

We now develop estimators for A. To do so we use the matrix exponential expression

$$F = e^{Ah} = \sum_{i=0}^{\infty} \frac{(Ah)^i}{i!} = I + Ah + H = I + Ah + O(h^2) \text{ as } h \to 0. \tag{3.8}$$

Rearranging terms we get

$$A = \frac{1}{h}(F - I) - \frac{1}{h}H = \frac{1}{h}(F - I) + O(h) \text{ as } h \to 0, \tag{3.9}$$

which suggest the following simple estimator of A

$$\hat{A} = \frac{1}{h}(\hat{F} - I), \tag{3.10}$$

where $\hat{F}$ is the OLS estimator of F. We now develop the asymptotic distribution for $\hat{A}$ and the bias in $\hat{A}$.

Theorem 3.1 *Assume X_t follows Model (2.1) and that all characteristic roots of the coefficient matrix A have negative real parts. Let $\{X_{th}\}_{t=1}^{n}$ be the available data and suppose A is estimated by (3.10) with $\hat{F}$ defined by (3.1). When h is fixed, as $n \to +\infty$, we have*

$$\hat{A} - A \xrightarrow{p} \frac{1}{h}(F - I - Ah) = \frac{1}{h}H = O(h) \text{ as } h \to 0, \tag{3.11}$$

where $H = F - I - Ah$, and

$$h\sqrt{n}Vec\left[\hat{A} - \frac{1}{h}(F - I)\right] \xrightarrow{d} N(0, \Gamma(0)^{-1} \otimes G), \tag{3.12}$$

where $\Gamma(0) = Var(X_t) = \sum_{i=0}^{\infty} F^i G F'^i$, $G = E(\varepsilon_t \varepsilon_t')$.

Theorem 3.2 *Assume that X_t follows Model (2.2) where $W(t)$ is a vector Brownian Motion with covariance matrix Σ and that all characteristic roots of the coefficient matrix A have negative real parts. Let $\{X_{th}\}_{t=1}^{n}$ be the available data and suppose A is estimated by (3.10) with $\hat{F}$ defined by (3.1). When h is fixed and $n \to \infty$, the bias formula is:*

$$BIAS(\hat{A}) = E(\hat{A} - A) = \frac{1}{h}H + \frac{-b}{T} + o(T^{-1}), \tag{3.13}$$

where $H = F - I - Ah$, and $T = nh$ is the time span of the data. If $B(\theta)$ is unknown, then

$$b = G[(I - C)^{-1} + C(I - C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I - \lambda C)^{-1}]\Gamma(0)^{-1}, \tag{3.14}$$

where $\Gamma(0) = Var(X_t) = \sum_{i=0}^{\infty} F^i \cdot G \cdot F'^i$, $G = E(\varepsilon_t \varepsilon_t')$, and $Spec(C)$ is the set of eigenvalues of C. If $B(\theta)$ is known, then

$$b = G[C(I - C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I - \lambda C)^{-1}]\Gamma(0)^{-1}. \tag{3.15}$$

Remark 3.2 *Expression (3.11) extends the result in equation (32) of Lo (1988) to the multivariate case. According to Theorem 3.2, the bias of the estimator (3.10) can be decomposed into two parts, the discretization bias and the estimation bias, which take the following forms:*

$$\text{discretization bias} = \frac{H}{h} = \frac{F - I - Ah}{h} = O(h) \text{ as } h \to 0, \tag{3.16}$$

$$\textit{estimation bias} = \frac{-b}{T} + o(T^{-1}). \tag{3.17}$$

It is difficult to determine the signs of the discretization bias and the estimation bias in a general multivariate case. However, in the univariate case, the signs are opposite to each other as shown in Section 4.2.

Higher order approximations are possible. For example, we may take the matrix exponential series expansion to the second order to produce a more accurate estimate using

$$\begin{aligned}
F = e^{Ah} &= \sum_{i=0}^{\infty} \frac{(Ah)^i}{i!} \\
&= I + Ah + \frac{Ah}{2}\left[(e^{Ah} - I) + \frac{-A^2h^2}{3!} + \frac{-2A^3h^3}{4!} + \ldots + \frac{-(n-2)A^{n-1}h^{n-1}}{n!} + \cdots\right] \\
&= I + Ah + \frac{Ah}{2}[F - I] + \eta \\
&= I + Ah + \frac{Ah}{2}[F - I] + O(h^3) \text{ as } h \to 0.
\end{aligned} \tag{3.18}$$

Consequently,

$$\begin{aligned}
A &= \frac{2}{h}(F - I)(F + I)^{-1} - \frac{2}{h}\eta(F + I)^{-1} = \frac{2}{h}(F - I)(F + I)^{-1} + \nu \\
&= \frac{2}{h}(F - I)(F + I)^{-1} + O(h^2) \text{ as } h \to 0.
\end{aligned} \tag{3.19}$$

After neglecting terms smaller than $O\left(h^2\right)$, we get the alternative estimator

$$\hat{A} = \frac{2}{h}(\hat{F} - I)(\hat{F} + I)^{-1}. \tag{3.20}$$

Theorem 3.3 *Assume that X_t follows Model (2.1) and that all characteristic roots of the coefficient matrix A have negative real parts. Let $\{X_{th}\}_{t=1}^n$ be the available data and A is estimated by (3.20) with $\hat{F}$ defined by (3.1). When h is fixed, $n \to +\infty$, we have*

$$\hat{A} - A \xrightarrow{p} \frac{2}{h}(F - I)(F + I)^{-1} - A = O(h^2) \textit{ as } h \to 0,$$

and

$$h\sqrt{n}Vec\left[\hat{A} - \frac{2}{h}(F - I)(F + I)^{-1}\right] \xrightarrow{d} N(0, \Psi),$$

where

$$\Psi = 16\Upsilon[\Gamma(0)^{-1} \otimes G]\Upsilon', \quad \Upsilon = (F' + I)^{-1} \otimes (F + I)^{-1}.$$

Theorem 3.4 *Assume that X_t follows (2.2) where $W(t)$ is a vector Brownian motion with covariance matrix Σ and that all characteristic roots of the coefficient matrix A have negative real parts. Let $\{X_{th}\}_{t=1}^n$ be the available data and suppose A is estimated by (3.20) with $\hat{F}$ defined by (3.1). When h is fixed, $n \to \infty$, and $T = hn$, the bias formula is:*

$$BIAS(\hat{A}) = -\nu - \frac{4}{T}(I + F)^{-1}b(I + F)^{-1} - \frac{4}{h}L(I + F)^{-1} + o(T^{-1}), \tag{3.21}$$

where $\nu = A - \frac{2}{h}(F - I)(F + I)^{-1}$, $\Delta = [I_M \otimes (I + F)^{-1}] \cdot \Gamma(0)^{-1} \otimes G \cdot [I_M \otimes (I + F)^{-1}]'$, and L is a $M \times M$ matrix whose ij^{th} element is given by

$$L_{ij} = \frac{1}{n}\sum_{s=1}^{M} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s}, \tag{3.22}$$

with e_i being a column vector of dimension M^2 whose i^{th} element is 1 and other elements are 0. If $B(\theta)$ is an unknown vector, then

$$b = G[(I-C)^{-1} + C(I-C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1}.$$

If $B(\theta)$ is a known vector, then

$$b = G[C(I-C^2)^{-1} + \sum_{\lambda \in Spec(C)} \lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1}.$$

Remark 3.3 *Theorem 3.4 shows that the bias of the estimator (3.20) can be decomposed into a discretization bias and an estimation bias as follows:*

$$\textit{discretization bias} \;=\; -\nu = \frac{2}{h}(F-I)(F+I)^{-1} - A = O(h^2) \textit{ as } h \to 0, \tag{3.23}$$

$$\textit{estimation bias} = -\frac{4}{T}(I+F)^{-1}b(I+F)^{-1} - \frac{4}{h}L(I+F)^{-1} + o(T^{-1}). \tag{3.24}$$

As before, it is difficult to determine the signs of the discretization bias and estimation bias in a general multivariate case. However, in the univariate case, the signs are opposite each other as reported in Section 4.2.

Remark 3.4 *The estimator (3.10) is based on a first order Taylor expansion whereas the estimator (3.20) is based on a second order expansion, so it is not surprising that (3.20) has a smaller discretization bias than (3.10). It is not as easy to compare the magnitudes of the two estimation biases. In the univariate case, however, we show in Section 4.2 that the estimator (3.20) has a larger estimation bias than the estimator (3.10).*

4 Relations to Existing Results

4.1 The Euler and Trapezoidal Approximations

The estimators given above include as special cases the two estimators obtained from the Euler approximation and the trapezoidal approximation. Consequently, both the asymptotic and the bias properties are applicable to these two approximation models and the simple framework above unifies some earlier theory on the estimation of approximate discrete time models.

The Euler approximate discrete time model is of the form:

$$X_t - X_{t-1} = AhX_{t-1} + Bh + u_t. \tag{4.1}$$

The OLS estimator of A is given by

$$[\widehat{I+Ah}, \widehat{Bh}] := \left[n^{-1}\sum_{t=1}^{n} X_t Z'_{t-1}\right]\left[n^{-1}\sum_{t=1}^{n} Z_{t-1}Z'_{t-1}\right]^{-1} =: [\hat{F}, \hat{g}]. \tag{4.2}$$

If B is known apriori and assumed zero without loss of generality, then the OLS estimator of A is

$$\widehat{[I+Ah]} = \left[n^{-1}\sum_{t=1}^{n} X_t X_{t-1}^{'}\right]\left[n^{-1}\sum_{t=1}^{n} X_{t-1} X_{t-1}^{'}\right]^{-1} =: [\hat{F}], \tag{4.3}$$

where Z_{t-1}, $\hat{F}$, $\hat{g}$ are defined in the same way as before. Hence,

$$\hat{A} = \frac{1}{h}[\hat{F} - I]. \tag{4.4}$$

This is precisely the estimator given by (3.10) based on a first order expansion of the matrix exponential $\exp(Ah)$ in h.

The trapezoidal approximate discrete time model is of the form

$$X_t - X_{t-1} = \frac{1}{2}Ah(X_t + X_{t-1}) + Bh + \nu_t. \tag{4.5}$$

If $B = 0$, the approximate discrete model becomes

$$X_t - X_{t-1} = \frac{1}{2}Ah(X_t + X_{t-1}) + \nu_t. \tag{4.6}$$

Note that (4.6) is a simultaneous equations model, as emphasized by Bergstrom (1966,1984). We show that the two stage least squares estimator of A from (4.5) is equivalent to the estimator given by (3.20) based on a second order expansion of $\exp(Ah)$ in h. To save space, we focus on the approximate discrete time model with known $B = 0$. The result is easily extended to the case of unknown B.

The two stage least squares estimator of Bergstrom (1984) takes the form

$$\hat{A} = \left[\sum_{t=1}^{n} \frac{1}{h}(X_t - X_{t-1})V_t'\right]\left[\sum_{t=1}^{n} \frac{1}{2}(X_t + X_{t-1})V_t'\right]^{-1}, \tag{4.7}$$

where

$$V_t = \frac{1}{2}(X_t^* + X_{t-1}), \tag{4.8}$$

$$X_t^* = \left[\sum_{t=1}^{n} X_t X_{t-1}'\right]\left[\sum_{t=1}^{n} X_{t-1} X_{t-1}'\right]^{-1} X_{t-1}. \tag{4.9}$$

Theorem 4.1 *The two stage least squares estimator suggested in Bergstrom (1984) has the following form*

$$\hat{A} = \frac{2}{h}[\hat{F} - I][\hat{F} + I]^{-1}, \tag{4.10}$$

and is precisely the same estimator as that given by (3.20) based on a second order expansion of $\exp(Ah)$ *in* h.

4.2 Bias in univariate models

The univariate diffusion model considered in this section is the OU process:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma dW(t), \; X(0) = 0, \tag{4.11}$$

where $W(t)$ is a standard scalar Brownian motion. The exact discrete time model corresponding to (4.11) is

$$X_t = \phi X_{t-1} + \mu(1 - e^{-\kappa h}) + \sigma\sqrt{\frac{1 - e^{-2\kappa h}}{2\kappa}}\epsilon_t, \tag{4.12}$$

where $\phi = e^{-\kappa h}$, $\epsilon_t \sim iid\, N(0,1)$ and h is the sampling interval.

The ML estimator of κ (conditional on X_0) is given by

$$\hat{\kappa} = -\ln(\hat{\phi})/h, \tag{4.13}$$

where

$$\hat{\phi} = \frac{n^{-1}\Sigma X_t X_{t-1} - n^{-2}\Sigma X_t \Sigma X_{t-1}}{n^{-1}\Sigma X_t^2 - n^{-2}(\Sigma X_{t-1})^2}, \tag{4.14}$$

and $\hat{\kappa}$ exists provided $\hat{\phi} > 0$. Tang and Chen (2009) analyzed the asymptotic properties and derived the finite sample variance formula and the bias formula, respectively,

$$Var(\hat{\kappa}) = \frac{1 - \phi^2}{Th\phi^2} + o(T^{-1}), \tag{4.15}$$

$$E(\hat{\kappa}) - \kappa = \frac{1}{T}\left(\frac{5}{2} + e^{\kappa h} + \frac{e^{2\kappa h}}{2}\right) + o(T^{-1}). \tag{4.16}$$

When μ is known (assumed to be 0), the exact discrete model becomes

$$X_t = \phi X_{t-1} + \delta\sqrt{\frac{1 - e^{-2\kappa h}}{2\kappa}}\epsilon_t, \tag{4.17}$$

and the ML estimator of κ is $\hat{\kappa} = -\ln(\hat{\phi})/h$, where $\hat{\phi} = \Sigma X_t X_{t-1}/\Sigma X_{t-1}^2$. In this case, Yu (2009) derived the following bias formula under stationary initial conditions

$$E(\hat{\kappa}) - \kappa = \frac{1}{2T}(3 + e^{2\kappa h}) - \frac{2(1 - e^{-2n\kappa h})}{Tn(1 - e^{-2\kappa h})} + o(T^{-1}). \tag{4.18}$$

When the initial condition is $X(0) = 0$, the bias formula becomes

$$E(\hat{\kappa}) - \kappa = \frac{1}{2T}(3 + e^{2\kappa h}) + o(T^{-1}). \tag{4.19}$$

Since the MLE is based on the exact discrete time model, there is no discretization bias in (4.12) and (4.17). The bias in $\hat{\kappa}$ is induced entirely by estimation and is always positive.

We may link our results for multivariate systems to the univariate model. For example, $\kappa = -A$ in (4.11) and the first order Taylor series expansion (i.e., the Euler method) gives the estimator

$$\widehat{\kappa}_1 = \frac{1}{h}[1 - \hat{\phi}]. \tag{4.20}$$

In this case the results obtained in Theorems 3.1 and 3.2 may be simplified as in the following two results.

Theorem 4.2 *Assuming $\kappa > 0$, when h is fixed, and $n \to \infty$, we have*

$$\hat{\kappa}_1 - \kappa \xrightarrow{p} -\frac{\exp(-\kappa h) - 1 + \kappa h}{h} = O(h) \text{ as } h \to 0, \tag{4.21}$$

and

$$h\sqrt{n}\left[\hat{\kappa}_1 - \frac{1-\exp(-\kappa h)}{h}\right] \xrightarrow{d} N(0, 1-\exp(-2\kappa h)). \tag{4.22}$$

For the OU process with an unknown mean,

$$BIAS(\hat{\kappa}_1) = -\frac{H}{h} + \frac{1+3\exp(-\kappa h)}{T} + o(T^{-1}), \tag{4.23}$$

For the OU process with a known mean,

$$BIAS(\hat{\kappa}_1) = -\frac{H}{h} + \frac{2\exp(-\kappa h)}{T} + o(T^{-1}), \tag{4.24}$$

where $\frac{1+3\exp(-\kappa h)}{T} + o(T^{-1})$ *and* $\frac{2\exp(-\kappa h)}{T} + o(T^{-1})$ *are the estimation biases in the two models, respectively. In both models, the discretization bias has the following form:*

$$\frac{-H}{h} = -\frac{\exp(-\kappa h) - 1 + \kappa h}{h}. \tag{4.25}$$

Remark 4.1 *From (4.22) the asymptotic variance for* $\hat{\kappa}_1$ *is*

$$AsyVar(\hat{\kappa}_1) = \frac{1-\exp(-2\kappa h)}{Th}. \tag{4.26}$$

Remark 4.2 *The estimation bias is always positive in both models. If* $\kappa h \in (0,3]$ *which is empirically realistic, the discretization bias may be written as*

$$\begin{aligned} \frac{-H}{h} &= -\kappa^2 h \sum_{i=2}^{\infty} \frac{(-\kappa h)^{i-2}}{i!} \\ &= -\kappa^2 h \sum_{j=2,4,\cdots} \frac{(-\kappa h)^{j-2}}{(j+1)!}(j+1-\kappa h) \\ &< 0. \end{aligned} \tag{4.27}$$

This means that the discretization bias has sign opposite to that of the estimation bias.

Remark 4.3 *For the unknown mean model, if* $T < h(1+3\phi)/(\kappa h + \phi - 1)$*, the estimation bias is larger than the discretization bias in magnitude because this condition is equivalent to*

$$\frac{1+3\exp(-\kappa h)}{T} > \frac{\kappa h + \exp(-\kappa h) - 1}{h}.$$

Further

$$\begin{aligned} h(1+3\phi)/(\kappa h + \phi - 1) &= \frac{h(1+3(1-\kappa h + O(h^2)))}{\frac{1}{2}\kappa^2 h^2 - \frac{1}{6}\kappa^3 h^3 + O(h^4)} \\ &= \frac{2}{\kappa^2 h}(4 - 3\kappa h + O(h^2)))\left(1 - \frac{1}{3}\kappa h + O(h^2)\right)^{-1} \\ &= \frac{2}{\kappa^2 h}(4 - 3\kappa h + O(h^2)))\left(1 + \frac{1}{3}\kappa h + O(h^2)\right) \\ &= \frac{8}{\kappa^2 h}\left(1 + O(h)\right). \end{aligned}$$

In empirically relevant cases, $8/(\kappa^2 h)$ *is likely to take very large values, thereby requiring very large values of* T *before the estimation bias is smaller than the discretization bias. For example, if* $\kappa = 0.1$ *and* $h = 1/12$, $T > 9,600$ *years are needed for the bias to be smaller. The corresponding result for the known mean case is* $2h\phi/(\kappa h + \phi - 1) = \left(4/(\kappa^2 h)\right)(1+O(h))$ *and again large values of* T *are required to reduce the relative magnitude of the estimation bias.*

Similarly, the second order expansion (i.e. the trapezoidal method) gives the estimator

$$\hat{\kappa}_2 = -\hat{A} = -\frac{2}{h}[\hat{F} - I][\hat{F} + I]^{-1} = \frac{2(1-\hat{\phi})}{h(1+\hat{\phi})}, \tag{4.28}$$

for which we have the following result.

Theorem 4.3 *Assuming $\kappa > 0$, when h is fixed, and $n \to \infty$, we have*

$$\hat{\kappa}_2 - \kappa \xrightarrow{p} \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))} - \kappa = O(h^2) \text{ as } h \to 0, \tag{4.29}$$

and

$$h\sqrt{n}\left[\hat{\kappa}_2 - \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))}\right] \xrightarrow{d} N\left(0, \frac{16(1-\exp(-\kappa h))}{(1+\exp(-\kappa h))^3}\right). \tag{4.30}$$

For the OU process with an unknown mean,

$$BIAS(\hat{\kappa}_2) = \nu + \frac{8}{T(1+\exp(-\kappa h))} + o(T^{-1}). \tag{4.31}$$

For the OU process with a known mean,

$$BIAS(\hat{\kappa}_2) = \nu + \frac{4}{T(1+\exp(-\kappa h))} + o(T^{-1}), \tag{4.32}$$

where $\frac{8}{T(1+\exp(-\kappa h))} + o(T^{-1})$ and $\frac{4}{T(1+\exp(-\kappa h))} + o(T^{-1})$ are the two estimation biases. In both models, the discretization bias has the form

$$\nu = -\kappa + \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))} = O(h^2). \tag{4.33}$$

Remark 4.4 *From (4.30) the asymptotic variance for $\hat{\kappa}_2$ is*

$$AsyVar(\hat{\kappa}_2) = \frac{16(1-\exp(-\kappa h))}{Th(1+\exp(-\kappa h))^3}. \tag{4.34}$$

Remark 4.5 *The estimation bias is always positive in both models. If $\kappa h \in (0, 2]$, the discretization bias may be written as*

$$\begin{aligned} \nu = -\kappa + \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))} &= \frac{-\kappa}{1+\exp(-\kappa h)} \sum_{i=3}^{\infty} \frac{(i-2)(-\kappa h)^{i-1}}{i!} \\ = \frac{-\kappa}{1+\exp(-\kappa h)} \sum_{j=3,5,\cdots} \frac{(-\kappa h)^{j-1}}{(j+1)!}&((j-2)(j+1) - \kappa h(j-1)) \\ < 0. \end{aligned} \tag{4.35}$$

Hence, the discretization bias has the opposite sign of the estimation bias.

Remark 4.6 *For the unknown mean model, if $T < 8h/\left(\kappa h(1+\phi) - 2(1-\phi)\right)$, the estimation bias is larger than the discretization bias in magnitude because this condition is equivalent to*

$$\frac{8}{T(1+\exp(-\kappa h))} > \kappa - \frac{2(1-\exp(-\kappa h))}{h(1+\exp(-\kappa h))}.$$

Further

$$\frac{8h}{\kappa h(1+\phi)-2(1-\phi)} = \frac{8h}{\kappa h(2-\kappa h+\frac{1}{2}\kappa^2h^2+O(h^3))-2(\kappa h-\frac{1}{2}\kappa^2h^2+\frac{1}{6}\kappa^3h^3+O(h^4))}$$
$$= 8h\left(\frac{1}{6}\kappa^3h^3+O(h^4)\right)^{-1} = \frac{48}{\kappa^3h^2}\left(1+O(h)\right)^{-1}$$
$$= \frac{48}{\kappa^3h^2}\left(1+O(h)\right).$$

Again, in empirically relevant cases, $48/(\kappa^3h^2)$ is likely to take very large values thereby requiring very large values of T before the estimation bias is smaller than the discretization bias. For example, if $\kappa = 0.1$ and $h = 1/12$, $T > 6,912,000$ years are needed for the bias to be smaller. Hence the estimation bias is inevitably much larger than the discretization bias in magnitude for all realistic sample spans T.

Remark 4.7 *It has been argued in the literature that ML should be used whenever it is available and the likelihood function should be accurately approximated when it is not available analytically; see Durham and Gallant (2002) and Aït-Sahalia (2002) for various techniques to accurately approximate the likelihood function. From the results in Theorems 4.2 and 4.3 we can show that the total bias of the MLE based on the exact discrete time model is bigger than that based on the Euler and the trapezoidal approximation. For example, for the estimator based on the trapezoidal approximation, considering $\nu = O(h^2)$ as $h \to 0$, when the model is the OU process with an unknown mean,*

$$\begin{aligned}
|BIAS(\hat{\kappa}_{ML})| - |BIAS(\hat{\kappa}_2)| &= \frac{5+2e^{\kappa h}+e^{2\kappa h}}{2T} - \left|\frac{8}{T(1+e^{-\kappa h})}+v\right| + o(T^{-1}) \\
&= \frac{5+2e^{\kappa h}+e^{2\kappa h}}{2T} - \frac{8}{T(1+e^{-\kappa h})} - v + o(T^{-1}) \\
&= \frac{(1-\phi)^2(1+5\phi)}{2T\phi^2(1+\phi)} - v + o(T^{-1}) \qquad (4.36) \\
&> 0.
\end{aligned}$$

Using the same method, it is easy to prove the result still holds for the OU process with an known mean. Similarly, one may show that

$$|BIAS(\hat{\kappa}_{ML})| - |BIAS(\hat{\kappa}_1)| > 0,$$

in both models.

Remark 4.8 *The two approximate estimators reduce the total bias over the exact ML and also the asymptotic variance when $\kappa > 0$. This is because*

$$AsyVar(\hat{\kappa}_{ML}) - AsyVar(\hat{\kappa}_1) = \frac{1-\phi^2}{Th\phi^2} - \frac{1-\phi^2}{Th} > 0. \qquad (4.37)$$

and

$$AsyVar(\hat{\kappa}_{ML}) - AsyVar(\hat{\kappa}_2) = \frac{1-\phi^2}{Th\phi^2} - \frac{16(1-\phi)}{Th(1+\phi)^3} \qquad (4.38)$$
$$= \frac{(1-\phi)^3}{Th\phi^2}\frac{\left(\phi^2+6\phi+1\right)}{(1+\phi)^3} > 0. \qquad (4.39)$$

In consequence, the two approximate methods are preferred to the exact ML for estimating the mean reversion parameter in the univariate setting. Of course, the two approximate methods do NOT improve the asymptotic efficiency of the MLE. This is because the asymptotic variance of the MLE is based on large T asymptotics whereas the asymptotic variance of $\hat{\kappa}_1$ and $\hat{\kappa}_2$ is based on large n asymptotics and the two approximate estimators are inconsistent with fixed h. Nevertheless, equations (4.22) and (4.30) seem to indicate that in finite (perhaps very large finite) samples, the inconsistent estimators may lead to smaller variances than the MLE, which will be verified by simulations.

Remark 4.9 *Comparing Theorem 4.2 and Theorem 4.3, it is easy to see the estimator (4.28) based on the trapezoidal approximation leads to a smaller discretization bias than the estimator (4.20) based on the Euler approximation. However, when $\kappa h > 0$ and hence $\phi = e^{-\kappa h} \in (0,1)$, the gain in the discretization error is earnt at the expense of an increase in the estimation error. For the OU process with an unknown mean,*

$$\begin{aligned} \textit{estimation bias } (\hat{\kappa}_2) - \textit{estimation bias } (\hat{\kappa}_1) &= \frac{8}{T(1+e^{-\kappa h})} - \frac{1+3e^{-\kappa h}}{T} + o(T^{-1}) \\ &= \frac{(1-\phi)(7+3\phi)}{T(1+\phi)} + o(T^{-1}) > 0. \end{aligned} \tag{4.40}$$

Similarly, for the OU process with a known mean,

$$\begin{aligned} \textit{estimation bias } (\hat{\kappa}_2) - \textit{estimation bias } (\hat{\kappa}_1) &= \frac{4}{T(1+e^{-\kappa h})} - \frac{2e^{-\kappa h}}{T} + o(T^{-1}) \\ &= \frac{(1-\phi)(4+2\phi)}{T(1+\phi)} + o(T^{-1}) > 0. \end{aligned} \tag{4.41}$$

Since the sign of the discretization bias is opposite to that of the estimation bias, and the trapezoidal rule makes the discretization bias closer to zero than the Euler approximation, we have the following result in both models.

$$|BIAS(\hat{\kappa}_2)| - |BIAS(\hat{\kappa}_1)| > 0.$$

Remark 4.10 *The estimator based on the Euler method leads not only to a smaller bias but also to a smaller variance than that based on the trapezoidal method when $\kappa > 0$. This is because*

$$\begin{aligned} AsyVar(\hat{\kappa}_2) - AsyVar(\hat{\kappa}_1) &= \frac{16(1-\phi)}{Th(1+\phi)^3} - \frac{1-\phi^2}{Th} \\ &= \frac{(1-\phi)^2(3+\phi)[4+(1+\phi)^2]}{Th(1+\phi)^3} > 0. \end{aligned} \tag{4.42}$$

In consequence, the Euler method is preferred to the trapezoidal method and exact ML for estimating the mean reversion parameter in the univariate setting.

5 Bias in General Univariate Models

5.1 Univariate square root model

The square root model, also known as the Cox, Ingersoll and Ross (1985, CIR hereafter) model, is of the form

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\sqrt{X(t)}dW(t). \tag{5.1}$$

If $2\kappa\mu/\sigma^2 > 1$, Feller (1951) showed that the process is stationary, the transitional distribution of cX_t given X_{t-1} is non-central $\chi^2_\nu(\lambda)$ with the degree of freedom $\nu = 2\kappa\mu\sigma^{-2}$ and the non-central component $\lambda = cX_{t-1}e^{-\kappa h}$, where $c = 4\kappa\sigma^{-2}(1-e^{-\kappa h})^{-1}$. Since the non-central χ^2-density function is an infinite series involving the central χ^2 densities, the explicit expression of the MLE for $\theta = (\kappa, \mu, \sigma)$ is not attainable.

To obtain a closed-form expression for the estimator of θ, we follow Tang and Chen (2009) by using the estimator of Nowman. The Nowman discrete time representation of the square root model is

$$X_t = \phi_1 X_{t-1} + (1-\phi_1)\mu + \sigma\sqrt{X_{t-1}\frac{1-\phi_1^2}{2\kappa}}\epsilon_t, \tag{5.2}$$

where $\phi_1 = e^{-\kappa h}$, $\epsilon_t \sim iid\ N(0,1)$ and h is the sampling interval. Hence, Nowman's estimator of κ is

$$\hat{\kappa}_{Nowman} = -\frac{1}{h}\ln(\hat{\phi}_1), \tag{5.3}$$

where

$$\hat{\phi}_1 = \frac{n^{-2}\sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1}^{-1} - n^{-1}\sum_{t=1}^n X_t X_{t-1}^{-1}}{n^{-2}\sum_{t=1}^n X_{t-1}\sum_{t=1}^n X_{t-1}^{-1} - 1}. \tag{5.4}$$

For the stationary square root process, Tang and Chen (2009) derived explicit expressions to approximate $E(\hat{\phi}_1 - \phi_1)$ and $Var(\hat{\phi}_1)$. Using the following relations,

$$E(\hat{\kappa}_{Nowman} - \kappa) = -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) - \frac{1}{2\phi_1^2}E(\hat{\phi}_1 - \phi_1)^2 + O(n^{-3/2})\right], \tag{5.5}$$

and

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}[Var(\hat{\phi}_1) + O(n^{-2})], \tag{5.6}$$

they further obtained the approximations to $E(\hat{\kappa}_{Nowman} - \kappa)$ and $Var(\hat{\kappa}_{Nowman})$. With a fixed h and $n \to \infty$ they derived the asymptotic distribution of $\sqrt{n}(\hat{\kappa}_{Nowman} - \kappa)$. The fact that the mean of the asymptotic distribution is zero implies that the Nowman method causes no discretization bias for estimating κ.

The estimator of κ based on the Euler approximation also has a closed form expression under the square root model. The Euler discrete time model is

$$X_t = \phi_2 X_{t-1} + (1-\phi_2)\mu + \sigma\sqrt{X_{t-1}h}\epsilon_t, \tag{5.7}$$

where $\phi_2 = (1-\kappa h)$. Hence, the Euler scheme estimator of κ is

$$\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_2 - 1), \tag{5.8}$$

where

$$\hat{\phi}_2 = \frac{n^{-2}\sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1}^{-1} - n^{-1}\sum_{t=1}^n X_t X_{t-1}^{-1}}{n^{-2}\sum_{t=1}^n X_{t-1}\sum_{t=1}^n X_{t-1}^{-1} - 1}. \tag{5.9}$$

Obviously $\hat{\phi}_2 = \hat{\phi}_1$. Hence, $\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_1 - 1)$. Considering $\phi_1 = e^{-\kappa h} = 1 - \kappa h + \sum_{i=2}^\infty (-\kappa h)^i/i!$, the finite sample bias for $\hat{\kappa}_{Euler}$ can be expressed as

$$E(\hat{\kappa}_{Euler} - \kappa) = -\frac{1}{h}E(\hat{\phi}_1 - \phi_1) - \frac{1}{h}H, \tag{5.10}$$

where

$$-\frac{1}{h}H = -\frac{1}{h}\sum_{i=2}^{\infty}(-\kappa h)^i/i! = O(h), \quad \text{as} \quad h \to 0, \tag{5.11}$$

which is the discretization bias caused by discretizing the dirft function. Since the asymptotic mean of $\sqrt{n}(\hat{\phi}_1 - \phi_1)$ and hence the asymptotic mean of $\sqrt{n}(\hat{\kappa}_{Euler} - \kappa + \frac{1}{h}H)$ is zero for a fixed h and $n \to \infty$, the Euler discretization of the diffusion function introduces no discretization bias to κ under the square root model.

Furthermore, the finite sample variance for $\hat{\kappa}_{Euler}$ is

$$Var(\hat{\kappa}_{Euler}) = \frac{1}{h^2}Var(\hat{\phi}_1). \tag{5.12}$$

If $\kappa > 0$, $\phi_1 = e^{-\kappa h} < 1$. When h is fixed, we have

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}\left[Var(\hat{\phi}_1) + O(n^{-2})\right] > \frac{1}{h^2}Var(\hat{\phi}_1) = Var(\hat{\kappa}_{Euler}), \tag{5.13}$$

leading to

$$\frac{Var(\hat{\kappa}_{Euler})}{Var(\hat{\kappa}_{Nowman})} = \phi_1^2 + O(n^{-1}) < 1. \tag{5.14}$$

According to (5.14), the Euler scheme always gains over Nowman's method in terms of variance. The smaller is ϕ_1, the larger the gain.

Tang and Chen (2009) obtained a bias formula of $E(\hat{\phi}_1 - \phi_1)$ for the Nowman estimator under the square root model. Unfortunately, the expression is too complex to be used to determine the sign of the bias analytically. However, the simulation results reported in the literature (Phillips and Yu, 2009, for example) and in our own simulations reported in Section 6 suggest that $E(\hat{\kappa}_{Euler} - \kappa) > 0$. Since $H > 0$, (5.10) implies that

$$E(\hat{\phi}_1 - \phi_1) < 0,$$

and the estimation bias $-\frac{1}{h}E(\hat{\phi}_1 - \phi_1)$ dominates the discretization bias $-\frac{1}{h}H$ in the Euler approximation. Consequently, the negative discretization bias $-\frac{1}{h}H$ reduces the total bias in the Euler method. Consequently, the bias in $\hat{\kappa}_{Nowman}$ is larger than that in $\hat{\kappa}_{Euler}$ because

$$\begin{aligned} E(\hat{\kappa}_{Nowman} - \kappa) &= -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) - \frac{1}{2\phi_1^2}E(\hat{\phi}_1 - \phi_1)^2 + O(n^{-3/2})\right] \\ &\geq -\frac{1}{h}\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) \\ &\geq -\frac{1}{h}E(\hat{\phi}_1 - \phi_1) - \frac{1}{h}H = E(\hat{\kappa}_{Euler} - \kappa). \end{aligned} \tag{5.15}$$

The Milstein scheme is another popular approximation approach. For the square root model, the discrete time model obtained by the Milstein scheme is given by

$$X_t = X_{t-1} + \kappa(\mu - X_{t-1})h + \sigma\sqrt{X_{t-1}h}\epsilon_t + \frac{1}{4}\sigma^2 h\left[\epsilon_t^2 - 1\right]. \tag{5.16}$$

Let $a = \sigma\sqrt{X_{t-1}h}$, $b = \frac{1}{4}\sigma^2 h$, $Y_t = X_t - X_{t-1} - \kappa(\mu - X_{t-1})h + \frac{1}{4}\sigma^2 h$, then Equation (5.16) can be represented by

$$Y_t = a\epsilon_t + b\epsilon_t^2 = b\left[\left(\epsilon_t + \frac{a}{2b}\right)^2 - \frac{a^2}{4b^2}\right]. \tag{5.17}$$

Since $\epsilon_t \sim iid\, N(0,1)$, $Z = \left(\epsilon_t + \frac{a}{2b}\right)^2$ follows a noncentral χ^2 distribution with 1 degree of freedom and noncentrality parameter $\lambda = \frac{a^2}{4b^2}$. Elerian (1998) showed that the density of Z may be expressed as

$$f(z) = \frac{1}{2}\exp\left\{-\frac{\lambda+z}{2}\right\}\left(\frac{z}{\lambda}\right)^{-1/4} I_{-1/2}\left(\sqrt{\lambda z}\right), \tag{5.18}$$

where

$$I_{-1/2}(x) = \sqrt{\frac{2}{x}}\sum_{i=0}^{\infty}\frac{(x/2)^{2i}}{i!\Gamma(j+0.5)} = \sqrt{\frac{1}{2\pi x}}\{\exp(x)+\exp(-x)\}.$$

This expression may be used to compute the log-likelihood function of the approximate model (5.16). Unfortunately, the ML estimator does not have a closed form expression and it is therefore difficult to examine the relative performance of the bias and the variance using analytic methods. The performance of the Milstein scheme is therefore compared to other methods in simulations.

5.2 Diffusions with linear drift

We consider the following general diffusion process with a linear drift

$$dX(t) = \kappa(\mu - X(t))dt + \sigma q(X(t);\psi)dW(t), \tag{5.19}$$

as a generalization to the Vasicek and the square root models, where $\sigma q(X(t);\psi)$ is a general diffusion function with parameters ψ, and $\theta = (\kappa, \mu, \sigma, \psi) \in R^d$ is the unknown parameter vector. This model include the well known Constant Elasticity of Variance (CEV) model, such as the Chan, et al (1992, CKLS) model, as a special case. In this general case, the transitional density is not analytically available.

The Nowman approximate discrete model is

$$X_t = \phi_1 X_{t-1} + (1-\phi_1)\mu + \sigma q(X_{t-1};\psi)\sqrt{\frac{1-\phi_1^2}{2\kappa}}\epsilon_t, \tag{5.20}$$

The Euler approximate discrete model is

$$X_t = \phi_2 X_{t-1} + (1-\phi_2)\mu + \sigma q(X_{t-1};\psi)\sqrt{h}\epsilon_t. \tag{5.21}$$

Theorem 5.1 *For Model (5.19), the MLE of κ based on the Nowman approximation is*

$$\hat{\kappa}_{Nowman} = -\frac{1}{h}\ln(\hat{\phi}_1), \tag{5.22}$$

where $\hat{\phi}_1$ is the ML estimator for ϕ_1 in (5.20). The MLE of κ based on the Euler approximation is

$$\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_2 - 1), \tag{5.23}$$

where $\hat{\phi}_2$ is the ML estimator for ϕ_2 in (5.21). Then we have

$$\hat{\phi}_2 = \hat{\phi}_1. \tag{5.24}$$

Remark 5.1 *The ML estimator of ϕ_1 does not have a closed-form expression. Neither does the ML estimator of ϕ_2. So numerical calculations are needed for comparisons. However, according to Theorem 5.1, even without a closed-form solution, we can still establish the equivalence of $\hat{\phi}_1$ and $\hat{\phi}_2$. After $\hat{\phi}_1$ and $\hat{\phi}_2$ are found numerically, one may find the estimators of κ by using the relations $\hat{\kappa}_{Nowman} = -\frac{1}{h}\ln(\hat{\phi}_1)$ and $\hat{\kappa}_{Euler} = -\frac{1}{h}(\hat{\phi}_2 - 1)$.*

To compare the magnitude of the bias in $\hat{\kappa}_{Nowman}$ to that of $\hat{\kappa}_{Euler}$, no general analytic result is available. However, under some mild conditions, comparison is possible. In particular, we make the following three assumptions. Assumption 1: $\hat{\phi}_1 - \phi_1 \sim O_p(n^{-1/2})$; Assumption 2: $E(\hat{\phi}_1 - \phi_1) < 0$; Assumption 3: $-\frac{1}{h}E(\hat{\phi}_1 - \phi_1) > -\frac{1}{h}H$, i.e., the estimation bias dominates the discretization bias in Euler approximation. Under Assumption 1, we get

$$E(\hat{\kappa}_{Nowman} - \kappa) = -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) - \frac{1}{2\phi_1^2}E(\hat{\phi}_1 - \phi_1)^2 + O(n^{-3/2})\right], \tag{5.25}$$

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}[Var(\hat{\phi}_1) + O(n^{-2})], \tag{5.26}$$

$$E(\hat{\kappa}_{Euler} - \kappa) = -\frac{1}{h}E(\hat{\phi}_1 - \phi_1) - \frac{1}{h}H, \tag{5.27}$$

and

$$Var(\hat{\kappa}_{Euler}) = \frac{1}{h^2}Var(\hat{\phi}_1), \tag{5.28}$$

where $H = \sum_{i=2}^{\infty}(-\kappa h)^i/i! = O(h^2)$.

If $\kappa > 0$, $\hat{\kappa}_{Euler}$ has a smaller finite sample variance than $\hat{\kappa}_{Nowman}$ because

$$Var(\hat{\kappa}_{Nowman}) = \frac{1}{h^2\phi_1^2}\left[Var(\hat{\phi}_1) + O(n^{-2})\right] \geq \frac{1}{h^2}Var(\hat{\phi}_1) = Var(\hat{\kappa}_{Euler}). \tag{5.29}$$

Under Assumptions 1, 2, 3, $\hat{\kappa}_{Euler}$ has a smaller bias than $\hat{\kappa}_{Nowman}$ because

$$\begin{aligned} E(\hat{\kappa}_{Nowman} - \kappa) &= -\frac{1}{h}\left[\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) - \frac{1}{2\phi_1^2}E(\hat{\phi}_1 - \phi_1)^2 + O(n^{-3/2})\right] \\ &\geq -\frac{1}{h}\frac{1}{\phi_1}E(\hat{\phi}_1 - \phi_1) \\ &\geq -\frac{1}{h}E(\hat{\phi}_1 - \phi_1) - \frac{1}{h}H = E(\hat{\kappa}_{Euler} - \kappa). \end{aligned} \tag{5.30}$$

6 Simulation Studies

6.1 Linear models

To examine the performance of the proposed bias formulae and to compare the two alternative approximation scheme in multivariate diffusions, we estimate $\kappa = -A$ in the bivariate model with a known mean:

$$dX_t = AX_t dt + \Sigma dW_t, \quad X_0 = 0, \tag{6.1}$$

where W_t is the standard bivariate Brownian motion whose components are independent, and

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}, \kappa = -A = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}.$$

Since A is triangular, the parameters are all identified. While keeping other parameters fixed, we let κ_{22} take various values over the interval (0,3], which covers empirically reasonable values of κ_{22} that apply for data on interest rates and volatilities. The mean reversion matrix is estimated with 10 years of monthly data. The experiment is replicated 10,000 times. Both the actual total bias and the actual standard deviation are computed across 10,000 replications. The actual total bias is split into two parts — discretization bias and estimation bias — as follows. The estimation bias is calculated as H/h and $-v$ as in (3.13) and (3.21) for the two approximate methods. The estimation bias is calculated as:

$$\text{estimation bias} = \text{actual total bias - discretization bias}$$

Figure 1 plots the biases of the estimate of each element in the mean reversion matrix κ, based on the Euler method, as a function of the true value of κ_{22}. Four biases are plotted, the actual total bias, the approximate total bias given by the formula in (3.13), the discretization bias H/h as in (3.13), and the estimation bias.

Several features are apparent in the figure. First, the actual total bias in all cases is large, especially when the true value of κ_{22} is small. Second, except for κ_{12} whose discretization bias is zero, the sign of the discretization bias for the other parameters is opposite to that of the estimation bias. Not surprisingly, in these cases, the actual total bias of estimator (3.10) is smaller than the estimation bias. The discretization bias for κ_{12} is zero because it is assumed that the true value is zero. In the bivariate set-up, however, it is possible that the sign of the discretization bias for the other parameters is the same as that of the estimation bias (for example when $\kappa_{12} = 5$ and $\kappa_{21} = -0.5$). Third, the bias in all parameters is sensitive to the true value of κ_{22}. Finally, the bias formula (3.13) generally works well in all cases.

Figure 2 plots the biases of the estimate of each element in the mean reversion matrix κ, based on the trapezoidal method, as a function of the true value of κ_{22}. Four biases are plotted, the actual total bias, the approximate total bias given by the formula in (3.21), the discretization bias $-\nu$ as in (3.21), and the estimation bias. In all cases, the discretization bias is closer to zero than that based on the Euler approximation. This suggests that the trapezoidal method indeed reduces the discretization bias. Moreover, the bias formula (3.21) generally works well in all cases.

The performance of the two approximation methods is compared in Figure 3, where the actual total bias of the estimators given by (3.10) and (3.20) is plotted. It seems that the bias of the estimator obtained from the trapezoidal approximation is larger than that from the Euler approximation for all parameters except κ_{12}. For κ_{12}, the performance of the two methods are very close with the Euler method being slightly worse when κ_{22} is large.

Figure 4 plots the actual standard deviations for the two approximate estimators, (3.10) and (3.20) as a function of κ_{22}. We notice that, for all the parameters, the standard deviation of the Euler method is smaller than that of the trapezoidal method. The percentage difference can be as high as 20%.

We also design an experiment to check the performance of the alternative estimators in the univariate case. Data are simulated from the univariate OU process with a known mean

$$dX(t) = -\kappa X(t)dt + \sigma dW(t), \quad X(0) = 0. \tag{6.2}$$

Figure 5 reports the bias in $\widehat{\kappa}$ obtained from the Euler method and the trapezoidal method in the OU process with a known mean. Three biases are plotted: the actual total bias, the estimation bias and the discretization bias. Figure 6 compares the bias in $\widehat{\kappa}$ obtained from the exact ML methods with that of the two approximate methods. Several conclusions may be drawn from these two Figures. First, our bias formula provides a good approximation to the actual total bias. Second, for the two approximate estimators, (4.20) and (4.28), the sign of the discretization bias is opposite to that of the estimation bias. Third, while the trapezoidal method leads to a smaller discretization bias than the Euler method, it has a larger estimation bias. Finally, the actual total bias for the Euler method is smaller than that of the trapezoidal method and both methods lead to a smaller total bias than the exact ML estimator (4.13).

Figure 7 reports the standard deviations for estimators (4.13), (4.20) and (4.28). It is easy to find that the standard deviations of estimator (4.20) is the smallest among those of all estimators. The standard deviations of estimator (4.28) are almost the same with those from the exact ML estimator (4.13), but smaller when κ is bigger than 1. Considering the sample size is 120, we can roughly say that, focusing on bias and standard deviation, the estimator (4.20) from the Euler approximation is better than the other estimators in comparatively small sample sizes.

6.2 Square root model

For the square root model, we designed an experiment to compare the performance of the various estimation methods, including the exact ML, the Euler scheme, the Nowman scheme and the Milstein scheme. In all cases we fix $h = 1/12$, $T = 120$, $\mu = 0.05$, $\sigma = 0.05$, but vary the value of κ from 0.05 to 0.5. These settings correspond to 10 years of monthly data in the estimation of κ. The experiment is replicated 10,000 times.

Table 1 reports the bias, the standard error (Std err), and the root mean square error (RMSE) of κ for all estimation methods, obtained across 10,000 replications. Several conclusions emerge from the table. First, all estimation methods suffer from a serious bias problem. Second, the Euler scheme performs best both in terms of bias and variance. Third, the ratios of the standard error of $\widehat{\kappa}_{Euler}$ and that of $\widehat{\kappa}_{Norman}$ are 0.9958, 0.9917, 0.9835, 0.9592 when κ is 0.05, 0.1, 0.2, 0.5, respectively. The ratio decreases as κ increases, as predicted in (5.14). Finally, although the bias for the Milstein method is larger than that for the Euler method, the variances for these two methods are very close.

7 Conclusions

This paper provides a framework for studying the implications of different discretization schemes in estimating the mean reversion parameter in both multivariate and univariate diffusion models with a linear drift function. The approach includes the Euler method and the trapezoidal method as special cases, an asymptotic theory is developed, and finite sample bias comparisons are conducted using analytic approximations. Bias is decomposed into a discretization bias and an estimation bias. It is shown that the discretization bias is of order $O(h)$ for the Euler method and $O(h^2)$ for the trapezoidal method, respectively, whereas the estimation bias is of the order of $O(T^{-1})$. Since in practical applications in finance it is very likely that h is much smaller than $1/T$, estimation bias is likely to dominate discretization bias.

Applying the multivariate theory to univariate models gives several new results. First, it is shown that in the Euler and trapezoidal methods, the sign of the discretization bias is opposite that of the estimation bias for practically realistic cases. Consequently, the bias in the two approximate method is smaller than the ML estimator based on the exact discrete time model. Second, although the trapezoidal method leads to a smaller discretization bias than the Euler method, the estimation bias is bigger. As a result, it is not clear if there is a gain in reducing the total bias by using a higher order approximation. When comparing the estimator based on the Euler method and the exact ML, we find that the asymptotic variance of the former estimator is smaller. As a result, there is clear evidence for preferring the estimator based on the Euler method to the exact ML in the univariate linear diffusion when the mean reversion is slow.

Simulations suggest the bias continues to be large in finite samples. It is also confirmed that for empirically relevant cases, the magnitude of the discretization bias in the two approximate methods is much smaller than that of the estimation bias. The two approximate methods lead to a smaller variance than exact ML. Most importantly for practical work, there is strong evidence that the bias formulae work well and so they can be recommended for analytical bias correction with these models.

For the univariate square root model, the Euler method is found to have smaller bias and smaller variance than the Nowman method. Discretizing the diffusion function both in the Euler method and the Nowman method causes no discretization bias on the mean reversion paramter. For the Euler method, we have derived an explicit expression for the discretization bias caused by discretizing the drift function. The simulation results suggest that the Euler method performs best in terms of both bias and variance.

The analytic and expansion results given in the paper are obtained for stationary systems. Bias analysis for nonstationary and explosive cases require different methods. For diffusion models with constant diffusion functions, it may be possible to extend recent finite sample and asymptotic expansion results for the discrete time AR(1) model (Phillips, 2010) to a continuous time setting. Such an analysis would involve a substantial extension of the present work and deserves treatment in a separate study.

Acknowledgement

Thanks go to the referee and the editor, the seminar participants at the SETA2010 meeting, the 2010 RMI Annual Conference, the 2010 Tsinghua Econometrics Summer Workshop for helpful comments on the original version. Phillips gratefully acknowledges support from the NSF under Grant No. SES 09-56687. Yu gratefully acknowledges support from the Singapore Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS. We would like to thank the comments from the editor and a referee.

Appendix

Proof of Lemma 3.3: Let $C = F^{'}$and then

$$\sum_{t=0}^{\infty} F^{'k} = (I - F^{'})^{-1} = (1 - C), \tag{7.1}$$

$$\begin{aligned} \sum_{k=0}^{\infty} F^{'k} tr(F^{k+1}) &= \sum_{k=0}^{\infty} F^{'k} \sum_{\lambda \in spec(F)} \lambda^{k+1} = \sum_{\lambda \in spec(F)} [\lambda \sum_{k=0}^{\infty} \lambda^k F^{'k}] \\ &= \sum_{\lambda \in spec(C)} [\lambda \sum_{k=0}^{\infty} \lambda^k C^k] = \sum_{\lambda \in spec(C)} \lambda (I - \lambda C)^{-1}, \end{aligned} \tag{7.2}$$

where $Spec(C)$ denotes the set of eigenvalues of C. Thus,

$$\sum_{k=0}^{\infty} F^{'2k+1} = \sum_{k=0}^{\infty} C^{2k+1} = C(I - C^2)^{-1}, \tag{7.3}$$

$$\Gamma(0) = Var(x_t) = \sum_{i=0}^{\infty} F^i \cdot G \cdot F^{'i} = D, \tag{7.4}$$

$$B_n = BIAS(\hat{F}) = E(\hat{F}) - F = -\frac{b}{n} + O(n^{-\frac{3}{2}}). \tag{7.5}$$

Proof of Theorem 3.1: By Lemma 3.1, for fixed h, as $n \to \infty$, $\hat{F} \xrightarrow{p} F$. Hence,

$$\hat{A} - A = \frac{1}{h}[\hat{F} - F] + \frac{1}{h}H \xrightarrow{p} \frac{1}{h}H.$$

From Equations (3.8), $\frac{1}{h}H = \frac{1}{h}[F - I - Ah] = O(h)$ as $h \to 0$, proving the first part.
(b) According to Lemma 3.1, fixed h, as $n \to \infty$,

$$\sqrt{n}\{Vec(\hat{F}) - Vec(F)\} \xrightarrow{d} N(0, (\Gamma(0))^{-1} \otimes G),$$

$$\begin{aligned} \sqrt{n} h Vec[\hat{A} - \frac{1}{h}(F - I)] &= \sqrt{n} Vec[\hat{A}h - (F - I)] \\ &= \sqrt{n} Vec[\hat{F} - F] \xrightarrow{d} N(0, (\Gamma(0))^{-1} \otimes G), \end{aligned}$$

giving the second part.

Proof of Theorem 3.2: According to formulae (3.8), (3.9) and Lemma 3.3,

$$\begin{aligned} E(\hat{A} - A) &= \frac{1}{h} E(\hat{F} - F) + \frac{1}{h}H = \frac{1}{h} E(\frac{-b}{n} + O(n^{-3/2})) + \frac{1}{h}H \\ &= -\frac{b}{T} + \frac{1}{h}H + o(T^{-1}). \end{aligned}$$

Proof of Theorem 3.3a: From formulae (3.19),

$$\begin{aligned}
\hat{A} - A &= \frac{2}{h}(\hat{F} - I)(\hat{F} + I)^{-1} - \frac{2}{h}(F - I)(F + I)^{-1} - \nu \\
&= \frac{2}{h}(\hat{F} + I - 2I)(\hat{F} + I)^{-1} - \frac{2}{h}(F - I)(F + I)^{-1} - \nu \\
&= \frac{2}{h}[I - 2(\hat{F} + I)^{-1}] - \frac{2}{h}[I - 2(F + I)^{-1}] - \nu \\
&= -\frac{4}{h}[(\hat{F} + I)^{-1} - (F + I)^{-1}] - \nu \\
&= \frac{4}{h}(I + F)^{-1}(\hat{F} - F)(I + \hat{F})^{-1} - \nu. \qquad (7.6)
\end{aligned}$$

As h is fixed, according Lemma 3.1, as $n \to \infty$, $\hat{F} \xrightarrow{p} F$, the first part of above equation goes to zero. And from formula (3.19),

$$\hat{A} - A \xrightarrow{p} -\nu = \frac{2}{h}(F - I)(F + I)^{-1} - A.$$

Proof of Theorem 3.3b: :

$$\begin{aligned}
Vec(\hat{A} - A + \nu) &= Vec[\hat{A} - \frac{2}{h}(F - I)(F + I)^{-1}] = \frac{4}{h}Vec[(I + F)^{-1}(\hat{F} - F)(I + \hat{F})^{-1}] \\
&= \frac{4}{h}\{(\hat{F}' + I)^{-1} \otimes (F + I)^{-1}\}Vec(\hat{F} - F).
\end{aligned}$$

Again when h is fixed, according to Lemma 3.1, as $n \to \infty$, $\sqrt{n}(\hat{F} - F) \xrightarrow{d} N(0, \Gamma(0)^{-1} \otimes G)$, and we get

$$h\sqrt{n}Vec[\hat{A} - \frac{2}{h}(F - I)(F + I)^{-1}] \xrightarrow{d} N(0, \Psi),$$

where

$$\Psi = 16\Upsilon[\Gamma(0)^{-1} \otimes G]\Upsilon', \quad \Upsilon = (F' + I)^{-1} \otimes (F + I)^{-1}$$

Proof of Theorem 3.4: From the proof of theorem 3.3, we have

$$\begin{aligned}
E[\hat{A}] - A &= -\frac{4}{h}E[(\hat{F} + I)^{-1} - (F + I)^{-1}] - \nu \\
&= -\frac{4}{h}E[(\hat{F} + I)^{-1}] + \frac{4}{h}(F + I)^{-1} - \nu.
\end{aligned}$$

For the first term, we note that

$$\begin{aligned}
(\hat{F} + I)^{-1} &= (I + F + \hat{F} - F)^{-1} = [(I + F)(I + (I + F)^{-1}(\hat{F} - F))]^{-1} \\
&= [I + (I + F)^{-1}(\hat{F} - F)]^{-1}(I + F)^{-1},
\end{aligned}$$

and

$$\begin{aligned}
[I + (I + F)^{-1}(\hat{F} - F)]^{-1} &= \sum_{i=0}^{\infty}(-1)^i[(I + F)^{-1}(\hat{F} - F)]^i \\
&= I - (I + F)^{-1}(\hat{F} - F) + [(I + F)^{-1}(\hat{F} - F)]^2 \\
&+ \sum_{i=3}^{\infty}(-1)^i[(I + F)^{-1}(\hat{F} - F)]^i.
\end{aligned}$$

By Lemma 3.1, we have

$$\sqrt{n}[Vec(\hat{F}) - Vec(F)] \xrightarrow{d} N(0, \Gamma(0)^{-1} \otimes G),$$

and so,

$$\hat{F}_{ij} - F_{ij} = O_P(n^{-\frac{1}{2}}).$$

Then,

$$[(I+F)^{-1}(\hat{F}-F)]^3 = O_p(n^{-\frac{3}{2}}) \text{ and } [(I+F)^{-1}(\hat{F}-F)]^i = o_p(n^{-\frac{3}{2}}),\ i \geq 3,$$

$$[I+(I+F)^{-1}(\hat{F}-F)]^{-1} = I - (I+F)^{-1}(\hat{F}-F) + [(I+F)^{-1}(\hat{F}-F)]^2 + O_p((n^{-\frac{3}{2}})),$$

and

$$\begin{aligned}
E[\hat{A}-A] &= -\frac{4}{h}E\{[I+(I+F)^{-1}(\hat{F}-F)^{-1}]\}(I+F)^{-1} + \frac{4}{h}(F+I)^{-1} + O(h^2) \\
&= \frac{4}{h}E\{(I+F)^{-1}(\hat{F}-F)(I+F)^{-1}\} - \frac{4}{h}E\{[(I+F)^{-1}(\hat{F}-F)]^2(I+F)^{-1}\} \\
&+ \frac{1}{h}O(n^{-\frac{3}{2}}) - \nu.
\end{aligned}$$

Now let $\hat{g} = [(I+F)^{-1}(\hat{F}-F)]$, so that

$$\sqrt{n}\cdot Vec[\hat{g}] = \sqrt{n}\cdot Vec[(I+F)^{-1}(\hat{F}-F)] = [I_M \otimes (I+F)^{-1}]\sqrt{n}Vec(\hat{F}-F) \xrightarrow{d} N(0,\Delta),$$

where $\Delta = [I_M \otimes (I+F)^{-1}]\cdot\Gamma(0)^{-1}\otimes G\cdot[I_M \otimes (I+F)^{-1}]'$. As a result,

$$Var(\sqrt{n}\cdot Vec(\hat{g})) = \Delta + o(1) \rightarrow Var[Vec(\hat{g})] = \frac{\Delta}{n} + o(n^{-1}),$$

and

$$\begin{aligned}
E[Vec(\hat{g})\cdot Vec(\hat{g})^T] &= Var[Vec(\hat{g})] + E[Vec(\hat{g})]\cdot E[Vec(\hat{g})]^T \\
&= \frac{\Delta}{n} + E[Vec(\hat{g})]\cdot E[Vec(\hat{g})]^T + o(n^{-1}).
\end{aligned}$$

From Lemma 3.3,

$$B_n = E(\hat{F}) - F = -\frac{b}{n} + O(n^{-\frac{3}{2}}).$$

When the exact discrete model involves an unknown $B(\theta)$ we have

$$b = G[(I-C)^{-1} + C(I-C^2)^{-1} + \sum_{\lambda\in Spec(C)}\lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1},$$

and when we have a prior knowledge that $B(\theta)=0$ in (2.2), we have

$$b = G[C(I-C^2)^{-1} + \sum_{\lambda\in Spec(C)}\lambda(I-\lambda C)^{-1}]\Gamma(0)^{-1}.$$

Then

$$\begin{aligned}
E[Vec(\hat{g})] &= E[(I_M \otimes (I+F)^{-1})Vec(\hat{F}-F)] \\
&= [I_M \otimes (I+F)^{-1}]E[Vec(\hat{F}-F)] \\
&= [I_M \otimes (I+F)^{-1}]Vec[E(\hat{F}-F)] \\
&= [I_M \otimes (I+F)^{-1}]Vec[-\frac{b}{n} + O(n^{-\frac{3}{2}})] = O(n^{-1}) \\
&\rightarrow E[Vec(\hat{g})Vec(\hat{g})^T] = \frac{\Delta}{n} + o(n^{-1}).
\end{aligned}$$

Here we assume $\hat{W} = [(I+F)^{-1}(\hat{F}-F)]^2 = \hat{g}\hat{g}$ and $\hat{W}_{ij} = \sum_{s=1}^{M} \hat{g}_{is}\hat{g}_{sj}$. It is easy to find that $\hat{g}_{is}$ is the $(M(s-1)+i)$th element of $Vec(\hat{g})$, and $\hat{g}_{is}\hat{g}_{sj}$ is the $(M(s-1)+i, M(j-1)+s)^{th}$ element of $Vec(\hat{g})Vec(\hat{g})$. Defining e_i to be the column vector of dimension M^2 whose i^{th} element is 1 and other elements are 0, we have

$$\begin{aligned} E[\hat{g}_{is}\hat{g}_{sj}] &= e'_{M(s-1)+i} E[Vec(\hat{g})Vec(\hat{g})]'] e_{M(j-1)+s} \\ &= \frac{1}{n} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s} + o(n^{-1}), \end{aligned}$$

$$\begin{aligned} E[\hat{W}_{ij}] &= \sum_{s=1}^{M} E[\hat{g}_{is}\hat{g}_{sj}] \\ &= \sum_{s=1}^{M} \frac{1}{n} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s} + o(n^{-1}). \end{aligned}$$

Next, define the matrix P with (i,j) element

$$P_{ij} = \frac{1}{n} \sum_{s=1}^{M} e'_{M(s-1)+i} \cdot \Delta \cdot e_{M(j-1)+s}.$$

Then

$$E\{[(I+F)^{-1}(\hat{F}-F)]^2\} = E(\hat{W}) = P + o(n^{-1}).$$

Again, using Lemma 3.3, the formula for the estimation bias is

$$\begin{aligned} E[\hat{A} - A] &= \frac{4}{h} E\{(I+F)^{-1}(\hat{F}-F)(I+F)^{-1}\} - \frac{4}{h} E\{[(I+F)^{-1}(\hat{F}-F)^2](I+F)^{-1}\} \\ &\quad + \frac{1}{h} O(n^{-\frac{3}{2}}) - \nu \\ &= \frac{4}{h}(I+F)^{-1}[-\frac{b}{n} + O(n^{-\frac{3}{2}})](I+F)^{-1} \\ &\quad - \frac{4}{h} \cdot W \cdot (I+F)^{-1} + \frac{1}{h} o(n^{-1}) + \frac{1}{h} O(n^{-\frac{3}{2}}) - \nu \\ &= -\frac{4}{T}(I+F)^{-1} \cdot b \cdot (I+F)^{-1} - \frac{4}{h} \cdot W \cdot (I+F)^{-1} - \nu + o(T^{-1}). \end{aligned}$$

Proof of Theorem 4.1: Using (4.8) and (4.9) in (4.7), we have

$$
\begin{aligned}
\sum_{t=1}^{n}\frac{1}{h}(X_t-X_{t-1})V_t' &= \frac{1}{2h}\sum_{t=1}^{n}X_tX_{t-1}' - \frac{1}{2h}\sum_{t=1}^{n}X_{t-1}X_{t-1}' \\
&+\frac{1}{2h}\sum_{t=1}^{n}X_tX_{t-1}'\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1}\sum_{t=1}^{n}X_{t-1}X_t' - \frac{1}{2h}\sum_{t=1}^{n}X_{t-1}X_t' \\
&= \frac{1}{2h}[\left(\sum_{t=1}^{n}X_tX_{t-1}'\right)\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1} - I \\
&+\left(\sum_{t=1}^{n}X_tX_{t-1}'\right)\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n}X_{t-1}X_t'\right)\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1} \\
&-\left(\sum_{t=1}^{n}X_{t-1}X_t'\right)\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1}]\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right) \\
&= \frac{1}{2h}[\hat{F}-I+\hat{F}\left(\sum_{t=1}^{n}X_{t-1}X_t'\right)\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1} \\
&-\left(\sum_{t=1}^{n}X_{t-1}X_t'\right)\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1}]\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right) \\
&= \frac{1}{2h}(\hat{F}-I)\left[I+\sum_{t=1}^{n}X_{t-1}X_t'\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)^{-1}\right]\sum_{t=1}^{n}X_{t-1}X_{t-1}' \\
&= \frac{1}{2h}(\hat{F}-I)\left[\sum_{t=1}^{n}X_{t-1}X_{t-1}'+\sum_{t=1}^{n}X_{t-1}X_t'\right] \\
&= \frac{1}{2h}(\hat{F}-I)\left(\sum_{t=1}^{n}X_{t-1}X_{t-1}'\right)(\hat{F}'+I). \qquad (7.7)
\end{aligned}
$$

By the same method, it is easy to obtain

$$
\left[\sum_{t=1}^{n}\frac{1}{2}(X_t+X_{t-1})V_t'\right]^{-1} = \left[\frac{1}{4}(\hat{F}+I)(\sum_{t=1}^{n}X_{t-1}X_{t-1}')(\hat{F}'+I)\right]^{-1} \qquad (7.8)
$$

Using the above two formulae in (4.7), the two stage least squares estimator is

$$
\hat{A} = \frac{2}{h}(\hat{F}-I)(\hat{F}+I)^{-1}. \qquad (7.9)
$$

Proof of Theorem 5.1: The Nowman approximate discrete time model yields the following transition function

$$
f(X_iX_{(i-1)}) = \frac{[(1-e^{-2\kappa h})/2\kappa]^{-1/2}}{\sqrt{2\pi}\sigma g(X_{i-1};\psi)}\exp\left\{-\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]^2}{2\sigma^2g^2(X_{i-1};\psi)(1-e^{-2\kappa h})/2\kappa}\right\}, \qquad (7.10)
$$

and the following log-likelihood function

$$
\begin{aligned}
\ell(\theta) = -\frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n}\ln[g(X_{i-1};\psi)] - \frac{n}{2}\ln(\frac{1-e^{-2\kappa h}}{2\kappa}) \\
-\sum_{i=1}^{n}\frac{[X_i-\phi_1X_{i-1}-(1-\phi_1)\mu]^2}{2\sigma^2g^2(X_{i-1};\psi)(1-e^{-2\kappa h})/2\kappa}. \qquad (7.11)
\end{aligned}
$$

The first order conditions are

$$\frac{\partial \ell(\theta)}{\partial \mu}=0 \quad \Rightarrow \quad \sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]}{g^2(X_{i-1};\psi)}=0, \tag{7.12}$$

$$\frac{\partial \ell(\theta)}{\partial \sigma^2}=0 \quad \Rightarrow \quad \sigma^2\left(\frac{1-e^{-2\kappa h}}{2\kappa}\right)-\frac{1}{n}\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)}=0, \tag{7.13}$$

$$\begin{aligned}\frac{\partial \ell(\theta)}{\partial \kappa}=0 \quad \Rightarrow \quad 0=&-\frac{n}{2}\left[\frac{2he^{-2\kappa h}}{1-e^{-2\kappa h}}-\frac{1}{\kappa}\right]-he^{-\kappa h}\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu](X_{i-1}-\mu)}{\sigma^2 g^2(X_{i-1};\psi)(1-e^{-2\kappa h})/2\kappa}\\ &-\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]^2}{2\sigma^2 g^2(X_{i-1};\psi)}\left[\frac{2(1-e^{-2\kappa h})-4\kappa h e^{-2\kappa h}}{(1-e^{-2\kappa h})^2}\right].\end{aligned} \tag{7.14}$$

and

$$\begin{aligned}\frac{\partial \ell(\theta)}{\partial \psi_j}=0 \quad \Rightarrow \quad 0=&\,\sigma^2\frac{1-e^{-2\kappa h}}{2\kappa}\sum_{i=1}^{n}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)}\\ &-\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)}.\end{aligned} \tag{7.15}$$

Taking Equation (7.13) into (7.14), the first term and the third term cancel and we obtain

$$\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu](X_{i-1}-\mu)}{g^2(X_{i-1};\psi)}=0. \tag{7.16}$$

Taking Equation (7.13) into (7.15), we have

$$\begin{aligned}0=&\,\frac{1}{n}\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)}\sum_{i=1}^{n}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)}\\ &-\sum_{i=1}^{n}\frac{[X_i-\phi_1 X_{i-1}-(1-\phi_1)\mu]^2}{g^2(X_{i-1};\psi)}\frac{\partial g(X_{i-1};\psi)/\partial\psi_j}{g(X_{i-1};\psi)}.\end{aligned} \tag{7.17}$$

Equations (7.12), (7.16) and (7.17) yield the ML estimators, $\hat{\phi}_1$, $\hat{\mu}$ and $\hat{\psi}$ and Equation (7.13) gives the ML estimator, $\hat{\sigma}^2$.

The Euler approximate discrete model yields the following log-likelihood function,

$$\ell(\theta)=-\frac{n}{2}\ln(\sigma^2)-\sum_{i=1}^{n}\ln[g(X_{i-1};\psi)]-\sum_{i=1}^{n}\frac{[X_i-\phi_2 X_{i-1}-(1-\phi_2)\mu]^2}{2\sigma^2 h g^2(X_{i-1};\psi)}. \tag{7.18}$$

It is easy to obtain the first order conditions, three of which are identical to those in (7.12), (7.16) and (7.17). Hence,

$$\hat{\phi}_2=\hat{\phi}_1. \tag{7.19}$$

References

Aït-Sahalia, Y., 1999, Transition Densities for Interest Rate and Other Non-linear Diffusions. Journal of Finance, 54, 1361–1395.

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

Aït-Sahalia, Y., 2008, Closed-Form Likelihood Expansions for Multivariate Diffusions. Annals of Statistics, 36, 906-937.

Aït-Sahalia, Y., and Kimmel R., 2007, Maximum Likelihood Estimation of Stochastic Volatility Models. Journal of Financial Economics, 83, 413–452.

Aït-Sahalia, Y., and Yu, J., 2006, Saddlepoint Approximations for Continuous-Time Markov Processes. Journal of Econometrics, 134, 507-551.

Bao, Y. and Ullah, A., 2009, On Skewness and Kurtosis of Econometric Estimators, Econometrics Journal, 12, 232-247.

Bergstrom, A. R., 1966, Nonrecursive Models as Discrete Approximations to Systems of Stochastic Differential Equations. Econometrica, 34, 173–182.

Bergstrom, A. R., 1984, Continuous Time Stochastic Models and Issues of Aggregation over Time. Handbook of Econometrics, 1146-1211

Bergstrom, A. R., 1990, Continuous Time Econometric Modelling. Oxford University Press, Oxford.

Black, F., and Scholes, M., 1973, The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81, 654-673.

Chan, K., Karolyi, F, Longstaff, F., and A. Sanders, 1992, An empirical comparison of alternative models of short term interest rates. Journal of Finance 47, 1209–1227.

Cox, J., Ingersoll, J., and Ross, S., 1985, A Theory of the Term Structure of Interest Rates. Econometrica, 53, 385–407.

Durham, G., and Gallant, A. R., 2002, Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes. Journal of Business and Economic Statistics, 20, 297–316.

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

Elerian, O., 1998, A Note on the Existence of a Closed-form Conditional Transition Density for the Milstein Scheme, Economics discussion paper 1998-W18, Nuffield College, Oxford.

Florens-Zmirou, D., 1989, Approximate Discrete-time Schemes for Statistics of Diffusion Processes. Statistics, 20, 547-557.

Feller, W., 1951, Two Singular Diffusion Problems. Annals of Mathematics, 54, 173-182.

Fuller, W. A., 1976, Introduction to Statistical Time Series. 20, New York: Wiley.

Hannan, E. J., 1970, Multiple Time Series. 20, New York: Wiley.

Hansen, L. P., and Sargent, T. J., 1983, The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities. Econometrica, 51, 377-388.

Lo, A. W., 1988, Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data. Econometric Theory, 4, 231-247.

Milstein, G. N., 1978, A Method of Second-Order Accuracy Integration of Stochastic Differential Equations. Theory of Probability and its Applications, 23, 396-401.

Nicholls, D. F., and Pope, A. L., 1988, Bias in the Estimation of Multivariate Autoregressions. Australian Journal of Statistics, 30, 296-309.

Nowman, K. B., 1997, Gaussian Estimation of Single-factor Continuous Time Models of the Term Structure of Interest Rates. Journal of Finance, 52, 1695-1703.

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.

Phillips, P. C. B., 1972, The Structural Estimation of Stochastic Differential Equation Systems. Econometrica, 40, 1021-1041.

Phillips, P. C. B., 1973, The Problem of Identification in Finite Parameter Continuous Time Models. Journal of Econometrics, 1, 351-362.

Phillips, P. C. B., 2010, Folklore theorems, implicit maps, and new unit root limit theory, Working paper, Yale University.

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

Phillips, P. C. B., and Yu, J., 2005b, Comment: A Selective Overview of Nonparametric Methods in Financial Econometrics. Statistical Science, 20, 338-343.

Phillips, P. C. B., and Yu, J., 2009a, Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. Handbook of Financial Time Series, 707-742.

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

Piazzesi, M., 2009, Affine Term Structure Models, in Handbook for Financial Econometrics, edt by Aït-Sahalia, Y. and L. Hansen, North-Holland.

Sargan, J. D., 1974, Some Discrete Approximations to Continuous Time Stochastic Models. Journal of the Royal Statistical Society, Series B, 36, 74-90.

Sundaresan, S.M., 2000, Continuous-Time Models in Finance: A review and an Assessment. Journal of Finance, 55, 1569-1622.

Tang, C. Y., and Chen, S. X., 2009, Parameter Estimation and Bias correction for Diffusion processes. Journal of Econometrics, 149, 65-81.

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

Ullah, A., Wang, Y., and Yu, J., 2010, Bias in the Mean Reversion Estimator in the Continuous Time Gaussian and Levy Processes. Working Paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University.

Yamamoto, T. and Kunitomo, N., 1984, Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models. Annals of the Institute of Statistical Mathematics, 36, 419-430.

Yu, J., 2009, Bias in the Estimation of Mean Reversion Parameter in Continuous Time Models. Working Paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University.

Table 1: Exact and approximate ML estimation of κ from the square root model using 120 monthly observations. The experiment is replicated 10,000 times.

Method	Exact	Euler	Nowman	Milstein
		$\kappa = 0.05$		
Bias	.1156	.1126	.1152	.1132
Std err	.2251	.2205	.2249	.2206
RMSE	.2531	.2476	.2526	.2480
		$\kappa = 0.1$		
Bias	.1392	.1342	.1387	.1350
Std err	.2670	.2590	.2668	.2592
RMSE	.3011	.2917	.3007	.2922
		$\kappa = 0.2$		
Bias	.1615	.1529	.1610	.1538
Std err	.3178	.3070	.3178	.3068
RMSE	.3565	.3430	.3562	.3432
		$\kappa = 0.5$		
Bias	.1869	.1625	.1862	.1639
Std err	.4210	.3999	.4209	.3993
RMSE	.4607	.4317	.4603	.4316

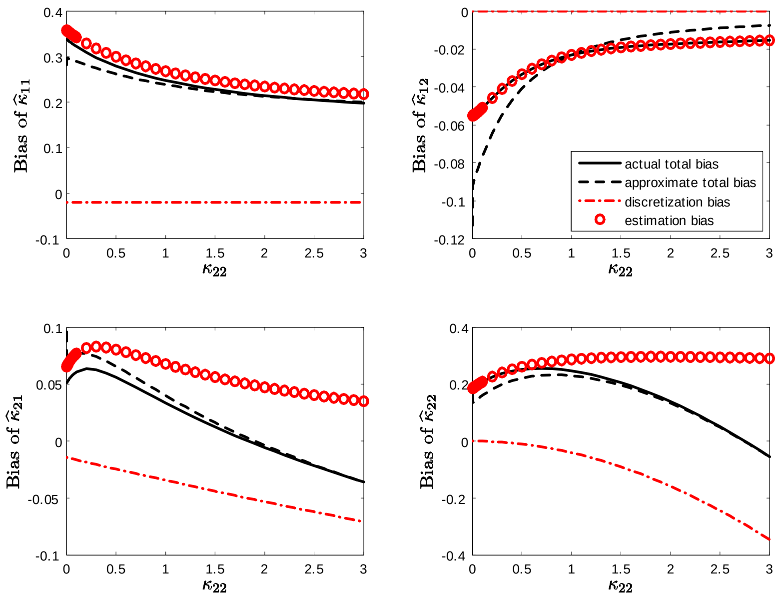

Figure 1: The bias of the elements in $\hat{A}$ in Model (6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the Euler method. The solid line is the actual total bias; the broken line is the approximate total bias according to the formula (3.13); the dashed line is the discretization bias H/h; the point line is the estimation bias. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

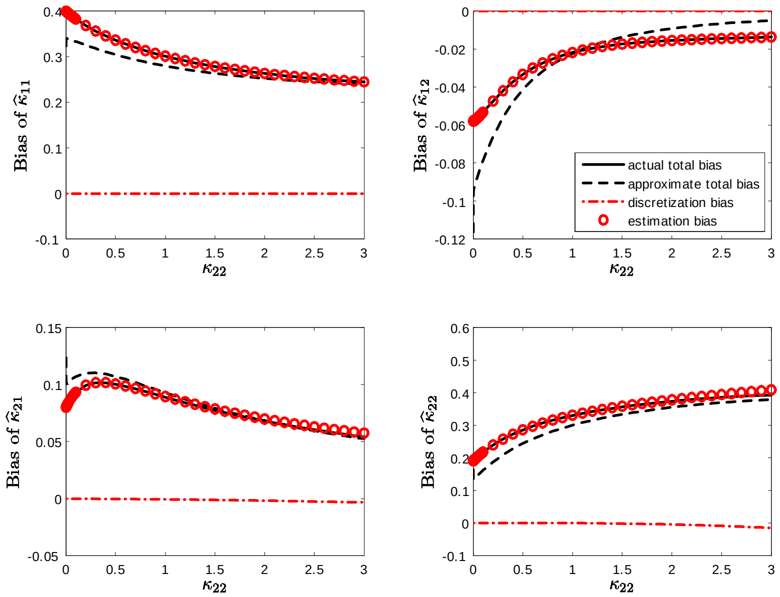

Figure 2: The bias of the elements in $\hat{A}$ in Model (6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the trapezoidal method. The solid line is the actual total bias; the broken line is the approximate bias according to the formula (3.13); the dashed line is the discretization bias $-v$; the point line is the estimation bias. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

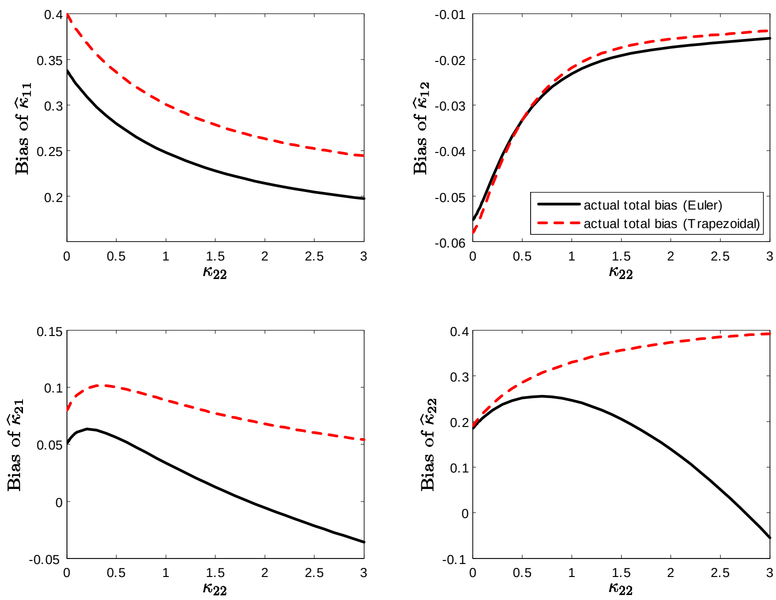

Figure 3: The bias of the elements in $\hat{A}$ in Model (6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the Euler and the trapezoidal methods, respectively. The solid line is the actual total bias for the Euler method; the broken line is the actual total bias for the trapezoidal method. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

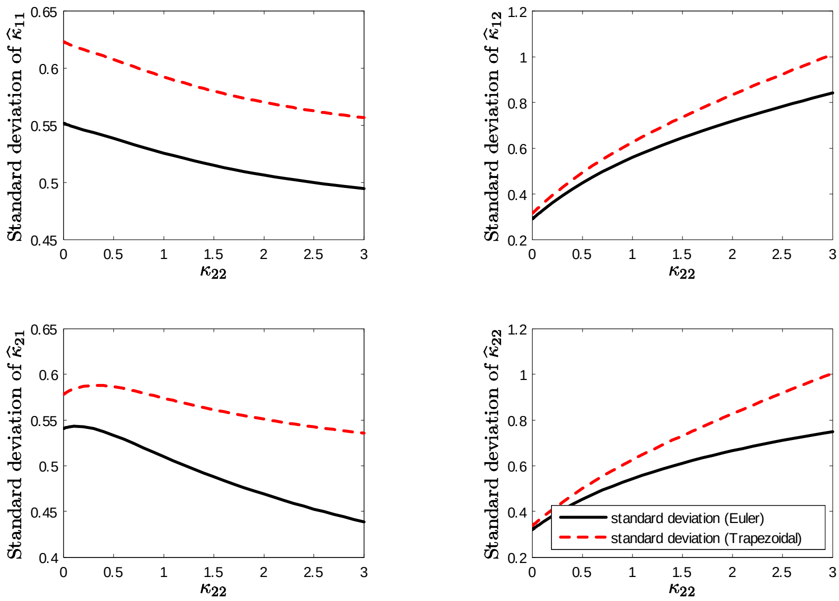

Figure 4: The standard deviation of the elements in $\hat{A}$ in Model (6.1) as a function of κ_{22} at the monthly frequency and $T = 10$. The estimates are obtained from the Euler and the trapezoidal methods, respectively. The solid line is the standard deviation for the Euler method; the broken line is the standard deviation for the trapezoidal method. The true value for κ_{11}, κ_{12}, and κ_{21} is 0.7, 0, and 0.5, respectively.

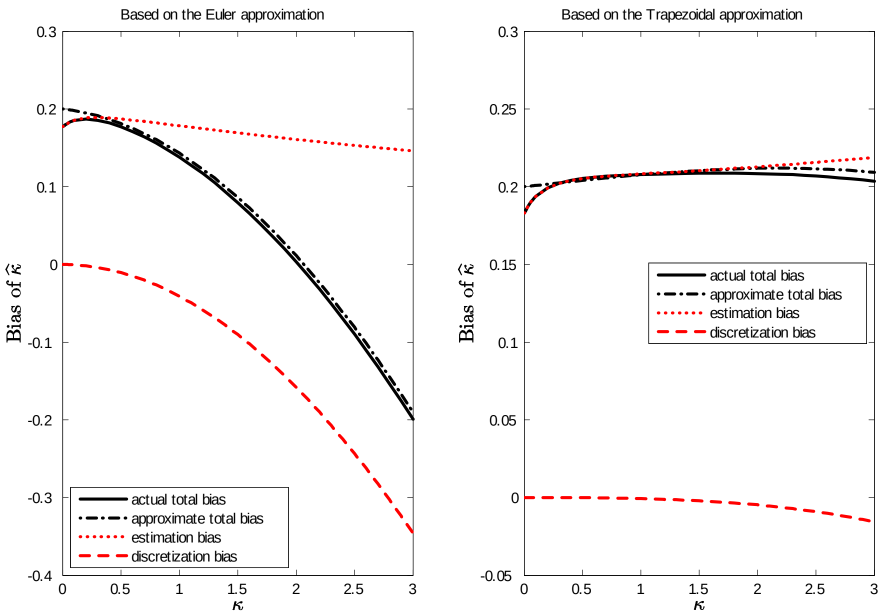

Figure 5: The bias of the κ estimates in the univariate model as a function of κ at the monthly frequency and $T = 10$ for the two approximate methods. The left panel is for the Euler method and the right panel is for the trapezoidal method. The solid line is the actual total bias; the dashed line is the approximate total bias; the dotted line is the estimation bias; the broken line is the discretization bias.

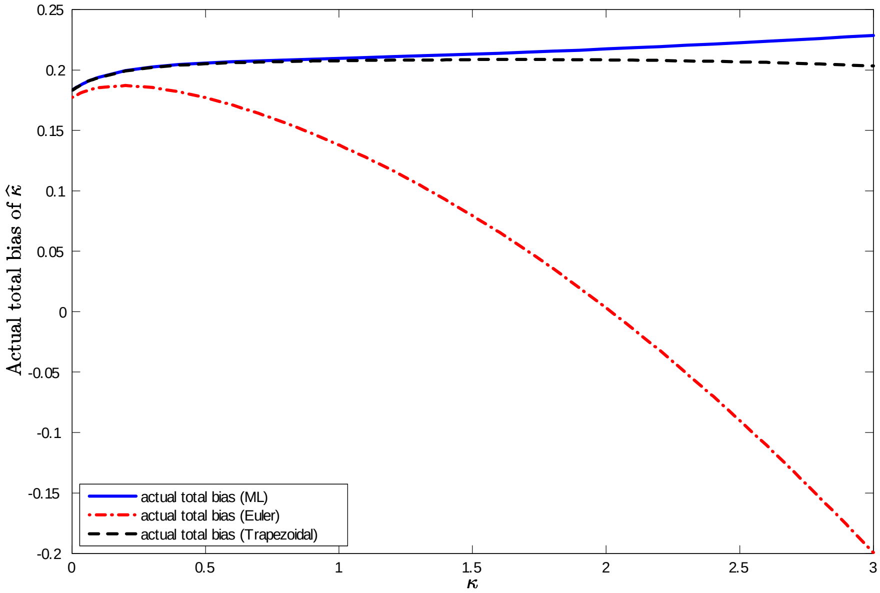

Figure 6: The actual total bias of the κ estimates in the univariate model as a function of κ at the monthly frequency and $T = 10$ for the two approximate methods and the exact ML. The solid line is for the exact ML; the dashed line is for the Euler method; the broken line is for the trapezoidal method.

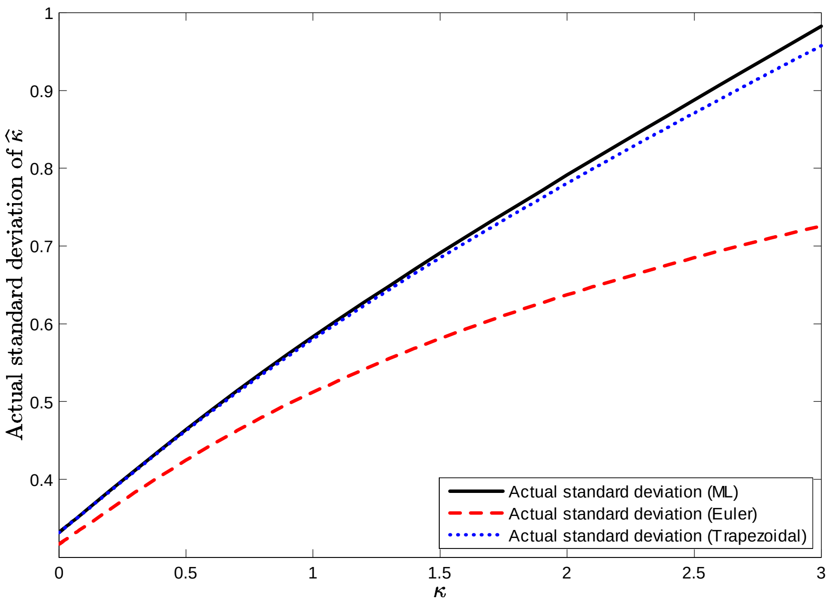

Figure 7: The standard deviation of the κ estimates in the univariate model as a function of κ at the monthly frequency and $T = 10$. The solid line is for the exact ML; the broken line is for the Euler method; the dotted line is for the trapezoidal method.