Finite Sample Theory in Continuous-Time Models

XIAOHU WANG *Fudan University*

Continuous-time models have broad applications in many core areas of economics and finance. This chapter first briefly introduces applications of continuous-time models in modeling the dynamics of short-term interest rates. The short-term rate is a fundamental factor in the valuation of nearly all of the derivatives in the financial markets. Over the past 40 years, many estimation methods have been proposed to estimate continuous-time models using discrete samples. However, almost all of those methods suffer from finite-sample bias. The bias problem is particularly severe for the mean-reversion parameter, which measures the persistence level of the interest-rate process. Moreover, this bias propagates and leads to considerable bias in the price calculations of interest rate contingent claims, such as bonds and bond options. The focus of this chapter is to conduct a detailed review of the bias issue. Two bias-correction methods are also discussed, namely, the jackknife method and the indirect inference method, which can effectively reduce the estimation bias of the mean-reversion parameter and the bias in pricing contingent claims. Simulation results are provided to illustrate the characteristics of the bias and investigate the performance of the two bias-correction methods

4.1 Introduction

Let $X(t)$ denote a stochastic process with t taking values continuously over the period of $[0, T]$, where T is the time span of the process. Stochastic differential equations (SDEs) driven by Brownian motion provide a convenient way to model the evolution of $X(t)$ over infinitesimal time intervals. The use of SDEs to model variables in economics and finance has many advantages. For example, SDEs can capture interactions among variables during the unit observation period and make a clear distinction between the treatment of stock variables and the treatment of flow variables. A detailed discussion can be found in Bergstrom (1966). Because of these advantages, continuous-time models are widely applied in many core areas of economics and finance, including neoclassical growth theory, capital accumulation, game theory, the asset pricing of financial derivatives, volatility modeling, and term structure theory. Certain examples are summarized in Bergstrom (1990), RC. (1992), Sundaresan (2000), and Piazzesi (2010).

This chapter focuses primarily on scalar, time-homogeneous SDEs that take the form of

$$dX(t) = a(X(t); \theta)\, dt + \sigma(X(t); \theta)\, dW(t), \tag{4.1.1}$$

where $dX(t)$ denotes the change of $X(t)$ over an infinitesimal instant, $a(\cdot; \theta)$ and $\sigma(\cdot; \theta)$ are the drift and the diffusion functions, θ is a vector of unknown parameters, and $W(t)$ denotes standard Brownian motion. The initial value of $X(t)$ at time $t = 0$ is denoted as $X(0) = O_p(1)$. The model in (4.1.1) indicates that $dX(t)$ follows a conditional normal distribution with mean $a(X(t); \theta)\, dt$ and variance $\{\sigma(X(t); \theta)\}^2 dt$. The term $dW(t)$ represents shocks to the system, whose effect on $dX(t)$ is scaled by the variance term $\sigma(X(t); \theta)$.

Although modeled in continuous time, observations of the process are typically available at discrete time points, for example, at $n = T/\Delta$ equally spaced points $\{t\Delta\}_{t=1}^{n}$, where Δ is the sampling interval. If an observation measures the value of $X(t)$ at a particular time point, it is referred to as stock data. In contrast, if an observation represents the value of $X(t)$ over a time interval, it is referred to as flow data. In addition, many ultrahigh-frequency stock data in finance are available at random and unequally spaced points (see, e.g., Aït-Sahalia et al. (2010)). The interface between models formulated in continuous time and data collected at discrete points is fundamental to the econometric treatment of continuous systems and leads to issues that are quite different from those typically encountered in a discrete time-series analysis. In this chapter, we only discuss the estimation issues for equally spaced stock data.

Under regular conditions, the maximum likelihood (ML) estimators of the parameters in the diffusion process (4.1.1) are consistent, asymptotically normal, and asymptotically efficient. Due to these optimal asymptotic properties, the ML approach is suitable for estimating the model whenever it is available (e.g., Lo (1988)). The likelihood function is the product of the transition densities. When the transition density is available in closed form, the ML estimation is easy to implement. However, except for some exceptional cases, the transition densities of discrete-time observations are unavailable in closed-form. Accordingly, various approximate ML methods have been proposed. A partial list includes methods based on approximate discrete-time models such as the Euler, trapezoidal, and Norman approximations proposed in Bergstrom (1966), Florens-Zmirou (1989), and Nowman (1997), respectively, along with the Hermit approximation method developed in Aït-Sahalia (1999, 2002, 2008), Aït-Sahalia and Kimmel (2007), and Li (2013), and the in-fill simulation method attributable to Pedersen (1995), Santa-Clara (1997), and Durham and Gallant (2002).

In addition to those approximate ML methods, a variety of moment-based estimation approaches have been developed, such as the generalized method of moments by Chan et al. (1992) and Hansen and Scheinkman (1995), the efficient method of moments by Gallant and Tauchen (1996), the empirical characteristic function method by Singleton (2001) and Knight and Yu (2002), and the Bayesian Markov Chain Monte Carlo methods by Eraker (2001) and Elerian et al. (2001).

The above methods have achieved significant success in estimating the SDEs consistently and efficiently in asymptotics. However, in virtually every one of these methods, the estimation of certain critical parameters suffers from severe finite-sample bias. Moreover, that bias does not disappear even when the sample size increases unboundedly under the asymptotic scheme of $\Delta \to 0$ with a fixed T. Thus, the availability of high-frequency data provides no help in reducing the magnitude of the bias.

Take the Ornstein–Uhlenbeck (OU) process as an example:

$$dX(t) = \kappa(\mu - X(t))\,dt + \sigma dW(t), \tag{4.1.2}$$

where κ, μ, and σ are three constants. For this simple process, the exact transition density of the discrete-time observations $\{X_{t\Delta}\}$ is available in closed form. Many studies in the literature point out that the ML estimator of κ, denoted by $\widehat{\kappa}_{ML}$, has substantial finite-sample bias; see, e.g., Ball and Torous (1996), Chapman and Pearson (2000), Phillips and Yu (2005, 2009), and Yu and Phillips (2001). In addition, Tang and Chen (2009) and Yu (2012) provide explicit expressions to approximate the bias. Their formulae reveal that the leading term of the bias is a function of T and not Δ, implying that the bias disappears only when T, the number of observation years, goes to infinity. However, in most empirical applications, data are available only for a small value of T. It is not realistic to expect time-series data in the real world to last for many years without any structural breaks.

Wang et al. (2011) argue that the quasi-ML estimator of κ based on the Euler approximation of the OU process outperforms $\widehat{\kappa}_{ML}$ and other estimators based on higher-order approximations in terms of finite-sample bias. The bias of the quasi-ML estimator consists of two parts, namely, the estimation bias and the discretization bias. Wang et al. (2011) prove that the two types of bias have opposite directions. Therefore, the estimation bias is offset by the discretization bias, making the total bias of the quasi-ML estimator smaller than that of the exact ML estimator. However, the bias of the Euler quasi-ML estimator remains large.

In finance, SDEs are often applied to model the dynamics of asset prices, exchange rates, volatility, and short-term interest rates. Thus, the prices of financial derivatives, such as bonds and options, are developed as functions of SDEs' parameters. These functions are often nonlinear and can be sensitive to certain parameters in SDEs. Well-known examples include Black and Scholes (1973), Merton (1973), and Hull and White (1987) for stock options, and Vasicek (1977), Cox et al. (1985), Chan et al. (1992), Heston (1993), and Duffie et al. (2000) for interest rate contingent claims. As a result, the finite-sample bias of the estimated parameters in SDEs will propagate to and generate bias in the price calculations of financial derivatives.

Phillips and Yu (2005) comprehensively study the finite-sample bias problem in pricing interest-rate contingent claims. Both univariate models and multivariate models are considered. Phillips and Yu (2009) extend their study to the bias issue of price calculations of stock options for models with and without stochastic volatility. These articles propose two alternative methods to reduce the finite-sample pricing bias: the jackknife method and the indirect-inference method. They further point out that due to the nonlinearity of pricing formulae, applying the bias-correction methods directly to the contingent claims works better in reducing pricing bias than the method of obtaining the bias-reduction estimates of the SDE parameters first and then plugging the estimates into the pricing formulae of contingent claims. Simulations and empirical studies by Phillips and Yu (2005, 2009) show that the two methods can achieve considerable bias reduction compared with the ML estimation in pricing financial derivatives.

The purpose of this chapter is to provide a systematic introduction to the finite-sample bias problem in estimating SDEs' parameters and pricing financial derivatives. In Section 2, the applications of SDEs to modeling short-term interest rates are briefly introduced. Furthermore, the OU process provides an example to show the connections between the prices of interest-rate contingent claims and the SDE parameters. Section 3 gives the bias formulae of the ML and quasi-ML estimators of the parameters in SDEs and discusses how the parameter bias translates into contingent

claim pricing. The two bias-correction techniques are studied in detail in Section 4. Simulation examples are provided in Sections 3 and 4 to show the gravity of the bias problem and the performance of the bias-correction methods. Section 5 concludes.

4.2 Applications in Term Structure Modeling

As application examples in finance, this section introduces the use of diffusion processes in the term structure literature. In Subsection 4.2.1, we briefly summarize the scalar diffusions applied in modeling the short-term risk-free interest rates. For the case in which the dynamics of the short-term rate follow the OU process defined in (4.1.2), the contingent claims based on interest rates have closed-form pricing formulae. Subsection 4.2.2 displays these pricing formulae to show the roles of SDE parameters in pricing contingent claims.

4.2.1 Modelling short-term interest rates

The short-term risk-free interest rate can be defined with the help of zero-coupon bonds. A bond that makes a terminal payoff without the risk of default and without paying any intermediate coupons is called a zero-coupon bond. Let us normalize the terminal payoff as \$1 and use $P(t,s)$ to denote the price of the bond that is issued at time t with maturity date s. Holding the bond until its maturity generates a risk-free per-period return as

$$\gamma_t^{(s)} = -\frac{\log P(t,s)}{s-t}. \tag{4.2.1}$$

$\gamma_t^{(s)}$ is often referred to as yield-to-maturity. The short-term risk-free interest rate is defined as the limit of $\gamma_t^{(s)}$ when $s \to t$:

$$r(t) = \lim_{s\to t} \gamma_t^{(s)} = \lim_{s\to t} -\frac{\log P(t,s)}{s-t}.$$

The short-term rate is a fundamental factor in the valuation of nearly all of the derivatives in the financial markets. In addition, interest-rate variability is a major source of risk for banks and other financial institutions. Unsurprisingly, therefore, problems related to short-term interest rates attract enormous attention in the economics and finance literature. One important branch of inquiry on this topic is concerned with how best to model the short-term rate dynamics. Over the past 50 years, many single-factor and multifactor interest-rate models specified by various SDEs have been proposed. Table 1 displays some prominent single-factor models that have achieved remarkable success in modeling and forecasting short-term rates.

Multifactor models assume that the short-term rate depends on a state vector $Z(t)$, i.e., $r(t) = f(Z(t))$, where $f(\cdot)$ is a real-valued function. Various multivariate diffusion processes have been applied to model the dynamics of $Z(t)$. Examples of multifactor models can be found in Duffie and Kan (1996), Gourieroux and Sufana (2006), and their references.

Another group of models for short-term interest rates is that of continuous-time models with stochastic volatility. Many empirical studies show that the short-term rate exhibits properties such as fat tails and persistent volatility patterns. Stochastic volatility models can better fit these properties. An incomplete list of references includes Hull and White (1987), Andersen and Lund (1997), Gallant and Tauchen (1998), Eraker (2001), Durham (2003), Trolle and Schwartz (2009).

4.2.2 Pricing formulae of contingent claims

The univariate OU process defined in (4.1.2) is identical to the Vasicek model in Table 1, with the relationship between the parameters being $\alpha_1 = \kappa\mu$, $\alpha_2 = -\kappa$, and $\beta_1 = \sigma$. Assuming that the short-term rate follows an OU process, Vasicek (1977) derives an explicit pricing formula for

Table 4.1: Alternative model specifications for short-term interest rates. In this table, $r := r(t)$ denotes the interest rate at time t, $W := W(t)$ is the standard Brownian motion, $\{\alpha_i\}_{i=1}^4$ and $\{\beta_i\}_{i=1}^4$ are unknown parameters. Moreover, GBM, BS (1980), CIR VR (1980), and Aït (1996) refers to Geometric Brownian motion, Brennan and Schwartz (1980), Cox et al. (1980), and Aït-Sahalia (1996), respectively.

Merton (1973)	$dr = \alpha_1 dt + \beta_1 dW$
Vasicek (1977)	$dr = (\alpha_1 + \alpha_2 r)\, dt + \beta_1 dW$
CIR (1985)	$dr = (\alpha_1 + \alpha_2 r)\, dt + \beta_2 r^{1/2} dW$
Dothan (1978)	$dr = \beta_2 r dW$
GBM	$dr = \alpha_2 r dt + \beta_2 r dW$
BS (1980)	$dr = (\alpha_1 + \alpha_2 r)\, dt + \beta_2 r dW$
CIR VR (1980)	$dr = \beta_2 r^{3/2} dW$
CKLS (1992)	$dr = (\alpha_1 + \alpha_2 r)\, dt + \beta_3 r^{\beta_4} dW$
Tauchen (1995)	$dr = \left(\alpha_1 + \alpha_2 r + \alpha_3 r^2 + \alpha_4 / r\right) dt + \beta_3 r^{\beta_4} dW$
Aït (1996)	$dr = \left(\alpha_1 + \alpha_2 r + \alpha_3 r^2 + \alpha_4 / r\right) dt + \left(\beta_1 + \beta_2 r + \beta_3 r^{\beta_4}\right) dW$

the zero-coupon bond. Jamshidian (1989) obtains the pricing formula for European options, with bonds as the underlying assets. To display these pricing formulae, we assume in this subsection that the short-term rate follows the OU process defined in (4.1.2) and specify it here again by letting $r(t) := X(t)$ for convenience:

$$dr(t) = \kappa(\mu - r(t))\, dt + \sigma dW(t). \tag{4.2.2}$$

By holding a zero-coupon bond to its maturity date s, a risk-free return is possible. However, the bond is tradable at random prices at any time before its maturity. Holding a bond from time t to period $t + h < s$ generates a holding-period return defined as

$$hpr_t^{t+h} = \frac{\log P(t+h, s) - \log P(t, s)}{h},$$

which is a random variable due to the uncertainty of the price $P(t+h, s)$. Hence, zero-coupon bonds are risky assets.

The bond price $P(t, s)$ depends on t and the short-term rate $r(t)$. Hence, by using the Ito's lemma, the bond price has the dynamics

$$\begin{aligned} dP(t,s) &= P_t dt + P_r dr + \frac{1}{2} P_{rr} (dr)^2 \\ &= \left(P_t + \kappa(\mu - r(t)) P_r + \frac{1}{2}\sigma^2 P_{rr}\right) dt + \sigma P_r dW(t), \end{aligned}$$

where P_t denotes the first-order derivative of the bond price with respect to t, P_r and P_{rr} are the first- and second-order derivatives, respectively, with respect to the short-term rate $r(t)$. It is further obtained that

$$\begin{aligned} \frac{dP(t,s)}{P(t,s)} &= \frac{\left(P_t + \kappa(\mu - r(t)) P_r + \frac{1}{2}\sigma^2 P_{rr}\right)}{P(t,s)} dt + \frac{\sigma P_r}{P(t,s)} dW(t) \\ &= \mu_p dt + \sigma_p dW(t). \end{aligned}$$

where μ_p represents the conditional expected return of the bond, and σ_p^2 is the conditional variance. Define the market price of risk as the expected excess return per unit risk, that is $\lambda := (\mu_p - r)/\sigma_p$. This definition yields the equation of

$$\mu_p - r = \lambda \sigma_p,$$

which shows that the expected excess return of the bond equals the market price of risk multiplying the standard deviation associated with the bond. Substituting for the representations of μ_p and σ_p gives a partial differential equation (PDE) of the bond price P:

$$P_t + \kappa\left(\mu - \frac{\lambda\sigma}{\kappa} - r(t)\right)P_r + \frac{1}{2}\sigma^2 P_{rr} - rP = 0. \tag{4.2.3}$$

The solution of the PDE with the boundary condition of $P(s,s) = 1$ gives the bond pricing formula derived in Vasicek (1977). We report this formula in Theorem 4.2.1.

Theorem 4.2.1 *Assume the short-term rate $r(t)$ follows the OU process given in (4.2.2). At time period t, the price of the zero-coupon bond with the terminal payoff \$1 and the maturity date s follows the formula of*

$$P(t,s) = A(t,s)\, e^{-B(t,s)r(t)}, \tag{4.2.4}$$

where

$$A(t,s) = \exp\left\{\left(\mu - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2}\right)[B(t,s) - s + t] - \frac{\sigma^2}{4\kappa}[B(t,s)]^2\right\}, \tag{4.2.5}$$

and

$$B(t,s) = \frac{1}{\kappa}\left(1 - e^{-\kappa(s-t)}\right). \tag{4.2.6}$$

The pricing formula in (4.2.4) makes the yield-to-maturity defined in (4.2.1) an affine function of the short-term rate:

$$\gamma_t^{(s)} = -\frac{\log P(t,s)}{s-t} = -\frac{\log A(t,s)}{s-t} + \frac{B(t,s)}{s-t}r(t).$$

Hence, the Vasicek model is a special case of the affine term structure models.

Another way to price bonds is with the help of a risk-neutral probability measure. Imagine that there is a probability measure, denoted by Q^*, under which the short-term rate has the following dynamics

$$dr(t) = \kappa(\mu^* - r(t))\,dt + \sigma dW^*(t),$$

where $\mu^* = \mu - (\lambda\sigma)/\kappa$, and $W^*(t)$ is a standard Brownian motion under Q^*. The application of Ito's lemma gives the dynamics of the bond price under Q^* as

$$dP(t,s) = \left(P_t + \kappa(\mu^* - r(t))P_r + \frac{1}{2}\sigma^2 P_{rr}\right)dt + \sigma P_r dW^*(t).$$

From Equation (4.2.3), it is obtained that the expected excess return of the bond under the measure Q^* is zero, i.e.,

$$\frac{\left(P_t + \kappa(\mu^* - r(t))P_r + \frac{1}{2}\sigma^2 P_{rr}\right)}{P(t,s)} - r = 0.$$

In other words, under the probability measure Q^*, investors are risk-neutral in the sense that they do not require excess returns for any risk they bear. This property gives the reason for Q^* being named as a risk-neutral probability measure. Furthermore, due to the risk-neutral attitude, the bond price can be calculated as the expected values of its further payoff discounted at the short-term rate, i.e.,

$$P(t,s) = E_t^*\left[\exp\left(-\int_t^{t+s} r(\tau)\,d\tau\right)\right],$$

where E_t^* denotes the expectation under the measure Q^* and the information available at time t. Providing a thorough discussion of the risk-neutral pricing theory and the connections between physical probability measures and the associated risk-neutral measures is beyond the scope of this chapter. Interested readers can find classic textbooks for more details, e.g., Shreve (2004).

We end this subsection by introducing the pricing formulae of bond options developed in Jamshidian (1989). A European option is a contract that gives the holders a right to buy or sell a bond at a particular time (expiration date) for a predetermined price. This predetermined price is called the strike price of the option. Consider a zero-coupon bond with a fixed maturity date s and a terminal payoff (also called principal value) L. The current price of this bond is determined by

$$P(t,s) = L \times A(t,s)\, e^{-B(t,s)r(t)}, \tag{4.2.7}$$

where $A(t,s)$ and $B(t,s)$ are given in (4.2.5) and (4.2.6), respectively. Jamshidian (1989) derived explicit pricing formulae for options written on this bond, which are reported in the following Theorem 4.2.2.

Theorem 4.2.2 *Consider a zero-coupon bond that gives a terminal payoff L at the maturity date s. (a) A European call option written on this bond with strike price K and expiration date τ $(t < \tau < s)$ has the pricing formula of*

$$C_{OP}(t,\tau,s,K) = LP(t,s)\,\Phi(h) - KP(t,\tau)\,\Phi(h-\delta_p); \tag{4.2.8}$$

(b) The price of a corresponding put option takes the form of

$$P_{OP}(t,\tau,s,K) = KP(t,\tau)\,\Phi(\delta_p - h) - LP(t,s)\,\Phi(-h),$$

where $P(t,s)$ is given in (4.2.7), $\Phi(\cdot)$ denotes the standard normal cumulative function,

$$\delta_p = \frac{\sigma}{\kappa}\left(1 - e^{-\kappa(s-\tau)}\right)\sqrt{\frac{1-e^{-2\kappa(\tau-t)}}{2\kappa}},$$

and

$$h = \delta_p^{-1}\log\left(\frac{P(t,s)\,L}{P(t,\tau)\,K}\right) + \frac{\delta_p}{2}.$$

4.3 Estimation and Finite-sample Bias

4.3.1 Vasicek model with an unknown mean

Although modeled by a continuous-time diffusion process, the short-term rate $r(t)$ can only be observed at discrete time points. Assume that observations of $r(t)$ are available at equally spaced discrete time points from time zero to time T with the sampling interval Δ. This means that the observations are collected at time points $\{t\Delta\}_{t=0}^{n}$ from $t=0$ to $t=n$, with $n := T/\Delta$ being the total number of observations. The short-term rate $r(t)$ is often taken as the normalized annual rate. Thus, in the term structure literature, if $r(t)$ is observed at daily (weekly or monthly) frequency over 10 years, it is often set as $\Delta = 1/252$ ($1/52$ or $1/12$), and $T = 10$ represents the number of observing years. T is referred to as the time span of the data. In the rest of this chapter, we use $\{r_{t\Delta}\}$ to represent the observations of $r(t)$, and we write the term simply as $\{r_t\}$ when there is no confusion.

To estimate the continuous-time diffusion processes, the ML estimators are asymptotically optimal under regular conditions; accordingly, they are often adopted when available. The simple Vasicek process in (4.2.2) has an equivalent discretization (see, e.g., Phillips (1972)):

$$r_t = e^{-\kappa\Delta} r_{t-1} + \mu\left(1 - e^{-\kappa\Delta}\right) + \varepsilon_t, \tag{4.3.1}$$

where the errors

$$\varepsilon_t = \sigma\int_{(t-1)\Delta}^{t\Delta} e^{-\kappa(t\Delta - s)} dW(s),$$

for $t = 1, 2, ..., n$, are a sequence of independent and identically distributed (i.i.d.) variables with the normal distribution $N\left(0, \sigma^2\left(1 - e^{-2\kappa\Delta}\right)/(2\kappa)\right)$. The transition density of the observations is readily obtained as

$$r_t | r_{t-1} \sim N\left(e^{-\kappa\Delta} r_{t-1} + \mu\left(1 - e^{-\kappa\Delta}\right), \sigma^2\left(1 - e^{-2\kappa\Delta}\right)/(2\kappa)\right).$$

Therefore, the log likelihood function

$$\mathcal{L}(\theta) = \sum_{t=1}^{n} \ln f\left(r_t | r_{t-1}\right),$$

has a simple, explicit expression, where $f\left(r_t | r_{t-1}\right)$ denotes the conditional density function of r_t given r_{t-1}. Tang and Chen (2009) provide explicit formulae of the ML estimators as

$$\widehat{\kappa}_n^{ML} = -\frac{1}{\Delta}\ln\left(\widehat{\varphi}_1\right) \text{ with } \widehat{\varphi}_1 = \frac{\sum_{t=1}^{n} r_t r_{t-1} - n^{-1}\sum_{t=1}^{n} r_t \sum_{t=1}^{n} r_{t-1}}{\sum_{t=1}^{n} r_{t-1}^2 - n^{-1}\left(\sum_{t=1}^{n} r_{t-1}\right)^2}, \tag{4.3.2}$$

$$\widehat{\mu}_n^{ML} = \widehat{\varphi}_2 \text{ with } \widehat{\varphi}_2 = \frac{n^{-1}\sum_{t=1}^{n}\left(r_t - \widehat{\varphi}_1 r_{t-1}\right)}{1 - \widehat{\varphi}_1},$$

and

$$\widehat{\sigma}_n^{ML} = \sqrt{\frac{2\widehat{\kappa}\widehat{\varphi}_3}{1 - \widehat{\varphi}_1^2}} \text{ with } \widehat{\varphi}_3 = n^{-1}\sum_{t=1}^{n}\left[r_t - \widehat{\varphi}_1 r_{t-1} - \widehat{\varphi}_2\left(1 - \widehat{\varphi}_1\right)\right]^2.$$

When $\kappa > 0$, the Vasicek process in (4.2.2) is stationary. Under stationarity, Tang and Chen (2009) prove the asymptotic normality of the ML estimators when $T \to \infty$ with a fixed Δ. In addition, they derive the bias formula for each ML estimator. Their results show that the biases of $\widehat{\mu}_n^{ML}$ and $\widehat{\sigma}_n^{ML}$ are negligible when the sample size n is reasonably large. However, the estimator $\widehat{\kappa}_n^{ML}$ suffers from substantial finite-sample bias even when $n \to \infty$, as long as the time span T is small. Theorem 4.3.1 reports the bias formula of $\widehat{\kappa}_n^{ML}$ developed in Tang and Chen (2009).

Theorem 4.3.1 *The ML estimator $\widehat{\kappa}_n^{ML}$ given in (4.3.2) has the bias formula of*

$$\text{Bias}\left(\widehat{\kappa}_n^{ML}\right) = E\left(\widehat{\kappa}_n^{ML}\right) - \kappa = \frac{1}{T}\left(\frac{5}{2} + e^{\kappa\Delta} + \frac{1}{2}e^{2\kappa\Delta}\right) + O\left(n^{-2}\right). \tag{4.3.3}$$

To examine the performance of the bias formula in (4.3.3), Figure 4.1 compares the values calculated from the bias formula with the simulated finite-sample bias. The range of κ is $(0, 3]$, which covers the reasonable values found in empirical studies of short-term interest rates. We simulate data from the Vasicek process in (4.2.2), with $T = 4$, $\Delta = 1/52$, and $\Delta = 1/252$, respectively, corresponding to 4 years of weekly data and daily data. For each simulated sample path, κ is estimated by the ML method. We replicate this experiment 1,000 times to obtain the simulated bias of $\widehat{\kappa}_n^{ML}$, which can provide an excellent approximation of the actual bias. The solid black line in Figure 4.1 draws the simulated bias of $\widehat{\kappa}_n^{ML}$. The dotted blue line reports the values calculated from the leading term of the bias formula in (4.3.3).

Several features appear in Figure 4.1. First, the actual bias of $\widehat{\kappa}_n^{ML}$ is a nonlinear function of κ, and the curvature is large when κ is small. Second, instead of going to zero, the bias increases sharply as $\kappa \to 0$. Thus, the bias percentage is very high when κ is low. Third, the bias magnitude has almost no change when the sampling frequency changes from weekly to daily, showing that the availability of high-frequency data cannot reduce the bias. Fourth, the bias formula in (4.3.3) provides a very good approximation of the actual bias when κ is large. However, when κ is small,

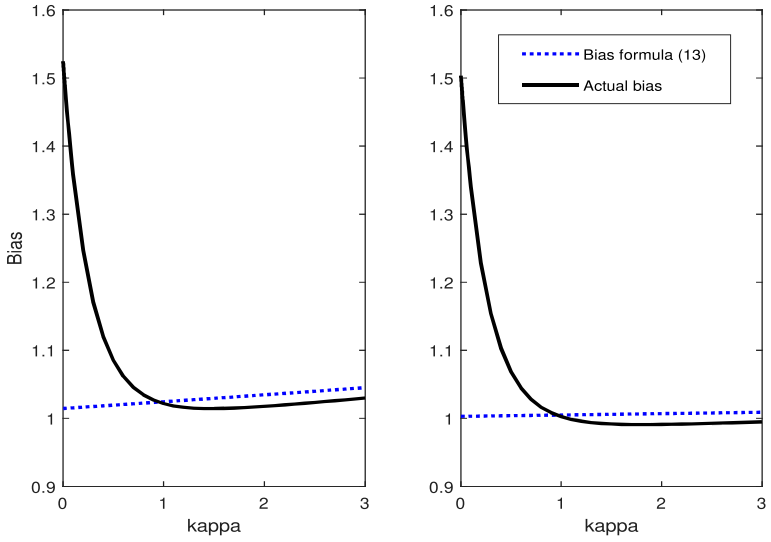

Figure 4.1: Bias of $\widehat{\kappa}_n^{ML}$ as a function of κ. From left to right, the two panels responds to weekly data ($\Delta = 1/52$) and daily data ($\Delta = 1/252$), respectively.

the bias formula (4.3.3) fails to capture the curvature of the actual bias; thus, it does not perform well in approximating the actual bias.

To better understand the bias property of $\widehat{\kappa}_n^{ML}$ and the bias formula in (4.3.3), we expand $(\widehat{\varphi}_1)$ about the point $\varphi_1 = \exp\{-\kappa\Delta\}$:

$$\begin{aligned}\widehat{\kappa}_n^{ML} &= -\frac{1}{\Delta}\ln(\widehat{\varphi}_1) \\ &= -\frac{1}{\Delta}\left\{\ln(\varphi_1) + \frac{1}{\varphi_1}(\widehat{\varphi}_1 - \varphi_1) - \frac{1}{2\varphi_1^2}(\widehat{\varphi}_1 - \varphi_1)^2 + \cdots\right\},\end{aligned}$$

The above Taylor expansion leads to the formula

$$E\left(\widehat{\kappa}_n^{ML}\right) = \kappa - \frac{1}{\Delta}\left\{\frac{1}{\varphi_1}E(\widehat{\varphi}_1 - \varphi_1) - \frac{1}{2\varphi_1^2}E(\widehat{\varphi}_1 - \varphi_1)^2 + \cdots\right\}. \tag{4.3.4}$$

Note that $\varphi_1 = \exp\{-\kappa\Delta\}$ is the autoregressive (AR) root of the discretization in (4.3.1), and $\widehat{\varphi}_1$, as defined in (4.3.2), is the ML estimation. Formula (4.3.4) reveals that the moments of $\widehat{\varphi}_1 - \varphi_1$ of every order contribute to the bias of $\widehat{\kappa}_n^{ML}$.

The finite sample properties of $\widehat{\varphi}_1$ have been well studied in the literature. Alternative bias formulae under various model settings have been developed. A partial list includes Hurvicz (1950), Kendall (1954), Marriott and Pope (1954), White (1961), Shenton and Johnson (1965), Yamamoto and Kunitomo (1984), and Vinod and Shenton (1996). The bias formula derived in Marriott and Pope (1954) is

$$E(\widehat{\varphi}_1 - \varphi_1) = -\frac{3\varphi_1 + 1}{n} + o\left(\frac{1}{n}\right).$$

Bartlett (1946) develops an approximation of the finite-sample variance:

$$Var(\widehat{\varphi}_1) = \frac{1 - \varphi_1^2}{n} + o\left(\frac{1}{n}\right).$$

The bias formula of $\widehat{\kappa}_n^{ML}$ given in (4.3.3) is obtained by substituting the bias formula of Marriott and Pope (1954) and the variance formula of Bartlett (1946) into the Taylor expansion (4.3.4) and ignoring all the higher-order moments of $\widehat{\varphi}_1 - \varphi_1$. Involving only the first two moments of

$\widehat{\varphi}_1$ omits the bias of $\widehat{\kappa}_n^{ML}$ contributed by the higher-order moments of $\widehat{\varphi}_1$, such as the skewness and the kurtosis. Yu (2012) argues that including higher-order moments of $\widehat{\varphi}_1$ can lead to a bias formula of $\widehat{\kappa}_n^{ML}$ that captures the nonlinearity of the actual bias when the values of κ are small. He proves this argument for a Vasicek model with $\mu = 0$, which we will introduce in the following subsection.

4.3.2 Vasicek model with $\mu = 0$

The Vasicek process with $\mu = 0$ has the expression

$$dr(t) = -\kappa r(t)\, dt + \sigma dW(t), \tag{4.3.5}$$

whose exact discretization is a first-order AR model without intercept:

$$r_t = \varphi r_{t-1} + \varepsilon_t,$$

where $\varphi = e^{-\kappa\Delta}$ and

$$\varepsilon_t \overset{i.i.d}{\sim} N\left(0, \sigma^2\left(1 - e^{-2\kappa\Delta}\right)/(2\kappa)\right).$$

The ML estimator of κ, in this case, takes the form of

$$\widehat{\kappa} = -\frac{1}{\Delta}\ln(\widehat{\varphi}) \quad \text{with} \quad \widehat{\varphi} = \left(\sum_{t=1}^{n} r_t r_{t-1}\right) / \sum_{t=1}^{n} r_{t-1}^2. \tag{4.3.6}$$

The Taylor expansion reveals the following relationship between the bias of $\widehat{\kappa}$ and the moments of $\widehat{\varphi}$:

$$E(\widehat{\kappa}) = \kappa - \frac{1}{\Delta}\left\{\frac{1}{\varphi}E(\widehat{\varphi} - \varphi) - \frac{1}{2\varphi^2}E(\widehat{\varphi} - \varphi)^2 + \cdots\right\}.$$

Yu (2012), at the beginning of the paper, provides a bias formula of $\widehat{\kappa}$ by only calculating the first two moments of $\widehat{\varphi} - \varphi$:

$$E(\widehat{\kappa}) - \kappa = \frac{2}{T}\left(1 + \frac{1}{n}\right) + \frac{1}{2T}\left(e^{2\kappa\Delta} - 1\right) + o\left(\frac{1}{T}\right). \tag{4.3.7}$$

The paper then argues that if $\kappa \to 0$, it has $e^{2\kappa\Delta} - 1 \to 0$, which makes the constant term $\frac{2}{T}\left(1 + \frac{1}{n}\right)$ the only one dominating in the bias approximation. Hence, the nonlinearity of the actual bias of $\widehat{\kappa}$ as $\kappa \to 0$ is not captured by the bias approximation in (4.3.7).

To better capture the nonlinearity of the actual bias when κ is small, Yu (2012) develops another bias approximation that involves the information from the higher-order moments of $\widehat{\varphi} - \varphi$. This bias approximation is presented in Theorem 4.3.2.

Theorem 4.3.2 *The estimator $\widehat{\kappa}$ given in (4.3.6) has the bias formula of*

$$\begin{aligned} E(\widehat{\kappa}) - \kappa \;=\; & \frac{2}{T}\left(1 + \frac{1}{n}\right) + \frac{1}{2T}\left(e^{2\kappa\Delta} - 1\right) \\ & - \frac{2\left(1 - e^{-2n\kappa\Delta}\right)}{Tn\left(1 - e^{-2\kappa\Delta}\right)}\left\{1 + \frac{2}{n} - \frac{1 - e^{-2n\kappa\Delta}}{n^2\left(1 - e^{-2\kappa\Delta}\right)}\right\} + o\left(\frac{1}{T}\right). \end{aligned} \tag{4.3.8}$$

Compared with (4.3.7), the bias formula given in Theorem 4.3.2 has the additional term

$$-\frac{2\left(1 - e^{-2n\kappa\Delta}\right)}{Tn\left(1 - e^{-2\kappa\Delta}\right)}\left\{1 + \frac{2}{n} - \frac{1 - e^{-2n\kappa\Delta}}{n^2\left(1 - e^{-2\kappa\Delta}\right)}\right\}.$$

When $\kappa \to 0$, it has

$$\frac{1 - e^{-2n\kappa\Delta}}{n\left(1 - e^{-2\kappa\Delta}\right)} \to 1,$$

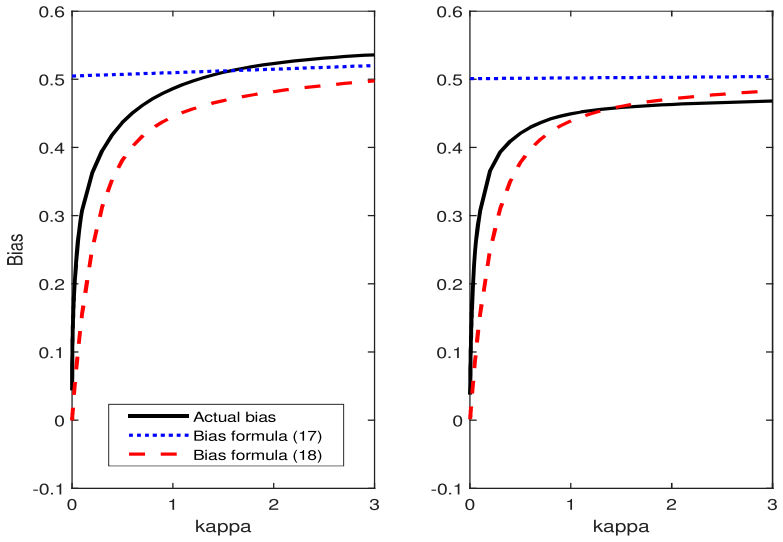

Figure 4.2: Bias of $\widehat{\kappa}$ as a function of κ. From left to right, the two panels responds to weekly data ($\Delta = 1/52$) and daily data ($\Delta = 1/252$), respectively.

and

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

Therefore, when $\kappa \to 0$, the additional term has the same order as the leading term in the bias formula, and cannot be ignored in the bias approximation. More importantly, the limit of the additional term will cancel out the leading term in the bias formula when $\kappa \to 0$. As a result, it is the middle term in the bias formula (4.3.8), i.e., $\frac{1}{2T}\left(e^{2\kappa\Delta}-1\right)$, which plays a dominating role as $\kappa \to 0$. This middle term can well capture the nonlinearity of the actual bias for small values of κ.

In addition, Yu (2012) proves that when $\Delta \to 0$ with a fixed time-span T, it has $n = T/\Delta \to \infty$, and

$$E\left(\widehat{\kappa}\right)-\kappa \approx \frac{2}{T}+\frac{1-e^{2\kappa T}}{T^2\kappa},$$

which clearly shows that the bias does not disappear when high-frequency observations are available.

Figure 4.2 compares the performance of formulae (4.3.7) and (4.3.8) in approximating the actual bias of $\widehat{\kappa}$. The dotted blue line and the dashed red line report the bias values calculated from formulae (4.3.7) and (4.3.8), respectively. The solid black line gives the actual bias from simulations that are replicated 1,000 times. The data are simulated based on the Vasicek process in (4.3.5) with $T = 4$, $\Delta = 1/52$, and $\Delta = 1/252$, respectively. For each simulated sample path, κ is estimated by the estimator $\widehat{\kappa}$ defined in (4.3.6).

Two critical features can be seen in Figure 4.2. First, the bias approximation from Formula (4.3.7), which uses only the information from the first two moments of $\widehat{\varphi}-\varphi$ can well approximate the actual bias when κ is large. In contrast, the bias approximation from Formula (4.3.8), which involves the information from higher-order moments of $\widehat{\varphi}-\varphi$, accurately approximates the actual bias for all concerned values of κ, and perfectly captures the nonlinearity of the bias when $\kappa \to 0$. Second, the actual bias of $\widehat{\kappa}$ goes to zero as $\kappa \to 0$. This feature sharply differs from the case where the intercept of the Vasicek process is unknown, as shown in Figure 4.1.

4.3.3 Bias of quasi-ML estimators

Wang et al. (2011) point out that although the exact ML approach is optimal asymptotically, the as-

sociated estimators may not perform the best in terms of the finite-sample properties. They derive the bias formulae of the quasi-ML estimators of κ that are obtained based on various approximate discretizations of the OU process. It is shown that the estimator based on the first-order Euler discretization has a smaller finite-sample bias than the estimators based on high-order approximations, which in turn perform better than the exact ML estimator. The reason for this finding is that the bias can be decomposed into two parts, the estimation bias and the discretization bias, which have opposite signs. A lower-order approximation generates a more significant discretization bias that helps reduce the magnitude of the total bias.

For the OU process with an unknown mean, as described in (4.2.2), integrating both sides of the equation over the interval from $(t-1)\Delta$ to $t\Delta$ leads to the equation

$$r_t - r_{t-1} = -\kappa \int_{(t-1)\Delta}^{t\Delta} r(s)\, ds + \kappa\mu\Delta + \varepsilon_t,$$

where $\varepsilon_t = \sigma\left(W_{t\Delta} - W_{(t-1)\Delta}\right)$ is an i.i.d. sequence of random variables with the normal distribution $N\left(0, \sigma^2\Delta\right)$. The Euler approximation is defined as

$$\int_{(t-1)\Delta}^{t\Delta} r(s)\, ds \approx r_{t-1} \cdot \Delta,$$

which leads to the following approximate discretization of the OU process:

$$r_t \approx (1 - \kappa\Delta)\, r_{t-1} + \kappa\mu\Delta + \varepsilon_t. \tag{4.3.9}$$

For convenience, Equation (4.3.9) is referred to as Euler discretization hereafter.

The Euler discretization in (4.3.9) presents an AR(1) regression with normally distributed errors. The ML estimator of the AR root, $\widehat{\varphi}_1$, is given by (4.3.2). From the relationship between κ and the AR root in the regression of (4.3.9), the quasi-ML estimator of κ based on the Euler discretization has the form of

$$\widehat{\kappa}_{Euler} = \frac{1}{\Delta}\left(1 - \widehat{\varphi}_1\right). \tag{4.3.10}$$

The Taylor expansion of $\varphi_1 = \exp\{-\kappa\Delta\}$ shows that

$$\varphi_1 = \sum_{i=0}^{\infty} \frac{(-\kappa\Delta)^i}{i!} = 1 - \kappa\Delta + H,$$

where $H = \sum_{i=2}^{\infty} \frac{(-\kappa\Delta)^i}{i!} = O\left(\Delta^2\right)$ as $\Delta \to 0$. Therefore,

$$\kappa = \Delta^{-1}\left(1 - \varphi_1\right) + \Delta^{-1} H$$

and

$$\widehat{\kappa}_{Euler} - \kappa = -\frac{1}{\Delta}\left(\widehat{\varphi}_1 - \varphi_1\right) - \frac{H}{\Delta}.$$

The bias of the Euler estimator $\widehat{\kappa}_{Euler}$ is then expressed as

$$\text{Bias}\left(\widehat{\kappa}_{Euler}\right) = E\left(\widehat{\kappa}_{Euler} - \kappa\right) = -\frac{1}{\Delta} E\left(\widehat{\varphi}_1 - \varphi_1\right) - \frac{H}{\Delta}.$$

It becomes clear that the bias of $\widehat{\kappa}_{Euler}$ has two components. The first component is $-\frac{1}{\Delta} E\left(\widehat{\varphi}_1 - \varphi_1\right)$. This component is called the estimation bias, which comes from estimating the AR root in the regression of (4.3.9). The second component is $-H/\Delta$, which is attributable to the truncation of the Taylor expansion of $\varphi_1 = \exp\{-\kappa\Delta\}$. This component is named the discretization bias. It has been well demonstrated in the discrete-time AR regression literature that the estimator $\widehat{\varphi}_1$ has a downward bias. Hence, the estimation bias of $\widehat{\kappa}_{Euler}$ is positive. In contrast, the representation of H suggests that the discretization bias $-H/\Delta$ should be negative when $\kappa\Delta$ is reasonably small. Therefore, the two bias components have opposite signs and partially cancel each other out. As a result, the magnitude of the total bias of $\widehat{\kappa}_{Euler}$ is reduced. Theorem 4.3.3 gives the explicit form of the bias formula of $\widehat{\kappa}_{Euler}$ developed in Wang et al. (2011).

Theorem 4.3.3 *Assume the short-term rate $r(t)$ follows the OU process in (4.2.2) with $\kappa > 0$. The estimator $\widehat{\kappa}_{Euler}$ has the bias of*

$$Bias\left(\widehat{\kappa}_{Euler}\right) = \frac{1+3\exp\{-\kappa\Delta\}}{T} - \frac{H}{\Delta} + o\left(\frac{1}{T}\right), \tag{4.3.11}$$

where $\left(1+3\exp\{-\kappa\Delta\}\right)/T > 0$ is the estimation bias, and $-\frac{H}{\Delta} < 0$ is the discretization bias.

Remark 4.3.4 *Wang et al. (2011) prove that as long as $T < 8\left(1+O(\Delta)\right)/\left(\kappa^2\Delta\right)$, the first term of the bias formula (4.3.11) dominates the second term in magnitude. In other words, the estimation bias of $\widehat{\kappa}_{Euler}$ has a greater magnitude than the discretization bias. Therefore, the total bias of $\widehat{\kappa}_{Euler}$ takes the same sign as the estimation bias, which is positive. However, because the discretization bias is negative, the total bias of $\widehat{\kappa}_{Euler}$ takes a smaller value than the estimation bias. Notably, $8/\left(\kappa^2\Delta\right)$ accounts for very large values in empirically relevant cases.*

To compare the Euler estimator $\widehat{\kappa}_{Euler}$ with that of the exact ML estimator $\widehat{\kappa}_n^{ML}$, we have

$$\begin{aligned}
&\text{Bias}\left(\widehat{\kappa}_n^{ML}\right) - \text{Bias}\left(\widehat{\kappa}_{Euler}\right) \\
=\ & \frac{5+2e^{\kappa\Delta}+e^{2\kappa\Delta}}{2T} - \frac{1+3e^{-\kappa\Delta}}{T} + \frac{H}{\Delta} + o\left(\frac{1}{T}\right) \\
=\ & \frac{3+2e^{\kappa\Delta}+e^{2\kappa\Delta}-6e^{-\kappa\Delta}}{2T} + \frac{H}{\Delta} + o\left(\frac{1}{T}\right) \\
=\ & \frac{3+2(1+\kappa\Delta)+(1+2\kappa\Delta)-6(1-\kappa\Delta)+O\left(\Delta^2\right)}{2T} + \frac{H}{\Delta} + o\left(\frac{1}{T}\right) \\
=\ & \frac{10\kappa\Delta+O\left(\Delta^2\right)}{2T} + \frac{H}{\Delta} + o\left(\frac{1}{T}\right) > 0.
\end{aligned}$$

Hence, the quasi-ML estimator $\widehat{\kappa}_{Euler}$ has a smaller finite-sample bias than the exact ML estimator when $\kappa > 0$ and Δ is small.

Moreover, from the equation in (4.3.10), the variance of $\widehat{\kappa}_{Euler}$ has the representation of

$$Var\left(\widehat{\kappa}_{Euler}\right) = \frac{1}{\Delta^2} Var\left(\widehat{\varphi}_1\right),$$

In contrast, the delta method suggests that the exact ML estimator $\widehat{\kappa}_n^{ML}$ has the variance of

$$Var\left(\widehat{\kappa}_n^{ML}\right) = \frac{1}{\Delta^2}\frac{1}{\varphi_1^2}\left(Var\left(\widehat{\varphi}_1\right) + O\left(n^{-3/2}\right)\right),$$

when $n \to \infty$ with a fixed Δ. When $\kappa > 0$, which makes $\varphi_1 = \exp\{-\kappa\Delta\} < 1$, we have

$$Var\left(\widehat{\kappa}_{Euler}\right) < Var\left(\widehat{\kappa}_n^{ML}\right).$$

Therefore, the Euler estimator $\widehat{\kappa}_{Euler}$ can perform better than the exact ML estimator in terms of finite-sample variance.

In addition to Euler discretization, Wang et al. (2011) examine the finite-sample performance of alternative quasi-ML estimators from a variety of approximation methods, including the trapezoidal approximation, the Milstein approximation, and the Norman approximation. Their results show that when a higher-order approximation method is adopted, the discretization bias of the estimator decreases, but the estimation bias increases. Because the discretization bias plays a role in offsetting the estimation bias, a higher-order approximation leads to an estimator with a more significant total bias. Therefore, the total bias of the Euler estimator is smaller than that of the estimator based on any higher-order approximation.

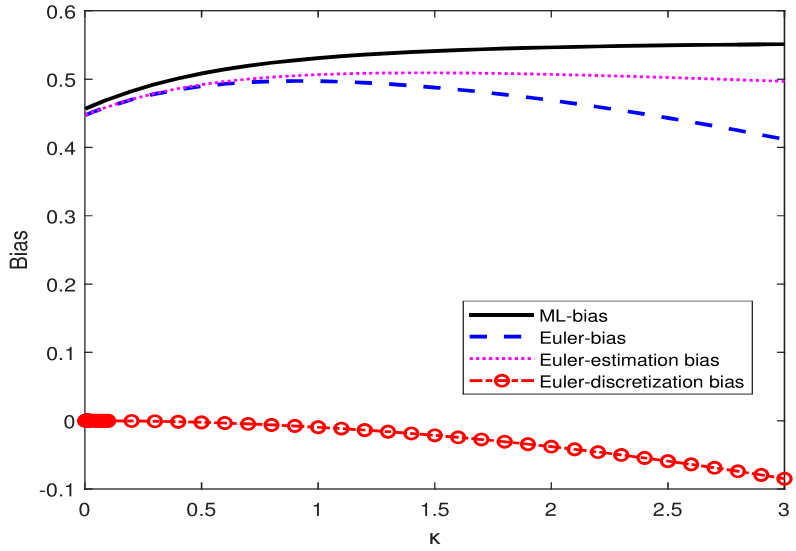

Figure 4.3: Bias of alternative estimators of κ for the OU process with a known mean when $\Delta = 1/52$ and $T = 4$.

Wang et al. (2011) further extend their results to include the OU process with a known mean, general univariate diffusion processes (such as the CIR process), and multivariate OU processes.

Figures 4.3-4.4 provide simulation results to compare the finite-sample performance of κ's alternative estimators. For simplicity, we simulate data from the OU process with a known mean of $\mu = 0$:

$$dr(t) = -\kappa r(t)\, dt + \sigma dW(t).$$

For each simulation, we let $T = 4$ and $\Delta = 1/52$, which corresponds to four years of weekly observations. We replicate the simulation 1,000 times to calculate the bias and the variance of the Euler estimator and the exact ML estimator of κ. Figure 4.3 reports the bias values. Figure 4.4 displays the simulated variance.

Figures 4.3-4.4 clearly show that the Euler estimator performs better than the exact ML estimator in terms of both finite-sample bias and variance. Figure 4.3 also plots the estimation bias and the discretization bias of the Euler estimator, showing that these two types of bias have opposite signs. Therefore, the estimation bias is offset partially by the discretization bias, reducing the magnitude of the total bias of the Euler estimator.

4.3.4 Bias for general diffusion processes

Estimation of general diffusion processes also encounters the issue of finite sample bias. Consider the following general diffusion process with a linear drift:

$$dr(t) = \kappa(\mu - r(t))\, dt + \sigma q(r(t); \psi)\, dW(t), \tag{4.3.12}$$

where ψ is the vector of unknown parameters, and $q(r(t); \psi)$ is a function of $r(t)$ and ψ. The diffusion process in (4.3.12) includes most of the models listed in Table 1 as special cases. The exact discretization of this general diffusion process takes the form of

$$r_t = e^{-\kappa\Delta} r_{t-1} + \mu\left(1 - e^{-\kappa\Delta}\right) + \int_{(t-1)\Delta}^{t\Delta} e^{\kappa(s-t\Delta)} \sigma q(r(s); \psi)\, dW(s).$$

The above discretization gives an AR(1) regression with heteroskedasticity. The estimation of the AR root suffers from finite-sample bias, which translates into the estimation of κ in the end.

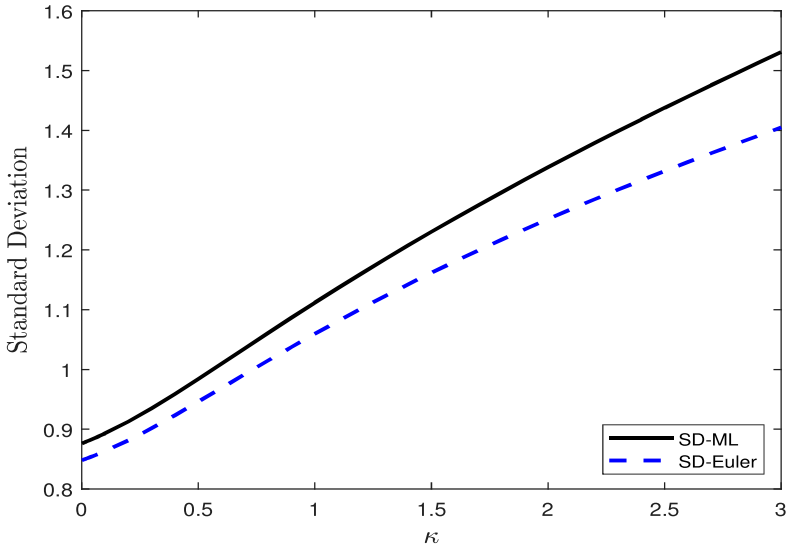

Figure 4.4: Standard deviations of alternative estimators of κ for the OU process with a known mean when $\Delta = 1/52$ and $T = 4$.

Due to the complexity of the function $q(r(t);\psi)$, the transitional density of the general diffusion process is not analytically available. Thus, quasi-ML methods based on various approximation methods are often adopted. For example, Nowman (1997) proposes the approximation method as

$$\begin{aligned}\int_{(t-1)\Delta}^{t\Delta} e^{\kappa(s-t\Delta)}\sigma q(r(s);\psi)\,dW(s) &\approx q(r_{t-1};\psi)\int_{(t-1)\Delta}^{t\Delta} e^{\kappa(s-t\Delta)}\sigma dW(s)\\ &:= q(r_{t-1};\psi)\,\varepsilon_t,\end{aligned}$$

which leads to the Nowman discretization

$$r_t \approx e^{-\kappa\Delta}r_{t-1} + \mu\left(1-e^{-\kappa\Delta}\right) + q(r_{t-1};\psi)\,\varepsilon_t, \tag{4.3.13}$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence with $N\left(0,\sigma^2\left(1-e^{-2\kappa\Delta}\right)/(2\kappa)\right)$ distribution. Then, the approximate transition density function, the log-likelihood function, and the associated quasi-ML estimators of parameters can be obtained. The Nowman discretization can provide an excellent approximation of the exact discretization when the sampling interval Δ is reasonably small.

Take the famous CIR process as an example, where $r(t)$ follows the general diffusion process (4.3.12) with $q(r(t);\psi) = \sqrt{r(t)}$:

$$dr(t) = \kappa(\mu - r(t))\,dt + \sigma\sqrt{r(t)}dW(t). \tag{4.3.14}$$

The Nowman method gives the following approximate discretization

$$r_t \approx e^{-\kappa\Delta}r_{t-1} + \mu\left(1-e^{-\kappa\Delta}\right) + \sqrt{r_{t-1}}\varepsilon_t.$$

Tang and Chen (2009) develop the corresponding quasi-ML estimator of κ as

$$\widehat{\kappa}_{CIR-Nowman} = -\frac{1}{\Delta}\ln\left(\widehat{\beta}_1\right),$$

where $\widehat{\beta}_1$ is the quasi-ML estimator of the AR root in the Nowman discretization that takes the form of

$$\widehat{\beta}_1 = \frac{n^{-2}\sum_{t=1}^{n} r_t \sum_{t=1}^{n} r_{t-1}^{-1} - n^{-1}\sum_{t=1}^{n} r_t r_{t-1}^{-1}}{n^{-2}\sum_{t=1}^{n} r_{t-1}\sum_{t=1}^{n} r_{t-1}^{-1} - 1}.$$

$\widehat{\beta}_1$ is a biased estimator, whose bias translates into $\widehat{\kappa}_{CIR-Nowman}$. Tang and Chen (2009) provide an explicit bias formulae for $\widehat{\kappa}_{CIR-Nowman}$, which we report in Theorem 4.3.5.

Theorem 4.3.5 *For a stationary CIR process as given in (4.3.14) with $2\kappa\mu/\sigma^2 > 2$, the quasi-ML estimator $\widehat{\kappa}_{CIR-Nowman}$ has the bias as*

(a) when $n \to \infty$ with a fixed Δ,

$$Bias\left(\widehat{\kappa}_{CIR-Nowman}\right) = \frac{1}{T} B_3\left(\theta; \Delta\right) + O\left(n^{-2}\right);$$

(b) when $n \to \infty$ with $T \to \infty$ and $\Delta \to 0$, and for some $c > 2$, $T\Delta^{1/c} \to \infty$,

$$Bias\left(\widehat{\kappa}_{CIR-Nowman}\right) = \frac{4}{T} + o\left(T^{-1}\right),$$

where $B_3\left(\theta; \Delta\right)$ takes a complicated form that can be found in Tang and Chen (2009), which is omitted here for space limit.

Wang et al. (2011) propose using the Euler approximation to estimate the CIR model and compare the finite-sample performance of the Euler estimator with that of the estimator from the Nowman approximation. Unlike the OU process, the quasi-ML estimator of the diffusion parameter σ^2 of the CIR process has a substantial finite-sample bias. Detailed analysis and the bias formula can be found in Tang and Chen (2009).

4.3.5 Bias in pricing contingent claims

As studied in Subsection 4.2.2, the pricing formulae of contingent claims, such as bonds and bond options, are nonlinear functions of the parameters in the diffusion process. Therefore, the bias in estimating SDEs' parameters translates into the price calculation of contingent claims. In addition, the estimation variance of the SDEs' parameters may significantly contributes to the bias in pricing contingent claims, because of the nonlinearity of the pricing formulae. In other words, even if all the parameter estimations have no bias, the estimation of the contingent claims prices can still suffer from severe bias problem.

To investigate the bias issue in pricing contingent claims, let us assume the short-term rate $r\left(t\right)$ follows the OU process defined in (4.2.2), and all the parameters are known except for κ. Use $P\left(\kappa\right)$, a function of κ, to denote the price of contingent claims, which is either the price of a zero-coupon bond as given in (4.2.4) or the price of a bond option as listed in (4.2.8). Let $\widehat{\kappa}_n^{ML}$ be the ML estimator of κ defined in (4.3.2). By virtue of the invariance principle, replacing κ in $P\left(\kappa\right)$ by $\widehat{\kappa}_n^{ML}$ generates the ML estimate of $P\left(\kappa\right)$, denoted by

$$\widehat{P}_n^{ML} = P\left(\widehat{\kappa}_n^{ML}\right).$$

The Taylor expansion about point κ leads to the following equation

$$\widehat{P}_n^{ML} = P\left(\kappa\right) + \frac{\partial P\left(\kappa\right)}{\partial \kappa}\left(\widehat{\kappa}_n^{ML} - \kappa\right) + \frac{\partial^2 P\left(\kappa\right)}{\partial \kappa^2}\left(\widehat{\kappa}_n^{ML} - \kappa\right)^2 + \cdots.$$

Taking expectations on both sides of the above equation, we have

$$E\left(\widehat{P}_n^{ML}\right) - P\left(\kappa\right) = \frac{\partial P\left(\kappa\right)}{\partial \kappa} E\left(\widehat{\kappa}_n^{ML} - \kappa\right) + \frac{\partial^2 P\left(\kappa\right)}{\partial \kappa^2} E\left(\widehat{\kappa}_n^{ML} - \kappa\right)^2 + \cdots. \tag{4.3.15}$$

Equation (4.3.15) reveals two important facts of the bias of $\widehat{P}_n^{ML}$. First, the bias of $\widehat{\kappa}_n^{ML}$ translates into $\widehat{P}_n^{ML}$, amplified or reduced by the multiplier $\partial P\left(\kappa\right)/\partial\kappa$. Second, the mean-squared-error (MSE) of $\widehat{\kappa}_n^{ML}$ contributes to the bias of $\widehat{P}_n^{ML}$. Its influence on the bias is scaled by

$\partial^2 P(\kappa)/\partial\kappa^2$. When the price formula $P(\kappa)$ has strong nonlinearity, which makes the magnitude of $\partial^2 P(\kappa)/\partial\kappa^2$ significant, the finite-sample bias of $\widehat{P}_n^{ML}$ can be very substantial even if $\widehat{P}_n^{ML}$ has no bias.

In a more realistic situation where all the parameters $\theta = (\kappa, \mu, \sigma)$ in the OU process are unknown, the finite-sample bias and variance of the estimation of each parameter contribute to the bias of the contingent-claim price $P(\theta)$. Define

$$\widehat{P}_n^{ML} := P\left(\widehat{\theta}_n^{ML}\right) = P\left(\widehat{\kappa}_n^{ML}, \widehat{\mu}_n^{ML}, \widehat{\sigma}_n^{ML}\right).$$

First expanding $P\left(\widehat{\theta}_n^{ML}\right)$ about the true value of the parameter vector θ, then taking the expectations, we can have

$$\begin{aligned} E\left(\widehat{P}_n^{ML}\right) \approx\ & P(\theta) + \left[\begin{array}{ccc} \frac{\partial P(\theta)}{\partial\kappa} & \frac{\partial P(\theta)}{\partial\mu} & \frac{\partial P(\theta)}{\partial\sigma} \end{array}\right] \left[\begin{array}{c} E\left(\widehat{\kappa}_n^{ML}\right) - \kappa \\ E\left(\widehat{\mu}_n^{ML}\right) - \mu \\ E\left(\widehat{\sigma}_n^{ML}\right) - \sigma \end{array}\right] \\ & + E\left\{ \left(\widehat{\theta}_n^{ML} - \theta\right) \left[\begin{array}{ccc} \frac{\partial^2 P(\theta)}{\partial\kappa^2} & \frac{\partial^2 P(\theta)}{\partial\kappa\partial\mu} & \frac{\partial^2 P(\theta)}{\partial\kappa\partial\sigma} \\ \frac{\partial^2 P(\theta)}{\partial\mu\partial\kappa} & \frac{\partial^2 P(\theta)}{\partial\mu^2} & \frac{\partial^2 P(\theta)}{\partial\mu\partial\sigma} \\ \frac{\partial^2 P(\theta)}{\partial\sigma\partial\kappa} & \frac{\partial^2 P(\theta)}{\partial\sigma\partial\mu} & \frac{\partial^2 P(\theta)}{\partial\sigma^2} \end{array}\right] \left(\widehat{\theta}_n^{ML} - \theta\right)' \right\}, \end{aligned}$$

where $\widehat{\theta}_n^{ML} - \theta$ is a 1×3 row vector and $(\cdot)'$ denotes the vector transpose. In this case, besides the bias and variance, the covariance of the estimators $\left(\widehat{\kappa}_n^{ML}, \widehat{\mu}_n^{ML}, \widehat{\sigma}_n^{ML}\right)$ also contributes to the bias of $\widehat{P}_n^{ML}$.

We provide some simulation results in Table 4.2 to further illustrate the bias problem in pricing contingent claims. The simulation settings are similar to that in Phillips and Yu (2005, 2009), where the bias problem has been systematically studied and more simulation results can be found.

The left panel of Table 4.2 reports the simulation results for the Vasicek model. The data are simulated from the OU process with $\kappa = 0.02$, $\mu = 0.12$, and $\sigma = 0.01$. Let κ be the only unknown parameter and $\widehat{\kappa}$ the ML estimator defined in (4.3.6). Use BP to denote the price of a three-year discount bond with the terminal payoff $L = \$100$, and OP as the price of a two-year call option written on the discount bond. Set the annual interest rate to be 6%, and the strike price of the option as $K = 105 \times \exp\{-3 \times 0.06\}$. The true values of BP and OP are calculated by using the pricing formulae given in (4.2.7) and (4.2.8), respectively. We assume the market price of risk to be $\lambda = 0$ for simplicity. Replacing κ in the pricing formulae by $\widehat{\kappa}$ yields the ML estimates of the bond price and the option price, denoted by $\widehat{BP}_n^{ML}$ and $\widehat{OP}_n^{ML}$, respectively. For each sample path, we simulate 5,000 data with $T = 30$ and $\Delta = 1/250$, corresponding to 30 years of daily observations. We replicate the simulation 5,000 times to calculate the bias of $\widehat{\kappa}$, $\widehat{BP}_n^{ML}$ and $\widehat{OP}_n^{ML}$.

Two features are apparent in the left panel of Table 4.2. First, the ML estimator $\widehat{\kappa}$ is severely upward biased. The bias percentage is 215.09%. Second, $\widehat{BP}_n^{ML}$ and $\widehat{OP}_n^{ML}$ have negative bias. Although economically significant, the bias percentage of the bond price is -0.909% only. In contrast, the bias percentage of the option price is much more substantial, reaching -7.54%.

The right panel of Table 4.2 reports simulation results for the Black-Scholes option pricing model, in which we can see that the parameter of the underlying diffusion process has almost no estimation bias, whereas the estimation of the option price has substantial bias. In the Black-Scholes model, the stock price $S(t)$ is assumed to be a process of geometric Brownian motion (GBM):

$$dS(t) = \alpha S(t)\, dt + \sigma S(t)\, dB(t), \tag{4.3.16}$$

where α and σ are two constants. Consider a European call option written on this stock with a strike price K and a maturity τ. The Black-Scholes option pricing formula takes the form of

$$OP = S(t)\,\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \tag{4.3.17}$$

Table 4.2: Bias in pricing contingent claims. The left panel reports the simulation results for the Vasicek model. $\widehat{BP}_n^{ML}$ is the estimate of the price of a three-year discount bond; $\widehat{OP}_n^{ML}$ represents the estimate of the price of a two-year call option written on the discount bond. The right panel gives the simulation results for the Black-Scholes model, where $\widehat{OP}_n^{ML}$ denotes the estimate of the price of a one-week call option written on a stock.

	Vasicek model			Black-Scholes	
	$\widehat{\kappa}_n^{ML}$	$\widehat{BP}_n^{ML}$	$\widehat{OP}_n^{ML}$	$\widehat{\sigma}_n^{2,ML}$	$\widehat{OP}_n^{ML}$
True	0.02	83.12	5.5098	0.45	5.44
Mean	0.063	82.37	5.0944	0.4478	6.25
Bias (in %)	215.09	-0.909	-7.54	-0.4791	14.96

where $d_1 = \left[\ln\left(S\left(t\right)/K\right) + \left(r + 0.5\sigma^2\right)\tau\right] / \left(\sigma\sqrt{\tau}\right)$, $d_2 = d_1 - \sigma\sqrt{\tau}$, $\Phi\left(\cdot\right)$ denotes the cumulative function of the standard normal distribution, and r is the averaged annual interest rate from time t to $t+\tau$. In this example, the only unknown parameter in the option price is σ^2. Lo (1988) advocates the following ML estimator of σ^2 with discrete-time price observations $\{S_t\}_{t=0}^n$:

$$\widehat{\sigma}_n^{2,ML} = \frac{1}{n}\sum_{t=0}^{n-1}\left(\ln\frac{S_{t+1}}{S_t} - n^{-1}\sum_{t=0}^{n-1}\ln\frac{S_{t+1}}{S_t}\right)^2. \tag{4.3.18}$$

Putting $\widehat{\sigma}_n^{2,ML}$ into the pricing formula yields the ML estimate of the option price $\widehat{OP}_n^{ML} = OP\left(\widehat{\sigma}_n^{2,ML}\right)$.

We simulate data from the GBM defined in (4.3.16) with $\alpha = 0$, $\sigma^2 = 0.45$, $\Delta = 1/250$, and $n = T/\Delta = 250$, corresponding to one year of daily observations. Assume that the initial price of the stock is $S\left(0\right) = 100$. Consider a deep out-of-money call option with $K = 1.4S\left(0\right)e^{r\tau}$, $r = 6\%$, and $\tau = 5/250$ (the option expires in one week). For each simulated sample path, calculate the ML estimates of $\widehat{\sigma}_n^{2,ML}$ and $\widehat{OP}_n^{ML}$. We replicate the experiment 5,000 times to obtain the bias of the ML estimates, and report them in the right panel of Table 4.2.

The simulation results show that the ML estimator of the parameter in the diffusion price, $\widehat{\sigma}_n^{2,ML}$, has almost no bias. In contrast, the estimate of the option price, $\widehat{OP}_n^{ML}$, is severely upward biased, with the bias percentage of 14.96%. The bias in pricing option is attributable to the nonlinearity of the price formula with respect to σ^2. In the same spirit of Equation (4.3.15), it is easy to get

$$\begin{aligned} & E\left(\widehat{OP}_n^{ML}\right) - OP\left(\sigma^2\right) \\ = \; & \frac{\partial OP\left(\sigma^2\right)}{\partial\sigma^2}E\left(\widehat{\sigma}_n^{2,ML} - \sigma^2\right) + \frac{\partial^2 OP\left(\sigma^2\right)}{\partial\left(\sigma^2\right)^2}E\left(\widehat{\sigma}_n^{2,ML} - \sigma^2\right)^2 + \cdots \end{aligned}$$

When $\widehat{\sigma}_n^{2,ML}$ has almost no bias, the contribution of the first term in the above equation to the option pricing bias becomes negligible. However, $\widehat{\sigma}_n^{2,ML}$ can have significant variance, and hence significant MSE. Phillips and Yu (2009) point out that the price of the deep-out-money call option is a strongly nonlinear function of σ^2. That means $\frac{\partial^2 OP\left(\sigma^2\right)}{\partial\left(\sigma^2\right)^2}$ has a large magnitude. Therefore, the MSE of $\widehat{\sigma}_n^{2,ML}$ is amplified by the multiplier $\frac{\partial^2 OP\left(\sigma^2\right)}{\partial\left(\sigma^2\right)^2}$, making $\widehat{OP}_n^{ML}$ has substantial finite-sample bias.

4.4 Bias-correction Methods

To reduce the estimation bias of κ in the OU process defined in (4.2.2), it is possible to simply apply the explicit bias formulae, such as those in (4.3.3), (4.3.8), and (4.3.11), to perform bias correction. However, no bias formulae are available when interest quantities are the parameters in either general diffusion processes or the prices of contingent claims, such as bonds and options. This section introduces two bias-correction methods that work well without knowing the bias formula: the jackknife method and the indirect inference method. Both methods can effectively reduce bias. The jackknife technique is easy to implement. In comparison, the indirect inference approach can simultaneously reduce bias and variance, at least in some situations.

4.4.1 Jackknife method

The jackknife method was proposed by Quenouille (1956) to reduce the finite sample bias in parametric estimation problems. It has many applications in discrete time models; see, for example, Efron (1982), Shao and Tu (1995), and Hahn and Newey (2004). Phillips and Yu (2005) implement this method in estimating continuous-time diffusion processes and contingent claims prices.

To illustrate the idea of the jackknife technique, let us consider the OU process (4.2.2) as an example. The exact ML estimator $\widehat{\kappa}_n^{ML}$ has been defined in (4.3.2), where $n = T/\Delta$ represents the sample size. Formula (4.3.3) describes the finite-sample bias of $\widehat{\kappa}_n^{ML}$, which is rewritten here for convenience:

$$E\left(\widehat{\kappa}_n^{ML}\right) = \kappa + \frac{c}{T} + O\left(n^{-2}\right) = \kappa + \frac{c}{n\Delta} + O\left(n^{-2}\right),$$

where $c = 5/2 + e^{\kappa\Delta} + e^{2\kappa\Delta}/2$. Now, decompose the whole sample into m consecutive subsamples, each of which has $l = n/m$ observations. Let $\widehat{\kappa}_{li}^{ML}$ denote the ML estimator of κ by using the observations in the i^{th} subsample, for $i = 1, 2, \ldots, m$. From the formula (4.3.3), $\widehat{\kappa}_{li}^{ML}$ has the finite-sample bias

$$E\left(\widehat{\kappa}_{li}^{ML}\right) = \kappa + \frac{m}{n\Delta}c + O\left(\left(\frac{n}{m}\right)^{-2}\right).$$

Define the jackknife estimator as

$$\widehat{\kappa}_{jack} = \frac{m}{m-1}\widehat{\kappa}_n^{ML} - \frac{\sum_{i=1}^{m}\widehat{\kappa}_{li}^{ML}}{m^2 - m}. \tag{4.4.1}$$

The expectation of the jackknife estimator is

$$\begin{aligned} & E\left(\widehat{\kappa}_{jack}\right) = \frac{m}{m-1}\left(\kappa + \frac{c}{n\Delta}\right) - \frac{\sum_{i=1}^{m}\left(\kappa + \frac{m}{n\Delta}c\right)}{m^2 - m} + O\left(n^{-2}\right) \\ = \; & \kappa + O\left(n^{-2}\right). \end{aligned}$$

The above formula clearly shows that the leading term of the bias of $\widehat{\kappa}_n^{ML}$ is canceled out by the counterparts of $\widehat{\kappa}_{li}^{ML}$. As a result, the bias of the jackknife estimator $\widehat{\kappa}_{jack}$ is reduced to the order of $O\left(n^{-2}\right)$.

To apply the jackknife technique, people often choose $m = 2$ for simplicity. However, choosing larger values of m can help reduce the finite-sample variance of the jackknife estimator. The variance of $\widehat{\kappa}_{jack}$ with small values of m is often greater than that of the ML estimator $\widehat{\kappa}_n^{ML}$. However, under broad conditions, these two estimators can have the same asymptotic variance. For the stationary OU process with $\kappa > 0$, Tang and Chen (2009) prove that $\sqrt{n}\left(\widehat{\kappa}_n^{ML} - \kappa\right)$ converges to a normally distributed random variable as $n \to \infty$ with a fixed Δ. From the formula in (4.4.1), we have

$$\sqrt{n}\left(\widehat{\kappa}_{jack} - \kappa\right) = \frac{m\sqrt{n}\left(\widehat{\kappa}_n^{ML} - \kappa\right)}{m-1} - \frac{1}{m-1}\left\{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\sqrt{l}\left(\widehat{\kappa}_{li}^{ML} - \kappa\right)\right\}.$$

The stationarity of the OU process makes $\widehat{\kappa}_{li}^{ML}$, for $i = 1, 2, \ldots, m$, a weakly dependent sequence. Hence, when $n \to \infty$ with $m \to \infty$ and $1/l + m/l \to 0$, the second term in the above equation goes to zero, making $\widehat{\kappa}_{jack}$ have the same limiting distribution as $\widehat{\kappa}_n^{ML}$.

When the quantity of interest becomes the price of the contingent claims, as in the case of bonds and options, plugging the jackknife estimator $\widehat{\kappa}_{jack}$ into the corresponding pricing formula may not lead to a better estimate of the price than the ML estimate in terms of finite-sample bias. This is because, as shown by Equation (4.3.15), both the bias and the MSE of $\widehat{\kappa}_{jack}$ contribute to the bias of the price estimation. The MSE part is critically important for pricing bias when price formulae have strong nonlinearity. The jackknife estimator $\widehat{\kappa}_{jack}$ has a smaller bias but a greater variance than $\widehat{\kappa}_n^{ML}$. More often than not, $\widehat{\kappa}_{jack}$ has a larger MSE than $\widehat{\kappa}_n^{ML}$. Accordingly, the plug-in estimator $P\left(\widehat{\kappa}_{jack}\right)$ may have a greater bias than the ML estimator $P\left(\widehat{\kappa}_n^{ML}\right)$.

To obtain a bias-reduced estimator for the contingent claims price, Phillips and Yu (2005) recommend applying the jackknife technique directly to the price estimation through the following procedure:

1. Obtain the ML estimate $\widehat{\kappa}_n^{ML}$ using the entire sample;
2. Calculate the price estimate $P\left(\widehat{\kappa}_n^{ML}\right)$;
3. Use subsamples to obtain the estimates $\left\{\widehat{\kappa}_{li}^{ML}\right\}_{i=1}^{m}$;
4. Calculate the price estimates $P\left(\widehat{\kappa}_{li}^{ML}\right)$, for $i = 1, \ldots, m$;
5. Obtain the final estimate of the price by using the jackknife estimator

$$\widehat{P}_{jack} = \frac{m}{m-1} P\left(\widehat{\kappa}_n^{ML}\right) - \frac{\sum_{i=1}^{m} P\left(\widehat{\kappa}_{li}^{ML}\right)}{m^2 - m}. \tag{4.4.2}$$

The estimation procedure above reveals another merit of the jackknife method: Implementing the jackknife technique does not require any knowledge of the analytical form of the bias. Moreover, provided that the bias of $P\left(\widehat{\kappa}_n^{ML}\right)$ can be expanded asymptotically in a series of increasing powers of n^{-1}, it can be proved that the jackknife estimator $\widehat{P}_{jack}$ has a bias of order $O\left(n^{-2}\right)$, not $O\left(n^{-1}\right)$.

Table 4.3 summarizes certain simulation results to reveal the performance of the jackknife technique in reducing the bias of option pricing. The simulation settings are the same as in Table 4.2. Again, we use OP to denote the price of a two-year call option written on a discount bond that expires in three years with the terminal payoff $L = \$100$. $\widehat{OP}_n^{ML}$ represents the ML estimator of the option price, and $\widehat{OP}_n^{Jack}$ is the jackknife estimator with $m = 2$. $\widehat{OP}_n^{II}$ is the indirect-inference estimator that will be introduced in Subsection 4.4.2.

Table 4.3 clearly shows that the jackknife technique can significantly reduce the estimation bias compared with the ML approach. The bias percentage of the ML estimator $\widehat{OP}_n^{ML}$ is -7.54%. The jackknife estimator $\widehat{OP}_n^{Jack}$ reduces that percentage to -4.92%. However, Table 4.3 also shows that the jackknife estimator has a greater standard error than the ML estimator and slightly increases the MSE.

4.4.2 Indirect inference method

Many models in economics and finance are too complex to permit the likelihood function to be constructed analytically, but they can readily be simulated. To estimate these models effectively, Smith Jr (1993) and Gourieroux et al. (1993) develop a simulation-based estimation method that is called the indirect inference (II) approach. This technique can also be applied to perform bias correction in estimation. Among many other applications, the II technique is used by Phillips and Yu (2009) to correct the bias in pricing contingent claims, Gouriéroux et al. (2010) to reduce bias

Table 4.3: Comparison of alternative estimators in option pricing. Data are simulated from the OU process in (4.2.2) with $\kappa = 0.02$. A two-year call option written on a discount bond that expires in three years with the terminal payoff $L = 100$ is considered. $\widehat{OP}_n^{ML}$, $\widehat{OP}_n^{Jack}$, and $\widehat{OP}_n^{SM}$ are the ML estimator, the jackknife estimator with $m = 2$, and the indirect inference estimator, respectively.

True value: $OP = 5.5098$			
Estimation Method	$\widehat{OP}_n^{ML}$	$\widehat{OP}_n^{Jack}$	$\widehat{OP}_n^{II}$
Mean	5.0944	5.2386	5.3923
Bias (in %)	-7.54	-4.9226	-2.134
Std err	0.7712	0.8517	0.7434
RMSE	0.8760	0.8939	0.7526

in estimating dynamic panel models, and Jiang et al. (2018) to reduce bias in structural change-point estimation.

This subsection takes the OU process defined in (4.2.2), with κ being the only unknown parameter, as an example to introduce the idea of the II method. We first consider the estimation of κ and then study the estimation of contingent claims prices.

Assume n is the number of observations available. Then, $\widehat{\kappa}_n^{ML}$ defined in (4.3.2) gives the ML estimate of κ. Empirical studies in the asset pricing literature suggest that reasonable values of κ fall into the interval $[0, 3]$. For any possible choice of $\kappa \in [0, 3]$, we can simulate data straightforwardly from the exact discretization of the OU process given in (4.3.1). Replicating the simulation K times generates K simulated data sets, denoted by $\widetilde{R}^{(j)} = \left\{\widetilde{r}_1^{(j)}, \widetilde{r}_2^{(j)}, \ldots, \widetilde{r}_n^{(j)}\right\}$, for $j = 1, 2, \ldots, K$. For the jth simulated data set, an ML estimate $\widetilde{\kappa}_n^{ML,j}$ can be obtained by using the same formula as in (4.3.2). Then, the indirect inference estimator of κ is defined as

$$\widehat{\kappa}_n^{II} = \arg\min_{\kappa\in[0,3]} \left| \widehat{\kappa}_n^{ML} - \frac{1}{K}\sum_{j=1}^{K} \widetilde{\kappa}_n^{ML,j} \right|. \tag{4.4.3}$$

The logic behind the indirect inference estimator is as follows. By construction, $\widetilde{\kappa}_n^{ML,j}$ should have the same finite sample properties as $\widehat{\kappa}_n^{ML}$. The sample mean of $\widetilde{\kappa}_n^{ML,j}$ can serve as a good approximation to the population expectation of $\widehat{\kappa}_n^{ML}$. In the case where $K \to \infty$, the law of large numbers implies that

$$\frac{1}{K}\sum_{j=1}^{K} \widetilde{\kappa}_n^{ML,j} \xrightarrow{p} E\left(\widetilde{\kappa}_n^{ML,j}\right) := b_n(\kappa),$$

where $\xrightarrow{p}$ denotes convergence in probability and $b_n(\kappa)$ is called the binding function. For each value of $\kappa \in [0, 3]$, $b_n(\kappa)$ gives the expectation of the ML estimator of κ. When $b_n(\kappa)$ is invertible, the II estimator becomes

$$\widehat{\kappa}_n^{II} = \arg\min_{\kappa\in[0,3]} \left| \widehat{\kappa}_n^{ML} - b_n(\kappa) \right| = b_n^{-1}\left(\widehat{\kappa}_n^{ML}\right).$$

This means that the II estimator aims at finding one value of κ that minimizes the distance between the ML estimate $\widehat{\kappa}_n^{ML}$ and its expectation. Note that any bias occurring in $\widehat{\kappa}_n^{ML}$ also happens in the binding function $b_n(\kappa)$. Hence, $\widehat{\kappa}_n^{II}$ can correct the bias appearing in $\widehat{\kappa}_n^{ML}$.

Moreover, the II estimator is possible to possess a smaller finite-sample variance than $\widehat{\kappa}_n^{ML}$. The Taylor expansion of $b_n^{-1}\left(\widehat{\kappa}_n^{ML}\right)$ about the point $b_n(\kappa)$ takes the form of

$$\widehat{\kappa}_n^{II} = b_n^{-1}\left(\widehat{\kappa}_n^{ML}\right)$$

$$= \kappa + \left(\frac{\partial b_n(\kappa)}{\partial \kappa}\bigg|_{\kappa=b_n(\kappa)} \right)^{-1} \left(\widehat{\kappa}_n^{ML} - b_n(\kappa) \right) + o_p\left(\left(\widehat{\kappa}_n^{ML} - b_n(\kappa) \right) \right),$$

which leads to the approximation

$$Var\left(\widehat{\kappa}_n^{II}\right) \approx \left(\frac{\partial b_n(\kappa)}{\partial \kappa}\bigg|_{\kappa=b_n(\kappa)} \right)^{-2} Var\left(\widehat{\kappa}_n^{ML}\right).$$

Therefore, when $\left|\partial b_n(\kappa)/\partial\kappa\right| > 1$, the slope of the binding function being greater than unity, we can have $Var\left(\widehat{\kappa}_n^{II}\right) < Var\left(\widehat{\kappa}_n^{ML}\right)$.

In addition to improving the finite-sample performance, the II estimator $\widehat{\kappa}_n^{II}$ can have the same good asymptotic properties as $\widehat{\kappa}_n^{ML}$. When $\widehat{\kappa}_n^{ML}$ is asymptotically unbiased with $\lim_{n\to\infty} E\left(\widehat{\kappa}_n^{ML}\right) = \kappa$, it has $\lim_{n\to\infty} b_n(\kappa) = \kappa$. Hence, $\widehat{\kappa}_n^{II}$ is asymptotically equivalent to $\widehat{\kappa}_n^{ML}$ as $n \to \infty$.

When the quantity of interest is either the bond price or the option price, constructing the II estimator needs to first set an interval containing possible price values. For any given price value in the interval, take the inverse of the pricing formula to obtain the corresponding value of the parameter in the diffusion process, then generate data by simulations. The concrete steps of applying the II technique to estimate the contingent claims price include:

1. Obtain the ML estimate of the price $\widehat{P}_n^{ML} = P\left(\widehat{\kappa}_n^{ML}\right)$ by using the real observations available.

2. Based on the estimate $\widehat{P}_n^{ML}$, set an interval for possible price values, denoted by Λ.

3. For any given choice of $p \in \Lambda$, invert the pricing formula, Formula (4.2.7) for the bond price and (4.2.8) for the option price, to obtain the corresponding value of the parameter $\kappa = P^{-1}(p)$.

4. Simulate data from the OU process in (4.2.2) with $\kappa = P^{-1}(p)$ to get the simulated sample paths $\widetilde{R}^{(j)} = \left\{\widetilde{r}_1^{(j)}, \widetilde{r}_2^{(j)}, \ldots, \widetilde{r}_n^{(j)}\right\}$, for $j = 1, 2, \ldots, K$. For each simulated sample path, obtain the ML estimate $\widetilde{\kappa}_n^{ML,j}$ and $\widetilde{P}_n^{ML,j}(p) = P\left(\widetilde{\kappa}_n^{ML,j}\right)$.

5. Calculate the sample average of $\widetilde{P}_n^{ML,j}(p)$ as a simulated binding function

$$\widetilde{b}_n(p) = \frac{1}{K}\sum_{j=1}^{K} \widetilde{P}_n^{ML,j}(p). \tag{4.4.4}$$

6. Obtain the II estimate of P by

$$\widehat{P}_n^{II} = \arg\min_{p\in\Lambda} \left| \widehat{P}_n^{ML} - \widetilde{b}_n(p) \right|. \tag{4.4.5}$$

Simulation results reported in Table 4.3 demonstrate the finite-sample performance of the II estimator in pricing bond options. As in Table 4.2, a two-year call option written on a discount bond that expires in three years with the terminal payoff $L = 100$ is considered. We use $\widehat{OP}_n^{II}$ to denote the II estimator of the option price, and replicate the simulation 5,000 times to calculate the binding function.

Three features are clear in Table 4.3. First, compared to the ML estimator $\widehat{OP}_n^{ML}$, the II estimator $\widehat{OP}_n^{II}$ significantly reduces the bias, from -7.54% to -2.13%. Second, the finite-sample variance and the RMSE of $\widehat{OP}_n^{II}$ are also smaller than those of the ML estimator $\widehat{OP}_n^{ML}$. The variance of $\widehat{OP}_n^{ML}$ is 0.7712. In contrast, the variance of $\widehat{OP}_n^{II}$ is only 0.7434. Third,

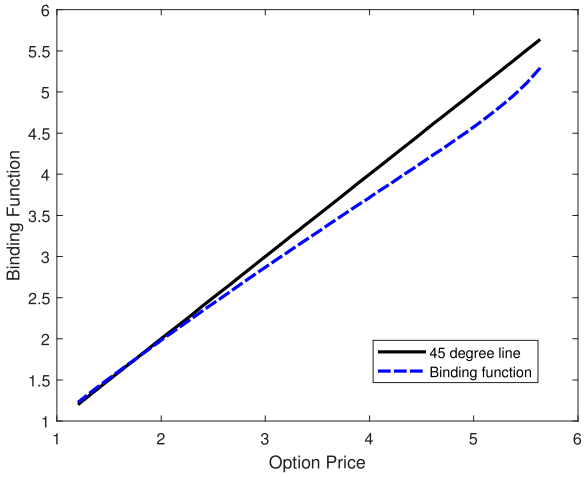

Figure 4.5: Binding function of the II estimator of the option price. The blue dash-dotted line draws the binding function $\widetilde{b}_n(p)$ defined in (4.4.4). The 45° solid line is plotted for comparison.

$\widehat{OP}_n^{II}$ performs better than the jackknife estimator $\widehat{OP}_n^{Jack}$ in terms of both finite-sample bias and variance. It is noticeable that the bias of $\widehat{OP}_n^{II}$ is less than half of the bias of $\widehat{OP}_n^{Jack}$.

To better understand the second feature mentioned above, namely, that $\widehat{OP}_n^{II}$ has a smaller finite-sample variance than $\widehat{OP}_n^{ML}$, Figure 4.5 plots the simulated binding function $\widetilde{b}_n(p)$ as defined in (4.4.4). It is clearly shown that the bonding function is nonlinear in the upper-right corner. The upper-right corner corresponds to a range of small values of κ, a region suggested by most empirical studies in the term structure literature. More importantly, the slope of the binding function in the upper-right corner is greater than unity. When the replication number increases, i.e., $K \to \infty$, it has

$$\lim_{K\to\infty} \widetilde{b}_n(p) = \lim_{K\to\infty} \frac{1}{K} \sum_{j=1}^{K} \widetilde{P}_n^{ML,j}(p) = E\left(\widetilde{P}_n^{ML,j}(p)\right) := b_n(p),$$

and

$$\widehat{OP}_n^{II} = \arg\min_{p\in\Lambda} \left|\widehat{OP}_n^{ML} - b_n(p)\right| = b_n^{-1}\left(\widehat{OP}_n^{ML}\right).$$

The Delta method implies that

$$Var\left(\widehat{OP}_n^{II}\right) \approx \left(\frac{\partial b_n(p)}{\partial p}\bigg|_{p=b_n(p)}\right)^{-2} Var\left(\widehat{OP}_n^{ML}\right).$$

Hence, the slope of the binding function being larger than unity makes the variance of $\widehat{OP}_n^{II}$ smaller than that of $\widehat{OP}_n^{ML}$.

4.5 Conclusion

This chapter briefly reviews recent developments in the finite-sample theory of continuous-time models. It consists of three main parts. The first part introduces the applications of univariate diffusion processes in modeling short-term interest rates. Taking the OU process as an example, this part builds up the connections between the diffusion process and the prices of contingent claims, including both bonds and bond options.

The second part studies the finite-sample properties of the estimation of diffusion processes. We focus on the persistency parameter in the OU process and note that although it enjoys asymptotic efficiency, the ML estimator suffers from severe finite-sample bias. Various bias formulae for the exact ML estimator are introduced. Then, the quasi-ML estimator based on the Euler approximation is presented, along with an explanation of why this quasi-ML approach can perform better than the exact ML method in terms of finite-sample bias and variance. Finally, the bias problems in estimating general diffusion processes and in pricing contingent claims are discussed.

The third part of the chapter advocates two bias-correction methods: the jackknife method and the indirect inference method. Both methods can effectively reduce the finite-sample bias when the explicit bias formula is unknown. The jackknife technique is easy to apply. In contrast, the indirect inference approach may reduce the finite-sample bias and the variance simultaneously.

Bibliography

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *The review of financial studies 9*(2), 385–426.

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

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica 70*(1), 223–262.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics 36*(2), 906–937.

Aït-Sahalia, Y., J. Fan, and D. Xiu (2010). High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association 105*(492), 1504–1517.

Aït-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of financial economics 83*(2), 413–452.

Andersen, T. G. and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of econometrics 77*(2), 343–377.

Ball, C. A. and W. N. Torous (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance 3*(2), 215–238.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society 8*(1), 27–41.

Bergstrom, A. R. (1966). Nonrecursive models as discrete approximations to systems of stochastic differential equations. *Econometrica: Journal of the Econometric Society 34*(1), 173–182.

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

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Brennan, M. J. and E. S. Schwartz (1980). Analyzing convertible bonds. *Journal of Financial and Quantitative analysis 15*(4), 907–929.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*(3), 1209–1227.

Chapman, D. A. and N. D. Pearson (2000). Is the short rate drift actually nonlinear? *The Journal of Finance 55*(1), 355–388.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1980). An analysis of variable rate loan contracts. *The Journal of Finance 35*(2), 389–403.

Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical finance 6*(4), 379–406.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica 68*(6), 1343–1376.

Durham, G. B. (2003). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics 70*(3), 463–487.

Durham, G. B. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics 20*(3), 297–338.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM.

Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica 69*(4), 959–993.

Eraker, B. (2001). Mcmc analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics 19*(2), 177–191.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics 20*(4), 547–557.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric theory 12*(4), 657–681.

Gallant, A. R. and G. Tauchen (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association 93*(441), 10–24.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of applied econometrics 8*(S1), S85–S118.

Gouriéroux, C., P. C. Phillips, and J. Yu (2010). Indirect inference for dynamic panel models. *Journal of Econometrics 157*(1), 68–77.

Gourieroux, C. and R. Sufana (2006). A classification of two-factor affine diffusion term structure models. *Journal of Financial Econometrics 4*(1), 31–52.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*(4), 1295–1319.

Hansen, L. P. and J. A. Scheinkman (1995). Back to the future: Generating moment implications for continuous-time markov processes. *Econometrica 63*(4), 767–804.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies 6*(2), 327–343.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *The journal of finance 42*(2), 281–300.

Hurvicz, L. (1950). Least square bias in time series. In *Statistical Inference in Dynamic Economic Models*, pp. 365–383. New York, Wiley.

Jamshidian, F. (1989). An exact bond option formula. *The Journal of Finance 44*(1), 205–209.

Jiang, L., X. Wang, and J. Yu (2018). New distribution theory for the estimation of structural break point in mean. *Journal of Econometrics 205*(1), 156–176.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika 41*(3-4), 403–404.

Knight, J. L. and J. Yu (2002). Empirical characteristic function in time series estimation. *Econometric Theory 18*(3), 691–721.

Li, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *The Annals of Statistics 41*(3), 1350–1380.

Lo, A. W. (1988). Maximum likelihood estimation of generalized itô processes with discretely sampled data. *Econometric Theory 4*(2), 231–247.

Marriott, F. and J. Pope (1954). Bias in the estimation of autocorrelations. *Biometrika 41*(3/4), 390–402.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science 4*(1), 141–183.

Nowman, K. B. (1997). Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *The journal of Finance 52*(4), 1695–1706.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian journal of statistics 22*(1), 55–71.

Phillips, P. and J. Yu (2009). Simulation-based estimation of contingent-claims prices. *The Review of Financial Studies 22*(9), 3669–3705.

Phillips, P. C. and J. Yu (2005). Jackknifing bond option prices. *The Review of Financial Studies 18*(2), 707–742.

Phillips, P. C. B. (1972). The structural estimation of a stochastic differential equation system. *Econometrica: Journal of the Econometric Society 40*(6), 1021–1041.

Piazzesi, M. (2010). Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques*, pp. 691–766. Elsevier.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika 43*(3/4), 353–360.

RC., M. (1992). *Continuous Time Finance*. Basel: Blackwell.

Santa-Clara, P. (1997). Simulated likeliehood estimation of diffusions with an application to the short tem interest rate. *UCLA: Finance*.

Shao, J. and D. Tu (1995). Applications to time series and other dependent data. In *The Jackknife and Bootstrap*, pp. 386–415. Springer.

Shenton, L. and W. L. Johnson (1965). Moments of a serial correlation coefficient. *Journal of the Royal Statistical Society: Series B (Methodological) 27*(2), 308–320.

Shreve, S. E. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer.

Singleton, K. J. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics 102*(1), 111–141.

Smith Jr, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*(S1), S63–S84.

Sundaresan, S. M. (2000). Continuous-time methods in finance: A review and an assessment. *The Journal of Finance 55*(4), 1569–1622.

Tang, C. Y. and S. X. Chen (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics 149*(1), 65–81.

Trolle, A. B. and E. S. Schwartz (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *The Review of Financial Studies 22*(5), 2007–2057.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics 5*(2), 177–188.

Vinod, H. D. and L. Shenton (1996). Exact moments for autor1egressive and random walk models for a zero or stationary initial value. *Econometric Theory 12*(3), 481–499.

Wang, X., P. C. Phillips, and J. Yu (2011). Bias in estimating multivariate and univariate diffusions. *Journal of Econometrics 161*(2), 228–245.

White, J. S. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika 48*(1-2), 85–94.

Yamamoto, T. and N. Kunitomo (1984). Asymptotic bias of the least squares estimator for multivariate autoregressive models. *Annals of the Institute of Statistical Mathematics 36*(3), 419–430.

Yu, J. (2012). Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics 169*(1), 114–122.

Yu, J. and P. C. Phillips (2001). A gaussian approach for continuous time models of the short-term interest rate. *The Econometrics Journal 4*(2), 210–224.