

ECONOMETRIC ANALYSIS OF CONTINUOUS TIME MODELS: A SURVEY OF PETER PHILLIPS'S WORK AND SOME NEW RESULTS

JUN YU

Econometric analysis of continuous time models has drawn the attention of Peter Phillips for 40 years, resulting in many important publications by him. In these publications he has dealt with a wide range of continuous time models and the associated econometric problems. He has investigated problems from univariate equations to systems of equations, from asymptotic theory to finite sample issues, from parametric models to nonparametric models, from identification problems to estimation and inference problems, and from stationary models to nonstationary and nearly nonstationary models. This paper provides an overview of Peter Phillips's contributions in the continuous time econometrics literature. We review the problems that have been tackled by him, outline the main techniques suggested by him, and discuss the main results obtained by him. Based on his early work, we compare the performance of three asymptotic distributions in a simple setup. Results indicate that the *in-fill* asymptotics significantly outperforms the *long-span* asymptotics and the *double* asymptotics.

JEL Classifications: C22, C32

1. INTRODUCTION

The history of continuous time modeling in economics and finance dates back to more than one hundred years ago when Bachelier (1900) first discussed the use of Brownian motion to analyze price movements and to evaluate contingent claims in financial markets. The use of continuous time models is now widely found in economics, especially in macroeconomics and financial economics.

Many continuous time models used in practice are expressed in terms of a stochastic differential equation (SDE):

$$dX(t) = \mu(X(t))dt + \zeta(dt), X(0) = X_0, \quad (1)$$

I gratefully acknowledge financial support from the Ministry of Education AcRF Tier 2 fund under Grant No. MOE2011-T2-2-096. I would like to thank Peter Phillips for extensive discussions on the subject, seminar participants at NZESG and Singapore Management University, Federico Bandi, Bruce Hansen, Xiaohu Wang, Qiankun Zhou, and especially a referee for helpful comments.

Sim Kee Boon Institute for Financial Economics, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903; email: yujun@smu.edu.sg.

where $X(t) = (X_1(t), \dots, X_M(t))'$ is a M -dimensional continuous time random process and $\zeta(dt)$ is a vector of white noise disturbances with covariance $\sigma(X(t))\sigma'(X(t))$. If $M = 1$, the model is univariate. If $\mu(X(t)) = \mu(X(t), \theta)$ and $\sigma(X(t)) = \sigma(X(t), \theta)$, the model is parametrically specified. Otherwise, the model is nonparametrically specified or semi-parametrically specified. When $\zeta(dt) = \sigma(X(t))dW(t)$, where $W(t)$ is a vector of standard Brownian motion, Model (1) is a diffusion process. In this case, $\mu(X(t))$ and $\sigma(X(t))$ are referred to as the drift term and the diffusion term, respectively. Econometric analysis of Model (1) is based on discretely sampled data, which are assumed to be recorded at $(0, h, 2h, \dots, nh(= T))$ in the time interval $[0, T]$. So $n + 1$ is the total number of observations, h the sampling interval, and T the time span of the data. While the data are assumed equispaced, such an assumption is made for convenience of presentation and may be relaxed. Macroeconomic variables are typically available at the annual and quarterly frequencies, and more recently, at the monthly frequency. As a result, $h = 1, 1/4, 1/12$, corresponds to the annual, quarterly, and monthly frequency, when measuring time in years. In this paper, we use $X(t)$ to represent a continuous time process and X_{ih} to represent a discrete time process. When there is no confusion, denote $X_i := X_{ih}$.

Three alternative sampling schemes have been employed in the literature for developing asymptotic theory, namely,

$$T \rightarrow \infty, \quad h \text{ is fixed, hence } n \rightarrow \infty \quad (\text{A.1})$$

$$T \rightarrow \infty, \quad h \rightarrow 0 \text{ and hence } n \rightarrow \infty \quad (\text{A.2})$$

$$h \rightarrow 0, \quad T \text{ is fixed and hence } n \rightarrow \infty \quad (\text{A.3})$$

Scheme (A1) is referred to as the *long-span* asymptotics, Scheme (A3) the *in-fill* asymptotics, and (A2) the *double* asymptotics. Obviously, (A2) is stronger than (A1) and (A3).

There are strong reasons why continuous time models appeal to economists and financial specialists as “the economy does not cease to exist in between observations” (Bartlett, 1946; Phillips, 1988). On aggregate levels, economic decision making almost always involves many agents and is typically done at different times. With the vast advancement in globalization, economic integration and information technology in the modern era, news arrives at shorter intervals and economic activities take place in a nonstop fashion (Bergstrom and Nowman, 2007). As a result, continuous time models may provide a reasonable approximation to the actual dynamics of economic behavior. Another important advantage with continuous time models is that they provide a convenient mathematical framework for the development of financial economic theory, enabling simple and often analytically tractable ways to price financial assets. Applications of continuous time models to price financial contingent claims have created a field called *mathematical finance*, which has undergone amazingly fast development in the last three decades.

There are other reasons why continuous time models are used in economics and finance. For example, economics, finance, and related fields often distinguish between quantities which are stocks and those which are flows. A stock variable is measured at one specific time, and represents a quantity existing at that point in time, which may have been accumulated in the past. A flow variable is measured over an interval of time. Therefore, a flow would be measured per unit of time. Continuous time models allow for separate treatments of these two types of variables. Another example would be that time aggregation is not an issue in the continuous time setup, whereas it may present obstacles in discrete time models, including some widely used specifications, such as GARCH models (Drost and Nijman, 1993).

One of the most important proponents of the use of continuous time econometric models in macroeconomics is Rex Bergstrom, a New Zealand economist, whose important work can be found in Bergstrom (1966, 1983, 1984, 1985a, 1985b, 1986, 1990) and Bergstrom and Wymer (1976). Phillips (1993) is a summary of the work of Bergstrom; see also Phillips (2010). The use of continuous time models in finance is best seen in two of the most influential papers in financial economics, Black and Scholes (1973) and Merton (1973). Merton (1990) contains the most important work of his in the area. Model (1) is useful not only in economics and finance but also in science.

Directly influenced by Bergstrom, Peter Phillips completed his Master's thesis in 1971 at the University of Auckland, under the supervision of Bergstrom, working on estimation issues of continuous time models. A research article from his Master's thesis appeared in *Econometrica* in 1972. This was a remarkable kickoff to an illustrious career. Since then, Phillips has spent 40 years working on continuous time models, leading to more than 20 publications in the area. His written contributions cover many important issues in the literature, from univariate equations to systems of equations, from asymptotic theory to finite sample issues, from parametric models to nonparametric models, from identification problems to estimation and inference problems, and from stationary models to nonstationary and nearly nonstationary models. Naturally, his research focus changed as the field evolved and over time the applications of continuous time models have shifted from macroeconomics to finance.

Like the contributions he made to other areas in econometrics, Peter Phillips has significantly raised the level of technicality and the level of rigor in the continuous time econometrics literature. Many contributions that he has made to this literature are fundamental and have long-lasting effects. He has been the frontrunner in estimation, identification, finite sample theory, and nonstationary and nearly nonstationary issues in the continuous time literature.

It is a great honor and privilege for me to have this opportunity to write a review article summarizing Peter Phillips's extraordinary contributions to the continuous time econometrics literature. There are several survey articles in the literature, namely, Aït-Sahalia (2007), Fan (2005), Phillips and Yu (2009a), and Sorensen (2004). All these surveys review only estimation methods. The present survey

focuses on Phillips's contributions to the literature but the scope goes beyond estimation.

The rest of the article is organized as follows. Section 2 reviews parametric methods, including the identification problem. Section 3 reviews nonparametric methods. Section 4 reviews his work in the near unit root continuous time model, where his primary concerns are about the finite sample issues of traditional approaches. Section 5 presents a new set of results in a simple setup and shows that the in-fill asymptotic theory is superior to the long-span asymptotic theory and the double asymptotic theory. Section 6 concludes. The Appendix contains the proof of a result in Section 5.

2. PARAMETRIC METHODS

2.1. Identification

The continuous time model, considered in Phillips (1972, 1974, 1976a, 1976b), is a special parametric case of Model (1) in which $\mu(X(t)) = A(\theta)X(t) + B(\theta)$ and $\zeta(dt)$ is a vector of white noise disturbances with covariance Σ . See Bergstrom (1984) for the definition of the white noise disturbances. It is important to remark that $\zeta(dt)$ is not necessarily a Gaussian process. The interpretation of (1) is that the random process $X(t)$ satisfies the following stochastic integral equation:

$$X(t) = X(0) + \int_0^t (A(\theta)X(s) + B(\theta))ds + \int_0^t \zeta(ds). \quad (2)$$

When (2) is estimated with discretely sampled data, one must ensure the set of parameters θ is identified. In general, unfortunately, θ is not identifiable from discrete time models in the multivariate context. This is the well known aliasing problem in statistics, signal processing, computer graphics and related disciplines, which refers to an effect that causes different continuous signals to become indistinguishable when sampled discretely. To the best of my knowledge, Phillips (1973) is the first serious attempt to address this identification problem in the continuous time econometrics literature. In my opinion, this is a fundamental contribution to the literature.

The idea of aliasing can be explained simply by using a sinusoid, a periodic function of time. Figure 1 plots a set of discrete samples whose sampling interval is 1 (see the arrows) and two different sinusoids (see the two lines). Obviously, the two functions could have produced the same samples. Hence, it is impossible to tell which function has produced the discrete sample. One solution to the aliasing problem is to collect discrete time sample data at a frequency higher than the Nyquist frequency. Although this may be a reasonable solution in natural sciences, it may not work in economics as one cannot typically control the sampling interval.

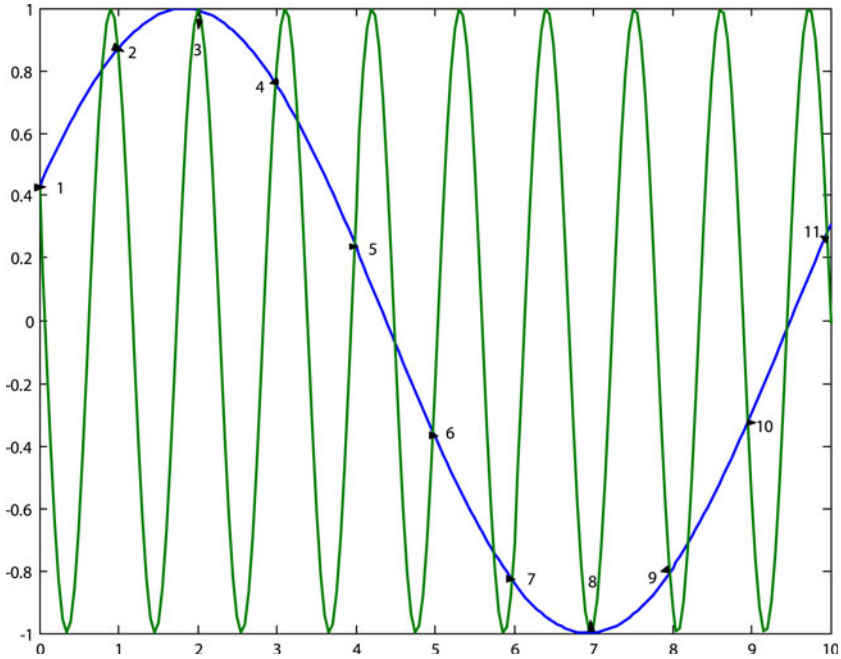


FIGURE 1. Aliasing problem. Arrows represent a set of discrete samples whose sampling interval is 1.

To illustrate the aliasing problem in continuous time econometrics models, consider the first-order specification in only stock variables as observable,

$$\begin{aligned}
 dX(t) &= A(\theta)X(t)dt + \zeta(dt), \quad X(0) = X_0, \quad E(\zeta(dt)) = 0, \\
 E(\zeta(dt)\zeta(dt)') &= \Sigma dt.
 \end{aligned}
 \tag{3}$$

The exact discrete time model is given by:

$$X_t = \exp(A(\theta)h)X_{t-1} + \varepsilon_t, \quad X_0 = X_0, \quad E(\varepsilon_t) = 0, \quad E(\varepsilon_t\varepsilon_t') = \Omega.
 \tag{4}$$

Phillips (1973) showed that:

$$\exp(Ah)\Sigma \exp(A'h) - \Sigma = A\Omega + \Omega A'.$$

So (A, Σ) is identifiable in Model (3), if and only if the matrix A is identifiable in $\exp(Ah)$. In general, however, equation

$$\exp(Ah) = B$$

does not have a unique solution for A . In particular, if some of the eigenvalues of A are complex, then by adding the imaginary numbers $\pm 2in\pi$ to each pair of

conjugate complex eigenvalues, the equation still holds true. In order to achieve the identification, certain restrictions have to be placed on A . Phillips (1973) derived a rank condition in the case where there are linear homogeneous relations between the elements of a row of A . Hansen and Sargent (1983) extended the result by showing that Ω provides extra identifying information about A . As a result, identifiability may be less difficult than one would think. In certain regions of the parameter space, there may be no identification problem at all, even if the sampling frequency is lower than the Nyquist frequency. Further contributions on this subject include Kessler and Rahbek (2004) and McCrorie (2003, 2009).

Since the 1980s, there has been a great deal of interest in unit roots and cointegration in econometrics. Phillips (1991) formulated error correction models and cointegrated systems in continuous time. A cointegrated system in continuous time takes the form:

$$X_1(t) = BX_2(t) + u_1(t) \quad (5)$$

$$DX_2(t) = u_2(t), \quad (6)$$

where $X_1(t)$ is a m_1 -vector process, $X_2(t)$ is a m_2 -vector process, $u_1(t)$ and $u_2(t)$ are both stationary continuous time residuals. Phillips (1991) showed that the exact discrete time model is given by:

$$X_i = \exp(-EA)X_{i-1} + \varepsilon_i = (I - EA)X_{i-1} + u_i \quad (7)$$

$$u_i = \varepsilon_i + \frac{1}{e}EAX_{i-1}, \quad (8)$$

where $X = [X_1', X_2']'$, $E = [I', 0']'$, $A = [I, -B]$, and u_i is stationary, because both ε_i and AX_{i-1} are stationary. As

$$\exp(-EA) = I - EA - \frac{1}{2!}(EA)^2 + \dots = I - \frac{e-1}{e}EA,$$

the relationship between B and the autoregressive coefficients $I - EA$ in the exact discrete time model is linear. Hence, there is no aliasing problem here. This result is in sharp contrast to the stationary continuous time models. Phillips (1991) proposed a frequency domain based estimation method and developed asymptotic distributions for the estimates. It turns out that the estimates of the long-run equilibrium coefficients converge at the rate $O_p(T^{-1})$, which is faster than that in the case of stationary models. This feature is consistent with that found in the discrete time framework. Kessler and Rahbek (2001, 2004) extended the results of Phillips (1991).

2.2. Estimation of Linear Models

The difficulty with estimating the parameters of the continuous time model (2) (i.e., θ) lies in the fact that only discretely observed data are available. To facilitate

estimation, Bergstrom (1966) proposed approximating the continuous time model (2) by the following discrete time model:

$$X_{i+1} - X_i = \left(\frac{1}{2} A(\theta) \{X_{i+1} + X_i\} + B(\theta) \right) h + u_{i+1}, \tag{9}$$

where the trapezoidal rule is used to approximate the integral, namely,

$$\int_{(i-1)h}^{ih} \mu(X(s); \theta) ds \approx \frac{h}{2} \{ \mu(X_i; \theta) + \mu(X_{i-1}; \theta) \}.$$

The discrete time model is then estimated by standard methods, such as the three-stage least squares (LS). As model (9) is merely an approximation to model (2), such an estimation approach inevitably suffers from the discretization bias that is induced by the trapezoidal rule. Obviously, the bias depends on the sampling interval, h , and does not disappear even if $T \rightarrow \infty$. The bigger the h , the larger the discretization bias.

The trapezoidal approximation is closely related to the Euler scheme approximation given by:

$$\int_{(i-1)h}^{ih} \mu(X(s); \theta) ds \approx h \mu(X_{i-1}; \theta),$$

which leads to the approximate discrete time model:

$$X_{i+1} - X_i = (A(\theta)X_i + B(\theta))h + v_{i+1}. \tag{10}$$

Obviously, (10) can be estimated by LS. As argued in Phillips and Yu (2009a), the two approximations are equivalent to $O(h)$. Obviously, Bergstrom’s approximation is an implicit method and in the multivariate case, it leads to a nonrecursive simultaneous equations model approximation to a system of recursive stochastic differential equations, whereas the Euler approximation is explicit and leads to a recursive simultaneous equations model.

Alternatively, one can obtain the exact discrete time model in the sense that the observations at the discrete points in time that are generated by (2) also satisfy the exact discrete time model. The main advantage in the exact discrete time model is that it avoids the discretization bias, no matter how big h is. This approach was taken seriously in Phillips (1972) and is a fundamental contribution.

To obtain the discrete time model of (2), one can first solve the SDE for $X(t)$. Assume that $A(\theta)$ has distinct eigenvalues $\lambda_1(\theta), \dots, \lambda_M(\theta)$, all with negative real parts. Consequently, there must exist a matrix $P(\theta)$ such that:

$$P(\theta)A(\theta)P(\theta)' = \text{diag}(\lambda_1(\theta), \dots, \lambda_M(\theta)) := \Lambda(\theta).$$

Bergstrom (1966) showed that the solution to SDE (2) is given by:

$$X(t) = \left[X(0) + A^{-1}(\theta)B(\theta) \right] \exp(A(\theta)t) - A^{-1}(\theta)B(\theta) + u(t), \tag{11}$$

where $\exp(A) := I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots$ and $u(t) = \int_0^t \exp(A(\theta)(t-s)) P \exp(A(\theta)(t-s)) P' \zeta(ds)$. As a result, the exact discrete time model is given by:

$$X_{i+1} = \exp(A(\theta)h)X_i - A^{-1}(\theta)(\exp(A(\theta)h) - I)B(\theta) + \varepsilon_{i+1}, \tag{12}$$

where

$$E(\varepsilon_i) = 0, E(\varepsilon_i \varepsilon_i') = \int_0^h \exp(A(\theta)s) \Sigma \exp(A(\theta)'s) ds := \Omega.$$

Phillips (1972) used a generalized LS procedure to estimate θ that is equivalent to maximum likelihood (ML) if the covariance matrix of the disturbances is known, and showed that the estimators are consistent and asymptotically efficient when $T \rightarrow \infty$ and a fixed h , i.e., the *long-span* asymptotic theory. With simulated data from a three-variable trade-cycle model, Phillips (1972) examined the finite sample properties of the estimates and found the superior finite sample performance of the exact discrete model relative to the approximate discrete model.

Sargan (1974) showed that the discretization bias of the three-stage LS estimators for the trapezoidal approximation is asymptotically $O(h^2)$ as $h \rightarrow 0$, whereas Bergstrom (1984) showed that the discretization bias of the LS estimator for the Euler approximation is asymptotically $O(h)$. So the *double* asymptotics is required. To illustrate the magnitude of the discretization error in the approximate model (10), consider the following univariate continuous time model:

$$dX(t) = \kappa(\mu - X(t))dt + \zeta(dt). \tag{13}$$

So $A = -\kappa$ and $B = \kappa\mu$. If $\kappa > 0$, $X(t)$ is stationary with μ as the long-run mean, κ as the speed of mean reversion, $1/(\kappa \ln 2)$ as the half-life of a shock. The Euler scheme leads to the following approximate model:

$$X_{i+1} = \kappa\mu h + (1 - \kappa h)X_i + v_{i+1}, \tag{14}$$

whereas the exact discrete time model is:

$$X_{i+1} = \mu(1 - e^{-\kappa h}) + e^{-\kappa h} X_i + \varepsilon_{i+1}. \tag{15}$$

Clearly, we use $\kappa\mu h$ to approximate $\mu(1 - e^{-\kappa h})$ and $1 - \kappa h$ to approximate $e^{-\kappa h}$. The order of the approximation error may be obtained from the following Taylor expansions:

$$\mu(1 - e^{-\kappa h}) = \kappa\mu h + O(h^2), \tag{16}$$

$$e^{-\kappa h} = 1 - \kappa h + O(h^2). \tag{17}$$

Consequently, if h is small, the Euler scheme should provide a good approximation to the exact discrete time model. However, if h is large, the Euler approximation can be poor. For a numerical example, consider the case where $\kappa = 2$

(an empirically realistic value) and $h = 1/12$ (i.e., monthly data) in which case $e^{-\kappa h}$ is 0.8465, whereas $1 - \kappa h$ is 0.8333 and the approximation is reasonable. But if $\kappa = 2$ and $h = 1$ (i.e., annual data), $e^{-\kappa h}$ is 0.1353 whereas $1 - \kappa h$ is -1 . The quality of the approximation is unsatisfactory with the sign of the coefficient flipped. The economic implications of these two values would be very different. Note that the autoregressive coefficient implied by the Euler model is always smaller than that implied by the exact model when $\kappa h \in (0, 1)$, because $e^{-\kappa h} > 1 - \kappa h$.

The model considered by Phillips (1972) in the Monte Carlo study is the three-variable trade-cycle model given by:

$$dC(t) = \alpha[(1-s)Y(t) + \alpha - C(t)]dt + \zeta_1(dt), \quad (18)$$

$$dY(t) = \lambda[C(t) + DK(t) - Y(t)]dt + \zeta_2(dt), \quad (19)$$

$$dK(t) = \gamma[vY(t) - K(t)]dt + \zeta_3(dt), \quad (20)$$

where C = consumption, Y = income, and K = capital. There are 6 parameters in the model. Phillips (1972) simulated 25 observations from the model and performed the generalized least square estimation to the exact discrete model and the three-stage least squares to the approximate discrete model. While the setup may seem easy from today's perspective, given the rapid development in computing technology and software in recent years, it was conceivably much harder in the 1960s to develop the computer program and do the computing.

As his first publication in econometrics, Phillips (1972) is filled with major conceptual advances, cutting edge technical innovation and sophisticated Monte Carlo exercise for the time, and perhaps most importantly, enormous practical relevance to empirical studies in economics and finance. It was masterfully put together with good knowledge of linear algebra, differential equations, computational mathematics, statistics, and economics. It is even more remarkable given the fact that it was based on his Master's thesis. Indeed, the argument of eliminating the discretization bias has had a long-run impact on the literature. Concern about discretization bias has motivated many researchers to introduce various new methods to estimate continuous time financial models over the last 2 decades (see, for example, Lo (1988), Pedersen (1995), Elerian, Chib, and Shephard (2001), and Aït-Sahalia (2002)).

Phillips (1974) generalized the results by allowing the presence of identities, higher order of derivatives, and exogenous variables. These augmentations are practically important in macroeconomics. For instance, often imposed into the macroeconomic model are well known identities such as the balance of payments identity and the national income identity. In the context of the exact discrete model, Phillips (1974) showed that the presence of identities, whether it is in the first-order model or in the higher order models, is unlikely to cause any estimation problems, whereas the introduction of exogenous variables represents

a more serious complication. To explain the complication, consider the following model:

$$dX(t) = (A(\theta)X(t) + B(\theta)Z(t))dt + \zeta(dt), \tag{21}$$

where $Z(t)$ is a vector of exogenous variables, observed at the same discrete point in time as for $X(t)$. The exact discrete time model of (21) is given by:¹

$$X_i = \exp(Ah)X_{i-1} - A^{-1}(\exp(Ah) - I)B + \int_0^h \exp(sA)BZ(ih - s)ds + \varepsilon_i, \tag{22}$$

where $\varepsilon_i = \int_{(i-1)h}^{ih} \exp(A(\theta)(ih - s))P \exp(A(\theta)(ih - s))P' \zeta(ds)$. If $Z(t)$ is not a simple integrable function of time, it cannot be integrated out analytically and hence a continuous record is needed for $Z(t)$ before the model can be estimated. Since in practice $Z(t)$ is always observable only at a grid of discrete points, we have to approximate the integral. Noting that polynomials are simple integrable functions, Phillips (1974) proposed to expand $Z(ih - s)$ in a second order Taylor series about $s = 0$ and to use the three-point Lagrange interpolation formula to approximate $Z(ih - s)$, namely:

$$\hat{Z}(ih - s) = Z_i - s(Z_{i-2} - 4Z_{i-1} + 3Z_i) + s^2(Z_i - 2Z_{i-1} + Z_{i-2}) / (2h^2).$$

Substituting out $Z(ih - s)$ in (22) by $\hat{Z}(ih - s)$ and integrating out the polynomials, we have:

$$X_{i+1} = \exp(A(\theta)h)X_i + E_2Z_i + E_3Z_{i-1} + E_4Z_{i-2} + \eta_{i+1}, \tag{23}$$

where the expressions for $E_j, j = 2, 3, 4$ are given in Phillips (1974). The generalized least squares procedure can then be applied to estimate the approximate model. In general, there is an approximation error in (23) which depends on h . The smaller is h , the smaller the discretization error. However, if elements of $Z(t)$ are polynomials in t of degree of at most two, there is no approximation error in (23), because in this case (23) is the exact discrete model. Phillips (1976a, 1976b) made further contributions along this line of research.

So far all the variables are assumed to be observed at specific points in time. This assumption is reasonable for stock variables. Flow variables, such as $C(t)$ and $Y(t)$ in the trade-cycle model given by (18), (19), and (20), can be observed only as the integrals, such as $\int_{(i-1)h}^{ih} C(s)ds$ and $\int_{(i-1)h}^{ih} Y(s)ds$. Let $X_i = \int_{(i-1)h}^{ih} C(s)ds$ when $C(t)$ is a flow variable. The exact discrete time model for X_i is different from that when $C(t)$ is assumed to be a stock variable. Phillips (1978) made several contributions in this context.

It is important to point out that the continuous time models discussed in this section have been widely applied to describe and forecast the movement of the economies of many industrial countries. For example, Bailey, Hall, and Phillips

(1987) used a continuous time model to make predictions of the New Zealand economy. Bergstrom (1996) provided a comprehensive list of the applications.

An important special case of Model (2) is when $\zeta(dt)$ is a Gaussian process. In the univariate case, the model takes the form:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma dW(t), X(0) = X_0. \tag{24}$$

This Ornstein-Uhlenbeck (OU) process is used by Vasicek (1977) to describe the movement of short term interest rates. From (12), the exact discrete model is of the form:

$$X_i = \mu \left(1 - e^{-\kappa h}\right) + e^{-\kappa h} X_{i-1} + \sigma \sqrt{(1 - e^{-2\kappa h})/(2\kappa)} \varepsilon_i, \tag{25}$$

where $\varepsilon_i \sim$ i.i.d. $N(0, 1)$ where i.i.d. denote independent and identically distributed. This is a Gaussian AR(1) process and can be estimated by LS/ML. As a result, the asymptotic theory developed by Phillips (1972) under stationarity assumptions (i.e., $\kappa > 0$) is applicable. The results in the multivariate Vasicek model can be established in the same manner.²

2.3. Estimation of Nonlinear Models

An important class of continuous time models is nonlinear diffusion processes. A general time homogenous parametric diffusion process takes the form:

$$dX(t) = \mu(X(t); \theta)dt + \sigma(X(t); \theta)dW(t), \tag{26}$$

where $\mu(X(t); \theta)$ is a nonaffine function of $X(t)$ or $\sigma(X(t); \theta) \neq \sigma$ or both and θ is a vector of unknown parameters. This class of parametric models has been widely used to characterize the temporal dynamics of financial variables, including stock prices, interest rates, exchange rates, and volatilities. Typically financial variables are measured in years and observed at higher frequencies than macroeconomic variables. If $X(t)$ is observed monthly (weekly or daily), we have $h = 1/12$ (1/52 or 1/252). However, ultrahigh frequency data have become available in recent years.

Arguably, the most important continuous time model in finance is the so-called affine model of Duffie and Kan (1996), where both $\mu(X(t); \theta)$ and $\sigma(X(t); \theta)\sigma(X(t); \theta)'$ are affine functions of $X(t)$. Vasicek's model is of course a simple but important special univariate case in the affine family.

Another important special case was proposed by Cox, Ingersoll, and Ross (1985, CIR, hereafter). This so-called square root process, often used to describe movements in short term interest rates and in volatility, takes the form:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma \sqrt{X(t)}dW(t), \quad X_0 \sim \Gamma\left(2\kappa/\sigma^2, 2\kappa\mu/\sigma^2\right), \tag{27}$$

where $\Gamma(\alpha, \beta)$ is a Gamma distribution with parameters α and β . The error term in the exact discrete model is non-Gaussian. The same model was used by Heston (1993) to describe movements in volatility.

One advantage of using affine continuous time models is that the formulation permits closed-form or nearly closed-form solutions to many important pricing formulae. For a general treatment, see Duffie and Kan (1996). In the context of the OU process, Vasicek (1977) derived the expression for bond prices and Jamshidian (1989) gave the corresponding formula for bond option prices. In the context of the square root process, CIR (1985) derived the expressions for bond prices and bond option prices.

While the affine models are analytically tractable in terms of asset pricing, they may not necessarily fit the data well. Chan, Karolyi, Longstaff, and Sanders (1992) (CKLS, hereafter) considered a univariate interest rate diffusion process:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma X^\gamma(t)dW(t), \tag{28}$$

where γ is a free parameter that is to be determined by data. Obviously, this model has a more flexible diffusion term than the affine models. Unfortunately, the exact discrete time model is not analytically available and the error term is not Gaussian unless $\gamma = 0$.

To enable a Gaussian discrete time representation of (28), Nowman (1997) assumed that the conditional volatility is constant between two consecutive observation points, $[(t - 1)h, th], t = 1, 2, \dots, n$. Namely, the CKLS model is approximated by the OU process locally, with the diffusion term $\sigma X^\gamma(t)$ replaced with $\sigma X_{(s-1)h}^\gamma$ when $t \in [(s - 1)h, sh)$. From Phillips (1972), the exact discrete time model of the approximate continuous time model is:

$$X_i = \mu \left(1 - e^{-\kappa h}\right) + e^{-\kappa h} X_{i-1} + \sigma X_{i-1}^\gamma \sqrt{(1 - e^{-2\kappa h}) / (2\kappa)} v_i, \tag{29}$$

where $v_i \sim$ i.i.d. $N(0, 1)$. Equation (29) is amenable to ML estimation, since the transition density is Gaussian. Since there is a discretization bias in (29), the *double* asymptotics is needed for a consistent estimation of μ and κ .

Motivated from the observation that (29) is not the exact discrete time model of (28), Yu and Phillips (2001) (see also a correction made in Phillips and Yu (2011)) obtained an exact discrete time model of (28) using the stopping time argument. The approach is based on the idea that any continuous time martingale can be written as a Brownian motion after a suitable time change ($M_t = W_{[M]_t}$). That is, if the chronological time in a local martingale M_t is adjusted to time based on the evolution of the quadratic variation process $[M]_t$ of M , we have the time change given by $T_t = \inf\{s|[M]_s > t\}$ and the process is transformed to a Brownian motion (called the DDS Brownian motion).

To see how this approach can be used to estimate equation (28), first write (28) as:

$$\begin{aligned}
 X(t+h) &= \mu \left(1 - e^{-\kappa h}\right) + e^{-\kappa h} X(t) \\
 &\quad + \sigma \int_0^h e^{-\kappa(h-s)} X^\gamma(t+s) dW(t+s), \quad \forall h > 0.
 \end{aligned}
 \tag{30}$$

Let

$$\begin{aligned}
 M_t(h) &= \sigma \int_0^h e^{-\kappa(h-s)} X^\gamma(t+s) dW(t+s) \\
 &= e^{-\kappa h} \int_0^h e^{\kappa s} X^\gamma(t+s) dW(t+s) = e^{-\kappa h} H_t(h),
 \end{aligned}$$

where $H_t(h) := \int_0^h \sigma e^{\kappa s} X^\gamma(t+s) dW(t+s)$ is a continuous martingale whose quadratic variation is given by:

$$[H_t]_s = \int_0^s \sigma^2 e^{2\kappa s} X^{2\gamma}(t+s) ds,$$

so that

$$X(t+h) = \mu \left(1 - e^{-\kappa h}\right) + e^{-\kappa h} X(t) + e^{-\kappa h} H_t(h).$$

To construct the DDS Brownian motion to represent $H_t(h)$, one may apply the DDS Theorem to H_t with timing constant a so that

$$\tilde{h}_{j+1} = \inf \left\{ s : [H_{t_j}]_s \geq a \right\} = \inf \left\{ s : \int_0^s \sigma^2 e^{2\kappa s} X^{2\gamma}(t_j+s) ds \geq a \right\}, \tag{31}$$

we have

$$X(t_{j+1}) = \mu \left(1 - e^{-\kappa \tilde{h}_{j+1}}\right) + e^{-\kappa \tilde{h}_{j+1}} X(t_j) + e^{-\kappa \tilde{h}_{j+1}} H_{t_j}(\tilde{h}_{j+1}),$$

where $t_{j+1} = t_j + \tilde{h}_{j+1}$. By the DDS Theorem, $H_{t_j}(\tilde{h}_{j+1}) \sim N(0, a)$. Since the step size t_j and stopping times \tilde{h}_{j+1} are both endogenous, the ordinary least squares or weighted least squares procedures are inconsistent. To consistently estimate μ and κ , we may use $(1, X(t_j))$ as the instrument and set up the estimating equations:

$$\sum_j \left(e^{\kappa \tilde{h}_{j+1}} X(t_{j+1}) - \mu \left(e^{-\kappa \tilde{h}_{j+1}} - 1 \right) - X(t_j) \right) X(t_j) = 0, \tag{32}$$

and

$$\sum_j \left(e^{\kappa \tilde{h}_{j+1}} X(t_{j+1}) - \mu \left(e^{-\kappa \tilde{h}_{j+1}} - 1 \right) - X(t_j) \right) = 0. \tag{33}$$

The instrumental variable (IV) estimators of μ and κ can be obtained from solving these two equations jointly. The idea of using the stopping time argument to induce normality was more recently employed in different contexts by Park (2008b) and Chang (2012).

While the SDE (26) is formulated in continuous time, the sample data are always collected at discrete points in time or over discrete intervals in the case of flow data. One may argue that for financial variables, the sampled data are so frequently observed as to be nearly continuously available.

If a continuous record from $[0, 1]$ is indeed available, Phillips (1987b) proposed to estimate κ in the model:

$$dX(t) = -\kappa X(t)dt + dW(t), \quad (34)$$

by the following ML estimator:

$$\tilde{\kappa} = -\frac{\int_0^1 X(t)dX(t)}{\int_0^1 X(t)^2 dt}. \quad (35)$$

This is because the log-likelihood function of $X(t), t \in [0, 1]$, has the following form:

$$\ell(\kappa) = \int_0^1 -\kappa X(t)dX_t - \frac{1}{2} \int_0^1 \kappa^2 X(t)^2 dt.$$

As a consequence of (35), the finite sample distribution of $\hat{\kappa} - \kappa$ is:

$$\tilde{\kappa} - \kappa = \frac{\int_0^1 X(t)dW(t)}{\int_0^1 X(t)^2 dt}. \quad (36)$$

Phillips and Yu (2009b) proposed an alternative method to estimate parameters in (26) based on the following two properties: (1) the diffusion term can be fully uncovered from a continuous record and (2) for a diffusion process with a known diffusion term, the likelihood function of the process is analytically available via the Girsanov theorem. The method of Phillips and Yu (2009b) contains two steps.

To fix the idea, consider the SDE:

$$dX(t) = \mu(X(t); \theta_1)dt + \sigma(X(t); \theta_2)dW(t). \quad (37)$$

In the first step, parameters in the diffusion term are estimated from the empirical quadratic variation process. In the continuous time econometrics literature, the empirical quadratic variation is known as the realized variance (RV). The approach is justified by the fact that RV is a natural consistent estimate of quadratic variation. Also, RV has convenient distributional characteristics that are determined asymptotically by the (functional) central limit theory (CLT), as derived by Jacod (1993) and Barndorff-Nielsen and Shephard (2002).

To proceed, assume that $X(t)$ is observed at the following time points:

$$t = \underbrace{h, 2h, \dots, M_h h \left(= \frac{T}{K} \right)}_{\text{subsample 1}}, \underbrace{(M_h + 1)h, \dots, 2M_h h \left(= \frac{2T}{K} \right), \dots, n_h h (= T)}_{\text{subsample } K},$$

where $n_h = KM_h$ with K a fixed and positive integer and $M_h = O(n_h)$. Phillips and Yu (2009b) constructed the nonoverlapping K subsamples:

$$((k - 1)M_h + 1)h, \dots, kM_h h, \text{ where } k = 1, \dots, K,$$

so that each subsample has M_h observations over the interval $((k - 1)\frac{T}{K}, k\frac{T}{K}]$.

As $h \rightarrow 0$ and $M_h \rightarrow \infty$,

$$\sum_{i=2}^{M_h} (X_{(k-1)M_h+i h} - X_{(k-1)M_h+(i-1)h})^2 \xrightarrow{P} [X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}}, \tag{38}$$

and

$$\frac{\ln\left(\sum_{i=2}^{M_h} (X_{(k-1)M_h+i h} - X_{(k-1)M_h+(i-1)h})^2\right) - \ln\left([X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}}\right) + \frac{1}{2}s_k^2}{s_k} \xrightarrow{d} N(0, 1), \tag{39}$$

where

$$s_k = \min \left\{ \sqrt{\frac{r_k^2}{\left(\sum_{i=2}^{M_h} (X_{(k-1)M_h+i h} - X_{(k-1)M_h+(i-1)h})^2\right)^2}}, \sqrt{\frac{2}{M_h}} \right\},$$

and

$$r_k = \sqrt{\frac{2}{3} \sum_{i=2}^{M_h} (X_{(k-1)M_h+i h} - X_{(k-1)M_h+(i-1)h})^4},$$

for $k = 1, \dots, K$. Note that $[X]_T$ is the quadratic variation of X which can be consistently estimated by the empirical counterpart $[X_h]_T$ defined as:

$$[X_h]_T = \sum_{i=2}^{n_h} (X_{ih} - X_{(i-1)h})^2 := RV.$$

The definition of quadratic variation gives the limit (38), while the CLT (39) is based on the asymptotic theory of Barndorff-Nielsen and Shephard (2005). Based on the CLT (39), θ_2 can be estimated in the first stage by running a nonlinear least squares regression of

$$\frac{\ln\left(\sum_{i=2}^{M_h} (X_{(k-1)M_h+i h} - X_{(k-1)M_h+(i-1)h})^2\right) + \frac{1}{2}s_k^2}{s_k} \tag{40}$$

on

$$\frac{\ln \left(\sum_{i=2}^M \sigma^2 \left(X_{(k-1)M_h+(i-1)h}; \theta_2 \right) h \right) - \frac{1}{2} s_k^2}{s_k} \tag{41}$$

for $k = 1, \dots, K$. This produces a consistent estimate $\hat{\theta}_2$ of θ_2 as $h \rightarrow 0$. This is the *in-fill* asymptotics. In the second stage, the approximate in-fill log-likelihood function is maximized with respect to θ_1 , i.e.,

$$\operatorname{argmax}_{\theta_1} \left\{ \sum_{i=2}^n \frac{\mu(X_{(i-1)h}; \theta_1)}{\sigma^2(X_{(i-1)h}; \hat{\theta}_2)} (X_{ih} - X_{(i-1)h}) - \frac{h}{2} \sum_{i=2}^n \frac{\mu^2(X_{(i-1)h}; \theta_1)}{\sigma^2(X_{(i-1)h}; \hat{\theta}_2)} \right\}. \tag{42}$$

This produces a consistent estimate $\hat{\theta}_1$ of θ_1 as $T \rightarrow \infty$ and $h \rightarrow 0$ (i.e., the *double* asymptotics). The asymptotic theory for $\hat{\theta}_1$ and $\hat{\theta}_2$ is fully developed in Phillips and Yu (2009b). Other studies that have used realized variance to estimate and compare parametric models include Corradi and Distaso (2006, 2010) and Kristensen (2010a).

3. NONPARAMETRIC METHODS

Parametric continuous time models have proven very useful for predicting future economic activities and for pricing financial assets. However, theory usually is silent about which parametric forms to use. Often, parametric specifications are adopted for mathematical convenience. Misspecification of the model by a specific parametric form might lead to erroneous decision making. For example, Aït-Sahalia (1996) showed that when pricing bond options, traditional parametric models can yield significant pricing errors. In this section, we will review various nonparametric methods proposed by Peter Phillips.

3.1. Estimation of Drift and Diffusion Terms

Nonparametric estimation of continuous time models was pioneered by Aït-Sahalia (1996). The model he investigated takes the form:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma(X(t))dW(t), \tag{43}$$

where the diffusion term $\sigma(X(t))$ is an unknown function. The linearity in the drift term is an identification restriction, through which the diffusion term can be identified from the marginal distribution, without assuming $h \rightarrow 0$. Assuming the process $X(t)$ is stationary and $\pi(x)$ is the marginal distribution of the process, the relation between the diffusion term and the marginal distribution is given by:

$$\sigma^2(x) = \frac{1}{\pi(x)} \int_0^x \kappa(\mu - s)\pi(s)ds.$$

Replacing $\pi(x)$ with a nonparametric density estimator would yield a nonparametric estimator for $\sigma(x)$. Aït-Sahalia (1996) used a kernel function to estimate

$\pi(x)$ and $\sigma(x)$, and established the asymptotic normality of the estimates under the *long-span* asymptotics. Two assumptions are critical to the development of the method, namely stationarity of the process and linearity of the drift term. It is important to remark that Aït-Sahalia (1996) developed the *long-span* asymptotic theory. Other nonparametric methods in the literature that rely on the *long-span* asymptotics include Chen, Hansen, and Scheinkman (2000, 2009) and Kristensen (2010a, 2010b).

Bandi and Phillips (2003) significantly extended the results by developing a nonparametric method to estimate both the drift and diffusion terms without imposing the stationarity assumption. This absence of the stationarity assumption is important in financial time series analysis, because many financial time series, such as interest rates, stock prices, exchanges rates, and volatility, may be better modeled by martingale processes or processes of other forms of nonstationarity. Bandi and Phillips used both the long-span and in-fill asymptotics to solve the identification problem and avoid the aliasing problem. To achieve identification of the drift, the condition of recurrence is further assumed, so that the process repeats itself. Asymptotic theory is developed using the chronological local time process, a standardized local time process that is defined in terms of pure time units. For further details about the chronological local time process, see Park and Phillips (2001) and Phillips and Park (1999).

To fix the idea, assume the time homogenous diffusion model is nonparametrically specified as:

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t). \tag{44}$$

Note that:

$$E[dX(t) | X(t)] = \mu(X(t))dt, \tag{45}$$

$$E\left[(dX(t))^2 | X(t)\right] = \sigma^2(X(t))dt. \tag{46}$$

Application of the Nadaraya-Watson kernel method to (45) and (46) gives rise to the nonparametric estimator of $\mu(x)$:

$$\hat{\mu}(x) = \frac{\sum_{i=1}^n K\left(\frac{X_{ih}-x}{b}\right) \tilde{\mu}(X_{ih})}{\sum_{i=1}^n K\left(\frac{X_{ih}-x}{b}\right)},$$

and the nonparametric estimator of $\sigma^2(x)$:

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{X_{ih}-x}{b}\right) \tilde{\sigma}^2(X_{ih})}{\sum_{i=1}^n K\left(\frac{X_{ih}-x}{b}\right)},$$

where:

$$\begin{aligned} \tilde{\mu}(X_{ih}) &= \frac{1}{m(ih)h} \sum_{j=0}^{m(ih)-1} [X_{t(ih)_{j+h}} - X_{t(ih)_j}], \\ \tilde{\sigma}^2(X_{ih}) &= \frac{1}{m(ih)h} \sum_{j=0}^{m(ih)-1} [X_{t(ih)_{j+h}} - X_{t(ih)_j}]^2, \\ m(ih) &= \sum_{j=1}^n \mathbf{1}_{[|X_{jh} - X_{ih}| \leq \varepsilon]}, \forall i \leq n, \end{aligned}$$

and b is the bandwidth. To develop the asymptotic distributions of $\hat{\mu}(x)$, it is assumed that $n \rightarrow \infty$, $T \rightarrow \infty$, $h = T/n \rightarrow 0$, and $b \rightarrow 0$ (i.e., the *double* asymptotics). Unlike $\hat{\mu}(x)$, $\hat{\sigma}^2(x)$ is consistently estimated without requiring $T \rightarrow \infty$ (i.e., the *in-fill* asymptotics). Bandi and Phillips (2003) developed the asymptotic distributions of $\hat{\sigma}^2(x)$ for the case when T is finite and also for the case when $T \rightarrow \infty$.

The nonparametric estimates have recently been applied to various contexts. Bandi (2002) used them to estimate the short term interest rate model. Corradi and Distaso (2010) employed it to design a test statistic to distinguish one factor models against two factor models. Jeffrey, Kristensen, Linton, Nguyen, and Phillips (2004) made use of it to estimate a multifactor Heath-Jarrow-Morton model. Bandi and Phillips (2007) used it to develop a simple and robust approach for the parametric estimation of scalar homogeneous SDEs, which we briefly discuss below.

Suppose a parametric continuous time model takes the form

$$dX(t) = \mu(X(t); \theta_1)dt + \sigma(X(t); \theta_2)dW(t). \tag{47}$$

The estimator of θ_1 can be obtained by

$$\min_{\theta_1 \in \Theta_1} \|\hat{\mu} - \mu(X(t); \theta_1)\|,$$

and the estimator of θ_2 can be obtained by

$$\min_{\theta_2 \in \Theta_2} \|\hat{\sigma}^2 - \sigma^2(X(t); \theta_2)\|,$$

where $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ are the nonparametric estimates defined above,

$$\|\hat{\mu} - \mu(\cdot; \theta_1)\| = \sum_{i=1}^n (\hat{\mu}(X_{ih}) - \mu(X_{ih}; \theta_1))^2,$$

and

$$\|\hat{\sigma}^2 - \sigma^2(\cdot; \theta_2)\| = \sum_{i=1}^n (\hat{\sigma}^2(X_{ih}) - \sigma^2(X_{ih}; \theta_2))^2$$

are the Euclidean distance. Bandi and Phillips (2007) developed the asymptotic theory for the estimates of θ_1 and θ_2 under different sets of assumption. In particular, it was also shown that the consistency of θ_2 can be guaranteed with $h \rightarrow 0$ but does not need $T \rightarrow \infty$, so the *in-fill* asymptotics is sufficient for the parameters in the diffusion function. On the other hand, the consistency of θ_1 requires the *long-span* asymptotics. The results are consistent with those noted by Merton (1980).

The estimators of Bandi and Phillips (2003) have been used in the literature in a similar manner. Examples include Brugiere (1993), Florens-Zmirou (1993), Geman (1979), and Stanton (1997). The work of Bandi and Phillips (2003) has been extended in a number of more recent studies, namely, Bandi and Nguyen (2003), Bandi and Moloche (2008), and Park (2008a). Bandi and Phillips (2009) reviewed the literature of nonstationary continuous time methods.

3.2. Estimation of Integrated Variance

Financial market volatility is a key concept in financial economics. In diffusion processes, the diffusion term corresponds to an important measure of volatility—spot volatility. In previous sections, we have seen a variety of ways to estimate the diffusion term. Another important measure of volatility is the integrated variance, defined by,

$$IV = \int_0^1 \sigma^2(t)dt. \tag{48}$$

Obviously, this is the quadratic variation of the SDE

$$dX^*(t) = \mu(X^*(t))dt + \sigma(t)dW(t),$$

over a unit interval $[0, 1]$, where X^* is an efficient equilibrium price. Let $0 = t_{0,m} < t_{1,m} < \dots < t_{m,m} = 1$ be a sequence of deterministic partitions of $[0, 1]$ and $h_{1,m} = \sup_i |t_{i,m} - t_{i-1,m}|$ is the grid size. A common assumption adopted in the literature is that the partition involves a simple grid of equispaced points $\{t_{i,m} = \frac{i}{m} : i = 0, \dots, m\}$, in which case $h_{1,m} = \frac{1}{m}$ and $h_{\ell,m} = \frac{\ell}{m}$.

An important nonparametric estimate of IV is the empirical quadratic variation of $X(t)$, defined by

$$\sum_{i=1}^m [X_{i,m}^* - X_{i-1,m}^*]^2 := RV^{(m)}(X^*).$$

As explained before, the quantity is called RV and this nonparametric estimate has received a great deal of attention in the continuous time literature in recent years. Pioneering work includes Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001), and Barndorff-Nielsen and Shephard (2002).

While the deterministic partitions and the equispaced partitions greatly facilitate the development of asymptotic theory of RV, in the real ultrahigh frequency

data, such assumptions may be too strong. In particular, the phenomenon of flat pricing is very common in stock market trading, leading to stochastic durations of trade intervals. Phillips and Yu (2009c) generalized the standard asymptotic theory of RV to the cases where flat trading is present, with and without microstructure noise.

Phillips and Yu (2009c) considered two mechanisms to generate flat trading. First, the flat trading is determined by a simple Bernoulli process, i.e.,

$$X_{i,m} = \begin{cases} X_{i,m}^* & \text{if } \zeta_i = 1 \\ X_{i-1,m} & \text{if } \zeta_i = 0 \end{cases}, \tag{49}$$

where $X_{i,m}$ is the observed price, ζ_i is a Bernoulli sequence independent of X^* with $E(\zeta_i = 1) = \pi$, $X_{0,m} = X_{0,m}^* = O_p(1)$, and

$$dX^*(t) = \sigma(t)dW(t). \tag{50}$$

Phillips and Yu (2009c) showed that, as $m \rightarrow \infty$

$$\sqrt{m} [RV^{(m)}(X) - IV] \xrightarrow{d} MN \left(0, \frac{4 - 2\pi}{\pi} \int_0^1 \sigma^4(t) dt \right), \tag{51}$$

where MN signifies mixed normal.

Second, the flat trading is determined by an autoregressive conditional duration (ACD) process, i.e., as $m \rightarrow \infty$,

$$E(D_{m,[ms]} | \mathcal{F}_{\tau_{[ms]-1}}) \rightarrow_p \mu_D(s), \quad E(D_{m,[ms]}^2 | \mathcal{F}_{\tau_{[ms]-1}}) \rightarrow_p \omega_D^2(s), \tag{52}$$

where $D_{m,j}$ measures the duration between observations (in units of the interval m^{-1}) and may be (partly) dependent on past prices. Phillips and Yu (2009c) showed that, as $m \rightarrow \infty$

$$\begin{aligned} & \sqrt{m} \left\{ \sum_{j=1}^{J_m} [X^*(\tau_j) - X^*(\tau_{j-1})]^2 - \int_0^1 \sigma^2(t) dt \right\} \\ & \Rightarrow MN \left(0, 2 \int_0^1 \sigma^4(t) \frac{\omega_D^2(t)}{\mu_D(t)} dt \right). \end{aligned} \tag{53}$$

Model (49) assumes that when there is a price change, the efficient price is observed with an error. This assumption may be too strong in very high frequencies. A more realistic model is

$$X_{i,m} = \begin{cases} X_{i,m}^* + u_{i,m} & \text{if } \zeta_i = 1 \\ X_{i-1,m} & \text{if } \zeta_i = 0 \end{cases}, \tag{54}$$

so that when prices do change a noise component $u_{i,m}$ perturbs the observed price. It is assumed that u_{τ_j} is *iid* with zero mean, variance σ_u^2 and finite fourth moment. Let $[\widehat{p}^*]$ be the two-stage estimator of Zhang, Mykland, and Ait-Sahalia (2005). Under certain conditions, Phillips and Yu (2009c) show that as $m \rightarrow \infty$

$$m^{1/6} \left\{ [\widehat{p}^*] - \int_0^1 \sigma^2(t) dt \right\} \Rightarrow \left\{ \frac{8}{c^2} \sigma_u^4 + c\eta^2 \right\}^{1/2} N(0, 1), \tag{55}$$

where

$$\eta^2 = \frac{4 \left\{ \int_0^1 (1-s)^2 \mu_D(1-s) ds \right\}}{\int_0^1 \mu_D(s) ds} \int_0^1 \sigma^4(t) dt, \tag{56}$$

$K = cm^{2/3}$, and K is the number of nonoverlapping grids.

4. FINITE SAMPLE ISSUES

In the continuous time finance literature, it has frequently been argued that ML should be the method to use for parameter estimation and statistical inference. The statistical justification is that ML estimates have good asymptotic properties and well developed asymptotic theory. Moreover, sample sizes in typical financial data applications are large, leading to a common belief that these good asymptotic properties hold true in finite samples.

However, recently it has been forcefully argued by Peter Phillips (see Phillips and Yu, 2005a and 2009d) that the finite sample performance of the ML estimator can be very poor from both statistical and economic perspectives. For example, ML estimates of parameters in some continuous time models may be badly biased even when the sample size is very large and the sampling interval is very small. This is especially the case in the commonly occurring situation of drift parameter estimation in models where the process is nearly a unit root process. Financial variables, such as interest rates and volatility, typically have a root near unity, indicating an important shortcoming of ML from a practical viewpoint.

In the context of Vasicek model with a known long-run mean, Yu (2012) showed that the bias of ML estimate of κ is upward and approximated by $2/T$. The author further derived analytical expressions to approximate the bias, and argued that, depending on the initial condition, a nonlinear term in the bias formula may be particularly important when the mean reversion parameter is close to zero. In the context of Vasicek model and CIR model with an unknown long-run mean, Tang and Chen (2009) showed that the bias of ML estimate of κ is approximated by $4/T$. If the true value of $\kappa = 0.1$ and $T = 10$ (10 years data), the percentage bias in the ML estimate of κ , implied by these two results, is 200% and 400%, respectively.

Both Yu (2012) and Tang and Chen (2009) are motivated from Phillips and Yu (2005a) where the authors intuitively explained why the ML estimator of κ is severely biased upward. Denote the autoregressive coefficient by $\phi = \exp(-\kappa h)$. Note that $\phi \approx 1 - \kappa h$ by the first-order approximation. Hence, the bias in $\hat{\kappa}$ is approximately the bias in $\hat{\phi}$ multiplied by $1/h$. It is well known that in the context of the AR(1) model with an intercept only, to a first-order approximation, the bias $\hat{\phi}$ is $-\frac{1+3\phi}{n}$. Since $T = nh$, the bias in κ is approximately $\frac{1+3\phi}{T}$. This is an upward bias, which is mainly determined by the data span, not the sample size.

Phillips and Yu (2005a,b) performed extensive Monte Carlo experiments to confirm the substantial percentage bias in the context of Vasicek and CIR models. For example, if $\kappa = 0.1$, $n = 600$, and $h = 1/52$ (i.e., more than 10 years weekly data are used), ML estimates κ with 391% bias.

This finite sample problem turns out to be of great importance in the practical use of econometric estimates in asset pricing. Phillips and Yu (2005a, 2009d) took seriously the economic implications of this problem. It has been shown that there is nonlinear dependence of the pricing functional on the parameter estimates, which may well exacerbate bias and make good bias correction more subtle. In particular, even if the parameter estimates are all unbiased, the plug-in estimate of the asset price is biased due to the nonlinearity. For example, if $\kappa = 0.1$, $n = 600$, and $h = 1/52$ (i.e., more than 10 years weekly data are used), ML estimates a one-year near-the-money European option written on a 3-year discount bond with 61% bias. The bias is conceivably even larger when the option is deeper out-of-the-money and the nonlinearity becomes more pronounced; see Phillips and Yu (2009d). It is important to emphasize that the finite sample problems are not unique to ML and they are applicable to most standard estimation methods, such as GMM, nonlinear least squares, and quasi-ML.

In this section, we describe two approaches to improve the finite sample performances of ML that Peter Phillips proposed. The first of these is based on Quenouille's (1956) jackknife that is a general and computationally inexpensive method of bias reduction. The second approach is simulation-based. It involves the indirect inference estimation idea of Smith (1993) and Gouriéroux, Monfort, and Renault (1993). The two methods were used in Phillips and Yu (2005a) and Phillips and Yu (2009d), respectively. While both the jackknife and indirect inference methods have been widely used to reduce the bias in parameter estimates, the novelty in Phillips and Yu (2005a, 2009d) is that they applied the bias correction methods to asset prices directly.

4.1. Jackknife Estimation

Under quite general conditions, one can show that for standard consistent estimates such as ML estimates, there exists some constant a_1 such that

$$E(\hat{\theta}_n) = \theta + \frac{a_1}{n} + O\left(\frac{1}{n^2}\right). \quad (57)$$

According to (57), the bias decreases with the sample size. Quenouille (1956) proposed the jackknife as a solution to finite sample bias problems in parametric estimation contexts such as discrete time autoregressions. To fix ideas, let n be the number of observations in the whole sample and decompose the sample into m consecutive subsamples, each with ℓ observations, so that $n = m \times \ell$. The jackknife estimator of θ utilizes the subsample estimates of θ to assist in the bias reduction process, giving the jackknife estimator

$$\hat{\theta}_{jack} = \frac{m}{m-1} \hat{\theta}_n - \frac{\sum_{i=1}^m \hat{\theta}_{li}}{m^2 - m}, \quad (58)$$

where $\hat{\theta}_n$ and $\hat{\theta}_{li}$ are the estimates of θ obtained by application of a given method like ML to the whole sample and the i th subsample, respectively. It is easy to show that the bias in the jackknife estimate $\hat{\theta}_{jack}$ is of order $O(n^{-2})$ rather than $O(n^{-1})$.

Phillips and Yu (2005a) proposed to use the jackknife method to the quantity of interest directly. For example, if one wishes to estimate a bond option price, $c(\theta)$, instead of using $c(\hat{\theta})$, one can use

$$\hat{c}_{jack} = \frac{m}{m-1} \hat{c}_n - \frac{\sum_{i=1}^m \hat{c}_{li}}{m^2 - m} \tag{59}$$

It turns out the direct application of the jackknife to the quantity of interest yields more desirable finite sample performances. This is not surprising, because the nonlinearity in the pricing relation is taken into account in (59).

The jackknife has several nice properties. The first advantage is analytical simplicity. Unlike many other bias reduction methods, the jackknife does not rely on the explicit form of bias formula. Hence, it is applicable in a broad range of model specifications and is particularly useful when it is difficult or impossible to derive the explicit form of bias formula. A second advantage is that the jackknife is computationally much cheaper to implement. In fact, this method is not much more time consuming than the initial estimation itself. A drawback with the jackknife is that it cannot completely remove the bias, as it is only designed to decrease the order of magnitude of the bias. However, Phillips and Yu (2005a, b) showed that with a careful choice of subsampling, the jackknife can substantially reduce the bias with only a marginal increase in variance, leading to a reduction in mean squared error. Chambers (2010) and Chambers and Kyriacou (2010) obtained the asymptotic distributions of the jackknife estimator in the context of stationary and unit root autoregressive models. The asymptotic variance can be obtained from the asymptotic distributions and compared to that of the ML estimator.

Phillips and Yu (2005a, b) also compared the price implications of three different biases, namely, the specification bias, the discretization bias, and the finite sample estimation bias, in the context of affine models. It was found that the finite sample estimation bias was the most important and the discretization bias the least harmful in the near unit root models used in finance. This finding is very surprisingly and extremely important, as the discretization bias and the specification analysis have received much more attention in the continuous time financial econometrics literature. In the context of diffusions with the linear drift function, Wang, Phillips, and Yu (2011) compared analytically the magnitude of the estimation bias relative to that of the discretization bias induced by several popular discretization methods. The dominance of the estimation bias is established under rather weak conditions.

4.2. Indirect Inference Estimation

The indirect inference (II) procedure is a simulation-based estimation procedure and can be understood as a generalization of the simulated method of moments

approach of Duffie and Singleton (1993). It was first introduced by Smith (1993) and the term was coined by Gouriéroux, Monfort, and Renault (1993). It is also closely related to the method proposed by Gallant and Tauchen (1996). This method was originally proposed to deal with situations where the moments or the likelihood function of the true model are difficult to deal with (and hence traditional methods such as GMM and ML are difficult to implement), but the true model is amenable to data simulation. Because most continuous time models are easy to simulate but present difficulties in the analytic derivation of moment functions and likelihood, the II procedure has some convenient advantages in working with continuous time models in finance.

A carefully designed II estimator can also have good small sample properties of parameter estimates, as shown by MacKinnon and Smith (1998), Monfort (1996), Gouriéroux, Renault, and Touzi (2000) in the time series context, and by Gouriéroux, Phillips, and Yu (2010) in the panel context. The reason as to why II can remove the bias goes as follows. Whenever a bias occurs in an estimate and from whatever source, this bias will also be present in the same estimate obtained from data, which are of the same structure of the original data, simulated from the model for the same reasons. Hence, the bias can be calculated via simulations. The method, therefore, offers some interesting opportunities for bias correction and the improvement of finite sample properties in continuous time parameter estimation, as shown in Phillips and Yu (2009a).

To fix the idea of II for parameter estimation, consider the OU process. Suppose we need to estimate the parameter κ in:

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

from observations $\{X_h, \dots, X_{nh}\}$. An initial estimator of κ can be obtained, for example, by applying the Euler scheme to $\{X_h, \dots, X_{nh}\}$ (call it $\hat{\kappa}_n$). Such an estimator is involved with the discretization bias (due to the use of the Euler scheme) and also with a finite sample estimation bias (due to the poor finite sample property of ML in the near-unit-root situation).

Given a parameter choice κ , we apply the Euler scheme with a much smaller step size than h (say $\delta = h/10$), which leads to

$$\tilde{X}_{t+\delta}^k = \kappa(\mu - \tilde{X}_t^k)h + \tilde{X}_t^k + \sigma(\tilde{X}_t^k)\sqrt{\delta}\varepsilon_{t+\delta},$$

where

$$t = \underbrace{0, \delta, \dots, h(= 10\delta)}_{\text{group 1}}, \underbrace{h + \delta, \dots, 2h(= 20\delta)}_{\text{group 2}}, 2h + \delta, \dots, nh.$$

This sequence may be regarded as a nearly exact simulation from the continuous time OU model for small δ . We then choose every $(h/\delta)^{th}$ observation to form the sequence of $\{\tilde{X}_{ih}^k\}_{i=1}^n$, which can be regarded as data simulated directly from the OU model with the (observationally relevant) step size h .³

Let $\{\tilde{X}_h^k, \dots, \tilde{X}_{nh}^k\}$ be data simulated from the true model, where $k = 1, \dots, K$ with K being the number of simulated paths. It should be emphasized that it is important to ensure that the number of simulated observations and the sampling

interval are equal to the number of observations and the sampling interval in the observed sequence, respectively, for the purpose of the bias calibration. Another estimator of κ can be obtained by applying the Euler scheme to $\{\tilde{X}_h^k, \dots, \tilde{X}_{nh}^k\}$ (call it $\tilde{\kappa}_n^k$). Such an estimator and hence the expected value of them across simulated paths is naturally dependent on the given parameter choice κ .

The central idea in II estimation is to match the parameter obtained from the actual data with that obtained from the simulated data. In particular, the II estimator of κ solves

$$\hat{\kappa}_n = \frac{1}{K} \sum_{h=1}^K \tilde{\kappa}_n^k(\kappa) \quad \text{or} \quad \hat{\kappa}_n = \hat{\rho}_{0.5}(\tilde{\kappa}_n^k(\kappa)), \tag{60}$$

where $\hat{\rho}_\tau$ is the τ th sample quantile. In the case where K tends to infinity, the II estimator solves

$$\hat{\kappa}_n = E\left(\tilde{\kappa}_n^k(\kappa)\right) \quad \text{or} \quad \hat{\kappa}_n = \rho_{0.5}\left(\tilde{\kappa}_n^k(\kappa)\right) \tag{61}$$

where $E\left(\tilde{\kappa}_n^k(\kappa)\right)$ is called the mean binding function or the mean bias function, and $\rho_{0.5}\left(\tilde{\kappa}_n^k(\kappa)\right)$ is the median binding function or the median bias function, i.e.,

$$b_n(\kappa) = E\left(\tilde{\kappa}_n^k(\kappa)\right), \quad \text{or} \quad b_N(\kappa) = \rho_{0.5}\left(\tilde{\kappa}_n^k(\kappa)\right).$$

It is a finite sample functional relating the bias to κ . In the case where b_n is invertible, the indirect inference estimator is given by:⁴

$$\hat{\kappa}_n^{II} = b_n^{-1}(\hat{\kappa}_n). \tag{62}$$

When a median binding function is used, the estimator is the median unbiased estimator of Andrews (1993). Typically, the binding functions cannot be computed analytically in either case. That is why II needs to calculate the binding functions via simulations. While the mean is often used in the literature for the binding function, the median has certain advantages over the mean. First, the median is more robust to outliers than the mean. Second, it is easier to obtain the unbiased property via the median. In particular, while the linearity of $b_n(\kappa)$ gives rise to the mean-unbiasedness in $\hat{\kappa}_n^{II}$, only monotonicity is needed for $b_n(\kappa)$ to ensure the median-unbiasedness (Phillips and Yu, 2009d).

There are several advantages in the II procedure relative to the jackknife procedure. First, indirect inference is more effective in removing the bias in parameter estimates. Phillips and Yu (2009a) provided evidence to support this superiority of indirect inference. Second, the bias reduction may be achieved often without an increase in variance. In extreme cases of root near unity, the variance of II can be even smaller than that of ML (Phillips and Yu, 2009a). To see this, note that equation (62) implies:

$$Var\left(\hat{\kappa}_n^{II}\right) = \left(\frac{\partial b_n}{\partial \kappa}\right)^{-1} Var\left(\hat{\kappa}_n^{ML}\right) \left(\frac{\partial b_n}{\partial \kappa'}\right)^{-1}.$$

When $\partial b_n / \partial \kappa > 1$, the II has a smaller variance than ML.

A disadvantage in the II procedure is the high computational cost. It is expected that with the continuing explosive growth in computing power, such a drawback will become less of a concern. Nevertheless, to reduce the computational cost, one can choose a fine grid of discrete points of κ and obtain the binding function on the grid. Then standard interpolation and extrapolation methods can be used to approximate the binding functions at any point.

As pointed out earlier, since prices of contingent-claims are always nonlinear transformations of the system parameters, the insertion of even unbiased estimators into the pricing formulae will not assure unbiased estimation of a contingent-claim price. The stronger the nonlinearity, the larger the bias. As a result, plugging-in the indirect inference estimates into the pricing formulae may still yield an estimate of the price with unsatisfactory finite sample performances. This feature was illustrated in the context of various continuous time models and contingent claims in Phillips and Yu (2009d). To improve the finite sample properties of the contingent price estimate, Phillips and Yu (2009d) generalized the II procedure so that it is applied to the quantity of interest directly.

To fix the idea, suppose θ is the scalar parameter in the continuous time model on which the price of a contingent claim, $P(\theta)$, is based. Denote by $\hat{\theta}_n^{ML}$ the MLE of θ that is obtained from the actual data, and write $\hat{P}_n^{ML} = P(\hat{\theta}_n^{ML})$ be the ML estimate of P . \hat{P}_n^{ML} involves finite sample estimation bias due to the nonlinearity of the pricing function P in θ , or the use of the biased estimate $\hat{\theta}_n^{ML}$, or both these effects. The II approach involves the following steps:

1. Given a value for the contingent-claim price p , compute $P^{-1}(p)$ (call it $\theta(p)$), where $P^{-1}(\cdot)$ is the inverse of the pricing function $P(\theta)$.
2. Let $\tilde{\mathbf{S}}^k(p) = \{\tilde{S}_1^k, \tilde{S}_2^k, \dots, \tilde{S}_T^k\}$ be data simulated from the time series model (26) given $\theta(p)$, where $k = 1, \dots, K$ with K being the number of simulated paths. As argued above, we choose the number of observations in $\tilde{\mathbf{S}}^k(p)$ to be the same as the number of actual observations in \mathbf{S} for the express purpose of finite sample bias calibration.
3. Obtain $\tilde{\phi}_n^{ML,k}(p)$, the MLE of θ , from the k th simulated path, and calculate $\tilde{P}_n^{ML,k}(p) = P(\tilde{\phi}_n^{ML,k}(p))$.
4. Choose p so that the average behavior of $\tilde{P}_n^{ML,k}(p)$ is matched with \hat{P}_n^{ML} to produce a new bias corrected estimate.

The procedure can be generalized to cases where θ is a K -dimensional vector and where θ is obtained from cross-sectional data; see Phillips and Yu (2009d) for detailed discussion. Phillips and Yu (2009d) performed extensive Monte Carlo studies, showing that II works well, not only relative to ML but also relative to the jackknife procedure. Another simulation-based procedure for bias correction is the bootstrap method considered in Tang and Chen (2009). The performance of the bootstrap method is found to be dominated by II.

5. SOME NEW RESULTS

In econometrics, asymptotic theory always relies on the fiction of a sample with an infinity of observations. As shown earlier, there are three ways to do asymptotics in continuous time models, *long-span* ($T \rightarrow \infty$), *in-fill* ($h \rightarrow 0$), and *double*, all leading to a sample with infinite observations.

In this section, we will compare the performance of these three alternative asymptotic distributions in the context of the OU process with a known long-run mean:

$$dX(t) = -\kappa X(t)dt + dW(t), \tag{63}$$

where κ , the mean reversion parameter, is the parameter of the interest and is assumed to be positive.

Data, namely, $\{X_{0h}, X_{1h}, \dots, X_{nh}\}$ with $nh = T$, are simulated from the exact discrete time model:

$$X_t = \exp(-\kappa h)X_{t-1} + \varepsilon_t$$

and κ is estimated by:

$$\hat{\kappa} = -\frac{\ln \left\{ \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2} \right\}}{h}, \tag{64}$$

which is the ML estimate.

Since $\kappa > 0$, when $T \rightarrow \infty$ and h is fixed, the standard, *long-span*, asymptotic theory implies that

$$\sqrt{n}(\hat{\kappa} - \kappa) \xrightarrow{d} N \left(0, \frac{\exp(2\kappa h) - 1}{h^2} \right). \tag{65}$$

Hence,

$$\hat{\kappa} \overset{a}{\sim} N \left(\kappa, \frac{\exp(2\kappa h) - 1}{hT} \right). \tag{66}$$

The asymptotic normality is not surprising as the process is stationary, ergodic, and asymptotically independent. Taking h to the limit of 0 in (66), we obtain the *double* asymptotic distribution:

$$\hat{\kappa} \overset{a}{\sim} N \left(\kappa, \frac{2\kappa}{T} \right). \tag{67}$$

If $h \rightarrow 0$ and T is fixed, the continuous record is observable and the *in-fill* log-likelihood function is given by:

$$\int_0^T -\kappa X(t)dX(t) - \int_0^T \frac{1}{2} \kappa^2 X(t)^2 dt. \tag{68}$$

From Phillips (1987b), the finite sample distribution of the ML estimator based on (68) is known. Hence, it is natural to use this distribution to approximate the distribution of $\hat{\kappa}$, namely,

$$\hat{\kappa} \overset{a}{\sim} \kappa - \frac{\int_0^T X(t)dW(t)}{\int_0^T X(t)^2 dt}. \tag{69}$$

Obviously, this *in-fill* limiting distribution is asymmetric and nonnormal. This compares interestingly with the two limiting distribution in (66) and (67).

To facilitate computation of the distribution in (69), we first prove a lemma. This result can be found in Lánska (1979) and the proof is given in the Appendix.

LEMMA. Assume $X(t)$ follows the SDE:

$$dX(t) = \mu(t, X(t), \theta)dt + dW(t). \tag{70}$$

Then the log-likelihood of $\{X(t)\}_{t=0}^T$ is given by:

$$F(T, X(T), \theta) - \int_0^T \left\{ f(t, X(t), \theta) + \frac{1}{2}\mu^2(t, X(t), \theta) \right\} dt, \tag{71}$$

where

$$F(t, X(t), \theta) = \int_0^{X(t)} \mu(t, y, \theta)dy,$$

and

$$f(t, X(t), \theta) = \frac{\partial F(t, X(t), \theta)}{\partial t} + \frac{1}{2} \frac{\partial \mu(t, X(t), \theta)}{\partial x}.$$

Applying Lemma to the model given in (63), we have:

$$F(t, X(t), \kappa) = \int_0^{X(t)} -\kappa y dy = -\frac{1}{2}\kappa X^2(t) \text{ and } f(t, X(t), \kappa) = -\kappa.$$

Hence, the log-likelihood can be rewritten as:

$$-\frac{1}{2}\kappa X(T)^2 - \int_0^T \left[-\frac{1}{2}\kappa + \frac{1}{2}\kappa^2 X(t)^2 \right] dt. \tag{72}$$

The finite sample distribution of the ML estimator that maximizes (72) is

$$\frac{-X(T)^2 + T}{2 \int_0^T X(t)^2 dt}. \tag{73}$$

This is the *in-fill* asymptotic distribution of $\hat{\kappa}$ defined in (64). Compared with the asymptotic distribution (69), (73) only needs to calculate a Riemann integral. Furthermore, (73) compares interestingly with the unit root limit distribution

$$\frac{W(1)^2 - 1}{2 \int_0^1 W(t)^2 dt}$$

obtained in Phillips (1987a).

To compare the performance of the three limiting distributions, (66), (67), and (73), in the context of financial time series, we simulate data from the OU model with various values for κ , h , and T . The actual finite sample distribution is obtained from 50,000 replications of the estimates of $\hat{\kappa}$ given by (64). The true values of κ are set at 0.1, 1, 10. The value of 0.1 is empirically realistic for interest rate data while the value of 1 is empirically realistic for volatility. These two values suggest a slow speed of mean reversion. While $\kappa = 10$ is not empirically very realistic for financial time series, we include it here for the purpose of comparison. The true values of h are set at 1/12 and 1/250, corresponding to monthly and daily frequencies, respectively. It is rare in the continuous time finance literature that one uses data at a frequency lower than monthly. At the same time, it is more and more common to acquire data at a frequency lower than daily in empirical work. So $h = 1/12$ is an upper bound in empirical financial studies. The true values of T are set at 2, 10, 50. A 50-year time series span is perhaps close to the maximum in empirical work.

Tables 1-3 report the 0.5%, 1%, 5%, 10%, 90%, 95%, 99%, and 99.5% quantiles of the four distributions, for $\kappa = 0.1, 1, 10$, respectively. A few results emerge

TABLE 1. Performance of three asymptotic distributions, (66), (67) and (73), when $\kappa = 0.1$. This table reports the 0.5%, 1%, 5%, 10%, 90%, 95%, 99%, and 99.5% quantiles of the exact distribution, the *in-fill* asymptotic distribution, the *double* asymptotic distribution, and the *long-span* asymptotic distribution. The true distribution is obtained via simulations based on 50,000 replications

h and T	Methods	0.5%	1%	5%	10%	90%	95%	99%	99.5%
$h = 1/12$ $T = 2$	exact	-0.8447	-0.7042	-0.3563	-0.2155	1.7071	2.6012	5.1676	6.5470
	in-fill	-0.8518	-0.7200	-0.4355	-0.3330	1.1327	1.9147	4.0532	5.0527
	double	-0.7145	-0.6357	-0.4201	-0.3053	0.5053	0.6201	0.8357	0.9145
	long-span	-0.7355	-0.6546	-0.4336	-0.3157	0.5157	0.6336	0.8546	0.9355
$h = 1/12$ $T = 10$	exact	-0.1427	-0.1099	-0.0329	0.0027	0.5885	0.7869	1.3212	1.5574
	in-fill	-0.1513	-0.1211	-0.0511	-0.0218	0.4831	0.6724	1.1543	1.3987
	double	-0.2643	-0.2290	-0.1326	-0.0812	0.2812	0.3326	0.4290	0.4643
	long-span	-0.2673	-0.2318	-0.1346	-0.0828	0.2828	0.3346	0.4318	0.4673
$h = 1/12$ $T = 50$	exact	0.0145	0.0214	0.0401	0.0521	0.2470	0.2987	0.4150	0.4700
	in-fill	0.0116	0.0181	0.0353	0.0461	0.2296	0.2806	0.3917	0.4453
	double	-0.0629	-0.0471	-0.0040	0.0189	0.1811	0.2040	0.2471	0.2629
	long-span	-0.0637	-0.0479	-0.0046	0.0185	0.1815	0.2046	0.2479	0.2637
$h = 1/250$ $T = 2$	exact	-0.7886	-0.6447	-0.3264	-0.1962	1.6350	2.4563	4.7549	5.8324
	in-fill	-0.8691	-0.7188	-0.4333	-0.3329	1.1297	1.9050	4.0505	5.1093
	double	-0.7145	-0.6357	-0.4201	-0.3053	0.5053	0.6201	0.8357	0.9145
	long-span	-0.7155	-0.6365	-0.4208	-0.3057	0.5057	0.6208	0.8365	0.9155

Table continues on overleaf

TABLE 1. *continued*

h and T	Methods	0.5%	1%	5%	10%	90%	95%	99%	99.5%
$h = 1/250$ $T = 10$	exact	-0.1345	-0.1064	-0.0313	0.0022	0.5793	0.7792	1.2916	1.5214
	in-fill	-0.1533	-0.1224	-0.0524	-0.0222	0.4829	0.6715	1.1653	1.4047
	double	-0.2643	-0.2290	-0.1326	-0.0812	0.2812	0.3326	0.4290	0.4643
	long-span	-0.2644	-0.2291	-0.1327	-0.0813	0.2813	0.3327	0.4291	0.4644
$h = 1/250$ $T = 50$	exact	0.0131	0.0204	0.0398	0.0514	0.2447	0.2987	0.4141	0.4656
	in-fill	0.0101	0.0164	0.0348	0.0457	0.2281	0.2804	0.3952	0.4432
	double	-0.0629	-0.0471	-0.0040	0.0189	0.1811	0.2040	0.2471	0.2629
	long-span	-0.0629	-0.0472	-0.0041	0.0189	0.1811	0.2041	0.2472	0.2629

TABLE 2. Performance of three asymptotic distributions, (66), (67), and (73), when $\kappa = 1$. This table reports the 0.5%, 1%, 5%, 10%, 90%, 95%, 99%, and 99.5% quantiles of the exact distribution, the *in-fill* asymptotic distribution, the *double* asymptotic distribution, and the *long-span* asymptotic distribution. The true distribution is obtained via simulations based on 50,000 replications

h and T	Methods	0.5%	1%	5%	10%	90%	95%	99%	99.5%
$h = 1/12$ $T = 2$	exact	-0.6044	-0.4450	0.0507	0.2873	4.2440	5.6903	9.3293	11.1353
	in-fill	-0.4827	-0.3312	0.0295	0.2044	3.4411	4.5685	7.4923	8.3460
	double	-1.5758	-1.3263	-0.6449	-0.2816	2.2816	2.6449	3.3263	3.5758
	long-span	-1.7448	-1.4789	-0.7527	-0.3656	2.3656	2.7527	3.4789	3.7448
$h = 1/12$ $T = 10$	exact	0.3105	0.3589	0.5059	0.6029	1.9321	2.2628	3.0417	3.4019
	in-fill	0.3190	0.3589	0.4949	0.5863	1.8033	2.1118	2.7756	3.0594
	double	-0.1519	-0.0404	0.2644	0.4269	1.5731	1.7356	2.0404	2.1519
	long-span	-0.2067	-0.0898	0.2294	0.3996	1.6004	1.7706	2.0898	2.2067
$h = 1/12$ $T = 50$	exact	0.5999	0.6324	0.7277	0.7826	1.3311	1.4394	1.6541	1.7341
	in-fill	0.6058	0.6394	0.7309	0.7844	1.3052	1.4040	1.6063	1.6902
	double	0.4848	0.5347	0.6710	0.7437	1.2563	1.3290	1.4653	1.5152
	long-span	0.4622	0.5142	0.6565	0.7324	1.2676	1.3435	1.4858	1.5378
$h = 1/250$ $T = 2$	exact	-0.4481	-0.2763	0.1281	0.3356	3.8512	4.9482	7.7366	9.0473
	in-fill	-0.5183	-0.3584	0.0232	0.2018	3.3855	4.3957	7.1799	8.3709
	double	-1.5758	-1.3263	-0.6449	-0.2816	2.2816	2.6449	3.3263	3.5758
	long-span	-1.5836	-1.3333	-0.6498	-0.285	2.2854	2.6498	3.3333	3.5836
$h = 1/250$ $T = 10$	exact	0.3441	0.3880	0.5297	0.6217	1.8825	2.1808	2.8451	3.1350
	in-fill	0.3255	0.3674	0.5002	0.5874	1.8062	2.0977	2.7630	3.0470
	double	-0.1519	-0.0404	0.2644	0.4269	1.5731	1.7356	2.0404	2.1519
	long-span	-0.1545	-0.0427	0.2628	0.4256	1.5744	1.7372	2.0427	2.1545
$h = 1/250$ $T = 50$	exact	0.6229	0.6494	0.7365	0.7915	1.3129	1.4098	1.6178	1.7071
	in-fill	0.6173	0.6447	0.7291	0.7833	1.3010	1.3971	1.6027	1.6873
	double	0.4848	0.5347	0.6710	0.7437	1.2563	1.3290	1.4653	1.5152
	long-span	0.4838	0.5338	0.6704	0.7432	1.2568	1.3296	1.4662	1.5162

TABLE 3. Performance of three asymptotic distributions, (66), (67), and (73), when $\kappa = 10$. This table reports the 0.5%, 1%, 5%, 10%, 90%, 95%, 99%, and 99.5% quantiles of the exact distribution, the *in-fill* asymptotic distribution, the *double* asymptotic distribution, and the *long-span* asymptotic distribution. The true distribution is obtained via simulations based on 50,000 replications

<i>h</i> and <i>T</i>	Methods	0.5%	1%	5%	10%	90%	95%	99%	99.5%
<i>h</i> = 1/12 <i>T</i> = 10	exact	5.7633	6.0789	7.0056	7.5923	13.6800	15.0377	18.1743	19.5199
	in-fill	7.0171	7.2564	7.9545	8.3715	12.0206	12.6204	13.9213	14.3905
	double	6.3572	6.7100	7.6738	8.1876	11.8124	12.3262	13.2900	13.6428
	long-span	4.1281	4.6968	6.2503	7.0785	12.9215	13.7497	15.3032	15.8719
<i>h</i> = 1/12 <i>T</i> = 50	exact	7.7717	7.9616	8.4936	8.7995	11.4508	11.9073	12.7915	13.1478
	in-fill	8.5167	8.6440	9.0163	9.2234	10.8662	11.1287	11.6198	11.8075
	double	8.3709	8.5287	8.9597	9.1895	10.8105	11.0403	11.4713	11.6291
	long-span	7.3828	7.6363	8.3287	8.6979	11.3021	11.6713	12.3637	12.6172
<i>h</i> = 1/250 <i>T</i> = 2	exact	4.6178	4.9565	6.1693	6.9406	15.6703	17.5964	21.7356	23.4049
	in-fill	4.5770	4.9339	6.0794	6.8284	15.1978	17.0108	20.9087	22.3676
	double	1.8545	2.6434	4.7985	5.9474	14.0526	15.2015	17.3566	18.1455
	long-span	1.6805	2.4863	4.6874	5.8608	14.1392	15.3126	17.5137	18.3195
<i>h</i> = 1/250 <i>T</i> = 10	exact	6.9876	7.2632	7.9743	8.3950	12.1621	12.8098	14.1273	14.6391
	in-fill	7.0017	7.2722	7.9799	8.3712	12.0663	12.6887	13.9985	14.5156
	double	6.3572	6.7100	7.6738	8.1876	11.8124	12.3262	13.2900	13.6428
	long-span	6.2824	6.6425	7.6260	8.1504	11.8496	12.3740	13.3575	13.7176
<i>h</i> = 1/250 <i>T</i> = 50	exact	8.5022	8.6453	9.0254	9.2301	10.8840	11.1439	11.6297	11.8125
	in-fill	8.5152	8.6540	9.0270	9.2267	10.8525	11.1079	11.5879	11.7759
	double	8.3709	8.5287	8.9597	9.1895	10.8105	11.0403	11.4713	11.6291
	long-span	8.3377	8.4987	8.9385	9.1730	10.8270	11.0615	11.5013	11.6623

from the tables. First and more importantly, the *in-fill* asymptotics almost always performs significantly better than *long-span* and *double* asymptotics, regardless of the value for κ , h , and T .⁵ It is remarkable that the superiority even holds true when $h = 1/12$ and $T = 50$ for $\kappa = 0.1$ of 1. This is the worst case scenario where the *long-span* asymptotics is favored over the *in-fill* asymptotics. Second, the larger is T , the better *long-span* asymptotics. Third, the smaller is h , the better *in-fill* asymptotics. Fourth, the larger is κ , the better the three asymptotic distributions. Fifth, the *long-span* asymptotic distribution seems to perform better in the left tail than in the right tail. However, both the exact and the *in-fill* asymptotic distributions are heavily skewed. That partly explains why the *in-fill* asymptotics outperforms the *long-span* and *double* asymptotics. Figures 1-2 plot the four densities when $\kappa = 0.1$, $h = 1/12$, and $T = 2$ and 10. The plot reinforces the results found in the tables.

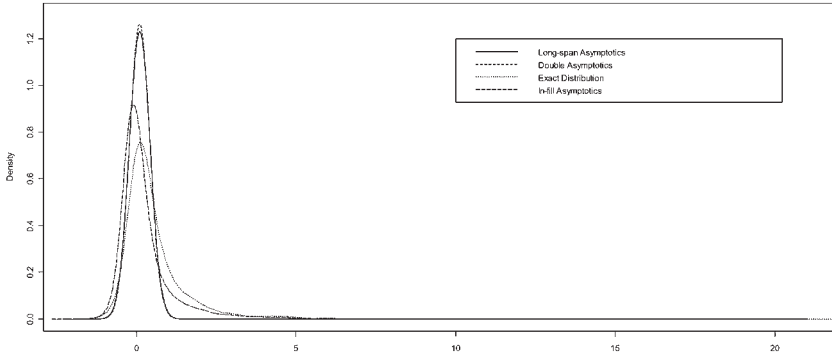


FIGURE 2. Four densities of $\hat{\kappa}$ when $\kappa = 0.1$, $h = 1/12$ and $T = 2$.

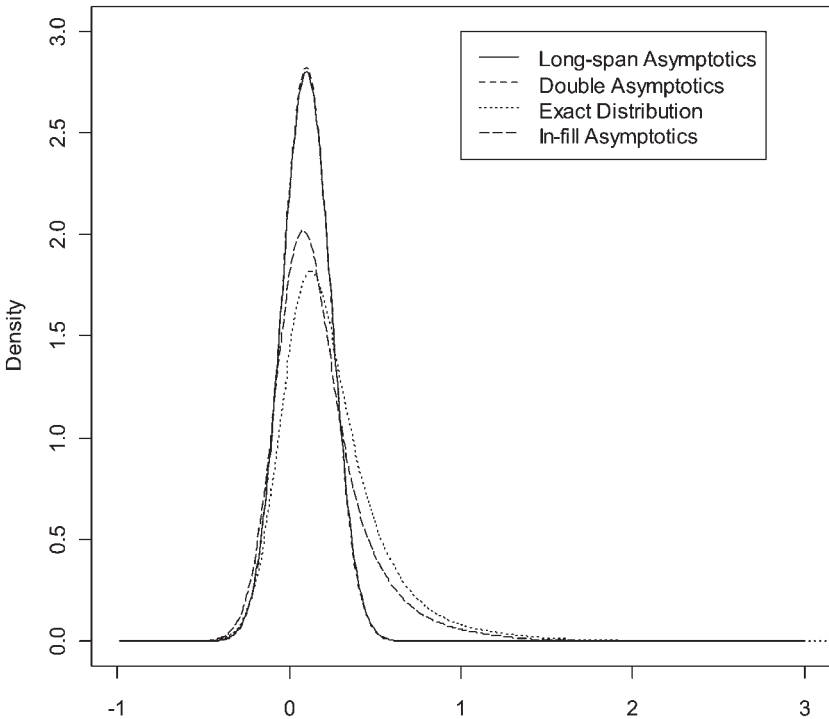


FIGURE 3. Four densities of $\hat{\kappa}$ when $\kappa = 0.1$, $h = 1/12$ and $T = 10$.

6. CONCLUSIONS

The theoretical development of econometric analysis of continuous time models has come a long way over recent decades. This paper has outlined some main developments in the methodology that Peter Phillips both initiated and undertook

to develop in the last 40 years. It is clear that while his contributions are mostly focused on theoretical aspects, his impact extends far beyond theory. Without any doubt, he will continue to make more insightful contributions in this important area, given the current level of his energy and devotion to this area of research.

NOTES

1. To simplify the notations, we write $A(\theta)$, $B(\theta)$, $P(\theta)$, and $\Lambda(\theta)$ by A , B , P , and Λ , respectively.
2. In the multivariate case, $\kappa > 0$ means that the eigenvalues of κ are all positive.
3. If the transition density of $X_{t+h}|X_t$ for the continuous time model is analytically available, exact simulation can be directly obtained. In this case, the Euler scheme at a finer grid is not necessary.
4. While the invertibility of b_n and b_N is not likely to be a strong assumption, as argued in Andrews (1993), the condition is difficult to verify in many interesting models.
5. The only exception is when $\kappa = 10$, $h = 1/12$ (hence, the model is very far away from the unit root), and T is large.

REFERENCES

- Aït-Sahalia, Y. (1996) Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.
- Aït-Sahalia, Y. (2002) Maximum likelihood estimation of discretely sampled diffusion: A closed-form approximation approach. *Econometrica* 70, 223–262.
- Aït-Sahalia, Y. (2007) Estimating continuous-time models using discretely sampled data. In R. Blundell, T. Persson, & W. K. Newey (eds.), *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*. Econometric Society Monographs. Cambridge University Press, 261–327.
- Andersen, T. & T. Bollerslev (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 4, 885–905.
- Andersen, T., T. Bollerslev, F.X. Diebold, & P. Labys (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.
- Andrews, D.W.K. (1993) Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* 61, 139–166.
- Bachelier, L. (1900) *Théorie de la spéculation*. Ph.D. thesis, Sorbonne, France.
- Bailey, R.W., V.B. Hall, & P.C.B. Phillips (1987) A model of output, employment, capital formation and inflation. In G. Gandolfo & F. Marzano (eds.), *Saggi in memoria di Victorio Marrama*. Giuffrè, pp. 703–776.
- Bandi, F. M. (2002) Short-term interest rate dynamics: A spatial approach. *Journal of Financial Economics* 65, 73–110.
- Bandi, F.M. & G. Moloche (2008) On the Functional Estimation of Multivariate Diffusion Processes. Working paper, University of Chicago.
- Bandi, F. M. & T. Nguyen (2003) On the functional estimation of jump-diffusion models. *Journal of Econometrics* 116, 293–328.
- Bandi, F. M. & P.C.B. Phillips (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.
- Bandi, F. M. & P.C.B. Phillips (2007) A simple approach to the parametric estimation of potentially nonstationary diffusions. *Journal of Econometrics* 137, 354–395.
- Bandi, F.M. & P.C.B. Phillips (2009) Nonstationary continuous-time processes. In Y. Aït-Sahalia & L.P. Hansen (eds.), *Handbook of Financial of Econometrics*. North-Holland.
- Barndorff-Nielsen, O. & N. Shephard (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253–280.

- Barndorff-Nielsen, O. & N. Shephard (2005) How accurate is the asymptotic approximation to the distribution of realized volatility? In D.W.K. Andrews, J. Powell, P. Ruud, & J. Stock (eds.), *Identification and Inference for Econometric Models*. Cambridge University Press, pp. 306–331.
- Bartlett, M.S. (1946) On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of Royal Statistical Society* 8, 27–41.
- Bergstrom, A.R. (1966) Nonrecursive models as discrete approximations to system of stochastic differential equations. *Econometrica* 34, 173–182.
- Bergstrom, A.R. & C.R. Wymer (1976) A model of disequilibrium neoclassical growth with its application to the United Kingdom. In A.R. Bergstrom (ed.), *Statistical Inference in Continuous Time Economic Models*. North-Holland Publishing Company.
- Bergstrom, A.R. (1983) Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica* 51, 117–152.
- Bergstrom, A.R. (1984) Continuous time stochastic models and issues of aggregation over time. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. II. Elsevier Science, pp. 1145–1212.
- Bergstrom, A.R. (1985a). The estimation of nonparametric functions in a Hilbert space. *Econometric Theory* 1, 7–26.
- Bergstrom, A.R. (1985b) The estimation of parameters in nonstationary higher-order continuous-time dynamic models. *Econometric Theory* 1, 369–385.
- Bergstrom, A.R. (1986) The estimation of open higher-order continuous time dynamic models with mixed stock and flow data. *Econometric Theory* 2, 350–373.
- Bergstrom, A.R. (1990) *Continuous Time Econometric Modelling*. Oxford University Press.
- Bergstrom, A.R. (1996) Survey of continuous time econometrics. In W.A. Barnett, G. Gandolfo, & C. Hillinger (eds.), *Dynamic Disequilibrium Modelling*. Cambridge University Press, pp. 3–25.
- Bergstrom, A.R. & B. Nowman (2007) *A Continuous Time Econometric Model of the United Kingdom with Stochastic Trends*. Cambridge University Press.
- Black, F. & M. Scholes (1973) The pricing of options and corporate liabilities. *Journal of Political Economics* 81, 637–654.
- Brugiere, P. (1993) Theoreme de Limite Centrale pour un Estimateur Nonpara-metrique de la Variance d'un Processus de Diffusion Multidimensionnelle. *Annales de l'Institut Henri Poincar'e* 29, 357–389.
- Chambers, M. (2010) Jackknife Estimation of Stationary Autoregressive Models. Working paper, Department of Economics, University of Essex.
- Chambers, M. & M. Kyriacou (2010) Jackknife Bias Reduction in the Presence of a Unit Root. Working paper, Department of Economics, University of Essex.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff, & A.B. Sanders (1992) An empirical comparison of alternative models of short term interest rates. *Journal of Finance* 47, 1209–1227.
- Chang, Y. (2012) Taking a New Contour: A Novel Approach to Panel Unit Root Tests. *Journal of Econometrics* 169, 15–28.
- Chen, X., L.P. Hansen, & J. Scheinkman (2000) Shape-preserving Estimation of Diffusions. Working paper, University of Chicago.
- Chen, X., L.P. Hansen, & J. Scheinkman (2009) Principal components and the long run. *Annals of Statistics* 37, 4279–4312.
- Corradi, V. & W. Distaso (2006) Semiparametric comparison of stochastic volatility models via realized measures. *Review of Economic Studies* 73, 635–667.
- Corradi, V. & W. Distaso (2010) Deterministic versus Stochastic Volatility. Manuscript, University of Warwick.
- Cox, J., Ingersoll, J., & S. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53, 385–407.
- Drost, F.C. & T.E. Nijman (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.
- Duffie, D. & R. Kan (1996) A yield-factor model of interest rate. *Mathematical Finance* 6, 379–406.

- Duffie, D. & K.J. Singleton (1993) Simulated moments estimation of Markov models of asset prices. *Econometrica* 61, 929–952.
- Elerian, O., Chib, S., & N. Shephard (2001) Likelihood inference for discretely observed non-linear diffusions. *Econometrica* 69, 959–993.
- Fan, J. (2005) A selective overview of nonparametric methods in financial econometrics (with discussion). *Statistical Science* 20, 317–357.
- Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790–804.
- Gallant, A. R. & G. Tauchen (1996) Which moments to match? *Econometric Theory* 12, 657–681.
- Geman, S.A. (1979) On a Common Sense Estimator for the Drift of a Diffusion. Manuscript, Brown University.
- Gouriéroux, C., A. Monfort, & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.
- Gouriéroux, C., P.C.B. Phillips & J. Yu (2010) Indirect inference for dynamic panel models. *Journal of Econometrics* 157, 68–77.
- Gouriéroux, C., E. Renault & N. Touzi (2000) Calibration by simulation for small sample bias correction. In R.S. Mariano, T. Schuermann, & M. Weeks (eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, pp. 328–358. Cambridge University Press.
- Hansen, L.P. & T.J. Sargent (1983) The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* 51, 377–387.
- Heston, S.L. (1993) A closed-form solution for options with stochastic volatility, with application to bond and currency options. *Review of Financial Studies* 6, 327–343.
- Jacod, J. (1993) Limit of random measures associated with the increments of a Brownian semimartingale. Mimeo, Laboratoire de Probabilités, Paris.
- Jamshidian, F. (1989) An exact bond option formula. *Journal of Finance* 44, 205–209.
- Jeffrey, A., D. Kristensen, O. Linton, T. Nguyen, & P.C.B. Phillips (2004) Nonparametric estimation of a multifactor Heath-Jarrow-Morton model: An integrated approach. *Journal of Financial Econometrics* 2, 251–289.
- Kessler, M. & A. Rahbek (2001) Asymptotic continuous time likelihood based cointegration inference. *Scandinavian Journal of Statistics* 28, 455–470.
- Kessler, M. & A. Rahbek (2004) Identification and inference for cointegrated and ergodic Gaussian diffusions. *Statistical Inference for Stochastic Processes* 36, 153–188.
- Kristensen, D. (2010a) Pseudo-maximum likelihood estimation in two classes of semiparametric diffusion models. *Journal of Econometrics* 156, 239–259.
- Kristensen, D. (2010b) Semi-nonparametric Estimation and Misspecification Testing of Diffusion Models. Discussion Paper, Department of Economics, University of Copenhagen.
- Lánska, V. (1979) Minimum contrast estimation in diffusion processes. *Journal of Applied Probability* 16, 65–75.
- Lo, A. W. (1988) Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory* 4, 231–247.
- MacKinnon, J.G. & A.A. Smith (1998) Approximate bias correction in econometrics. *Journal of Econometrics* 85, 205–230.
- McCrorie, J.R. (2003) The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematicae* 79, 9–16.
- McCrorie, J.R. (2009) Estimating continuous time models on the basis of discrete data via an exact discrete analog. *Econometric Theory* 25, 1120–1137.
- Merton, R.C. (1973) Theory of rational option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.
- Merton, R. C. (1980) On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323–361.
- Merton, R.C. (1990) *Continuous-time Finance*. Blackwell.
- 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.

- Park, J. (2008a) The Spatial Analysis of Time Series. Manuscript, Department of Economics, Texas A&M University.
- Park, J. (2008b) Martingale Regression and Time Change. Manuscript, Department of Economics, Texas A&M University.
- Park, J. & P.C.B. Phillips (2001) Nonlinear regression with integrated time series. *Econometrica* 69, 117–161.
- 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 a stochastic differential equation system. *Econometrica* 6, 1021–1041.
- Phillips, P.C.B. (1973) The problem of identification in finite parameter continuous time models. *Journal of Econometrics* 4, 351–262.
- Phillips, P.C.B. (1974) The estimation of some continuous time models. *Econometrica* 42, 803–823.
- Phillips, P.C.B. (1976a) The estimation of linear stochastic differential equations with exogenous variables. In A.R. Bergstrom (ed.), *Statistical Inference in Continuous Time Economic Models*. North-Holland, pp. 135–174.
- Phillips, P.C.B. (1976b) Some computations based on observed data series of the exogenous variable component in continuous systems. In A.R. Bergstrom (ed.), *Statistical Inference in Continuous Time Economic Models*. North-Holland, pp. 175–214.
- Phillips, P.C.B. (1978) The treatment of flow data in the estimation of continuous time systems. In A.R. Bergstrom, A.J.L. Catt & M. Preston (eds.), *Stability and Inflation: A Volume of Essays to Honour the Memory of A. W. H. Phillips*, Wiley, pp. 257–274.
- Phillips, P.C.B. (1987a) Time series regression with a unit root. *Econometrica* 55, 277–301.
- Phillips, P.C.B. (1987b) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
- Phillips, P.C.B. (1988) An interview with Professor A. R. Bergstrom. *Econometric Theory* 4, 301–328.
- Phillips, P.C.B. (1991) Error correction and long run equilibria in continuous time. *Econometrica* 59, 967–980.
- Phillips, P.C.B. (1993) Rex Bergstrom's career and research. In P.C.B. Phillips (ed.), *Models, Methods and Applications of Econometrics: Essays in Honor of A. R. Bergstrom*. Basil Blackwell, pp. 3–8.
- Phillips, P.C.B. (2010) Two New Zealand pioneer econometricians. *New Zealand Economic Papers* 44, 1–26.
- Phillips, P.C.B. & J.Y. Park (1999) Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15, 269–298.
- Phillips, P.C.B. & J. Yu (2005a) Jackknifing bond option prices. *Review of Financial Studies* 18, 707–742.
- Phillips, P.C.B. & J. Yu (2005b) Comment: A selective overview of nonparametric methods in financial econometrics. *Statistical Science* 20, 338–343.
- Phillips, P.C.B. & J. Yu (2009a) Maximum likelihood and Gaussian estimation of continuous time models in finance. *Handbook of Financial Time Series*, pp. 497–530.
- Phillips, P.C.B. & J. Yu (2009b) A two-stage realized volatility approach to estimation of diffusion processes with discrete data. *Journal of Econometrics* 150, 139–150.
- Phillips, P.C.B. & J. Yu (2009c) Information Loss in Volatility Measurement with Flat Price Trading. Working paper, Cowles Foundation, Yale University.
- Phillips, P.C.B. & J. Yu (2009d) Simulation-based estimation of contingent-claims prices. *Review of Financial Studies* 22, 3669–3705.
- Phillips, P.C.B. & J. Yu (2011) Corrigendum to “A Gaussian Approach for Continuous Time Models of the Short Term Interest Rate”. *Econometrics Journal* 14, 126–129.
- Quenouille, M.H. (1956) Notes on bias in estimation. *Biometrika* 43, 353–360.
- Sargan, J. D. (1974) Some discrete approximations to continuous time stochastic models. *Journal of the Royal Statistical Society, Series B* 36, 74–90.

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

Sorensen, H. (2004) Parametric inference for diffusion processes observed at discrete points in time: A survey. *International Statistical Review* 72, 337–354.

Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.

Tang, C.Y. & S.X. Chen (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.

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

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

Yu, J. & P.C.B. Phillips (2001) A Gaussian approach for continuous time models of the short term interest rate. *The Econometrics Journal* 4, 211–225.

Zhang, L., P.A. Mykland, & Y. Ait-Sahalia (2005) A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.

APPENDIX

Proofs of Lemma. From the definition of $F(t, X(t), \theta)$, we have:

$$\frac{\partial F(t, X(t), \theta)}{\partial x} = \mu(t, X(t), \theta) \quad \text{and} \quad \frac{\partial^2 F(t, X(t), \theta)}{\partial x^2} = \frac{\partial \mu(t, X(t), \theta)}{\partial x}.$$

Applying Ito’s lemma to $F(t, X(t), \theta)$, we get:

$$\begin{aligned} dF(t, X(t), \theta) &= \left(\frac{\partial F}{\partial t} + \mu(t, X(t), \theta) \frac{\partial F}{\partial x} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} \right) dt + \frac{\partial F}{\partial x} dW(t) \\ &= \left(\frac{\partial F}{\partial t} + \frac{\mu^2(t, X(t), \theta)}{2} + \frac{1}{2} \frac{\partial \mu}{\partial x} \right) dt + \mu(t, X(t), \theta) dW(t) \\ &\quad + \frac{\mu^2(t, X(t), \theta)}{2} dt. \end{aligned}$$

Hence,

$$\begin{aligned} F(T, X(T), \theta) &= \int_0^T \left(\frac{\partial F}{\partial t} + \frac{\mu^2(t, X(t), \theta)}{2} + \frac{1}{2} \frac{\partial \mu}{\partial x} \right) dt + \int_0^T \mu(t, X(t), \theta) dW(t) \\ &\quad + \int_0^T \frac{\mu^2(t, X(t), \theta)}{2} dt. \end{aligned}$$

By the Girsanov theorem, the log-likelihood of $\{X(t)\}_{t=0}^T$ is:

$$\begin{aligned} \ell(\theta) &= \int_0^T \mu(t, X(t), \theta) dX(t) - \int_0^T \frac{1}{2} \mu^2(t, X(t), \theta) dt \\ &= \int_0^T \mu(t, X(t), \theta) (dX(t) - \mu(t, X(t), \theta) dt) + \int_0^T \frac{1}{2} \mu^2(t, X(t), \theta) dt \end{aligned}$$

$$\begin{aligned}
&= \int_0^T \mu(t, X(t), \theta) dW(t) + \int_0^T \frac{1}{2} \mu^2(t, X(t), \theta) dt \\
&= F(T, X(T), \theta) - F(0, X(0), \theta) + \int_0^T \mu(t, X(t), \theta) dW(t) \\
&\quad + \int_0^T \frac{1}{2} \mu^2(t, X(t), \theta) dt \\
&= F(T, X(T), \theta) - \int_0^T \left(\frac{\partial F}{\partial t} + \frac{\mu^2(t, X(t), \theta)}{2} + \frac{1}{2} \frac{\partial \mu}{\partial x} \right) dt,
\end{aligned}$$

where the last equality is obtained from the expression of $F(T, X(T), \theta)$. This proves the lemma. ■