8-2014

Econometric analysis of continuous time models: A survey of Peter Phillips' work and some new results

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

Citation

YU, Jun. Econometric analysis of continuous time models: A survey of Peter Phillips' work and some new results. (2014). *Econometric Theory*. 30, (4), 737-774.
Available at: https://ink.library.smu.edu.sg/soe_research/1425

Published in Econometric Theory, 2014, 30 (4), pp. 737-774.
https://doi.org/10.1017/S0266466613000467
Pre-print version

Econometric Analysis of Continuous Time Models: A Survey of Peter Phillips' Work and Some New Results[1]

Jun Yu[2]

April 1, 2009

[1]I gratefully acknowledge financial support from the Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS. I would like to thank Peter Phillips for extensive discussions on the subject and Qiankun Zhou for comments.

[2]School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; email: yujun@smu.edu.sg.

Abstract

Econometric analysis of continuous time models has drawn the attention of Peter Phillips for nearly 40 years, resulting in many important publications by him. In these publications he has dealt with a wide range of continuous time models and econometric 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, from stationary models to nonstationary and nearly nonstationary models. This paper provides an overview of Peter Phillips' 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 two asymptotic distributions in a simple setup. Results indicate that the in-fill asymptotics significantly outperforms the long-span 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. There is strong reason why continuous time models appeal to economists and financial specialists as "the economy does not cease to exist in between observations" (Bartlett (1946) and Phillips (1988)). On aggregate levels, economic decision makings almost always involve many agents and are typically done at different times. With the vast advancement in globalization, economic integration and information technology in modern era, news arrives at shorter intervals and economic activities take place in a non-stop 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 continuous time econometric models in macroeconomics is Reg 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). 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 many important work of his in the area.

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 nearly 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, 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 has 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 rigor of continuous time literature since 1970s. Many contributions that he 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 Phillips' extraordinary contributions to the continuous time econometrics literature. I have organized this article as follows. Section 2 reviews a variety of methods that he proposed to estimate parametric continuous time models. Section 3 reviews the identification problem that he pointed out in a class of continuous time models and the solutions that he suggested. Section 4 reviews various nonparametric methods he developed in recent years. Section 5 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 6 presents a new set of results in a simple setup and show that the in-fill asymptotic theory is superior to the long-span asymptotic theory. Section 7 concludes.

2 Estimating Parametric Continuous Time Models

2.1 Estimating Continuous Time Macroeconometric Models

The continuous time model, considered in Phillips (1972, 1974, 1976a, 1976b), takes the form of a stochastic differential equation (SDE):

$$dX(t) = (A(\theta)X(t) + B(\theta))dt + \zeta(dt), X(0) = X_0. \tag{1}$$

where $X(t) = (X_1(t), \cdots, X_M(t))'$ is a M-dimensional continuous time random process, $A(\theta)$ ($B(\theta)$) is an $M \times M$ matrix ($n \times 1$ vector) whose elements depend on unknown parameters $\theta = (\theta_1, \cdots, \theta_K)$ that need to be estimated, $\zeta(dt)$ is a vector of white noise disturbances with covariance Σ (namely $E(\zeta(dt)\zeta(dt)') = \Sigma dt$). See Bergstrom (1984) for the definition of the white noise disturbances. A special case of the white noise disturbance is $dW(t)$ where $W(t)$ is a vector of standard Browian motion. 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:

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

The observed data are assumed to be recorded discretely at $(0, h, 2h, \cdots, nh(= T))$ in the time interval $[0, T]$. So $n + 1$ is the total number of observations, 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 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, respectively. In this paper, we use $X(t)$ to represent a continuous time process and X_t to represent a discrete time process. When there is no confusion, we simply write X_{th} as X_t.

The difficulty with estimating the parameters of the continuous time model (1) lies in the fact that only discretely observed data are available. To facilitate estimation, Bergstrom (1966) proposed approximating the continuous time model (1) by the following discrete time model:

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

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

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

The discrete time model is then estimated by standard methods, such as the three-stage least squares. As model (2) is merely an approximation to model (1), 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 $T \to \infty$. The bigger the h, the larger the discretization bias.

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

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

which leads to the approximate discrete time model:

$$X_{t+1} - X_t = (A(\theta)X_t + B(\theta))h + v_{t+1}. \tag{3}$$

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 non-recursive 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 time points in time that are generated by (1) 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 (1), one can first solve the SDE for $X(t)$. Assume that $A(\theta)$ has distinct eigenvalues $\lambda_1(\theta), \ldots, \lambda_M(\theta)$, all with negative real parts. Consequently, there must exist a matrix $P(\theta)$ such that:

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

Bergstrom (1966) showed that the solution to SDE (1) is given by:

$$X(t) = \left[X(0) + A^{-1}(\theta)B(\theta)\right]\exp(A(\theta)t) - A^{-1}(\theta)B(\theta) + u_t, \tag{4}$$

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

$$X_{t+1} = \exp(A(\theta)h)X_t - A^{-1}(\theta)(\exp(A(\theta)h) - I)B(\theta) + \varepsilon_{t+1}, \tag{5}$$

where

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

Phillips (1972) used a generalized least squares 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 $T \to \infty$. Using the terminology of Peter Phillips, this form of asymptotics is called the *long-span* asymptotics. 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.

To illustrate the magnitude of the discretization error in the approximate model (3), consider the following univariate continuous time model:

$$dX(t) = \kappa(\mu - X(t))dt + \zeta(dt). \tag{6}$$

So $A = -\kappa$, $B = \kappa\mu$. 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:

$$X_{t+1} = \kappa\mu h + (1 - \kappa h)X_t + v_{t+1}, \tag{7}$$

whereas the exact discrete time model is:

$$X_{t+1} = \mu\left(1 - e^{-\kappa h}\right) + e^{-\kappa h}X_t + \varepsilon_{t+1}. \tag{8}$$

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:

$$\mu(1 - e^{-\kappa h}) = \kappa\mu h + O(h^2), \tag{9}$$

$$e^{-\kappa h} = 1 - \kappa h + O(h^2). \tag{10}$$

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:

$$dC(t) = \alpha[(1-s)Y(t) + \alpha - C(t)]dt + \zeta_1(dt), \tag{11}$$

$$dY(t) = \lambda[C(t) + DK(t) - Y(t)]dt + \zeta_2(dt), \tag{12}$$

$$dK(t) = \gamma[\nu Y(t) - K(t)]dt + \zeta_3(dt), \tag{13}$$

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 softwares in recent years, it is 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 advance, cutting edge technical innovation, sophisticated Monte Carlo exercise, 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. The concern about the 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 (1987), Pedersen (1995), Chib, Elerian 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 macroeconometric 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:

$$dX(t) = (A(\theta)X(t) + B(\theta)Z(t))dt + \zeta(dt), \tag{14}$$

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 (14) is given by:[1]

$$X_t = \exp(Ah)X_{t-1} - A^{-1}(\exp(Ah) - I)B + \int_0^h \exp(sA)BZ(th - s)ds + \varepsilon_t, \tag{15}$$

[1]To simplify the notations, we write $A(\theta)$, $B(\theta)$, $P(\theta)$, and $\Lambda(\theta)$ by A, B, P, and Λ, respectively.

where $\varepsilon_t = \int_{(t-1)h}^{th} \exp(A(\theta)(t-s))P \exp(\Lambda(t-s))P'\zeta(ds)$. 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 polynomial is simple integrable functions, Phillips (1974) proposed to expand $Z(th-s)$ in a second order Taylor series about $s=0$ and to use the three-point Lagrange interpolation formula to approximate $Z(th-s)$, namely:

$$\hat{Z}(th-s) = Z_t - s(Z_{t-2} - 4Z_{t-1} + 3Z_t) + s^2(Z_t - 2Z_{t-1} + Z_{t-2})/(2h^2).$$

Substituting out $Z(th-s)$ in (15) by $\hat{Z}(th-s)$ and integrating out the polynomials, we have:

$$X_{t+1} = \exp(A(\theta)h)X_t + E_2 Z_t + E_3 Z_{t-1} + E_4 Z_{t-2} + \eta_{t+1}, \tag{16}$$

where the expressions for $E_i, i = 2, 3, 4$ 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 (16) which depends on h. The smaller the 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 (16) because in this case (16) 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 (11), (12) and (13), can be observed only as the integrals, such as $\int_{(t-1)h}^{th} C(s)ds$ and $\int_{(t-1)h}^{th} Y(s)ds$. Let $X_t = \int_{(t-1)h}^{th} C(s)ds$ when $C(t)$ is a flow variable. The exact discrete time model for X_t 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.

2.2 Estimation Continuous Time Financial Econometric Models

An important class of continuous time models used in finance is diffusion processes, where the randomness is driven by the Brownian motion. A general time homogenous parametric diffusion process takes the form:

$$dX(t) = \mu(X(t);\theta)dt + \sigma(X(t);\theta)dW(t), \tag{17}$$

where $W(t)$ is a (vector)-standard Brownian motion, $\sigma(X(t);\theta)$ is a given diffusion function, $\mu(X(t);\theta)$ is a given drift function, 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 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, ultra high 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 $X(t)$. There are two important special univariate cases in the affine family.

First, Vasicek (1977) proposed the following Ornstein-Uhlenbeck (OU) process to describe the movement of short term interest rates:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\, dW(t), X_0 \sim N(\mu, \sigma^2/(2\kappa)) \tag{18}$$

This model is a special case of (1) and hence, the exact discrete model is of the form:

$$X_{th} = \mu\left(1 - e^{-\kappa h}\right) + e^{-\kappa h} X_{(t-1)h} + \sigma\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}\varepsilon_t, \tag{19}$$

where $\varepsilon_t \sim$ i.i.d. $N(0,1)$. This is a Gaussian AR(1) process and can be estimated by ML. As a result, the asymptotic theory developed by Phillips (1972) under stationarity assumptions (i.e., $\kappa > 0$) is applicable.[2]

Second, Cox, Ingersoll and Ross (1985, CIR hereafter) proposed the square root process to describe movements in short term interest rates:[3]

$$dX(t) = \kappa(\mu - X(t))dt + \sigma\sqrt{X(t)}\, dW(t), X_0 \sim \Gamma(2\kappa/\sigma^2, 2\kappa\mu/\sigma^2)), \tag{20}$$

where $\Gamma(\alpha, \beta)$ is a Gamma distribution with parameters α and β. The error term in the exact discrete model is non-Gaussian.

One advantage of using the 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, they may not necessarily fit the data well. Chan, Karolyi, Longstaff, and Sanders (1992) (CKLS, hereafter) considered a univariate interest rate diffusion process:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma X^{\gamma}(t)dW(t), \tag{21}$$

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

$$X_{th} = \mu\left(1 - e^{-\kappa h}\right) + e^{-\kappa h} X_{(t-1)h} + \sigma X^{\gamma}_{(t-1)h}\sqrt{(1 - e^{-2\kappa h})/(2\kappa)}v_t, \tag{22}$$

[2] In the multivariate case, $\kappa > 0$ means that the eigenvalues of κ are all positive.

[3] Heston (1993) used the same specification for volatility.

where $v_t \sim$ i.i.d. $N(0,1)$. Equation (22) is amenable to ML estimation since the transition density is Gaussian.

Motivated from the observation that the Gaussian model (22) is not the exact discrete time model of (21), Yu and Phillips (2001) developed an exact Gaussian method to estimate continuous time models with a linear drift function of the following form:

$$dX(t) = \kappa(\mu - X(t))dt + \sigma(X(t);\theta)dW(t), \tag{23}$$

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 (23), first write (23) as:

$$X(t+\delta) = \mu\left(1 - e^{-\kappa h}\right) + e^{-\kappa h}X(t) + \int_0^\delta e^{-\kappa(\delta-\tau)}\sigma(X(t+\tau);\theta)dW(\tau), \forall\delta > 0. \tag{24}$$

Define $M(\delta) = \int_0^\delta e^{-\kappa(\delta-\tau)}\sigma(X(t+\tau);\theta)dW(\tau)$ and its quadratic variation process by:

$$[M]_\delta = \sigma^2 \int_0^\delta e^{-2\kappa(\delta-\tau)}\sigma(X(t+\tau);\theta)d\tau. \tag{25}$$

To construct the DDS Brownian motion to represent $M(\delta)$, one can use the stopping time. For any fixed constant $a > 0$, let

$$\delta_{j+1} = \inf\{s|[M_j]_s \geq a\} = \inf\{s|\sigma^2 \int_0^s e^{-2\kappa(s-\tau)}\sigma^2(X(t_j+\tau);\theta)d\tau \geq a\}. \tag{26}$$

and construct a sequence of time points $\{t_j\}$ using the iterations $t_{j+1} = t_j + \delta_{j+1}$ with t_1 assumed to be 0. Evaluating equation (24) at $\{t_j\}$, we have

$$X_{t_{j+1}} = \mu\left(1 - e^{-\kappa\delta_{j+1}}\right) + e^{-\kappa\delta_{j+1}}X_{t_j} + M(\delta_{j+1}). \tag{27}$$

where $M(\delta_{j+1}) = W_{[M]_{\delta_{j+1}}} \sim N(0,a)$ is the DDS Brownian motion. Hence, equation (27) is an exact discrete model with Gaussian disturbances and can be estimated directly by ML conditional on the sequence of time changes. The asymptotic distribution can be obtained via the Fisher information theory when $T \to \infty$.

While the SDE (17) 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:

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

by the following ML estimator:

$$\tilde{\kappa} = -\frac{\int_0^1 X(t)dX(t)}{\int_0^1 X(t)^2 dt}. \tag{29}$$

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 (29), the finite sample distribution of $\hat{\kappa} - \kappa$ is:

$$\tilde{\kappa} - \kappa = \frac{\int_0^1 X(t)dW(t)}{\int_0^1 X(t)^2 dt}. \tag{30}$$

Phillips and Yu (2009b) proposed an alternative method to estimate parameters in (17) based on the following two properties: (1) the diffusion term can be fully uncovered from a continuous record; (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:

$$dX(t) = \mu(X(t);\theta_1)dt + \sigma(X(t);\theta_2)dW(t), \tag{31}$$

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 (functional) central limit theory (CLT), as derived by Jacod (1994) and Barndorff-Nielsen and Shephard (2002).

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

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

where $n_h = KM_h$ with K a fixed and positive integer, and $M_h = O(n_h)$. Phillips and Yu constructed the non-overlapping K subsamples:

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

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

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

$$\sum_{i=2}^{M_h}(X_{(k-1)M_h+ih} - X_{(k-1)M_h+(i-1)h})^2 \xrightarrow{p} [X]_{k\frac{T}{K}} - [X]_{(k-1)\frac{T}{K}}, \tag{32}$$

and

$$\frac{\ln(\sum_{i=2}^{M_h}(X_{(k-1)M_h+ih}-X_{(k-1)M_h+(i-1)h})^2-\ln([X]_{k\frac{T}{K}}-[X]_{(k-1)\frac{T}{K}})+\frac{1}{2}s_k^2}{s_k}\xrightarrow{d}N(0,1), \tag{33}$$

where

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

and

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

for $k=1,\cdots,K$. 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 (32) while the CLT (33) is based on the asymptotic theory of Barndorff-Nielsen and Shephard (2005). Based on the CLT (33), θ_2 can be estimated in the first stage by running a nonlinear least squares regression of

$$\frac{\ln\left(\sum_{i=2}^{M_h}(X_{(k-1)M_h+ih}-X_{(k-1)M_h+(i-1)h})^2\right)+\frac{1}{2}s_k^2}{s_k} \tag{34}$$

on

$$\frac{\ln\left(\sum_{i=2}^{M}\sigma^2\left(X_{(k-1)M_h+(i-1)h};\theta_2\right)h\right)-\frac{1}{2}s_k^2}{s_k} \tag{35}$$

for $k=1,\cdots,K$. This produces a consistent estimate $\hat{\theta}_2$ of θ_2 as $h\to 0$. Using the terminology of Phillips, 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.,

$$\text{argmax}_{\theta_1}\left\{\sum_{i=2}^{n}\frac{\mu(X_{(i-1)h};\theta_1)}{\sigma^2(X_{(i-1)h};\hat{\theta}_2)}(X_{ih}-X_{(i-1)h})-\frac{h}{2}\sum_{i=2}^{n}\frac{\mu^2(X_{(i-1)h};\theta_1)}{\sigma^2(X_{(i-1)h};\hat{\theta}_2)}\right\}. \tag{36}$$

This produces a consistent estimate $\hat{\theta}_2$ of θ_2 as $T\to\infty$ and $h\to 0$ (i.e., both the in-fill and long-span asymptotics). The asymptotical theory for $\hat{\theta}_1$ and $\hat{\theta}_2$ is fully developed in Phillips and Yu (2009b).

3 Identification in Continuous Time Models

When continuous time models are estimated, one must ensure the set of parameters θ is identified from discrete time models. In general, unfortunately, θ is not identifiable from discrete time models

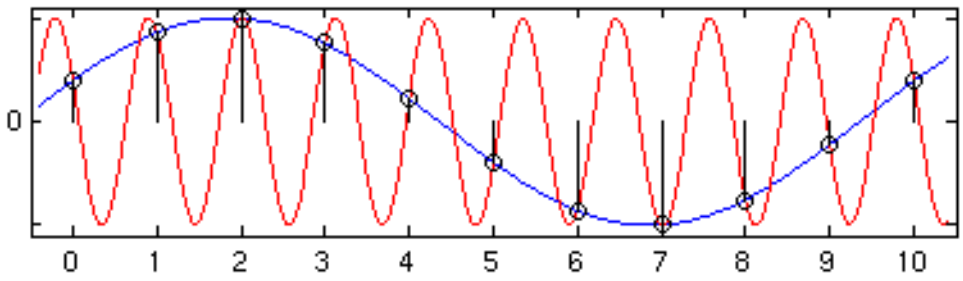

Figure 1: Aliasing problem. Circle points represent a set of discrete samples whose sampling interval is 1.

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 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 circle points) and two different sinusoids (see the two lines). Obviously the two functions could have produced the same samples. Hence, it impossible to tell which function has produced the discrete sample. The usual solution to overcome the aliasing problem is to collect discrete time sample at the 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.

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

$$dX(t) = A(\theta)X(t)dt + \zeta(dt), X(0) = X_0, E(\zeta(dt)) = 0, E(\zeta(dt)\zeta(dt)') = \Sigma dt. \tag{37}$$

The exact discrete time model is given by:

$$X_t = \exp(A(\theta)h)X_{t-1} + \varepsilon_t, X_0 = X_0, E(\varepsilon_t) = 0, E(\varepsilon_t\varepsilon_t') = \Omega. \tag{38}$$

Phillips (1973) showed that:

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

So (A, Σ) is identifiable in model (37) 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 Sargen (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.

Since the 1980's, 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:

$$\begin{aligned} X_1(t) &= BX_2(t) + u_1(t) && (39) \\ DX_2(t) &= u_2(t) && (40) \end{aligned}$$

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:

$$\begin{aligned} X_t &= \exp(-EA)X_{t-1} + \varepsilon_t = (I - EA)X_{t-1} + u_t && (41) \\ u_t &= \varepsilon_t + \frac{1}{e}EAX_{t-1} && (42) \end{aligned}$$

where $X = [X_1', X_2']'$, $E = [I', 0']'$, $A = [I, -B]$, and u_t is stationary because both ε_t and AX_{t-1} are stationary. As

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

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 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.

4 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. Misspeficiation 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.

4.1 Estimation of Drift and Diffusion Terms

Nonparametric estimation of continuous time models was pioneered by Aït-Sahala (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 $h \to 0$. 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.

Bandi and Phillips (2003) significantly extended the results by developing a nonparametric method to estimate both the drift and diffusion terms without imposing 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 modelled 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:

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

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

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:

$$\tilde{\mu}(X_{ih}) = \frac{1}{m(ih)h}\sum_{j=0}^{m(ih)-1}[X_{t(ih)_j+h} - X_{t(ih)_j}],$$

$$\tilde{\sigma}^2(X_{ih}) = \frac{1}{m(ih)h}\sum_{j=0}^{m(ih)-1}[X_{t(ih)_j+h} - X_{t(ih)_j}]^2,$$

$$m(ih) = \sum_{j=1}^{n}\mathbf{1}_{[|X_{jh}-X_{ih}|\leq\varepsilon]}, \forall i \leq n,$$

and b is the bandwidth. To develop the asymptotic distributions of $\hat{\mu}(x)$, it is assumed that $n \to \infty$, $T \to \infty$, $h = T/n \to 0$, and $b \to 0$. Unlike $\hat{\mu}(x)$, $\hat{\sigma}^2(x)$ is consistently estimated without requiring $T \to \infty$. 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 $T \to \infty$.

The nonparametric estimates has recently been applied to various contexts. Bandi (2004) used them to estimate the short term interest rate model. Corradi and Disto (2007) employed it to design a test statistic to distinguish one factor models against two factor models. Jeffrey, Kristensen, Linton, Nguyen, Phillips (2004) made use it to estimate a multifactor Heath-Jarrow-Morton model. Bandi and Phillips (2007) used it to developed 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 $\|\cdot\|$ is the Euclidian distance, $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ are the nonparametric estimates defined above. Bandi and Phillips (2007) developed the asymptotic theory for the parametric estimates. Bandi and Phillips (2007) showed that the consistency of θ_2 does not require $T \to \infty$, whereas the consistency of θ_1 requires $T \to \infty$. The results are consistent with those noted by Merton (1980).

4.2 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 $[0,1]$. Let $0 = t_{0,m} < t_{1,m} < \cdots < t_{m,m} = 1$ 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 equi-spaced points $\{t_{i,m} = \frac{i}{m} : i = 0, ..., m\}$, 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

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

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 equi-spaced partitions greatly facilitate the development of asymptotic theory of RV, in the real ultra high 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.,

$$p_{i,m} = \begin{cases} p^*_{i,m} & \text{if } \xi_i = 1 \\ p_{i-1,m} & \text{if } \xi_i = 0 \end{cases}, \tag{49}$$

where ξ_i is a Bernoulli sequence independent of p^* with $E(\xi_i = 1) = \pi$, $p_{0,m} = p^*_{0,m} = O_p(1)$, and

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

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

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

where MN signifies mixed normal.

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

$$\mathrm{E}\left(D_{m,[ms]}|\mathcal{F}_{\tau_{[ms]-1}}\right) \to_p \mu_D(s), \quad \mathrm{E}\left(D^2_{m,[ms]}|\mathcal{F}_{\tau_{[ms]-1}}\right) \to_p \omega^2_D(s), \tag{52}$$

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 $m \to \infty$

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

5 Finite Sample Issues

In 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 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 (2008) 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 a nonlinear term in the bias formula is 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 (2007) 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 (2008) and Tang and Chen (2007) are motivated from Phillips and Yu (2005a) where the authors intuitively explained why the ML estimator of κ is severely biased upward. Denote the autoregression 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 the finite sample problems. It has been shown that there is nonlinear dependence of the pricing functional on the parameter estimates, which may well exacerbate bias and makes good bias correction more subtle. In particular, even if the parameter estimates are all unbiased, the plug-in estimate of 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 larger when the option is deeper out-of-the-money and the nonlinearity becomes more pronounced; see Phillips and Yu (2009d). It it 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 and involves the indirect inference estimation idea of Smith (1993) and Gourieroux, 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.

5.1 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

$$E(\hat{\theta}_n) = \theta + \frac{a_1}{n} + O(\frac{1}{n^2}). \tag{54}$$

According to (54), 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

$$\hat{\theta}_{jack} = \frac{m}{m-1}\hat{\theta}_n - \frac{\sum_{i=1}^{m}\hat{\theta}_{li}}{m^2 - m}, \tag{55}$$

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 sub-sample, 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 $c(\hat{\theta})$, she can use

$$\hat{c}_{jack} = \frac{m}{m-1}\hat{c}_n - \frac{\sum_{i=1}^{m}\hat{c}_{li}}{m^2-m}. \tag{56}$$

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 (56).

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 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, jackknife can substantially reduce the bias with only a marginal increase in variance, leading to a reduction in mean squared error.

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.

5.2 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 (1996), Monfort (1996), Gouriéroux, Renault, Touzi (2000) in the time series context and by Gouriéroux, Phillips and Yu (2007) 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 $\{X_h, \cdots, X_{nh}\}$. An initial estimator of κ can be obtained, for example, by applying the Euler scheme to $\{X_h, \cdots, X_{nh}\}$ (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

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

where

$$t = \underbrace{0, \delta, \cdots, h(=10\delta)}, \underbrace{h+\delta, \cdots, 2h(=20\delta)}, 2h+\delta, \cdots, nh.$$

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 $\{\tilde{X}^k_{ih}\}_{i=1}^n$, which can be regarded as data simulated directly from the OU model with the (observationally relevant) step size h.[4]

Let $\{\tilde{X}^k_h, \cdots, \tilde{X}^k_{nh}\}$ be data simulated from the true model, where $k = 1, \cdots, K$ 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 $\{\tilde{X}^k_h, \cdots, \tilde{X}^k_{nh}\}$ (call it $\tilde{\kappa}^k_n$). 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

$$\hat{\kappa}_n = \frac{1}{K}\sum_{h=1}^{K} \tilde{\kappa}^k_n(\kappa) \text{ or } \hat{\kappa}_n = \hat{\rho}_{0.5}(\tilde{\kappa}^k_n(\kappa)), \tag{57}$$

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

$$\hat{\kappa}_n = E(\tilde{\kappa}^k_n(\kappa)) \text{ or } \hat{\kappa}_n = \rho_{0.5}(\tilde{\kappa}^k_n(\kappa)) \tag{58}$$

[4]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.

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

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

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:

$$\hat{\kappa}_n^{II} = b_n^{-1}(\hat{\kappa}_n). \tag{59}$$

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 (59) implies:

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

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 non-linear 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 $\widehat{P}_n^{ML} = P(\hat{\theta}_n^{ML})$ be the ML estimate of P. $\widehat{P}_n^{ML}$ involves finite sample estimation bias due to the non-linearity 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 $\widetilde{\mathbf{S}}^k(p) = \{\tilde{S}_1^k, \tilde{S}_2^k, \cdots, \tilde{S}_T^k\}$ be data simulated from the time series model (17) given $\theta(p)$, where $k = 1, \ldots, K$ with K being the number of simulated paths. As argued above, we choose the number of observations in $\widetilde{\mathbf{S}}^k(p)$ 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 $\widetilde{P}_n^{ML,k}(p) = P(\tilde{\phi}_n^{ML,k}(p))$.

4. Choose p so that the average behavior of $\widetilde{P}_n^{ML,k}(p)$ is matched with $\widehat{P}_n^{ML}$ 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 the proposed procedure works well, not only relative to ML but also relative to the jackknife procedure.

6 Some New Results

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

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

$$dX(t) = -\kappa X(t)dt + dW(t), \tag{60}$$

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

Data, namely, $\{X_{0h}, X_{1h}, \cdots, X_{nh}\}$ 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:

$$\hat{\kappa} = -\frac{\ln\left\{\frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}\right\}}{h}, \tag{61}$$

which is the ML estimate.

Since $\kappa > 0$, when $T \to \infty$ and h is fixed, the standard asymptotic theory implies that

$$\sqrt{n}(\hat{\kappa} - \kappa) \xrightarrow{d} N\left(0, \frac{\exp(2\kappa h) - 1}{h^2}\right). \tag{62}$$

Hence,

$$\hat{\kappa} \overset{a}{\sim} N\left(\kappa, \frac{\exp(2\kappa h) - 1}{hT}\right). \tag{63}$$

The asymptotic normality is not surprising as the process is stationary, ergodic and asymptotically independent.

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

$$\int_0^T -\kappa X(t) dX(t) - \int_0^T \frac{1}{2}\kappa^2 X(t)^2 dt. \tag{64}$$

From Phillips (1987b) the finite sample distribution of the ML estimator based on (64) is known. Hence, it is natural to use this distribution to approximate the distribution of $\hat{\kappa}$, namely,

$$\hat{\kappa} \overset{a}{\sim} \kappa - \frac{\int_0^T X(t) dW(t)}{\int_0^T X(t)^2 dt}. \tag{65}$$

Obviously, this limiting distribution is asymmetric and non-normal. This compares interestingly with the limiting distribution in (63).

To facilitate computation of the distribution in (65), 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:*

$$dX(t) = \mu(t, X(t), \theta)dt + dW(t). \tag{66}$$

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

$$F(T, X(T), \theta) - \int_0^T \left\{ f(t, X(t), \theta) + \frac{1}{2}\mu^2(t, X(t), \theta) \right\} dt, \tag{67}$$

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 (60), 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:

$$-\frac{1}{2}\kappa X(T)^2 - \int_0^T \left[-\frac{1}{2}\kappa + \frac{1}{2}\kappa^2 X(t)^2 \right] dt. \tag{68}$$

The finite sample distribution of the ML estimator that maximizes (68) is

$$\frac{-X(T)^2 + T}{2\int_0^T X(t)^2 dt}. \tag{69}$$

This is the in-fill asymptotic distribution of $\hat{\kappa}$ defined in (61). Compared with the asymptotic distribution (65), (69) only needs to calculate a Riemann integral. Furthermore, (69) 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 two limiting distributions, (63) and (69), 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 (61). The true values of κ are set at 0.1,1,10. The value of 0.1 is empirically realistic for interest data while the value of 1 is empirically realistic for volatility. These two values suggest the slow speed for mean reversion. While $\kappa = 10$ is not empirically realistic for financial time series, we include it for the purpose of comparison. The true values of h are set at 1/12, 1/250, corresponding to monthly and daily frequencies, respectively. It is now rare to see in continuous time literature to employ 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 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 three distributions, for $\kappa = 0.1, 1, 10$, respectively. A few results emerge from the tables. First and more importantly, the in-fill asymptotics almost always performs significantly better than the long-span 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 for and the in-fill asymptotics is favored against. Second, the larger the T, the better the long-span asymptotics. Third, the smaller the h, the better the in-fill asymptotics. Fourth, the larger the κ, the better the two 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 asymptotics.

7 Conclusions

The theoretical development of econometric analysis of continuous time models has come a long way. 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

[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.

mostly focused on theoretical aspects, his impact extends far beyond mere theory. Without any doubt, he will continue to make some more insightful contributions in this important area, given the current level of his energy and devotion to this area of research.

Appendix: Proof of Lemma

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

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

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

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

Hence,

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

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

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

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

References

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

[2] Aït-Sahalia, Y., 2002, Maximum likelihood estimation of discretely sampled diffusion: A closed-form approximation approach. *Econometrica*, 70, 223–262.

[3] Andersen, T., T. Bollerslev, 1998, Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, 4, 885-905.

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

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

[6] Bailey, R. W., V. B. Hall, and P.C.B. Phillips, 1987, A Model of Output, Employment, Capital Formation and Inflation, in G. Gandolfo and F. Marzano (eds.), *Saggi in memoria di Victorio Marrama*, Milano: Giuffrè 1987.

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

[8] Bandi, F. M. and P.C.B. Phillips, 2003, Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71, 241–283.

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

[10] Bandi, F. M. and P.C.B. Phillips, 2009, Nonstationary Continuous-Time Processes, in *Handbook of Financialof Econometrics*, eds Aït-Sahalia, Y. and L.P. Hansen, forthcoming.

[11] Barndorff-Nielsen, O. and N. Shephard 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society, Series B*, 64, 253-280.

[12] Barndorff-Nielsen, O. and N. Shephard, 2005, How accurate is the asymptotic approximation to the distribution of realized volatility?. In *Identification and Inference for Econometric Models*, eds by D.W.K. Andrews, J. Powell, P. Ruud and J. Stock, Cambridge University Press.

[13] 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,

[14] Bergstrom, Albert R. 1966. Nonrecursive models as discrete approximations to system of stochastic differential equations, *Econometrica*, 34, 173-82

[15] Bergstrom, Albert R. and C. R. Wymer, A model of disequilibrium neoclassical growth with its application to the United Kingdom, in *Statistical Inference in Continuous Time Economic Models* (edited by A. R. Bergstrom), Amsterdam: North-Holland Publishing Company.

[16] Bergstrom, Albert R. 1983. Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica* 51, 117–152.

[17] Bergstrom, Albert R. 1984. Continuous time stochastic models and issues of aggregation over time. In Z. Griliches and M.D. Intriligator, editors, *Handbook of Econometrics*. Vol. II (Elsevier Science, Amsterdam).

[18] Bergstrom, Albert R. 1985a. The estimation of nonparametric functions in a Hilbert space *Econometric Theory*, 1, 7-26

[19] Bergstrom, Albert R. 1985b. The estimation of parameters in nonstationary higher-order continuous-time dynamic models. *Econometric Theory* 1, 369–385.

[20] Bergstrom, Albert R. 1986. The estimation of open higher-order continuous time dynamic models with mixed stock and flow data. *Econometric Theory* 2, 350–373.

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

[22] Bergstrom, Albert R. 1996, Survey of continuous time econometrics, in *Dynamic Disequilibrium Modelling*, edited by W.A. Barnett, G. Gandolfo, and C. Hillinger, Cambridge: Cambridge University Press, 3-25.

[23] Bergstrom, Albert R. and B. Nowman 2007, *A Continuous Time Econometric Model of the United Kingdom with Stochastic Trends*, Cambridge University Press.

[24] Black, F. and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economics*, 81, 637-54.

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

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

[27] Drost, F.C., and Nijman, T.E., 1993. Temporal aggragation of Garch processes, *Enometrica*, 61, 909-927.

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

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

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

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

[32] Gallant, A. R., and G. Tauchen, 1996, Which Moments to Match? *Econometric Theory*, 12, 657–681.

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

[34] Gouriéroux, C., Phillips, P.C.B. and J. Yu, 2007, Indirect inference for dynamic panel models, *Journal of Econometrics*, forthcoming.

[35] Gouriéroux, C., E. Renault, and N. Touzi, 2000, Calibration by simulation for small sample bias correction, in R.S. Mariano, T. Schuermann and M. Weeks (eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, Cambridge University Press, p328-358.

[36] Hansen, L.P. and T.J. Sargent, 1983, The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities, *Econometrica* 51, 377-387.

[37] 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.

[38] Jacod, J. 1993, Limit of random measures associated with the increments of a Brownian semimartingale, Laboratoire de Probabilités, Paris (mimeo).

[39] Jamshidian, F., 1989, An Exact Bond Option Formula, *Journal of Finance*, 44, 205-209.

[40] 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.

[41] Lánska, V., 1979, Minimum contrast estimation in diffusion processes. *Journal of Applied Probability*, 16, 65–75.

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

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

[44] Merton, R. C., 1973, Theory of Rational Option Pricing, *Bell Journal of Economics and Management Science*, 4, 141-183.

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

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

[47] 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.

[48] Park, J. and P.C.B. Phillips, 2001, Nonlinear Regression with Integrated Time Series, *Econometrica*, 69, 117-161.

[49] 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.

[50] Phillips, P.C.B., 1972, The Structural Estimation of a Stochastic Differential Equation System, *Econometrica*, 6, 1021-1041.

[51] Phillips, P.C.B., 1973, The Problem of Identification in Finite Parameter Continuous Time Models, *Journal of Econometrics*, 4, 351-262.

[52] Phillips, P.C.B., 1974, The Estimation of Some Continuous Time Models, *Econometrica*, 42, 803-823.

[53] 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, 135-174.

[54] 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, 175-214.

[55] 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 and M. Preston (eds.), *Stability and Inflation: Essays in Memory of A. W. Phillips*, New York, 257-274.

[56] Phillips, P.C.B., 1987a, Time Series Regression with a Unit Root, *Econometrica*, 55, 277-301.

[57] Phillips, P.C.B., 1987b, Towards a Unified Asymptotic Theory for Autoregression, *Biometrika*, 74, 535-547.

[58] Phillips, P.C.B., 1988, An Interview with Professor A. R. Bergstrom, *Econometric Theory*, 4, 301-328.

[59] Phillips, P.C.B., 1991, Error Correction and Long Run Equilibria in Continuous Time, *Econometrica*, 59, 967-980.

[60] 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*, Oxford: Basil Blackwell.

[61] Phillips, P.C.B. and J. Y. Park, 1999, Asymptotics for Nonlinear Transformations of Integrated Time Series, *Econometric Theory*, 15, 269-298

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

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

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

[65] Phillips, P.C.B., and J. Yu, 2009b, A Two-Stage Realized Volatility Approach to Estimation of Diffusion Processes with Discrete Data, *Journal of Econometrics*, forthcoming.

[66] Phillips, P.C.B. and J. Yu, 2009c, Information Loss in Volatility Measurement with Flat Price Trading, Working paper, Cowles Foundation, Yale University.

[67] Phillips, P.C.B. and J. Yu, 2009d, Simulation-based Estimation of Contingent-claims Prices, *Review of Financial Studies*, forthcoming.

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

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

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

[71] Yu, J., 2008, Bias in the Estimation of Mean Reversion Parameter in Continuous Time Models, Working Paper, Singapore Management University.

[72] Yu, J. and P.C.B. Phillips, 2001, A Gaussian Approach for Continuous Time Models of the Short Term Interest Rate, *The Econometrics Journal*, 4, 211-225.

h and T	Methods	0.5%	1%	5%	10%	90%	95%	99%	99.5%
$h = 1/12$	exact	-0.8447	-0.7042	-0.3563	-0.2155	1.7071	2.6012	5.1676	6.5470
$T = 2$	in-fill	-0.8518	-0.7200	-0.4355	-0.3330	1.1327	1.9147	4.0532	5.0527
	long-span	-0.7355	-0.6546	-0.4336	-0.3157	0.5157	0.6336	0.8546	0.9355
$h = 1/12$	exact	-0.1427	-0.1099	-0.0329	0.0027	0.5885	0.7869	1.3212	1.5574
$T = 10$	in-fill	-0.1513	-0.1211	-0.0511	-0.0218	0.4831	0.6724	1.1543	1.3987
	long-span	-0.2673	-0.2318	-0.1346	-0.0828	0.2828	0.3346	0.4318	0.4673
$h = 1/12$	exact	0.0145	0.0214	0.0401	0.0521	0.2470	0.2987	0.4150	0.4700
$T = 50$	in-fill	0.0116	0.0181	0.0353	0.0461	0.2296	0.2806	0.3917	0.4453
	long-span	-0.0637	-0.0479	-0.0046	0.0185	0.1815	0.2046	0.2479	0.2637
$h = 1/250$	exact	-0.7886	-0.6447	-0.3264	-0.1962	1.6350	2.4563	4.7549	5.8324
$T = 2$	in-fill	-0.8691	-0.7188	-0.4333	-0.3329	1.1297	1.9050	4.0505	5.1093
	long-span	-0.7155	-0.6365	-0.4208	-0.3057	0.5057	0.6208	0.8365	0.9155
$h = 1/250$	exact	-0.1345	-0.1064	-0.0313	0.0022	0.5793	0.7792	1.2916	1.5214
$T = 10$	in-fill	-0.1533	-0.1224	-0.0524	-0.0222	0.4829	0.6715	1.1653	1.4047
	long-span	-0.2644	-0.2291	-0.1327	-0.0813	0.2813	0.3327	0.4291	0.4644
$h = 1/250$	exact	0.0131	0.0204	0.0398	0.0514	0.2447	0.2987	0.4141	0.4656
$T = 50$	in-fill	0.0101	0.0164	0.0348	0.0457	0.2281	0.2804	0.3952	0.4432
	long-span	-0.0629	-0.0472	-0.0041	0.0189	0.1811	0.2041	0.2472	0.2629

Table 1: Performance of two asymptotic distributions 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 and the long-span asymptotic distribution. The true distribution is obtained via simulations based on 10,000 replications.

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

Table 2: Performance of two asymptotic distributions 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 and the long-span asymptotic distribution. The true distribution is obtained via simulations based on 10,000 replications.

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

Table 3: Performance of two asymptotic distributions 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 and the long-span asymptotic distribution. The true distribution is obtained via simulations based on 10,000 replications.