Simulation-Based Estimation Methods for Financial Time Series Models

Jun Yu

Abstract This chapter overviews some recent advances on simulation-based methods of estimating financial time series models that are widely used in financial economics. The simulation-based methods have proven to be particularly useful when the likelihood function and moments do not have tractable forms and hence the maximum likelihood (ML) method and the generalized method of moments (GMM) are difficult to use. They are also useful for improving the finite sample performance of the traditional methods. Both frequentist and Bayesian simulation-based methods are reviewed. Frequentist's simulation-based methods cover various forms of simulated maximum likelihood (SML) methods, simulated generalized method of moments (SGMM), efficient method of moments (EMM), and indirect inference (II) methods. Bayesian simulation-based methods cover various MCMC algorithms. Each simulation-based method is discussed in the context of a specific financial time series model as a motivating example. Empirical applications, based on real exchange rates, interest rates and equity data, illustrate how to implement the simulation-based methods. In particular, we apply SML to a discrete time stochastic volatility model, EMM to estimate a continuous time stochastic volatility model, MCMC to a credit risk model, the II method to a term structure model.

1 Introduction

Relative to other fields in economics, financial economics has a relatively short history. Over the last half century, however, there has been an explosion of theoretical work in financial economics. At the same time, more and more complex

J. Yu (✉)
School of Economics and Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore 178903
e-mail: yujun@smu.edu.sg

J.-C. Duan et al. (eds.), *Handbook of Computational Finance*, Springer Handbooks of Computational Statistics, DOI 10.1007/978-3-642-17254-0_15,

financial products and services have been created. The size of financial markets has exponentially increased and the quality of database is hugely advanced. The major developments in theoretical finance and the availability of high quality data provide an extremely rich framework for empirical work in financial economics.

How to price financial assets has been a driving force for much of the research on financial asset pricing. With the growth in complexity in financial products and services, the challenges faced by the financial economists naturally grow accordingly, one of which is the computing cost. Another driving force for research in financial economics is to bring finance theory to data. Empirical analysis in financial economics often involves calculating the likelihood function or solving a set of moment conditions.

Traditional econometric methods for analyzing models in financial economics include maximum likelihood (ML), quasi-ML, generalized method of moments (GMM), and classical Bayesian methods. When the model is fully specified and the likelihood function has a tractable form, ML and Bayesian methods provide the full likelihood-based inference. Under mild regularity conditions, it is well recognized that the ML estimator (MLE) is consistent, asymptotically normally distributed and asymptotically efficient. Due to the invariance principle, a function of MLE is a MLE and hence inherits all the nice asymptotic properties (e.g., Zehna 1966). These features greatly facilitate applications of ML in financial economics. When the model is not fully specified but certain moments exist, GMM can be applied. Relative to ML, GMM trades off efficiency with robustness.

Financial data are typically available in the time series format. Consequently, time series methods are of critical importance to empirical research in financial economics. Historically, financial economists restricted themselves to a small class of time series models so that the setups were simple enough to permit an analytical solution for asset prices. Moreover, empirical analysis was often done based a small set of financial assets, so that the computational cost is kept low. The leading example is perhaps the geometric Brownian motion, which was used by Black and Scholes to price European options (Black and Scholes 1973) and by Merton to price corporate bonds (Merton 1974). In recent years, however, many alternative models and financial products have been proposed so that asset prices do not have analytical solutions any more. As a result, various numerical solutions have been proposed, one class of which is based on simulations. Although the use of simulation-based methods for asset pricing is sufficient important and merits a detailed review, it is beyond the scope of the present chapter. We refer readers to McLeish (2005) for a textbook treatment on asset pricing via simulation methods.

Even if the pricing formula of a financial asset has a tractable form, estimation of the underlying time series model is not always feasible by standard econometric methods. For many important financial time series models, the likelihood function or the moment conditions cannot be evaluated analytically and may be numerically formidable so that standard econometric methods, such as ML, GMM and Bayesian, are not feasible. For example, Heston (1993) derived a closed-form expression for the European option price under the square root specification for volatility. It is known that the ML estimation of Heston's stochastic volatility (SV) model from stock prices is notoriously difficult. For more complicated models where asset prices do not have a closed-form expression, it is almost always the case that standard estimation methods are difficult to use.

Parameter estimation is important for asset pricing. For example, in order to estimate the theoretical price of a contingent claim implied by the underlying time series model, one has to estimate the parameters in the time series model and then plug the estimates into the pricing formula. In addition, parameter estimates in financial time series models are necessary inputs to many other financial decision makings, such as asset allocation, value-at-risk, forecasting, estimation of the magnitude of microstructure noise, estimation of transaction cost, specification analysis, and credit risk analysis. For example, often alternative and sometimes competing time series specifications co-exist. Consequently, it may be important to check the validity of a particular specification and to compare the relative performance of alternative specifications. Obviously, estimation of these alternative specifications is an important preliminary step to the specification analysis. For another example, in order to estimate the credit spread of a risky corporate bond over the corresponding Treasury rate and the default probability of a firm, the parameters in the underlying structural model have to be estimated first.

In some cases where ML or GMM or Bayesian methods are feasible but financial time series are highly persistent, classical estimators of certain parameters may have poor finite sample statistical properties, due to the presence of a large finite sample bias. The bias in parameter estimation leads to a bias in other financial decision making. Moreover, the large finite sample bias often leads to a poor approximation to the finite sample distribution by the asymptotic distribution. As a result, statistical inference based on the asymptotic distribution may be misleading. Because many financial variables, such as interest rates and volatility, are highly persistence, this finite sample problem may be empirically important.

To overcome the difficulties in calculating likelihood and moments and to improve the finite sample property of standard estimators, many simulation-based estimation methods have been proposed in recent years. Some of them are methodologically general; some other are specially tailored to deal with a particular model structure. In this chapter, we review some simulation-based estimation methods that have been used to deal with financial time series models.

Stern (1997) is an excellent review of the simulation-based estimation methods in the cross-sectional context while Gouriéroux and Monfort (1995) reviewed the simulation-based estimation methods in the classical framework. Johannes and Polson (2009) reviewed the Bayesian MCMC methods used in financial econometrics. Our present review is different from these reviews in several important aspects. First, our review covers both the classical and Bayesian methods whereas Johannes and Polson (2009) only reviewed the Bayesian methods. Second, relative to Stern (1997) and Gouriéroux and Monfort (1995), more recently developed classical methods are discussed in the present chapter. Moreover, only our review discuss the usefulness of simulation-based methods to improve finite sample performances.

We organize the rest of this chapter by collecting the methods into four categories: simulation-based ML (SML), simulation-based GMM (SGMM), Bayesian

Markov chain Monte Carlo (MCMC) methods, and simulation-based resampling methods. Each method is discussed in the context of specific examples and an empirical illustration is performed using real data correspondingly. Section 15.2 overviews the classical estimation methods and explains why they may be difficult to use in practice. Section 15.3 discusses discrete time stochastic volatility models and illustrates the implementation of a SML method. Section 15.4 discusses continuous time models and illustrates the implementation of EMM. Section 15.5 discusses structure credit risk models and illustrates the implementation of a Bayesian MCMC method. Section 15.6 discusses continuous time models with a linear and persistent drift function and illustrates the implementation of the indirect inference (II) method in the context of Vasicek model for the short term interest rate. Finally, Sect. 15.7 concludes.

2 Problems with Traditional Estimation Methods

In many cases the likelihood function of a financial time series model can be expressed as:[1]

$$L(\theta) = p(\mathbf{X};\theta) = \int p(\mathbf{X}, \mathbf{V};\theta)d\mathbf{V}, \tag{1}$$

where $\mathbf{X} = (X_1, \cdots, X_n) := (X_h, \cdots, X_{nh})$ is the data observed by econometricians,[2] h the sampling interval, $p(\mathbf{X})$ the joint density of $\mathbf{X}$, $\mathbf{V}$ a vector of latent variables, θ a set of K parameters that econometricians wish to estimate. As $X(t)$ often represents the annualized data, when daily (weekly or monthly) data are used, h is set at 1/252 (1/52 or 1/12). Assume $T = nh$ is the time span of the data and the true values for θ is θ_0.

MLE maximizes the log-likelihood function over θ in a certain parameter space:

$$\hat{\theta}_n^{ML} := \text{argmax}_{\theta\in\Theta}\ell(\theta)),$$

where $\ell(\theta) = \ln L(\theta) = \ln p(\mathbf{X};\theta)$. The first order condition of the maximization problem is:

$$\frac{\partial \ell}{\partial \theta} = 0.$$

Under mild regularity conditions, the ML estimator (MLE) has desirable asymptotic properties of consistency, normality and efficiency. Moreover, the invariance property of MLE ensures that a smoothed transformation of MLE is a MLE of the same transformation of the corresponding parameters (Zehna 1966). This property has proven very useful in financial applications.

[1] Specific examples can be found below.

[2] When there is no confusion, we will use X_t and X_{th} interchangeably.

Unfortunately, when the integration in (1) is not analytically available and the dimension of $\mathbf{V}$ is high, numerical evaluation of (1) is difficult. If $p(\mathbf{X};\theta)$ is difficult to calculate, ML is not easy to implement.

Instead of maximizing the likelihood function, Bayesian methods update the prior density to the posterior density using the likelihood function, based on the Bayes theorem:

$$p(\theta|\mathbf{X}) \propto p(\mathbf{X};\theta)p(\theta),$$

where $p(\theta)$ is the prior density and $p(\theta|\mathbf{X})$ the posterior distribution. As in ML, if $p(\mathbf{X};\theta)$ is difficult to calculate, the posterior density $p(\theta|\mathbf{X})$ is generally difficult to evaluate.

Unlike ML or Bayesian methods that rely on the distributional assumption of the model, GMM only requires a set of moment conditions to be known. Let g be a set of q moment conditions, i.e.

$$E[g(\mathbf{X};\theta_0)] = 0$$

GMM minimizes a distance measure, i.e.

$$\hat{\theta}_n^{GMM} := \text{argmin}_{\theta\in\Theta} \left(\frac{1}{n}\sum_{t=1}^{n} g(X_t;\theta)\right)' W_n \left(\frac{1}{n}\sum_{t=1}^{n} g(X_t;\theta)\right)',$$

where W_n is a certain positive definite weighting matrix of $q \times q$-dimension ($q \geq K$), which may depend on the sample but not θ. Obviously, the implementation of GMM requires the moments to be known analytically or easy to calculate numerically. Since a fixed set of moments contain less information than a density, in general GMM uses less information than ML and hence is statistically less efficient. In the case where the moment conditions are selected based on the score functions (in which case $q = K$), GMM and ML are equivalent. However, sometimes moment conditions are obtained without a distributional assumption and hence GMM may be more robust than the likelihood-based methods. Under mild regularity conditions, Hansen (1982) obtained the asymptotic distributions of GMM estimators. Unfortunately, many financial time series models do not have an analytical expression for moments and moments are difficult to evaluate numerically, making GMM not trivial to implement.

Even if ML is applicable, MLE is not necessarily the best estimator in finite sample. Phillips and Yu (2005a,b, 2009a,b) have provided numerous examples to demonstrate the poor finite sample properties of MLE. In general there are three reasons for this. First, many financial variables (such as interest rates and volatility) are very persistent. When a linear time series model is fitted to these variables, ML and GMM typically lead to substantial finite sample bias for the mean reversion parameter even in very large samples. For example, when 2,500 daily observations are used to estimate the square root model of the short term interest rate, ML estimates the mean reversion parameter with nearly 300% bias. Second, often financial applications involve non-linear transformation of estimators of the system parameters. Even if the system parameters are estimated without any bias, insertion of even unbiased estimators into the nonlinear functions will not assure unbiased estimation of the quantity of interest. A well known example is the MLE of a deep out-of-money option which is highly nonlinear in volatility. In general, the more pronounced the nonlinearity, the worse the finite sample performance is. Third, even if a long-span sample is available for some financial variables and hence asymptotic properties of econometric estimators is more relevant, full data sets are not always employed in estimation because of possible structural changes in long-span data. When short-span samples are used in estimation, finite sample distributions can be far from the asymptotic theory.

A natural way to improve the finite sample performance of classical estimators is to obtain the bias in an analytical form and then remove the bias from the biased estimator, with the hope that the variance of the bias-corrected estimator does not increase or only increases slightly so that the mean square error becomes smaller. Unfortunately, the explicit analytical bias function is often not available, except in very simple cases.

When the likelihood function and moments are difficult to calculate or traditional estimators perform poorly in finite sample, one can resort to simulation methods. There has been an explosion of theoretical and empirical work using simulation methods in financial time series analysis over the last 15 years. In the following sections we will consider some important examples in financial economics and financial econometrics. Simulated-based methods are discussed in the context of these examples and an empirical illustration is provided in each case.

3 Simulated ML and Discrete Time SV Models

To illustrate the problem in ML, we first introduce the basic lognormal (LN) SV model of Taylor (1982) defined by

$$\begin{cases} X_t &= \sigma e^{h_t/2}\epsilon_t, \ t = 1, \ldots, n, \\ h_{t+1} &= \phi h_t + \gamma \eta_t, \ t = 1, \ldots, n-1, \end{cases} \tag{2}$$

where X_t is the return of an asset, $|\phi| < 1$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $\eta_t \overset{iid}{\sim} N(0,1)$, $corr(\epsilon_t, \eta_t) = 0$, and $h_1 \sim N(0, \gamma^2/(1-\phi^2))$. The parameters of interest are $\theta = (\sigma, \phi, \gamma)'$. This model is proven to be a powerful alternative to ARCH-type models (Geweke 1994; Danielsson 1994). Its continuous time counterpart has been used to pricing options contracts (Hull and White 1987).

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ and $\mathbf{V} = (h_1, \ldots, h_n)'$. Only $\mathbf{X}$ is observed by the econometrician. The likelihood function of the model is given by

$$p(\mathbf{X}; \theta) = \int p(\mathbf{X}, \mathbf{V}; \theta) d\mathbf{V} = \int p(\mathbf{X}|\mathbf{V}; \theta) p(\mathbf{V}; \theta) d\mathbf{V}. \tag{3}$$

To perform the ML estimation to the SV model, one must approximate the high-dimensional integral (3) numerically. Since a typical financial time series has at least several hundreds observations, using traditional numerical integration methods, such as quadratures, to approximate the high-dimensional integral (3) is numerically formidable. This is the motivation of the use of Monte Carlo integration methods in much of the SV literature.

The basic LN-SV model has been found to be too restrictive empirically for many financial time series and generalized in various dimensions to accommodate stylized facts. Examples include the leverage effect (Harvey and Shephard 1996; Yu 2005), SV-t (Harvey et al. 1994), super-position (Pitt and Shephard 1999b), jumps (Duffie et al. 2000), time varying leverage effect (Yu 2009b). An widely used specification, alternative to the LN-SV model, is the Heston model (Heston 1993).

In this section, we will review several approaches to do simulated ML estimation of the basic LN-SV model. The general methodology is first discussed, followed by a discussion of how to use the method to estimate the LN-SV model and then by an empirical application.

3.1 Importance Sampler Based on the Laplace Approximation

Taking the advantage that the integrand is a probability distribution, a widely used SML method evaluates the likelihood function numerically via simulations. One method matches the integrand with a multivariate normal distribution, draws a sequence of independent variables from the multivariate normal distribution, and approximates the integral by the sample mean of a function of the independent draws. Namely, a Monte Carlo method is used to approximate the integral numerically and a carefully selected multivariate normal density is served as an importance function in the Monte Carlo method. The technique in the first stage is known as the Laplace approximation while the technique in the second stage is known as the importance sampler. In this chapter the method is denoted LA-IS.

To fix the idea, in Stage 1, we approximate $p(\mathbf{X}, \mathbf{V}; \theta)$ by a multivariate normal distribution for $\mathbf{V}$, $N(\cdot; \mathbf{V}^*, -\Omega^{-1})$, where

$$\mathbf{V}^* = \arg\max_{\mathbf{V}} \ \ln p(\mathbf{X}, \mathbf{V}; \theta) \tag{4}$$

and

$$\Omega = \frac{\partial^2 \ln p(\mathbf{X}, \mathbf{V}^*; \theta)}{\partial \mathbf{V} \partial \mathbf{V}'}. \tag{5}$$

For the LN-SV model $\mathbf{V}^*$ does not have the analytical expression and hence numerical methods are needed. For example, Shephard and Pitt (1997), Durham (2006), Skaug and Yu (2007) proposed to use Newton's method, which involves recursive calculations of $\mathbf{V} = \mathbf{V}_{-} - \Omega^{-1}\mathbf{V}_{-}$, based on a certain initial vector of log-volatilities, $\mathbf{V}_0$.

Based on the Laplace approximation, the likelihood function can be written as

$$p(\mathbf{X};\theta) = \int p(\mathbf{X},\mathbf{V};\theta)d\mathbf{V} = \int \frac{p(\mathbf{X},\mathbf{V};\theta)}{N(\mathbf{V};\mathbf{V}^*,-\Omega^{-1})} N(\mathbf{V};\mathbf{V}^*,-\Omega^{-1})d\mathbf{V}. \quad (6)$$

The idea of importance sampling is to draw samples $\mathbf{V}^{(1)},\ldots,\mathbf{V}^{(S)}$ from $N(\cdot;\mathbf{V}^*,-\Omega^{-1})$ so that $p(\mathbf{X};\theta)$ is approximated by

$$\frac{1}{S}\sum_{s=1}^{S} \frac{p(\mathbf{X},\mathbf{V}^{(s)};\theta)}{N(\mathbf{V}^{(s)};\mathbf{V}^*,-\Omega^{-1})}. \quad (7)$$

After the likelihood function is obtained, a numerical optimization procedure, such as the quasi Newton method, can be applied to obtain the ML estimator.

The convergence of (7) to the likelihood function $p(\mathbf{X};\theta)$ with $S \to \infty$ is ensured by Komogorov's strong law of large numbers. The square root rate of convergence is achieved if and only if the following condition holds

$$Var\left(\frac{p(\mathbf{X},\mathbf{V}^{(s)};\theta)}{N(\mathbf{V}^{(s)};\mathbf{V}^*,-\Omega^{-1})}\right) < \infty.$$

See Koopman et al. (2009) for further discussions on the conditions and a test to check the convergence.

The idea of the LA-IS method is quite general. The approximation error is determined by the distance between the integrant and the multivariate normal distribution and the size of S. The Laplace approximation does not have any error if $p(\mathbf{X},\mathbf{V};\theta)$ is the Gaussianity in $\mathbf{V}$. In this case, $S = 1$ is big enough to obtain the exact value of the integral. The further $p(\mathbf{X},\mathbf{V};\theta)$ away from Gaussian in $\mathbf{V}$, the less precise the Laplace approximation is. In this case, a large value is needed for S.

For the LN-SV model, the integrand in (3) can be written as

$$p(\mathbf{X},\mathbf{V};\theta) = N\left(h_1, 0, \frac{\gamma^2}{1-\phi^2}\right)\prod_{t=2}^{n} N\left(h_t,\phi h_{n-1},\gamma^2\right)\prod_{t=1}^{n} N\left(X_t, 0, \sigma^2 e^{h_t}\right), \quad (8)$$

and hence

$$\ln p(\mathbf{X},\mathbf{V};\theta) = \ln N\left(h_1, 0, \frac{\gamma^2}{1-\phi^2}\right) + \sum_{t=2}^{n} \ln N\left(h_t,\phi h_{n-1},\gamma^2\right) + \sum_{t=1}^{n} \ln N\left(X_t, 0, \sigma^2 e^{h_t}\right). \quad (9)$$

It is easy to show that

$$\frac{\partial N(x;\mu,\sigma^2)/\partial x}{N(x;\mu,\sigma^2)} = -\frac{x-\mu}{\sigma^2}, \frac{\partial N(x;\mu,\sigma^2)/\partial \mu}{N(x;\mu,\sigma^2)} = -\frac{\mu-x}{\sigma^2},$$

$$\frac{\partial N(x;\mu,\sigma^2)/\partial \sigma^2}{N(x;\mu,\sigma^2)} = -\frac{1}{\sigma^2}\left(1-\frac{(x-\mu)^2}{\sigma^2}\right),$$

Using these results, we obtain the gradient of the log-integrand:

$$\begin{pmatrix} \frac{\partial \ln p(\mathbf{X},\mathbf{V};\theta)}{\partial h_1} \\ \frac{\partial \ln p(\mathbf{X},\mathbf{V};\theta)}{\partial h_2} \\ \vdots \\ \frac{\partial \ln p(\mathbf{X},\mathbf{V};\theta)}{\partial h_{n-1}} \\ \frac{\partial \ln p(\mathbf{X},\mathbf{V};\theta)}{\partial h_n} \end{pmatrix} = \begin{pmatrix} \frac{\phi h_2 - h_1}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_1^2 \\ \frac{\phi h_3 - \phi^2 h_2 + \phi h_1}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_2^2 \\ \vdots \\ \frac{\phi h_n - \phi^2 h_{n-1} + \phi h_{T-2}}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_{n-1}^2 \\ \frac{h_n - \phi h_{n-1}}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_n^2 \end{pmatrix}, \tag{10}$$

and the Hessian matrix of the log-integrand:

$$\Omega = \begin{pmatrix} -\frac{1}{\gamma^2} - \frac{1}{2}\epsilon_1^2 & \frac{\phi}{\gamma^2} & \cdots & 0 & 0 \\ \frac{\phi}{\gamma^2} & -\frac{1+\phi^2}{\gamma^2} - \frac{1}{2}\epsilon_2^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\frac{1+\phi^2}{\gamma^2} - \frac{1}{2}\epsilon_{n-1}^2 & \frac{\phi}{\gamma^2} \\ 0 & 0 & \cdots & \frac{\phi}{\gamma^2} & -\frac{1}{\gamma^2} - \frac{1}{2}\epsilon_n^2 \end{pmatrix}. \tag{11}$$

Durham (2006, 2007), Koopman et al. (2009), Skaug and Yu (2007) and Yu (2009b) applied the SML method to estimate generalized SV models and documented the reliable performance in various contexts.

3.2 *Monte Carlo Likelihood Method*

Durbin and Koopman (1997) proposed a closely related SML method which is termed Monte Carlo likelihood (MCL) method. MCL was originally designed to evaluate the likelihood function of a linear state-space model with non-Gaussian errors. The basic idea is to decompose the likelihood function into the likelihood of a linear state-space model with Gaussian errors and that of the remainder. It is known that the likelihood function of a linear state-space model with Gaussian errors can be calculated by the Kalman filter. The likelihood of the remainder is calculated by simulations using LA-IS.

To obtain the linear state-space form for the LN-SV model, one can apply the log-squared transformation to X_t:

$$\begin{cases} Y_t &= \ln X_t^2 = \ln \sigma^2 + h_t + \varepsilon_t,\ t = 1, \ldots, n, \\ h_{t+1} &= \phi h_t + \gamma \eta_t,\ t = 1, \ldots, n-1, \end{cases} \tag{12}$$

where $\varepsilon_t \overset{iid}{\sim} \ln \chi^2_{(1)}$ (i.e. no-Gaussian), $\eta_t \overset{iid}{\sim} N(0,1)$, $corr(\varepsilon_t, \eta_t) = 0$, and $h_1 \sim N(0, \gamma^2/(1-\phi^2))$. For any linear state-space model with non-Gaussian measurement errors, Durbin and Koopman (1997) showed that the log-likelihood function can be expressed as

$$\ln p(\mathbf{X}; \theta) = \ln L_G(\mathbf{X}; \theta) + \ln E_G \left[\frac{p_\varepsilon(\varepsilon; \theta)}{p_G(\varepsilon; \theta)} \right], \tag{13}$$

where $\ln L_G(\mathbf{X}; \theta)$ is the the log-likelihood function of a carefully chosen approximating Gaussian model, $p_\varepsilon(\varepsilon; \theta)$ the true density of $\varepsilon (:= (\varepsilon_1, \ldots, \varepsilon_n)')$, $p_G(\varepsilon; \theta)$ the Gaussian density of the measurement errors of the approximating model, E_G the expectation with respect to the importance density in connection to the approximating model.

Relative to (3), (13) has the advantage that simulations are only needed to estimate the departure of the likelihood from the Gaussian likelihood, rather than the full likelihood. For the LN-SV model, $\ln L_G(\mathbf{X}; \theta)$ often takes a much larger value than $\ln E_G \left[\frac{p_\varepsilon(\varepsilon;\theta)}{p_G(\varepsilon;\theta)} \right]$. As a result, MCL is computationally efficient than other simulated-based ML methods because it only needs a small number of simulations to achieve the desirable accuracy when approximating the likelihood. However, the implementation of the method requires a linear non-Gaussian state-space representation. Jungbacker and Koopman (2007) extended the method to deal with nonlinear non-Gaussian state-space models. Sandmann and Koopman (1998) applied the method to estimate the LN-SV model and the SV-t model. Broto and Ruiz (2004) compared the performance of alternative methods for estimating the LN-SV model and found supporting evidence for of the good performance of MCL.

3.3 Efficient Importance Sampler

Richard and Zhang (2007) developed an alternative simulated ML method. It is based on a particular factorization of the importance density and termed as Efficient Importance Sampling (EIS). Relative to the two SML methods reviewed in Sects 3.1 and 3.2, EIS minimizes locally the Monte Carlo sampling variance of the approximation to the integrand by factorizing the importance density. To fix the idea, assume $g(\mathbf{V}|\mathbf{X})$ is the importance density which can be constructed as

$$g(\mathbf{V}|\mathbf{X}) = \prod_{t=1}^{n} g(h_t | h_{t-1}, \mathbf{X}) = \prod_{t=1}^{n} \left\{ C_t e^{c_t h_t + d_t h_t^2} p(h_t | h_{t-1}) \right\}, \tag{14}$$

where c_t, C_t and d_t depend on $\mathbf{X}$ and h_{t-1} with $\{C_t\}$ be a normalization sequence so that g is a normal distribution. The sequences $\{c_t\}$ and $\{d_t\}$ should be chosen to match $p(\mathbf{X}, \mathbf{V}; \theta)$ and $g(\mathbf{V}|\mathbf{X})$ which, as we shown in Sect. 15.3.1, requires a high-dimensional non-linear regression. The caveat of EIS is to match each component in $g(\mathbf{V}|\mathbf{X})$ (i.e. $C_t e^{c_t h_t + d_t h_t^2} p(h_t|h_{t-1})$), to the corresponding element in the integrand $p(\mathbf{X}; \mathbf{V})$ (ie $p(X_t|h_t)p(h_t|h_{t-1})$) in a backward manner, with $t = n, n-1, \cdots, 1$. It is easy to show that C_t depends only on h_{t-1} but not on h_t. As a result, the recursive matching problem is equivalent to running the following linear regression backward:

$$\ln p(X_t|h_t^{(s)}) - \ln C_{t+1} = a + c_t h_t^{(s)} + d_t (h_t^{(s)})^2, \; s = 1, \cdots, S, \tag{15}$$

where $h_t^{(1)}, \ldots, h_t^{(S)}$ are drawn from the importance density and $h_t^{(s)}$ and $(h_t^{(s)})^2$ are treated as the explanatory variables in the regression model with $C_{n+1} = 1$.

The method to approximate the likelihood involves the following procedures:

1. Draw initial $\mathbf{V}^{(s)}$ from (2) with $s = 1, \cdots, S$.
2. Estimate c_t and d_t from (15) and do it backward with $C_{n+1} = 1$.
3. Draw $\mathbf{V}^{(s)}$ from importance density $g(\mathbf{V}|\mathbf{X})$ based on c_t and d_t.
4. Repeat Steps 2-3 until convergence. Denote the resulting sampler by $\mathbf{V}^{(s)}$.
5. Approximate the likelihood by

$$\frac{1}{S} \sum_{s=1}^{S} \left\{ \prod_{t=1}^{n} \frac{p(X_t|h_t^{(s)})}{C_t \exp\left(c_t h_t^{(s)} + d_t (h_t^{(s)})^2\right)} \right\}.$$

The EIS algorithm relies on the user to provide a problem-dependent auxiliary class of importance samplers. An advantage of this method is that it does not rely on the assumption that the latent process is Gaussian. Liesenfeld and Richard (2003, 2006) applied this method to estimate a number of discrete SV models while Kleppe et al. (2009) applied this method to estimate a continuous time SV model. Lee and Koopman (2004) compared the EIS method with the LA-IS method and found two methods are comparable in the context of the LN-SV model and the SV-t model. Bauwens and Galli (2008) and Bauwens and Hautsch (2006) applied EIS to estimate a stochastic duration model and a stochastic conditional intensity model, respectively.

3.4 An Empirical Example

For the purposes of illustration, we fit the LN-SV model to a widely used dataset (namely svpd1.txt). The dataset consists of 945 observations on daily pound/dollar exchange rate from 01/10/1981 to 28/06/1985. The same data were used in Harvey et al. (1994), Shephard and Pitt (1997), Meyer and Yu (2000), and Skaug and Yu (2007).

Table 1 SMLE of the LN-SV model

	σ	γ	ϕ	Log-likelihood
$S = 32$	0.6323	0.1685	0.9748	917.845
$S = 64$	0.6305	0.1687	0.9734	917.458

Matlab code (namely LAISLNSV.m) is used to implement the LA-IS method. Table 1 reports the estimates and the likelihood when $S = 32$. In Skaug and Yu (2007) the same method was used to estimate the same model but S was set at 64. The estimates and the log-likelihood value based on $S = 32$ are very similar to those based on $S = 64$, suggesting that a small number of random samples can approximate the likelihood function very well.

4 Simulated GMM and Continuous Time Models

Many models that are used to describe financial time series are written in terms of a continuous time diffusion $X(t)$ that satisfies the stochastic differential equation

$$dX(t) = \mu(X(t);\theta)dt + \sigma(X(t);\theta)dB(t), \tag{16}$$

where $B(t)$ is a standard Brownian motion, $\sigma(X(t);\theta)$ a diffusion function, $\mu(X(t);\theta)$ a drift function, and θ a vector of unknown parameters. The target here is to estimate θ from a discrete sampled observations, $X_h, \ldots, X_{nh}$ with h being the sampling interval. This class of parametric models has been widely used to characterize the temporal dynamics of financial variables, including stock prices, interest rates, and exchange rates.

Many estimation methods are based on the construction of the likelihood function derived from the transition probability density of the discretely sampled data. This approach is explained as follows. Suppose $p(X_{ih}|X_{(i-1)h},\theta)$ is the transition probability density. The Markov property of model (16) implies the following log-likelihood function for the discrete sample

$$\ell(\theta) = \sum_{i=1}^{n} \ln(p(X_{ih}|X_{(i-1)h},\theta)). \tag{17}$$

To perform exact ML estimation, one needs a closed form expression for $\ell(\theta)$ and hence $\ln(p(X_{ih}|X_{(i-1)h},\theta))$. In general, the transition density p satisfies the forward equation:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}$$

and the backward equation:

$$\frac{\partial p}{\partial s} = -\frac{1}{2}\frac{\partial^2 p}{\partial x^2},$$

where $p(y, t|x, s)$ is the transition density. Solving the partial differential equation numerically at $y = X_{ih}, x = X_{(i-1)h}$ yields the transition density. This approach was proposed by Lo (1988).

Unfortunately, only in rare cases, does the transition density $p(X_{ih}|X_{(i-1)h}, \theta)$ have a closed form solution. Phillips and Yu (2009) provide a list of examples in which $\ln(p(X_{ih}|X_{(i-1)h}, \theta))$ have a closed form analytical expression. These examples include the geometric Brownian Motion, Ornstein-Uhlenbeck (OU) process, square-root process, and inverse square-root process. In general solving the forward/backward equations is computationally demanding.

A classical and widely used estimation method is via the Euler scheme, which approximates a general diffusion process such as equation (16) by the following discrete time model

$$X_{ih} = X_{(i-1)h} + \mu(X_{(i-1)h}, \theta)h + \sigma(X_{(i-1)h}, \theta)\sqrt{h}\epsilon_i, \tag{18}$$

where $\epsilon_i \sim$ i.i.d. $N(0, 1)$. The transition density for the Euler discrete time model (18) has the following closed form expression:

$$X_{ih}|X_{(i-1)h} \sim N\left(X_{(i-1)h} + \mu(X_{(i-1)h}, \theta)h, \sigma^2(X_{(i-1)h}, \theta)h\right). \tag{19}$$

Obviously, the Euler scheme introduces a discretization bias. The magnitude of the bias introduced by Euler scheme is determined by h, which cannot be controlled econometricians. In general, the bias becomes negligible when h is close to zero. One way to use the full likelihood analysis is to make the sampling interval arbitrarily small by partitioning the original sampling interval so that the new subintervals are sufficiently fine for the discretization bias to be negligible. By making the subintervals smaller, one inevitably introduces latent variables between the two original consecutive observations $X_{(i-1)h}$ and X_{ih}. While our main focus is SGMM in this section, SML is possible and is discussed first.

4.1 SML Methods

To implement ML estimation, one can integrate out these latent observations.[3] When the partition becomes finer, the discretization bias is approaching 0 but the required

[3] Alternative to simulation-based approaches, one can use closed-form sequences to approximate the transition density itself, thereby developing an approximation to the likelihood function. Two different approximation mechanisms have been proposed in the literature. One is based on Hermite polynomial expansions (Aït-Sahalia 1999, 2002, 2008) whereas the other is based on the saddlepoint approximation (Aït-Sahalia and Yu 2006).

integration becomes high dimensional. In general, the integral does not have a closed-form expression and hence simulation-based methods can be used, leading to simulated ML estimators. To fix the idea, suppose that $M-1$ auxiliary points are introduced between $(i-1)h$ and ih, i.e.

$$((i-1)h \equiv)\tau_0, \tau_1, \cdots, \tau_{M-1}, \tau_M (\equiv ih).$$

Thus

$$\begin{aligned} p(X_{ih}|X_{(i-1)h};\theta) &= \int \cdots \int p(X_{\tau_M}, X_{\tau_{M-1}}, \cdots, X_{\tau_1}|X_{\tau_0};\theta) dX_{\tau_1} \cdots dX_{\tau_{M-1}} \\ &= \int \cdots \int \prod_{m=1}^{M} p(X_{\tau_m}|X_{\tau_{m-1}};\theta) dX_{\tau_1} \cdots dX_{\tau_{M-1}}. \end{aligned} \tag{20}$$

The second equality follows from the Markov property. The idea behind the simulated ML method is to approximate the densities $p(X_{\tau_m}|X_{\tau_{m-1}};\theta)$ (step 1), evaluate the multidimensional integral using importance sampling techniques (step 2) and then maximize the likelihood function numerically. To the best of my knowledge, Pedersen (1995) was the first study that suggested the idea in this context.

Pedersen's method relies on the Euler scheme, namely, approximates the latent transition densities $p(X_{\tau_m}|X_{\tau_{m-1}};\theta)$ based on the Euler scheme and approximates the integral by drawing samples of $(X_{\tau_{M-1}}, \cdots, X_{\tau_1})$ via simulations from the Euler scheme. That is, the importance sampling function is the mapping from $(\epsilon_1, \epsilon_2, \cdots, \epsilon_{M-1}) \mapsto (X_{\tau_1}, X_{\tau_2}, \cdots, X_{\tau_{M-1}})$ given by the Euler scheme:

$$X_{\tau_{m+1}} = X_{\tau_m} + \mu(X_{\tau_m};\theta)h/M + \sigma(X_{\tau_m},\theta)\sqrt{h/M}\epsilon_{m+1},\ m = 0, \cdots, M-2,$$

where $(\epsilon_1, \cdots, \epsilon_{M-1})$ is a multivariate standard normal.

Durham and Gallant (2002) noted two sources of approximation error in Pedersen's method, the discretization bias in the Euler scheme and the errors due to the Monte Carlo integration. A number of studies have provided methods to reduce these two sources of error. For example, to reduce the discretization bias in step 1, Elerian (1998) used the Milstein scheme instead of the Euler scheme while Durham and Gallant advocated using a variance stablization transformation, i.e. applying the Lamperti transform to the continuous time model. Certainly, other methods that can reduce the discretization bias may be used. Regarding step 2, Elerian et al. (2001) argued that the importance sampling function of Pedersen ignores the end-point information, X_{τ_M}, and Durham and Gallant (2002) showed that Pedersen's importance function draws most samples from regions where the integrand has little mass. Consequently, Pedersen's method is simulation-inefficient.

To improve the efficiency of the importance sampler, Durham and Gallant (2002) considered the following importance sampling function

$$X_{\tau_{m+1}} = X_{\tau_m} + \frac{X_{ih} - X_{\tau_m}}{ih - \tau_m} h/M + \sigma(X_{\tau_m}, \theta)\sqrt{h/M}\epsilon_{m+1},\ m = 0, \cdots, M-2,$$

where $(\epsilon_1, \cdots, \epsilon_{M-1})$ is a multivariate standard normal. Loosing speaking, this is a Brownian bridge because it starts from $X_{(i-1)h}$ at $(i-1)h$ and is conditioned to terminate with X_{ih} at ih. Another importance sampling function proposed by Durham and Gallant (2002) is to draw $X_{\tau_{m+1}}$ from the density $N(X_{\tau_m} + \tilde{\mu}_m h/M, \tilde{\sigma}_m^2 h/M)$ where $\tilde{\mu}_m = (X_{\tau_M} - X_{\tau_m})/(ih - \tau_m)$, $\tilde{\sigma}_m^2 = \sigma^2(X_{\tau_m})(M-m-1)/(M-m)$. Elerian et al. (2001) suggested the following tied-down process:

$$p(X_{\tau_1}, \cdots, X_{\tau_{M-1}} | X_{\tau_0}, X_{\tau_M}),$$

as the importance function and proposed using the Laplace approximation to the tied-down process. Durham and Gallant (2002) compared the performance of these three importance functions relative to Pedersen (1995) and found that all these methods deliver substantial improvements.

4.2 Simulated GMM (SGMM)

Not only is the likelihood function for (16) difficult to construct, but also the moment conditions; see, for example, Duffie and Singleton (1993) and He (1990).[4] While model (16) is difficult to estimate, data can be easily simulated from it. For example, one can simulate data from the Euler scheme at an arbitrarily small sampling interval. With the interval approaches to zero, the simulated data can be regarded as the exact simulation although the transition density at the coarser sampling interval is not known analytically. With simulated data, moments can be easily constructed, facilitating simulation-based GMM estimation. Simulated GMM (SGMM) methods have been proposed by McFadden (1989), Pakes and Pollard (1989) for iid environments, and Lee and Ingram (1991), Duffie and Singleton (1993) for time series environments.

Let $\{\widetilde{\mathbf{X}}_t^{(s)}(\theta)\}_{t=1}^{\mathcal{N}(n)}$ be the data simulated from (16) when parameter is θ using random seed s. Therefore, $\{\widetilde{X}_t^{(s)}(\theta_0)\}$ is drawn from the same distribution as the original data $\{\mathbf{X}_t\}$ and hence share the same moment characteristic. The parameter θ is chosen so as to "match moments", that is, to minimize the distance between

[4] However, when $X(t)$ is observed, Hansen and Scheinkman (1995) showed that there exist forward and reverse-time generators for stationary continuous time models and explained how to use these generators to construct moment conditions.

sample moments of the data and those of the simulated data. Assuming H represents K-moments, SGMM estimator is defined as:

$$\hat{\theta}_n^{SGMM} := \operatorname{argmin}_{\theta \in \Theta} \left(\frac{1}{n} \sum_{t=1}^{n} g(X_t) - \frac{1}{\mathcal{N}(n)} \sum_{t=1}^{\mathcal{N}(n)} g(\tilde{X}_t^{(s)}; \theta) \right)' W_n$$
$$\left(\frac{1}{n} \sum_{t=1}^{n} g(X_t) - \frac{1}{\mathcal{N}(n)} \sum_{t=1}^{\mathcal{N}(n)} g(\tilde{X}_t^{(s)}; \theta) \right)',$$

where W_n is a certain positive definite weighting matrix of $q \times q$-dimension ($q \geq K$), which may depend on the sample but not θ, $\mathcal{N}(n)$ is the number of number of observations in a simulated path. Under the ergodicity condition,

$$\frac{1}{\mathcal{N}(n)} \sum_{t=1}^{\mathcal{N}(n)} g(\tilde{X}_t^{(s)}; \theta_0) \xrightarrow{p} E(g(X_t; \theta_0))$$

and

$$\frac{1}{n} \sum_{t=1}^{n} g(X_t) \xrightarrow{p} E(g(X_t; \theta_0)),$$

justifying the SGMM procedure.

The SGMM procedure can be made optimal with a careful choice of the weighting function, given a set of moments. However, the SGMM estimator is in general asymptotically less efficient than SML for the reason that moments are less informative than the likelihood. Gallant and Tauchen (1996a,b) extended the SGMM technique so that the GMM estimator is asymptotically as efficient as SML. This approach is termed efficient method of moments (EMM), which we review below.

4.3 Efficient Method of Moments

EMM is first introduced by Gallant and Tauchen (1996a,b) and has now found many applications in financial time series; see Gallant and Tauchen (2001a,c) for the detailed account of the method and a review of the literature. While it is closely related to the general SGMM, there is one important difference between them. Namely, GMM relies on an ad hoc chosen set of moment conditions, EMM is based on a judiciously chosen set of moment conditions. The moment conditions that EMM is based on are the expectation of the score of an auxiliary model which is often referred to as the score generator.

For the purpose of illustration, let a SV model be the structural model. The SV model is the continuous time version of the Box-Cox SV model of Yu et al. (2006), which contains many classical continuous SV models as special cases, and is of the form:

$$dS(t) = \alpha_{10} S(t)dt + S(t)[1 + \delta(\beta_{10} + \beta_{12} h(t))]^{1/(2\delta)} dB_1(t),$$
$$dh(t) = -\alpha_{22} h(t)dt + dB_2(t).$$

Let the conditional density of the structural model (the Box-Cox SV model in this case) is defined by

$$p_t(X_t|Y_t, \theta),$$

where $X_t = \ln S(t)$, the true value of θ is θ_0, $\theta_0 \in \Theta \subset \Re^{\ell_\theta}$ with ℓ_θ being the length of θ_0 and Y_t is a vector of lagged X_t. Denote the conditional density of an auxiliary model by

$$f_t(X_t|Y_t, \beta), \beta \in R \subset \Re^{\ell_\beta}.$$

Further define the expected score of the auxiliary model under the structural model as

$$m(\theta, \beta) = \int \cdots \int \frac{\partial}{\partial \beta} \ln f(x|y, \beta) p(x|y, \theta) p(y|\theta) dx dy.$$

Obviously, in the context of the SV model, the integration cannot be solved analytically since neither $p(x|y, \theta)$ nor $p(y|\theta)$ has a closed form expression. However, it is easy to simulate from an SV model so that one can approximate the integral by Monte Carlo simulations. That is

$$m(\theta, \beta) \approx m_N(\theta, \beta) \equiv \frac{1}{N} \sum_{\tau=1}^{N} \frac{\partial}{\partial \beta} \ln f(\hat{X}_\tau(\theta)|\hat{Y}_\tau(\theta), \beta),$$

where $\{\hat{X}_\tau, \hat{Y}_\tau\}$ are simulated from the structural model. The EMM estimator is a minimum chi-squared estimator which minimizes the following quadratic form,

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} m'_N(\theta, \hat{\beta}_n)(I_n)^{-1} m_N(\theta, \hat{\beta}_n),$$

where $\hat{\beta}_n$ is a quasi maximum likelihood estimator of the auxiliary model and I_n is an estimate of

$$I_0 = \lim_{n \to \infty} Var\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{\frac{\partial}{\partial \beta} \ln f_t(x_t|y_t, \beta^*)\right\}\right)$$

with β^* being the pseudo true value of β. Under regularity conditions, Gallant and Tauchen (1996a,b) show that the EMM estimator is consistent and has the following asymptotic normal distribution,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{\partial}{\partial\theta} m(\theta_0, \beta^*)(I_0)^{-1} \frac{\partial}{\partial\theta'} m(\theta_0, \beta^*)\right).$$

For specification testing, we have

$$J_n = n m_N'(\hat{\theta}_n, \hat{\beta}_n)(I_n)^{-1} m_N(\hat{\theta}_n, \hat{\beta}_n) \xrightarrow{d} \chi^2_{\ell_\beta - \ell_\theta}$$

under the null hypothesis that the structural model is correct. When a model fails the above specification test one may wish to examine the quasi-t-ratios and/or t-ratios to look for some suggestion as to what is wrong with the structural model. The quasi-t-ratios are defined as

$$\hat{T}_n = S_n^{-1} \sqrt{n} m_N(\hat{\theta}_n, \hat{\beta}_n),$$

where $S_n = [diag(I_n)]^{1/2}$. It is well known that the elements of $\hat{T}_n$ are downward biased in absolute value. To correct the bias one can use the t-ratios defined by

$$\tilde{T}_n = Q_n^{-1} \sqrt{n} m_N(\hat{\theta}_n, \hat{\beta}_n),$$

where

$$Q_n = \left(diag\{I_n - \frac{\partial}{\partial\theta'} m_N(\hat{\theta}_n, \hat{\beta}_n)[m_N'(\hat{\theta}_n, \hat{\beta}_n)(I_n)^{-1} m_N(\hat{\theta}_n, \hat{\beta}_n)]^{-1} \frac{\partial}{\partial\theta} m_N(\hat{\theta}_n, \hat{\beta}_n)\}\right)^{1/2}.$$

Large quasi-t-ratios and t-ratios reveal the features of the data that the structural model cannot approximate.

Furthermore, Gallant and Tauchen (1996a,b) show that if the auxiliary model nests the data generating process, under regularity conditions the EMM estimator has the same asymptotic variance as the maximum likelihood estimator and hence is fully efficient. If the auxiliary model can closely approximate the data generating process, the EMM estimator is nearly fully efficient (Gallant and Long 1997; Tauchen 1997).

To choose an auxiliary model, the seminonparametric (SNP) density proposed by Gallant and Tauchen (1989) can be used since its success has been documented in many applications. As to SNP modeling, six out of eight tuning parameters are to be selected, namely, L_u, L_g, L_r, L_p, K_z, and K_y. The other two parameters, I_z and I_x, are irrelevant for univariate time series and hence set to be 0. L_u determines the location transformation whereas L_g and L_r determine the scale transformation. Altogether they determine the nature of the leading term of the Hermite expansion. The other two parameters K_z and K_y determine the nature of the innovation. To search for a good auxiliary model, one can use the Schwarz BIC criterion to move along an upward expansion path until an adequate model is found, as outlined in

Table 2 EMM estimate of the continuous time box-cox SV model

α_{10}	α_{22}	β_{10}	β_{12}	δ	χ^2_6
0.4364	0.5649	−0.1094	0.2710	0.1367	13.895

Bansal et al. (1995). To preserve space we refer readers to Gallant and Tauchen (2001b) for further discussion about the role of the tuning parameters and how to design an expansion path to choose them.

While EMM has found a wide range of applications in financial time series, Duffee and Stanton (2008) reported finite sample evidence against EMM when financial time series is persistent. In particular, in the context of simple term structure models, they showed that although EMM has the same asymptotic efficiency as ML, the variance of EMM estimator in finite sample is too large and cannot be accepted in practice.

4.4 An Empirical Example

For the purposes of illustration, we fit the continuous time Box-Cox SV model to daily prices of Microsoft. The stock price data consist of 3,778 observations on the daily price of a share of Microsoft, adjusted for stock split, for the period from March 13, 1986 to February 23, 2001. The same data have been used in Gallant and Tauchen (2001a) to fit a continuous time LN-SV model. For this reason, we use the same sets of tuning parameters in the SNP model as in Gallant and Tauchen (2001a), namely,

$$(L_u, L_g, L_r, L_p, K_z, I_z, K_y, I_y) = (1, 1, 1, 1, 6, 0, 0, 0).$$

Fortran code and the date can be obtained from an anonymous ftp site at ftp.econ.duke.edu. A EMM User Guide by Gallant and Tauchen (2001a) is available from the same site. To estimate the Box-Cox SV model, we only needed to change the specification of the diffusion function in the subroutine difuse in the fortran file emmuothr.f, i.e. "tmp1 = DEXP(DMIN1 (tmp1,bnd))" is changed to "tmp1 = (1+ delta* DMIN1 (tmp1,bnd))**(0.5/delta)". Table 2 reports the EMM estimates. Obviously, the volatility of Microsoft is very persistent since the estimated mean reversion parameter is close to zero and the estimate value of δ is not far away from 0, indicating that the estimated Box-Cox SV is not very different from the LN-SV model model.

5 Bayesian MCMC and Credit Risk Models

Credit derivatives market had experienced a fantastic growth before the global financial meltdown in 2007. The size of the market had grew so much and the credit risk management had been done so poorly in practice that the impact of the financial crisis is so big. Not surprisingly, how to estimate credit risk has received an increasing attention from academic researchers, industry participants, policy makers and regulators.

A widely used approach to credit risk modelling in practice is the so-called structural method. All structural credit risk models specify a dynamic structure for the underlying firm's asset and default boundary. Let V be the firm's asset process, r the risk-free interest rate, F the face value of a zero-coupon debt that the firm issues with the time to maturity T. Merton (1974) is the simplest structural model where V_t is assumed to follow a geometric Brownian motion:

$$d \ln V_t = (\mu - \sigma^2/2)dt + \sigma dB_t,\ V_0 = c, \tag{21}$$

The exact discrete time model, sampled with the step size h, is

$$\ln V_{t+1} = (\mu - \sigma^2/2)h + \ln V_t + \sigma\sqrt{h}\epsilon_t,\ V_0 = c, \tag{22}$$

which contains a unit root.

There are two types of outstanding claims faced by a firm that is listed in a stock exchange, an equity and a zero-coupon debt whose face value is F maturing at T. The default occurs at the maturity date of debt in the event that the issuer's assets are less than the face value of the debt (i.e. $V_T < F$). Under the assumption of (21) the firm's equity can be priced with the Black-Scholes formula as if it is a call option on the total asset value V of the firm with the strike price of F and the maturity date T. Namely, the equity claim, denoted by S_t, is

$$S_t \equiv S(V_t;\sigma) = V_t\Phi(d_{1t}) - Fe^{-r(T-t)}\Phi(d_{2t}), \tag{23}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal variate,

$$d_{1t} = \frac{\ln(V_t/F) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},$$

and

$$d_{2t} = \frac{\ln(V_t/F) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}.$$

Merton's model can be used to evaluate private firm credit risk and the credit spread of a risk corporate bond over the corresponding Treasure rate. The credit spread is given by

$$C(V_t;\theta) = -\frac{1}{T-\tau_t}\ln\left(\frac{V_t}{F}\Phi(-d_{1t}) + e^{-r(T-\tau_t)}\Phi(d_{2t})\right) - r. \tag{24}$$

The default probability is given by

$$P(V_t;\theta) = \Phi\left(\frac{\ln(F/V_t) - (\mu - \sigma^2/2)(T - \tau_t)}{\sigma\sqrt{T - \tau_t}}\right). \tag{25}$$

At a reasonably high frequency, S_t may be observed with errors due to the presence of various market microstructure effects. This observation motivates Duan and Fulop (2009) to consider the following generalization to Merton's model:

$$\ln S_t = \ln S(V_t;\sigma) + \delta v_t, \;\; v_t \sim N(0,1). \tag{26}$$

In a state-space framework, (26) is an observation equation and (22) is a state equation. Unfortunately, the Kalman filter is not applicable here since the observation equation is nonlinear.

Let $\mathbf{X} = (\ln S_1, \cdots, \ln S_n)'$, $\mathbf{V} = (\ln V_1, \cdots, \ln V_n)'$, and $\theta = (\mu, \sigma, \delta)'$. The likelihood function of (26) is given by

$$p(\mathbf{X};\theta) = \int p(\mathbf{X},\mathbf{V};\theta)d\mathbf{V} = \int p(\mathbf{X}|\mathbf{V};\mu)p(\mathbf{V};\theta)d\mathbf{V}. \tag{27}$$

In general this is a high-dimensional integral which does not have closed form expression due to the non-linear dependence of $\ln S_t$ on $\ln V_t$. Although in this section, our main focus is the Bayesian MCMC methods, SML is possible. Indeed all the SML methods discussed in Sect. 15.3 are applicable here. However, we will discuss a new set of SML methods – particle filters.

5.1 SML via Particle Filter

It is known that Kalman filter is an optimal recursive data processing algorithm for processing series of measurements generated from a linear dynamic system. It is applicable any linear Gaussian state-space model where all relevant conditional distributions are linear Gaussians. Particle filters, also known as sequential Monte Carlo methods, extend the Kalman filter to nonlinear and non-Gaussian state space models.

In a state space model, two equations have to be specified in the fully parametric manner. First, the state equation describes the evolution of the state with time. Second, the measurement equation relates the noisy measurements to the state. A recursive filtering approach means that received data can be processed sequentially rather than as a batch so that it is not necessary to store the complete data set nor to reprocess existing data if a new measurement becomes available. Such a filter consists of essentially two stages: prediction and updating. The prediction stage uses the system model to predict the state density forward from one measurement time to the next. Since the state is usually subject to unknown disturbances, prediction generally translates, deforms, and spreads the state density. The updating operation uses the latest measurement to modify the prediction density. This is achieved using Bayes theorem, which is the mechanism for updating knowledge about the target state in the light of extra information from new data. When the model is linear and Gaussian, the density in both stages is Gaussian and Kalman filter gives analytical expressions to the mean and the co-variance. As a byproduct, the full conditional distribution of measurements is available, facilitating the calculation of the likelihood.

For nonlinear and non-Gaussain state space models, the density in neither stage is not Gaussian any more and the optimal filter is not available analytically. Particle filter is a technique for implementing a recursive filter by Monte Carlo simulations. The key idea is to represent the required density in connection to prediction and updating by a set of random samples (known as "particles") with associated weights and to compute estimates based on these samples and weights. As the number of samples becomes very large, this simulation-based empirical distribution is equivalent the true distribution.

To fix the idea, assume that the nonlinear non-Gaussian state space model is of the form,

$$\begin{cases} Y_t &= H(X_t, e_t) \\ X_t &= F(X_{t-1}, u_t), \end{cases} \tag{28}$$

where X_t is a k-dimensional state vector,[5] u_t is a l-dimensional white noise sequence with density $q(u)$, v_t is a l-dimensional white noise sequence with density $r(v)$ and assumed uncorrelated with $\{u_s\}_{s=1}^t$, H and F are possibly nonlinear functions. Let $v_t = G(Y_t, X_t)$ and G' is the derivative of G as a function of Y_t. The density of the initial state vector is assumed to be $p_0(x)$. Denote $Y_{1:k} = \{Y_1, \cdots, Y_k\}$. The objective of the prediction is to obtain $p(X_t|Y_{1:t})$. It can be seen that

$$p(X_t|Y_{1:t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Y_{1:t-1})dX_{t-1}. \tag{29}$$

At time step t, when a new measurement Y_t becomes available, it may be used to update the predictive density $p(X_t|Y_{1:t-1})$ via Bayes rule in the updating stage,

$$p(X_t|Y_{1:t}) = \frac{p(Y_t|X_t)p(X_t|Y_{1:t-1})}{p(Y_t|Y_{1:t-1})}. \tag{30}$$

Unfortunately, for the nonlinear non-Gaussian state-space model, the recursive propagation in both stages is only a conceptual solution and cannot be determined analytically. To deal with this problem, particle filtering algorithm consists of recursive propagation of the weights and support points when each measurement is received sequentially so that the true densities can be approximated by the corresponding empirical density.

[5] In Merton's model, $X_t = \ln V_t$, $Y_t = \ln S_t$, $e_t = \sigma\sqrt{h}\epsilon_t$, $u_t = \delta v_t$.

Various versions of particle filters have been proposed in the literature. In this chapter we only summarize all the steps involved in Kitagawa's algorithm (Kitagawa 1996):

1. Generate M l-dimensional particles from $p_0(x)$, $f_0^{(j)}$ for $j = 1, \ldots, M$.
2. Repeat the following steps for $t = 1, \ldots, n$.
   (a) Generate M l-dimensional particles from $q(u)$, $u_t^{(j)}$ for $j = 1, \ldots, M$.
   (b) Compute $p_t^{(j)} = F(f_{t-1}^{(j)}, u_t^{(j)})$ for $j = 1, \ldots, M$.
   (c) Compute $\alpha_t^{(j)} = r(G(Y_t, p_t^{(j)}))$ for $j = 1, \ldots, M$.
   (d) Re-sample $\{p_t^{(j)}\}_{j=1}^{M}$ to get $\{f_t^{(j)}\}_{j=1}^{M}$ with probabilities proportional to $\{r(G(Y_t, p_t^{(j)})) \times |G'(Y_t, p_t^{(j)})|\}_{j=1}^{M}$.

Other particle filtering algorithms include sampling importance resampling filter of Gordon et al. (1993), auxiliary sampling importance resampling filter of Pitt and Shephard (1999a), and regularized particle filter (Musso et al. 2001).

To estimate the Merton's model via ML, Duan and Fulop employed the *particle filtering* method of Pitt (2002). Unlike the method proposed by Kitagawa (1995) which samples a point $X_t^{(m)}$ when the system is advanced, Duan and Fulop sampled a pair $(V_t^{(m)}, V_{t+1}^{(m)})$ at once when the system is advanced. Since the resulting likelihood function is not smooth with respect to the parameters, to ensure a smooth surface for the likelihood function, Duan and Fulop used the smooth bootstrap procedure for resampling of Pitt (2002).

Because the log-likelihood function can be obtained as a by-product of the filtering algorithm, it can be maximized numerically over the parameter space to obtain the SMLE. If $M \to \infty$, the log-likelihood value obtained from simulations should converge to the true likelihood value. As a result, it is expected that for a sufficiently large number of particles, the estimates that maximize the approximated log-likelihood function are sufficiently close to the true ML estimates.

5.2 *Bayesian MCMC Methods*

The structure in the state-space model ensures the pivotal role played by Bayes theorem in the recursive propagation. Not surprisingly, the requirement for the updating of information on receipt of new measurements are ideally suited for the Bayesian approach for statistical inference. In this chapter, we will show that Bayesian methods provide a rigorous general approach to the dynamic state estimation problem. Since many models in financial econometrics have a state-space representation, Bayesian methods have received more and more attentions in statistical analysis of financial time series.

The general idea of the Bayesian approach is to perform posterior computations, given the likelihood function and the prior distribution. MCMC is a class of algothims which enables one to obtain a correlated sample from a Morkov chain whose stationary transition density is the same as the posterior distribution. There are certain advantages in the Bayesian MCMC method. First, as a likelihood-based method, MCMC matches the efficiency of ML. Second, as a by-product of parameter estimation, MCMC provides smoothed estimates of latent variables because it augments the parameter space by including the latent variables. Third, unlike the frequentist's methods whose inference is almost always based on asymptotic arguments, inferences via MCMC are based on the exact posterior distribution. This advantage is especially important when the standard asymptotic theory is difficult to derive or the asymptotic distribution does not provide satisfactory approximation to the finite sample distribution. As a trade-off, one has to specify the prior distribution. In addition, with MCMC it is straightforward to obtain the exact posterior distribution of any transformation (linear or nonlinear) of model parameters and latent variables, such as the credit spread and the default probability. Therefore, the exact finite sample inference can easily be made in MCMC, whereas the ML method necessitates the delta method to obtain the asymptotic distribution. When the asymptotic distribution of the original parameters does not work well, it is expected that the asymptotic distribution yielded by the delta method may not work well. Fourth, numerical optimization is not needed in MCMC. This advantage is of practical importance when the likelihood function is difficult to optimize numerically. Finally, the proposed method lends itself easily to dealing with flexible specifications.

There are three disadvantages of the MCMC method. First, in order to obtain the filtered estimate of the latent variable, a separate method is required. This is in contrast with the ML method of Duan and Fulop (2009) where the filtered estimate of the latent variable is obtained as a by-product. Second, with the MCMC method the model has to be fully specified whereas the MLE remains consistent even when the microstructure noise is nonparametrically specified, and in this case, ML becomes quasi-ML. However, in recent years, semiparametric MCMC methods have appeared in the literature. For example, the flexibility of the error distribution may be accommodated by using a Dirichelt process mixture (DPM) prior (see Ferguson (1973) for the detailed account of DMP, and Jensen and Maheu (2008) for an application of DMP to volatility modeling). Finally, prior distributions have to be specified. In some cases, prior distributions may have important influences on the posterior analysis but it is not so obvious to specify the prior distributions.

From the Bayesian viewpoint, we understand the specification of the structural credit risk model as a hierarchical structure of conditional distributions. The hierarchy is specified by a sequence of three distributions, the conditional distribution of $\ln S_t \,|\, \ln V_t, \delta$, the conditional distribution of $\ln V_t \,|\, \ln V_{t-1}, \mu, \sigma$, and the prior distribution of θ. Hence, our Bayesian model consists of the joint prior distribution of all unobservables, here the three parameters, μ, σ, δ, and the unknown states, $\mathbf{V}$, and the joint distribution of the observables, here the sequence of contaminated log-equity prices $\mathbf{X}$. The treatment of the latent state variables $\mathbf{V}$ as the additional unknown parameters is the well known data-augmentation technique originally proposed by Tanner and Wong (1987) in the context of MCMC. Bayesian inference is then based on the posterior distribution of the unobservables given the data. In the sequel, we will denote the probability density function of a random variable θ by $p(\theta)$. By successive conditioning, the joint prior density is

$$p(\mu, \sigma, \delta, \mathbf{V}) = p(\mu, \sigma, \delta) p(\ln V_0) \prod_{t=1}^{n} p(\ln V_t | \ln V_{t-1}, \mu, \sigma). \tag{31}$$

We assume prior independence of the parameters μ, δ and σ. Clearly $p(\ln V_t | \ln V_{t-1}, \mu, \sigma)$ is defined through the state equations (22). The likelihood $p(\mathbf{X}|\mu, \sigma, \delta, \mathbf{V})$ is specified by the observation equations (26) and the conditional independence assumption:

$$p(\mathbf{X}|\mu, \sigma, \delta, \mathbf{V}) = \prod_{t=1}^{n} p(\ln S_t | \ln V_t, \delta). \tag{32}$$

Then, by Bayes' theorem, the joint posterior distribution of the unobservables given the data is proportional to the prior times likelihood, i.e.

$$p(\mu, \sigma, \delta, \mathbf{V}|\mathbf{X}) \propto p(\mu) p(\sigma) p(\delta) p(\ln V_0) \prod_{t=1}^{n} p(\ln V_t | \ln V_{t-1}, \mu, \sigma) \prod_{t=1}^{n} p(\ln S_t | \ln V_t, \delta). \tag{33}$$

Without data augmentation, we need to deal with the intractable likelihood function $p(\mathbf{X}|\theta)$ which makes the direct analysis of the posterior density $p(\theta|\mathbf{V})$ difficult. The particle filtering algorithm of Duan and Fulop (2009) can be used to overcome the problem. With data augmentation, we focus on the new posterior density $p(\theta, \mathbf{V}|\mathbf{X})$ given in (33). Note that the new likelihood function is $p(\mathbf{X}|\theta, \mathbf{V})$ which is readily available analytically once the distribution of ϵ_t is specified. Another advantage of using the data-augmentation technique is that the latent state variables $\mathbf{V}$ are the additional unknown parameters and hence we can make statistical inference about them.

The idea behind the MCMC methods is to repeatedly sample from a Markov chain whose stationary (multivariate) distribution is the (multivariate) posterior density. Once the chain converges, the sample is regarded as a correlated sample from the posterior density. By the ergodic theorem for Markov chains, the posterior moments and marginal densities can be estimated by averaging the corresponding functions over the sample. For example, one can estimate the posterior mean by the sample mean, and obtain the credible interval from the marginal density. When the simulation size is very large, the marginal densities can be regarded to be exact, enabling exact finite sample inferences. Since the latent state variables are in the parameter space, MCMC also provides the exact solution to the smoothing problem of inferring about the unobserved equity value.

While there are a number of MCMC algorithms available in the literature, we only use the Gibbs sampler which samples each variate, one at a time, from the full conditional distributions defined by (33). When all the variates are sampled in a cycle, we have one sweep. The algorithm is then repeated for many sweeps with the variates being updated with the most recent samples. With regularity conditions, the draws from the samplers converge to draw from the posterior distribution at a geometric rate. For further information about MCMC and its applications in econometrics, see Chib (2001) and Johannes and Polson (2003).

Defining $\ln V_{-t}$ by $\ln V_1, \ldots, \ln V_{t-1}, \ln V_{t+1}, \ldots, \ln V_n$, the Gibbs sampler is summarized as:

1. Initialize θ and $\mathbf{V}$.
2. Sample $\ln V_t$ from $\ln V_t \,|\, \ln V_{-t}, \mathbf{X}$.
3. Sample $\sigma|\mathbf{X}, \mathbf{V}, \mu, \delta$.
4. Sample $\delta|\mathbf{X}, \mathbf{V}, \mu, \sigma$.
5. Sample $\mu|\mathbf{X}, \mathbf{V}, \sigma, \delta$.

Steps 2–5 forms one cycle. Repeating steps 2–5 for many thousands of times yields the MCMC output. To mitigate the effect of initialization and to ensure the full convergence of the chains, we discard the so-call burn-in samples. The remaining samples are used to make inference.

It is easy to implement the Gibbs sampling for the credit risk model defined above. One can make use of the all purpose Bayesian software package WinBUGS. As shown in Meyer and Yu (2000) and Yu et al. (2006), WinBUGS provides an idea framework to perform the Bayesian MCMC computation when the model has a state-space form, whether it is nonlinear or non-Gaussian or both. As the Gibbs sampler updates only one variable at a time, it is referred as a single-move algorithm.

In the stochastic volatility literature, the single-move algorithm has been criticized by Kim et al. (1998) for lacking simulation efficiency because the components of state variables are highly correlated. More efficient MCMC algorithms, such as multi-move algorithms, can be developed for estimating credit risk models. In fact, Shephard and Pitt (1997), Kim et al. (1998), Chib et al. (2002), Liesenfeld and Richard (2006) and Omori et al. (2007) have developed various multi-move algorithms to estimate univariate and multivariate SV models. The idea of the multi-mover algorithms is to sample the latent vector $\mathbf{V}$ in a single block.

5.3 An Empirical Application

For the purposes of illustration, we fit the credit risk model to daily prices of AA a company from the Dow Jones Industrial Index. The daily equity values are obtained from the CRSP database over year 2003 (the logarithmic values are contained in a file named AAlogS.txt). The initial maturity of debt is 10 years. The debt is available from the balance sheet obtained from the Compustat annual file. It is compounded for 10 years at the risk-free rate to obtain F. The risk-free rate is obtained from the US Federal Reserve. Duan and Fulop fitted the same model to the same data using SML via particle filter and approximated the variance using the Fisher information matrix. Following Huang and Yu (2009), we use the following independent prior

Table 3 MCMC and SML estimates of the credit risk model

	μ		σ		$\delta \times 100$	
	Mean	Std err	Mean	Std err	Mean	Std err
Bayesian	0.3154	0.1689	0.1686	0.0125	0.5673	0.1225
SML	0.3130	0.1640	0.1589	0.0181	0.6820	0.2082

for the three system parameters: $\mu \sim N(0.3, 4)$, $\delta \sim IG(3, 0.0001)$, and $\sigma \sim IG(2.5, 0.025)$ where IG is the inverse-gamma distribution.

WinBugs code (aa.odc) is used to implement the MCMC method based on 55,000 sweeps of which the first 5,000 sweeps are thrown away. Table 3 reports the estimates (the posterior means) and the standard errors (the posterior standard errors). For the purpose of comparison, the SML estimates and their asymptotic standard errors, obtained directly from Duan and Fulop (2009, Table 1), are also reported. While the two sets of estimates are close to each other, their standard errors are further away.

6 Resampling Methods and Term Structure Models

It is well known dynamic models are estimated with bias by standard estimation methods, such as least squares (LS), maximum likelihood (ML) or generalized method of moments (GMM). The bias was developed by Hurwicz (1950) for the autoregressive parameter in the context of dynamic discrete time models. The percentage bias of the corresponding parameter, i.e. the mean reversion parameter, is much more pronounced in continuous time models than their discrete time counterparts. On the other hand, estimation is fundamentally important for many practical applications. For example, it provides parameter estimators which are used directly for estimating prices of financial assets and derivatives. For another example, parameter estimation serves as an important stage for the empirical analysis of specification and comparative diagnostics. Not surprisingly, it has been found in the literature that the bias in the mean reversion estimator has important implications for the specification analysis of continuous time models (Pritsker 1998) and for pricing financial assets (Phillips and Yu 2005a, 2009b). For instance, when the true mean reversion parameter is 0.1 and 600 weekly observations (i.e. just over 10 years of data) are available to estimate a one-factor square-root term structure model (Cox et al. 1985), the bias in the ML estimator of the mean reversion parameter is 391.2% in an upwards direction. This estimation bias, together with the estimation errors and nonlinearity, produces a 60.6% downward bias in the option price of a discount bond and 2.48% downward bias in the discount bond price. The latter figures are comparable in magnitude to the estimates of bias effects discussed in Hull (2000, Chap. 21.7). The biases would be even larger when less observations are available and do not disappear even when using long spans of data that are currently available. For example, when the true mean reversion parameter is 0.1 and

600 monthly observations (i.e. 50 years of data) are available to estimate the square-root diffusion model, the bias in the ML estimator of the mean reversion parameter is 84.5% in an upwards direction. This estimation bias implies a 24.4% downward bias in the option price of a discount bond and a 1.0% downward bias in the discount bond price.

In recent years, there have been interesting advances in developing analytical formulae to approximate the bias in certain model specifications. This is typically obtained by estimating higher order terms in an asymptotic expansion of the bias. For example, in the Vasicek term structure model with a known μ,

$$dX_t = \kappa(\mu - X_t)dt + \sigma dB_t, X_0 \sim N(\mu, \sigma^2/(2\kappa))$$

Yu (2009a,b) showed that the bias in the MLE of κ can be approximated by

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

When μ has to be estimated in the Vasicek model, Tang and Chen (2009) showed that the bias in the MLE of κ can be approximated by

$$E(\widehat{\kappa}) - \kappa = \frac{1}{2T}(e^{2\kappa h} + 2e^{\kappa h} + 5).$$

Interestingly, the same bias formula applies to a QML estimate of κ, developed by Nowman (1997), under the CIR model, as shown in Tang and Chen (2009).

For more complicated models, unfortunately, the approximate bias formula is not available. To reduce this bias in parameter estimation and in pricing contingent claims, Phillips and Yu (2005a) proposed a new jackknife procedure. Phillips and Yu (2005a) show that the jackknife method always trades off the gain that may be achieved in bias reduction with a loss that arises through increased variance.

The bootstrap method of Efron (1979) is another way to reduce the bias via simulation. It was shown to be an effective method for bias correction (Hall 1992) and was illustrated in the parameter estimation in the context of continuous time model in Tang and Chen (2009). Relative to the jackknife method, it does not significantly increase the variance. Relative to the two simulation-based procedures that will be discussed below, however, bootstrap seems to use less information and hence is expected to be less efficient.

6.1 Indirect Inference and Median Unbiased Estimation

Resampling methods may achieve bias reduction as well as variance reduction. In this chapter, two simulation-based resampling methods are discussed, indirect inference (II) and median unbiased estimation (MUE).

II and MUE are simulation-based estimation procedures and can be understood as a generalization of the simulated method of moments approach of Duffie and Singleton (1993). MUE was first introduced by Andrews (1993). II was first introduced by Smith (1993) and coined with the term by Gouriéroux et al. (1993). II 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 many continuous time models are easy to simulate but difficult to obtain moment and likelihood functions, the II procedure has some convenient advantages in working with continuous time models in finance.

The II and MUE procedures can have good small sample properties of parameter estimates, as shown by Andrews (1993), MacKinnon and Smith (1996), Monfort (1996), Gouriéroux et al. (2000) in the time series context and by Gouriéroux et al. (2005) in the panel context. The idea 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/MUE for parameter estimation, consider the Vasicek model which is typically used to describe the movement of the short term interest rate. 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) as well as 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 $\delta = h/100$), 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

$$t = \underbrace{0, \delta, \cdots, h(= 100\delta)}, \underbrace{h + \delta, \cdots, 2h(= 200\delta)}, 2h + \delta, \cdots, nh.$$

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 $\{\tilde{X}^k_{ih}\}^n_{i=1}$, which can be regarded as data simulated directly from the OU model with the (observationally relevant) step size h.[6]

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 choose the number of simulated observations and the sampling interval to be the same as the number of observations and the sampling interval in the observed sequence for the purpose of the bias calibration. Another estimator of κ can be obtained by applying the Euler scheme to $\{X^k_h, \cdots, X^k_{nh}\}$ (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/MUE is to match the parameter obtained from the actual data with that obtained from the simulated data. In particular, the II estimator and median unbiased estimator of κ solve, respectively,

$$\hat{\kappa}_n = \frac{1}{K}\sum_{h=1}^{K} \tilde{\kappa}^k_n(\kappa) \text{ or } \hat{\kappa}_n = \hat{\rho}_{0.5}(\tilde{\kappa}^k_n(\kappa)), \tag{34}$$

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

$$\hat{\kappa}_n = E(\tilde{\kappa}^k_n(\kappa)) \text{ or } \hat{\kappa}_n = \rho_{0.5}(\tilde{\kappa}^k_n(\kappa)), \tag{35}$$

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

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

It is a finite sample functional relating the bias to κ. In the case where b_n is invertible, the II estimator and median unbiased estimator are given by:

$$\hat{\kappa}^{II}_n = b^{-1}_n(\hat{\kappa}_n). \tag{36}$$

Typically, the binding functions cannot be computed analytically in either case. That is why II/MUE needs to calculate the binding functions via simulations. While often used in the literature for the binding function is the mean, 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 of the mean-unbiasedness in $\hat{\kappa}^{II}_n$, only monotonicity is needed for $b_n(\kappa)$ to ensure the median-unbiasedness (Phillips and Yu 2009b).

[6] 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.

There are several advantages in the II/MUE procedure relative to the jackknife procedure. First, II is more effective on removing the bias in parameter estimates. Phillips and Yu (2009a) provided evidence to support this superiority of II. Second, the bias reduction may be achieved often without an increase in variance. In extreme cases of root near unity, the variance of II/MUE can be even smaller than that of ML (Phillips and Yu 2009a). To see this, note that (36) 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/MUE estimator has a smaller variance than MLE. Gouriéroux et al. (2000) discussed the relationship among II, MUE and bootstrap in the context of bias correction.

A disadvantage in the II/MUE procedure is the high computational cost. It is expected that with the continuing explosive growth in computing power, such a drawback is of less 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 before, since prices of contingent-claims are always non-linear transformations of the system parameters, 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 II/MUE estimates into the pricing formulae may still yield an estimate of the price with unsatisfactory finite sample performances. This feature was illustrated in a 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 (2009b) generalized the II/MUE 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/MUE 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 (16) 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.

Table 4 ML, II and median unbiased estimates of κ in the Vasicek model

	MLE	II	MUE
$\hat{\kappa}$	0.2613	0.1358	0.1642

3. Obtain $\tilde{\phi}_n^{ML,k}(p)$, the MLE of θ, from the kth 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.

6.2 *An Empirical Application*

This empirical application compares the ML method and the simulation-based methods for estimating the mean reversion parameter in a context of Vasicek term structure model. The dataset of a short term interest rate series involves the Federal fund rate and is available from the H-15 Federal Reserve Statistical Release. It is sampled monthly and has 432 observations covering the period from January 1963 to December 1998. The same data were used in Ait-Sahalia (1999) and are contained in a file named ff.txt.

Matlab code, simVasicek.m, is used to obtain the ML, II and median unbiased estimates of κ in the Vasiecek model. Table 4 reports these estimates. The ML estimate is about twice as large as the II estimate. The II estimate is similar to the median unbiased estimate.

7 Conclusions

Simulation-based estimation of financial time series model has been ongoing in the financial econometric literature and the empirical finance literature for more than one decade. Some new developments have been made and some existing methods have been refined with the increasing complexity in models. More and more attention have been paid to the simulation-based methods in recent years. Researchers in empirical finance have sought to use these methods in practical applications in an increasing scale. we expect the need for these methods to grow further as the financial industry continues to expand and data sets become richer.

Acknowledgements I gratefully acknowledge financial support from the Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS. Data and program code used in this paper can be download from my website at http://www.mysmu.edu/faculty/yujun/research.html.

References

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

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

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

Aït-Sahalia, Y., & Yu, J. (2006). Saddlepoint approximations for continuous- ime markov processes. *Journal of Econometrics*, *134*, 507–551.

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit Root models. *Econometrica*, *61*, 139–166.

Bansal, R., Gallant, A. R., Hussey, R. & Tauchen, G. (1995). Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics*, *66*, 251–287.

Bauwens, L., & Galli, F. (2008). Efficient importance sampling for ML estimation of SCD models. *Computational Statistics and Data Analysis*, *53*, 1974–1992.

Bauwens, L., & Hautsch, N. (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics*, *4*, 450–493.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*, 637–659.

Broto, C., & Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, *18*, 613–649.

Chib, S. (2001). Markov Chain Monte Carlo methods: Computation and inference. In J.J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3569–3649). Amsterdam: North-Holland

Chib, S., Nardari, F., & Shephard, N. (2002). Markov Chain Monte Carlo methods stochastic volatility models. *Journal of Econometrics*, *108*, 281–316.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics*, *64*, 375–400.

Duan, J. C., & Fulop, A. (2009). Estimating the structural credit risk model when equity prices are contaminated by trading noises. *Journal of Econometrics*, *150*, 288–296.

Duffie, D., & Singleton, K. J. (1993). Simulated moments estimation of markov models of asset prices. *Econometrica*, *61*, 929–952.

Duffie, D., & Stanton, R. Evidence on simulation inference for near unit-root processes with
AQ1 implications for term structure estimation. *Journal of Financial Econometrics*, *6*, 108–142.

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

Durbin, J., & Koopman, S. J. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives (with discussion). *Journal of the Royal Statistical Society Series B*, *62*, 3–56.

Durham, G. (2006). Monte carlo methods for estimating, moothing, and filtering one and two-factor stochastic volatility models. *Journal of Econometrics*, *133*, 273–305.

Durham, G. (0000). SV mixture models with application to S&P 500 index returns. *Journal of
AQ2 Financial Economics*, *85*, 822–856.

Durham, G., & Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time iffusion processes. *Journal of Business and Economic Statistics*, *20*, 297–316.

Efron, B. (1982). *The Jackknife, the Bootstrap and other resampling method*. Philadephia: SIAM

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.

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

Ferguson, T. (1973). A bayesian analysis of some nonparametric problems. *Annals of Statistics*, *1*, 209–230.

Gallant, A. R., & Long, L. R. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, *84*, 125–141.

Gallant, A. R., & Tauchen, G. (1996a). Which moments to match?. *Econometric Theory*, *12*, 657–681.

Gallant, A. R., & Tauchen, G. (1996b). Which moments to match?. *Econometric Theory*, *12*, 657–681.

Gallant, A. R., & Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica 57*, 1091–1120.

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

Gallant, A. R., & Tauchen, G. (2001a). EMM: A program for efficient method of moments estimation. User's Guide, Department of Economics, University of North Carolina.

Gallant, A. R., & Tauchen, G. (2001b). SNP: A program for nonparametric time series analysis. User's Guide, Department of Economics, University of North Carolina.

Gallant, A. R., & Tauchen, G. (2001c). Efficient method of moments. Working paper, Department of Economics, University of North Carolina.

Geweke, J. (1994). Bayesian comparison of econometric models. Working Paper, Federal Reserve Bank of Minneapolis, Minnesota.

Gordon, N. J., Salmond, D. J., & Smith, A. E. M. (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEEE-Proceedings F*, *140*, 107–133.

Gouriéroux, C., & Monfort, A. (1995). *Simulation based econometric methods*. London: Oxford University Press.

Gouriéroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, *8*, S85–S118.

Gouriéroux, C., Renault, E. & Touzi, N. (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). London: Cambridge University Press.

Gouriéroux, C., Phillips, P. C. B., & Yu, J. (2010). Indirect inference for dynamic panel models.
AQ3 *Journal of Econometrics*, forthcoming.

Hall, P. (1992). *The bootstrap and edgeworth expansion*. Berlin: Springer.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.

Harvey, A. C., & Shephard, N. (1996). The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, *14*, 429–434.

Harvey, A. C., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–264.

He, H. (1990). Moment Approximation and Estimation of Diffusion Models of Asset Prices, Working paper, University of California at Berkeley.

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.

Huang, S. J., & Yu, J. (2009). Bayesian analysis of structural credit risk models with microstructure
AQ4 noises. *Journal of Economic Dynamics and Control*, forthcoming.

Hull, J., & White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, *42*, 281–300.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics*, *12*, 371–417.

Jensen, M. J., & Maheu, J. M. (2008). Bayesian semiparametric stochastic volatility modeling, Working Paper No. 2008-15, Federal Reserve Bank of Atlanta.

Johannes, M., & Polson, N. (2009). MCMC methods for continuous time asset pricing models. In
AQ5 Aït-Sahalia & Hansen (Eds.), *Handbook of financial econometrics* forthcoming.

Jungbacker, B., & Koopman, S. J. (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*, *94*, 827–839.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Kitagawa, G. (1996). Monte Carlo filter and smoother for Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, *5*, 1–25.

Kleppe, T., Skaug, H., & Yu, J. (2009). Simulated Maximum Likelihood estimation of continuous time stochastic volatility models. Working paper, Singapore Management University.

Koopman, S. J., Shephard, N., & Creal, D. (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics*, *149*, 2–11.

Lee, B. S., & Ingram, B. F. (1991). Simulation estimation of time-series models. *Journal of Econometrics*, *47*, 197–205.

Lee, K. M., & Koopman, S. J. (2004). Estimating stochastic volatility models: A comparison of two importance samplers. *Studies in Nonlinear Dynamics and Econometrics*, *8*, 1–15.

Liesenfeld, R., & Richard, J. F. (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance*, *10*, 505–531.

Liesenfeld, R., & Richard, J. F. (2006). Classical and bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews*, *25*, 335–360.

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

MacKinnon, J. G., & Smith, A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, *85*, 205–230.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, *57*, 995–1026.

McLeish, D. (2005). *Monte Carlo simulation and finance*. NY: Wiley.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, *29*, 449–470.

Meyer, R., & Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal*, *3*, 198–215.

Meyer, R., Fournier, D. A., & Berg, A. (2003). Stochastic volatility: Bayesian computation using automatic differentiation and the extended Kalman filter. *Econometrics Journal*, *6*, 408–420.

Monfort, A. (1996). A reappraisal of misspecified econometric models. *Econometric Theory*, *12*, 597–619.

Musso, C., Oudjane, N., & Le Gland, F. (2001). Improving regularized particle filters, In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice* (pp. 247–271). New York: Springer.

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.

Omori, Y., Chib, S., Shephard, N., & Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, *140*, 425–449.

Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, *57*(5), 1027–1057.

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., & Yu, J. (2005a). Jackknifing bond option prices. *Review of Financial Studies*, *18*, 707–742.

Phillips, P. C. B., & Yu, J. (2005b). Comments: A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, *20*, 338–343.

Phillips, P. C. B., & Yu, J. (2009a). Maximum likelihood and gaussian estimation of continuous time models in finance. *Handbook of financial time series*, 497–530.

Phillips, P. C. B., & Yu, J. (2009b). Simulation-based estimation of contingent-claims prices. *Review of Financial Studies*, *22*, 3669–3705.

Pitt, M. (2002). Smooth particle filters likelihood evaluation and maximisation, Working Paper, University of Warwick.

Pitt, M., & Shephard, N. (1999a). Filtering via simulation: Auxiliary particle filter. *The Journal of the American Statistical Association*, *94*, 590–599.

Pitt, M., & Shephard, N. (1999b). Time varying covariances: A factor stochastic volatility approach. In: J. M. Bernardo, J. O. Berger, A. P. David, & A. F. M. Smith (Eds.), *Bayesian statistics 6* (pp. 547–570). Oxford: Oxford University Press.

Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies*, *11*, 449–487.

Richard, J. F., & Zhang, W. (2006). Efficient high-dimensional importance. *Journal of Econometrics*, *141*, 1385–1411.

Sandmann, G., & Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, *87*, 271–301.

Shephard, N., & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, *84*, 653–667.

Skaug, H., & Yu, J. (2007). Automated Likelihood Based Inference for Stochastic Volatility Models, Working Paper, Singapore Management University.

Stern, S. (1997). Simulation-based estimation. *Journal of Economic Literature*, *35*, 2006–2039.

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

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes – a study of the daily sugar prices 1961–75. In O. D. Anderson (Ed.), *Time series analysis: Theory and practice, 1* (pp. 203–226). Amsterdam: North-Holland.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*, 177–186.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, 127, 165–178.

Yu, J. (2009a). Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models, Working Paper, Singapore Management University.

Yu, J. (2009b). Semiparametric Stochastic Volatility Model, Working Paper, Singapore Management University.

Yu, J., & Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, *25*, 361–384.

Yu, J., Yang, Z., & Zhang, X. B. (2006). A class of nonlinear stochastic volatility models and its implications on pricing currency options, *Computational Statistics and Data Analysis*, *51*, 2218–2231.

Zehna, P., (1966). Invariance of maximum likelihood estimation. *Annals of Mathematical Statistics*, *37*, 744–744.