10-2010

Simulation-based Estimation Methods for Financial Time Series Models

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

Simulation-based Estimation Methods for Financial Time Series Models*

Prepared for *Handbook of Computational Finance*

Jun Yu†

February 16, 2009

Abstract

This paper overviews some recent advances on simulation-based methods of estimating time series models and asset pricing models that are widely used in finance. 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 method (MLE) and the generalized method of moments (GMM) are difficult to use. They can also be useful for improving the finite sample performance of the traditional methods when financial time series are highly persistent and when the quantity of interest is a highly nonlinear function of system parameters.

The simulation-based methods are classified in this paper, based on the frequentist/Bayesian split. Frequentist's simulation-based methods cover simulated generalized method of moments (SMM), efficient method of moments (EMM), indirect inference (II), various forms of simulated maximum likelihood methods (SMLE). Asymptotic properties of these methods are discussed and asymptotic efficiency is compared. 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. The list of discussed financial time series models cover continuous time diffusion models, latent variable models, term structure models, asset pricing models, and structural models for credit risk.

Finite sample problems of the exact maximum likelihood method, such as finite sample bias, are also discussed. Simulation-based bias correction methods, such as indirect inference, simulation-based median unbiased estimation, and bootstrap methods are reviewed. A nice property about these simulation-based bias correction methods is that they retains the good asymptotic properties of maximum likelihood estimation while reducing finite sample bias.

*Yu gratefully acknowledge financial support from the Ministry of Education AcRF fund under Grant No. T206B4301-RS.

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

Empirical applications, based on real exchange rates, interest rates and equity data, illustrate how to implement the simulation based methods. In particular, we apply EMM to estimate a continuous time stochastic volatility model, MCMC to a structural model for credit risk, SMLE to a discrete time stochastic volatility model, II method to the Black-Scholes option pricing model, median unbiased estimation method to a one-factor bond option pricing model. Computer code and data are provided.

Keywords: Generalized method of moments, Maximum likelihood, MCMC, Indirect Inference, Bootstrap, Median Unbiased, Option pricing, Credit risk, Stock price, Exchange rate, Interest rate.

1 Introduction

Relative to other fields in economics, financial economics has a relative shorter history. However, over the last half century, there has been an explosion of theoretical work in financial economics. At the same time, interestingly, more and more complex financial products and services have been created. The size of financial markets has exponentially increased and the quality of database is hugely advanced.

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 finance is to bring finance theory to data. Empirical problems in financial economics almost always involve calculating a likelihood function or solving a set of moment conditions.

Traditional estimation methods include maximum likelihood (ML), quasi-ML, generalized method of moments (GMM), and Bayesian. When the model is fully specified and the likelihood function has a tractable form, ML and Bayesian provide the full likelihood-based inference. Under mild regularity conditions, it is well recognized that ML estimators (MLE) are consistent, asymptotic normally distributed and asympotically 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). This feature greatly facilitate financial applications of ML. When the model is not fully specified and certain moments exist, GMM can be applied. Relative to ML, GMM may be less efficient but more robust.

Financial data are typically available in time series format. Consequently, financial time series methods are of critical importance to empirical research in finance. Traditionally, financial economists restrict themselves on a small class of time series models and a small set of financial assets, so that the setups are simple enough to permit analytical solutions for asset prices. The leading example is perhaps the geometric Brownian motion. It was used by Black and Scholes to price an European option price (Black and Scholes, 1973) and by Merton to price corporate bonds (Merton, 1974). In recent years, however, many alternative models and many financial assets have been proposed so that asset prices do not have analytical solutions. As a result, various numerical solutions have been proposed, one of which is simulation-based. Although this problem is sufficient important and the solutions merit a detailed review (see McLeish (2005) for a textbook treatment), it is beyond the scope of the present chapter.

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 statistical methods. For many important financial time series models, the likelihood function or the moment conditions cannot be evaluated analytically and is numerically formidable so that classical statistical methods, such as ML, GMM and Bayesian, are not feasible. For example, Heston derived a closed-form expression for the European option price under the square root specification of stochastic volatility (SV) (Heston,

1993), but ML Estimation of Heston's SV model is notoriously difficult. For more complicated time series models where asset prices do not have closed-form expressions, it is almost always the case that standard estimation methods are difficult to use.

Other than to asset pricing, parameter estimates in financial time series models are inevitable 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, alternative time series specifications are available in many cases. Consequently, it may be important to check the validity of a particular specification and to compare the performance of alternative specifications. Obviously, estimation of these alternative specifications is an important preliminary step to the specification analysis. For another example, to estimate the theoretical price of a contingent claim implies by the underlying time series model, one has estimate the parameters in the time series model and then plug the estimates into the pricing formula.

In some case where ML or GMM or Bayesian methods are feasible, when financial time series are highly persistent, classical estimators of parameters in the underlying time series models may have poor finite sample statistical properties, due to the presence of large finite sample bias. The bias in parameter estimation leads to a bias in other financial decision making. Because many financial variables, such as interest rates and volatility, are highly persistence, this finite sample problem is empirically relevant.

To overcome the difficulties in calculating likelihood and moments or improving the finite sample property of standard estimators, many simulation-based estimation methods have been proposed in recent years. Some of them are methodologically general in principle; some other are specially designed to deal with a particular model structure. In this chapter, we review some simulation-based estimation methods for 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) is an excellent review of the simulation-based estimation methods in the classical framework. Johannes and Polson (2009) reviews 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. Second, relative to Stern (1997) and Gouriéroux and Monfort (1995), more recently developed classical methods are discussed in the present paper.

We organize the review by collecting the methods into four categories: simulation-based ML, simulation-based GMM, Bayesian Markov chain Monte Carlo (MCMC) methods, and indirect inference methods. All these four methods all discussed in the context of a specific example. An empirical illustration is performed using real data in each case. Section 2 overviews the classical estimation methods, explains why sometimes the classical estimation methods may be difficult to use. Section 3 discusses discrete time stochastic volatility models, reviews various simulation-based ML methods, and illustrate the implementation of one particular simulation-based ML method in the context of a basic discrete time SV model. Section 4 discusses continuous time models,

review various simulation-based GMM, and illustrate the implementation of efficient method of moments (EMM) in the context of a continuous time SV model. Section 5 discusses structure credit risk models, review Bayesian MCMC methods, and illustrate the implementation of the Gibb-sampler MCMC method in the context of a simple credit risk model. Section 6 discusses continuous time models with a persistent, linear drift term, reviews indirect inference methods that have been proposed to improve the finite sample performances of traditional methods, illustrate the implementation of indirect inference methods to price a deep out-of-money option in the context of Black-Scholes model. Finally, Section 7 concludes.

2 Classical Estimation Methods

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

$$\ell(\theta) = f(\mathbf{X};\theta) = \int f(\mathbf{X},\mathbf{h};\theta)d\mathbf{h}, \tag{1}$$

where $\mathbf{X} = (X_1,\cdots,X_n) := (X_h,\cdots,X_{nh})$ is the data observed by econometricians,[2] h is the sampling interval, $f(\mathbf{X})$ the joint density of $\mathbf{X}$, $\mathbf{h}$ a vector of some latent variables, θ a set of K parameters one econometricians wish to estimate. As $X(t)$ are often annualized, 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. Unfortunately, when the integration in (7) is not analytically available and the the dimension in $\mathbf{h}$ is high, numerical evaluation of (7) would be difficult.

ML needs to maximize (7) over θ in a certain parameter space, i.e.,

$$\hat{\theta}_n^{ML} = \text{argmax}_{\theta\in\Theta}\ell(\theta) = f(\mathbf{X};\theta)$$

The first order condition of the maximization problem is

$$g(\mathbf{X};\theta) := \frac{\partial f}{\partial\theta} = 0.$$

Under mild regularity conditions, ML estimator (MLE) has desirable asymptotic properties of consistency, normality and efficiency. Moreover, due to the invariance property of maximum likelihood (e.g, Zehna, 1966), a nonlinear transformation of MLE is a MLE of the nonlinear transformation of the corresponding parameters. This property has proven very useful in asset pricing and credit risk analysis. Since $f(\mathbf{X};\theta)$ is difficult to calculate, ML is not easy to implement.

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

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

[1] Specific examples can be found below.

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

where $f(\theta)$ is the prior density and $f(\theta|\mathbf{X})$ the posterior distribution. In the case where $f(\mathbf{X};\theta)$ is difficult to calculate, the posterior density $f(\theta|\mathbf{X})$ is generally difficult to calculate.

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

$$Eg(\mathbf{X};\theta_0) = 0$$

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 g functions 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 distributional assumptions and hence GMM may be more robust than the full likelihood methods. Under mild regularity conditions, Hansen (1982) obtained the asymptotic distributions for GMM estimators. Unfortunately, in many financial time series models, the g function in the moment conditions do not have analytical expression and is difficult to calculate numerically, making GMM not easy to implement.

Even if ML is applicable, MLE is not necessarily the best estimator in finite samples. Phillips and Yu (2005a, 2005a, 2009a, 2009b) provided numerous examples to demonstrate the poor finite sample property. There are three reasons for this. First, many financial variables (such as interest rates and volatility) are very persistent. When linear time series models are fitted to these variables, the MLE (and many other estimators alike) of the system parameters typically involves substantial finite sample bias even in very large samples. For example when 600 weekly observations are used to estimate the square root model of CIR, ML estimates the mean reversion parameter with nearly 300% bias. Second, often financial applications involve non-linear transformation of estimators of the system parameters. Two examples are asset pricing and the evaluation of default probability. 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. An well known example is the MLE of deep out-of-money option. In general, the more pronounced the nonlinear, the worse the finite sample performance. 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. One way to improve the finite sample performance of classical

estimators is to obtain the bias and the remove the bias from those estimator, with the hope that the variance of bias-corrected estimator does not increase or only increase slightly so that the mean square error becomes smaller. Unfortunately, explicit analytic evaluation of the bias function is almost always infeasible in time series models.

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

3 Discrete time SV models

We first focus on the so-called basic lognormal (LN) SV model of Taylor (1982) which is defined by

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

where X_t is the return of an asset, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $\eta_t \overset{iid}{\sim} N(0,1)$, $corr(\epsilon_t, \eta_t) = 0$, and $h_1 \sim N(0, \gamma^2/(1-\phi^2))$. The parameters of interest are $\theta = (\sigma, \phi, \gamma)'$. This model is a powerful alternative to ARCH-type models (Geweke (1994) and Danielsson (1994)).

Let $\mathbf{X} = (X_1, \ldots, X_T)'$ and $\mathbf{h} = (h_1, \ldots, h_T)'$. The likelihood function of the basic LN-SV model is given by

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

To perform 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 generalized in various dimensions to accommodate stylized facts in financial data. Leverage, SV-t, super-position, jumps. Alternative to LN-SV is Heston.

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 finally by an empirical application.

3.1 Laplace importance sampler (LA-IS)

One of the widely received Monte Carlo methods is to first match the integrand with a multivariate normal distribution and second draw a sequence of independent variables from the multivariate

normal distribution. That is, the carefully selected multivariate normal density is served as an importance function. The technique in Stage 1 is known as the Laplace approximation while the technique in Stage 2 is known as the importance sampler. In this paper we call the method LA-IS.

To fix the idea, in Stage 1, we match $p(\mathbf{X}, \mathbf{h}; \theta)$ and a multivariate normal distribution for $\mathbf{h}$ as closely as possible so that the mean and co-variance in the multivariate normal are taken to be $\mathbf{h}^*$ and $-\Omega^{-1}$, respectively, where

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

and

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

Then the Laplace approximation to the integrand (3) is

$$p(\mathbf{X}, \mathbf{h}; \theta) \approx N(\mathbf{h}; \mathbf{h}^*, -\Omega^{-1}), \tag{6}$$

where $N(\cdot; \mu, \Sigma)$ represents the density of a (multivariate) normal distribution with mean of μ and co-variance of Σ.

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

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

$$p(\mathbf{X}; \theta) = \int p(\mathbf{X}, \mathbf{h}; \theta) d\mathbf{h} = \int \frac{p(\mathbf{X}, \mathbf{h}; \theta)}{N(\mathbf{h}; \mathbf{h}^*, -\Omega^{-1})} N(\mathbf{h}; \mathbf{h}^*, -\Omega^{-1}) d\mathbf{h}. \tag{7}$$

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

$$\frac{1}{S} \sum_{s=1}^{S} \frac{p(\mathbf{X}, \mathbf{h}^{(s)}; \theta)}{N(\mathbf{h}^{(s)}; \mathbf{h}^*, -\Omega^{-1})}. \tag{8}$$

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 (8) 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

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

See Koopman, Shephard and Creal (2009) for further discussions on the conditions and a test for the convergence.

The LA-IS method is quite general. The approximation error is determined by how closely the integrant can be matched by the multivariate normal distribution. The Laplace approximation has no approximation error when $p(\mathbf{X}, \mathbf{h}; \theta)$ is Gaussian in $\mathbf{h}$. However, when $p(\mathbf{X}, \mathbf{h}; \theta)$ is far away from Gaussian in $\mathbf{h}$, it is expected the performance of Laplace approximation is not good.

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

$$p(\mathbf{X}, \mathbf{h}; \theta) = N\left(h_1, 0, \frac{\gamma^2}{1-\phi^2}\right) \prod_{t=2}^{T} N\left(h_t, \phi h_{t-1}, \gamma^2\right) \prod_{t=1}^{T} N\left(X_t, 0, \sigma^2 e^{h_t}\right), \tag{9}$$

and hence

$$\ln p(\mathbf{X}, \mathbf{h}; \theta) = \ln N\left(h_1, 0, \frac{\gamma^2}{1-\phi^2}\right) + \sum_{t=2}^{T} \ln N\left(h_t, \phi h_{t-1}, \gamma^2\right) + \sum_{t=1}^{T} \ln N\left(X_t, 0, \sigma^2 e^{h_t}\right). \tag{10}$$

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, denoted as G,

$$G = \begin{pmatrix} \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_1} \\ \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_2} \\ \vdots \\ \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_{T-1}} \\ \frac{\partial \ln p(\mathbf{X},\mathbf{h};\theta)}{\partial h_T} \end{pmatrix} = \begin{pmatrix} \frac{\phi h_2 - h_1}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_1^2 \\ \frac{\phi h_3 - \phi^2 h_2 + \phi h_1}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_2^2 \\ \vdots \\ \frac{\phi h_T - \phi^2 h_{T-1} + \phi h_{T-2}}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_{T-1}^2 \\ \frac{h_T - \phi h_{T-1}}{\gamma^2} - \frac{1}{2} + \frac{1}{2}\epsilon_T^2 \end{pmatrix} \tag{11}$$

Furthermore, the Hessian matrix of the log-integrand, denoted as Ω, is

$$\Omega = \begin{pmatrix} -\frac{1}{\gamma^2} - \frac{1}{2}\epsilon_1^2 & \frac{\phi}{\gamma^2} & \cdots & 0 & 0 \\ \frac{\phi}{\gamma^2} & -\frac{1+\phi^2}{\gamma^2} - \frac{1}{2}\epsilon_2^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -\frac{1+\phi^2}{\gamma^2} - \frac{1}{2}\epsilon_{T-1}^2 & \frac{\phi}{\gamma^2} \\ 0 & 0 & \cdots & \frac{\phi}{\gamma^2} & -\frac{1}{\gamma^2} - \frac{1}{2}\epsilon_T^2 \end{pmatrix} \tag{12}$$

where $\epsilon_t = X_t \exp(-0.5h_t)/\sigma$.

A closed related simulated ML method is proposed by Durbin and Koopman (1997). It is designed to evaluate the log-likelihood function of a linear state-space model with non-Gaussian errors. The basic idea of Durbin and Koopman (1997) is to use Laplace approximation to decompose the likelihood into the likelihood of a linear state-space model with Gaussian errors and the residual. It is known that the likelihood function of a linear state-space model with Gaussian errors can be calculated by applying Kalman filter. The residual is calculated by simulation. Sandmann and Koopman (1998) applied the method to estimate the LN-SV model and the SV-t model. To obtain the linear state-space form for the LN-SV model, one can apply the log-squared transformation to X_t:

$$\left\{ \begin{array}{lcl} Y_t & = & \ln X_t^2 = \ln sigma^2 + h_t + \varepsilon_t,\ t = 1, \ldots, T, \\ h_{t+1} & = & \phi h_t + \gamma \eta_t,\ t = 1, \ldots, T-1, \end{array} \right. \tag{13}$$

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 $h_1 \sim N(0, \gamma^2/(1-\phi^2))$.

3.2 Efficient importance sampler (EIS)

Richard and Zhang (2007) developed an alternative simulated ML method. It is based on a particular factorization of the importance density. It is also based on importance sampling and termed as Efficient Importance Sampling (EIS) procedure. Relative to the ML method reviewed in Section 2.1, EIS minimizes the Monte Carlo sampling variance of the approximation to the integrand by factorizing the importance density. To fix the idea, assume $g(\mathbf{h}|\mathbf{X})$ is the importance density which can be constructed as

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

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 $p(\mathbf{X}, \mathbf{h}; \theta)$ and $g(\mathbf{h}|\mathbf{X})$ which, as we shown in the last section, requires a high-dimensional non-linear regression. The caveat of EIS is to match each component in $g(\mathbf{h}|\mathbf{X})$ (ie $C_t e^{c_t h_t + d_t h_t^2} p(h_t|h_{t-1})$), to the corresponding element in the integrand $p(\mathbf{X}; \mathbf{h})$ (ie $p(X_t|h_t)p(h_t|h_{t-1})$) in a backward manner, with $t = T, T-1, \cdots, 1$. 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 $h_t^{(i)}$ and $h_t^{(i)}$ are treated as the explanatory variables in the regression model with $C_{T+1} = 1$.

The method to approximate the likelihood involves the following procedures:

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

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

The EIS algorithm relies on the user to provide a problem-dependent auxiliary class of importance samplers. Does not need the assumption that the latent process is Gaussian.

3.3 An empirical example

For the purposes of illustration, we fit the LN-SV model to a widely used dataset. 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 Shephard and Pitt (1997).

Matlab code (lais-lnsv.m) is used to implement the LA-IS method. Table 1 reports the estimates and the likelihood when $S = 32$. The code are implemented with 20 different random seeds to obtain the simulation standard errors.

4 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, $\sigma(X(t);\theta)$ is some specified diffusion function, $\mu(X(t);\theta)$ is a given drift function, and θ is a vector of unknown parameters. One wishes to obtain estimates of θ from a discrete sampled observations, $X_h, ..., X_{nh}$ with h being the sampling interval. This class of parametric model 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 a 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

$$\ell_{TD}(\theta) = \ln(p(X_{ih}|X_{(i-1)h},\theta)). \tag{17}$$

To perform exact ML estimation, one needs a closed form expression for $\ell_{TD}(\theta)$ 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 is approach proposed by Lo (1988).

Unfortunately, only in rare cases, do the transition density $p(X_{ih}|X_{(i-1)h},\theta)$ have closed form solutions. Phillips and Yu (2009) provide a list of examples in which $\ln(p(X_{ih}|X_{(i-1)h},\theta))$ have closed form analytical expressions, including 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.

An older estimation method is using 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 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 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}.

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 closer to 0 but the required integration becomes high dimensional and hence simulation-based methods can be used, leading to simulated ML estimators. To fix ideas, suppose $M-1$ auxiliary points are introduced between $(i-1)h$ and ih, i.e.,

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

The Markov property implies that

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

The idea behind the simulated ML method is to approximate the densities $p(X_{\tau_m}|X_{\tau_{m-1}};\theta)$ (step 1) and then evaluate the multidimensional integral using importance sampling techniques (step 2). Among the class of simulated ML methods that have been suggested in this context, Pedersen (1995) is one of the earliest contributions.

Pedersen suggested approximating the latent transition densities $p(X_{\tau_m}|X_{\tau_{m-1}};\theta)$ based on the Euler scheme and approximating 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 $(\epsilon_1, \epsilon_2, \cdots, \epsilon_{M-1})$ is a multivariate standard normal. implementations have been computationally burdensome.

As noted in Durham and Gallant (2002), there are two sources of approximation error in Pedersen's method. One is the (albeit reduced) discretization bias in the Euler scheme. The second is due to the Monte Carlo integration. These two errors can be further reduced by increasing the number of latent infill points and the number of simulated paths, respectively. However, the corresponding computational cost will inevitably be higher.

[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). Aït-Sahalia (2002) provided evidence that the closed-form approximation based on Hermite polynomials is more accurate and faster than the simulation-based approaches review here.

In order to reduce the discretization bias in step 1, Elerian (1998) suggested replacing the Euler scheme with the Milstein scheme while Durham and Gallant advocated using a variance stablization transformation, i.e., applying the Lamperti transform to the continuous time model. Certainly, any method that reduces the discretization bias can 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 $(\epsilon_1, \epsilon_2, \cdots, \epsilon_{M-1})$ 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) proposed a more efficient importance function which is based on the following tied-down process:

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

In particular, they proposed using the Laplace approximation (c.f., Phillips, 1984; Tierney and Kadane, 1986) to the tied-down process. That is, they used the distributional approximation $(X_{\tau_1}, \cdots, X_{\tau_{M-1}}) \sim N(\mathbf{x}^*, \Sigma^*)$ where

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \ln p(X_{\tau_1}, \cdots, X_{\tau_{M-1}} | X_{\tau_0}, X_{\tau_M})$$

$$\Sigma^2 = -\left[\frac{\partial^2 \ln p(X^*_{\tau_1}, \cdots, X^*_{\tau_{M-1}} | X_{\tau_0}, X_{\tau_M})}{\partial \mathbf{x}' \partial \mathbf{x}}\right]^{-1},$$

where $\mathbf{x} = (X_{\tau_1}, \cdots, X_{\tau_{M-1}})'$.

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

Not only is the likelihood function for (16) difficult to construct, but also the moment conditions. Prasao, Rao, Sorensen, Hansen and Scheikman, But if $X(t)$ is multivariate with some latent variables, moment conditions are not known in closed-form in general. $S(t)$ is a stock price, which is

assumed to follow the stochastic volatility (SV) model [Hull and White (1987)]:

$$dS(t) = \sigma_S S(t)\sigma(t)dB_1(t),$$
$$d\ln\sigma^2(t) = -\kappa\ln\sigma^2(t)dt + \gamma dB_2(t),$$

and $\{S_t\}_{t=1}^n$ is again a sample of equispaced time series observations on $S(t)$ with sampling interval h. Under certain assumptions, Hull and White (1987) showed that the value of a European call option is the Black-Scholes price integrated over the distribution of the mean volatility. Unfortunately, the option price does not have a closed-form solution. A flexible way for calculating option prices is via Monte Carlo simulations. For example, Hull and White (1987) designed an efficient procedure of carrying out the Monte Carlo simulation to calculate a European call option. In general, the price depends on κ, σ_S, and γ. For the SV model, it is well known that the likelihood function has no closed-form expression [Durham and Gallant (2002), and Kim, Shephard, and Chib (1998)].

While (16) is difficult to estimate, it is amenable to data simulation. 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 GMM estimation.

Simulated GMM methods have been proposed by McFadden (1989), Pakes and Pollard (1989) for iid environments, and Lee and Ingram, Duffie and Singleton (1993) for time series environments.

Let $\{\tilde{\mathbf{X}}_t^{(s)}(\theta)\}_{t=1}^{\mathfrak{T}(T)}$ be the data simulated from (16) when parameter is θ using random seed s. Therefore, $\{\tilde{X}_t^{(s)}(\theta_0)\}$ 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 sample moments of the data and those of the simulated data. Assuming H represents K-moments, SGMM estimator is defined as:

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

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, $\mathfrak{T}(T)$ is the number of number of observation in a simulated path. Under the ergodicity condition, we have

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

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 by 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 less likelihood. Gallant and Tauchen (1996) extended the SGMM technique so that the estimator is asymptotically as efficient as SML. This estimator is termed efficient method of moments (EMM), which we review below.

4.3 EMM

EMM is first introduced by Gallant and Tauchen (1996) and has now found many applications in financial time series; see Gallant and Tauchen (2001a) for a brief review of the literature. It is closely related to SGMM. An important difference between them is that while 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 EMM employs is the expectation of the score of an auxiliary model which is often referred to as the score generator.

Let the SV model of interest be the structural model. This is the continuous time version of the Box-Cox SV model of Yu, Yang and Zhang (2006).

$$dS(t) = \sigma_S S(t)[g(h_t, \delta)]^{1/2} dB_1(t),$$
$$dh(t) = -\kappa h(t)dt + \gamma dB_2(t),$$

The conditional density of the structural model is defined by

$$p_t(x_t|y_t, \theta),$$

where the true value of θ is θ_0 and $\theta_0 \in \Theta \subset \Re^{\ell_\theta}$ with ℓ_θ being the length of θ_0. Denote the conditional density of an auxiliary model by

$$f_t(x_t|y_t, \beta), \beta \in R \subset \Re^{\ell_\beta}$$

where y_t is a vector of lagged x_t. 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 closed form. 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

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

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

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

For specification testing, we have

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

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 (1996) 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) and 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 Bansal et al (1995).

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 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 Gallant and Tauchen (2001).

Fortran code (lais-lnsv.m) is used to implement the EMM method. Table 2 reports the estimates.

5 Stochastic duration models

6 Credit risk models

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, F the face value of a zero-coupon debt that the firm issues with the time to maturity T. Merton (1974) assumed that V_t evolves according to a geometric Brownian motion:

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

where $W(t)$ is a standard Brownian motion which is the driving force of the uncertainty in V_t, and c is a constant. The exact discrete time model is

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

where $\epsilon_t \sim N(0,1)$ and h is the sampling interval. Obviously there is a unit root in $\ln V_t$.

The firm is assumed to have two types of outstanding claims, namely, 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 (ie $V_T < F$). Since V_t is assumed to be a log-normal diffusion, 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. Similarly, one can derive pricing formulae for the corporate bond (Merton, 1974) and spreads of credit default swaps, although these formulae will not be used in this paper.

Assuming the risk-free interest rate is r, 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

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

When the firm is listed in an exchange, one may assume that S_t is observed at discrete time points, say $t = \tau_1, \cdots, \tau_n$. When there is no confusion, we simply write $t = 1, \cdots, n$. Since the joint density of $\{V_t\}$ is specified by (22), the joint density of $\{S_t\}$ can be obtained from Equation (23) by the change-of-variable technique. As S is analytically available, the Jacobian can be obtained, facilitating the ML estimation of θ (Duan, 1994).

The above approach requires the equilibrium equity prices be observable. This assumption appears to be too strong when data are sampled at a reasonably high frequency because the presence of various market microstructure effects contaminate the equilibrium price process. The presence of market microstructure noises motivates Duan and Fulop (2008) to consider the following generalization to Merton's model:

$$\ln S_t = \ln S(V_t;\sigma) + \delta v_t, \tag{24}$$

where $\{v_t\}$ is a sequence of iid standard normal variates. Equation (22) and Equation (24) form the basic credit risk model with microstructure noises which was studied by Duan and Fulop (2008).

One of the most compelling reasons for obtaining the estimates for the model parameters and the latent equity values is for credit applications. For example, Moody's KMV Corporation has successfully developed a structural model, by combining financial statement and equity market-based information, to evaluate private firm credit risk. Another practically important quantity is the credit spread of a risk corporate bond over the corresponding Treasure rate.

Using the notations of Duan and Fulop (2008), the credit spread is given by

$$C(V_n;\theta) = -\frac{1}{T-\tau_n}\ln\left(\frac{V_n}{F}\Phi(-d_{1n}) + e^{-r(T-\tau_n)}\Phi(d_{2n})\right) - r, \tag{25}$$

where the expressions for d_{1n} and d_{2n} were given in Section 2. The default probability is given by

$$P(V_n;\theta) = \Phi\left(\frac{\ln(F/V_n) - (\mu-\sigma^2/2)(T-\tau_n)}{\sigma\sqrt{T-\tau_n}}\right). \tag{26}$$

Putting the model in a state-space framework, Equation (24) is an observation equation and Equation (22) is a state equation. Unfortunately, the Kalman filter is not applicable here since the observation equation is nonlinear.

Let $\mathbf{X} = (\ln S_1, \cdots, \ln S_n)'$, $\mathbf{h} = (\ln V_1, \cdots, \ln V_n)'$, and $\theta = (\mu, \sigma, \delta)'$. The likelihood function of (24) is given by

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

where $p(\cdot)$ means the probability density function. 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 $\ln V_t$.

6.1 SML via particle filter

The transition density is difficult to evaluate.

particle filtering which is applicable to a broad class of nonlinear non-Gaussian multi-dimensional state space models of the form,

$$\begin{cases} y_t &= H(x_t, u_t) \\ x_t &= F(x_{t-1}, v_t), \end{cases} \tag{28}$$

where x_t is a k-dimensional state vector (here, $x_t = h_t$ is the one-dimensional log-volatility), v_t is a l-dimensional white noise sequence with density $q(v)$, u_t is a one dimensional white noise sequence with density $r(u)$ and assumed uncorrelated with $\{v_s\}_{s=1}^n$, H and F are possibly nonlinear functions. Let $u_t = G(y_t, x_t)$ and G' is the derivative of G as a function of y_t. The density of the initial state vector is assumed to be $p_0(x)$. We now summarize all the steps involved in Kitagawa's algorithm:

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 $q(v)$, $v_t^{(j)}$ for $j = 1, \ldots, M$.
   (b) Compute $p_t^{(j)} = F(f_{t-1}^{(j)}, v_t^{(j)})$ 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 $\{r(G(y_t, p_t^{(j)})) \times |G'(y_t, p_t^{(j)})|\}_{j=1}^M$.

To estimate the model via ML, built upon the work of Pitt and Shephard (1999) and Pitt (2002), Duan and Fulop developed a *particle filtering* method. Particle filter is an alternative to the Extended Kalman filter (EKF) with the advantage that, with sufficient samples, it approaches the true ML estimate. Hence, it can be made more accurate than the EKF. Like the idea in many other simulation based methods, particle filter essentially approximates the target distribution by the corresponding empirical distribution, based on a weighted set of particles. To avoid the

variance of importance weight to grow over time, it is important to perform the resampling step in the particle filtering.

6.2 Bayesian MCMC Methods

The two algorithms discussed in Section **??** are the classical ML methods. We now discuss how the Laplace approximation can be employed to perform a Bayesian MCMC analysis. The goal of MCMC methods is to sample from posterior densities. There are a number of different ways in which Bayesian MCMC can be applied to SV models. The first one samples from the posterior $p(\theta|\mathbf{X}) \propto p(\mathbf{X}|\theta)p(\theta)$. However, since the marginal likelihood $p(\mathbf{X}|\theta)$, given in terms of an integral, does not have a closed form expression, standard Bayesian analysis is not trivial. Alternatively, one can augment the parameter vector by the vector of latent variables and obtain the joint posterior $p(\theta, \mathbf{h}|\mathbf{X}) \propto p(\mathbf{X}, \theta|\mathbf{h})p(\mathbf{h}|\theta)p(\theta)$. Consequently, evaluation of $p(\mathbf{X}|\theta)$ becomes unnecessary. From the joint posterior, one can find the marginal distribution $p(\theta|\mathbf{X})$ to make inference about the model parameters and the marginal distribution $p(\mathbf{h}|\mathbf{X})$ to make inference about the log-volatilities. MCMC can be used to draw (correlated) samples from the high dimensional $(T+p)$ posterior density.

In the basic SV model, one way of sampling σ_X, ϕ, σ and $\mathbf{h}$ is to update each element of σ_X, ϕ, σ and each element in $\mathbf{h}$ one at a time (i.e. a single-mover). This so-called Gibbs sampling algorithm was suggested by Shephard (1993) and Jacquier et al. (1994). For SV models, the consecutive states are often highly dependent, rendering inefficient mixing and slow convergence of the Markov chain to the equilibrium distribution. To improve the simulation efficiency, Shephard and Pitt (1997), Kim et al. (1998) and Liesenfeld and Richard (2006) suggested MCMC methods which sample the vector $\mathbf{h}$ in a single block (i.e. a multi-mover).

In the present paper, we suggest an alternative Bayesian MCMC algorithm which achieves high simulation efficiency. The idea is as follows. First, an approximation to the likelihood function $p(\mathbf{X}|\theta)$ is obtained via the Laplace approximation (6), making augmentation of the parameter vector redundant. Second, we use the MH algorithm, developed by Metropolis et al. (1953) and Hastings (1970), to obtain a MCMC sample from the posterior of θ. To use the MH algorithm we first fit the model by maximizing the approximated marginal likelihood (6). Denote by $\hat{\theta}$ the resulting estimate of θ, and by $\hat{\Sigma}_{\hat{\theta}}$ its covariance matrix based on the observed Fisher information. The MH-proposal density is taken to be a multivariate normal, centered at the current parameter value, with covariance matrix $\hat{\Sigma}_{\hat{\theta}}$. Note further that for each value of θ proposed by the MH-algorithm the Laplace approximation is invoked via equation (6). While our algorithm is strongly related to that developed in Meyer et al. (2003), it differs in that Meyer et al. used a sequential Laplace approximation.

As the posterior sampling is done in a much lower dimensional space in the proposed integration sampler, it is natural to expect several advantages of our method relative to the MCMC algorithms that require sampling of the latent process. First, the integration sampler should have higher simulation efficiency. Second, while it is typically difficult to check the convergence of simulated

chains of latent variables as the sample size T grows, it is trivial to do so in our integration sampler, regardless of the sample size.

A major advantage of the MCMC algorithms that sample the latent process is that they provide an integrated framework for parameter estimation and latent variable smoothing. Indeed, the smoothed latent variable $E(h_t|\mathbf{X})$, and the associated posterior variance, is a by-product of the MCMC output (Jacquier et al. 1994). There are two sources of variation contributing to the posterior uncertainty in $\mathbf{h}$: uncertainty in $\mathbf{h}$ given θ and uncertainty about θ. It is worth pointing out that MCMC, by providing samples from $p(\mathbf{h}|\mathbf{X})$, directly accounts for both sources of uncertainty.

6.3 An empirical application

7 Simulation Methods for Improving Finite Sample Performances

7.1 indirect inference

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

A carefully designed II estimator can also have good small sample properties of parameter estimates, as shown by MacKinnon and Smith (1996), Monfort (1996), Gouriéroux, Renault, Touzi (2000) in the time series context and by Gouriéroux, Phillips and Yu (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 for parameter estimation, consider the OU process. Suppose we need to estimate the parameter κ in:

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

from observations $\{X_h, \cdots, X_{nh}\}$. An initial estimator of κ can be obtained, for example, by applying the Euler scheme to $\{X_h, \cdots, X_{nh}\}$ (call it $\hat{\kappa}_n$). Such an estimator is involved with the

discretization bias (due to the use of the Euler scheme) and also with a finite sample estimation bias (due to the poor finite sample property of ML in the near-unit-root situation).

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

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

where

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

This sequence may be regarded as a nearly exact simulation from the continuous time OU model for small δ. We then choose every $(h/\delta)^{th}$ observation to form the sequence of $\{\tilde{X}^k_{ih}\}_{i=1}^n$, which can be regarded as data simulated directly from the OU model with the (observationally relevant) step size h.[4]

Let $\{\tilde{X}^k_h, \cdots, \tilde{X}^k_{nh}\}$ be data simulated from the true model, where $k = 1, \cdots, K$ with K being the number of simulated paths. It should be emphasized that it is important to 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 estimation is to match the parameter obtained from the actual data with that obtained from the simulated data. In particular, the II estimator of κ solves

$$\hat{\kappa}_n = \frac{1}{K}\sum_{h=1}^{K} \tilde{\kappa}^k_n(\kappa) \text{ or } \hat{\kappa}_n = \hat{\rho}_{0.5}(\tilde{\kappa}^k_n(\kappa)), \tag{29}$$

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

$$\hat{\kappa}_n = E(\tilde{\kappa}^k_n(\kappa)) \text{ or } \hat{\kappa}_n = \rho_{0.5}(\tilde{\kappa}^k_n(\kappa)) \tag{30}$$

where $E(\tilde{\kappa}^k_n(\kappa))$ is called the mean binding function or the mean bias function, and $\rho_{0.5}(\tilde{\kappa}^k_n(\kappa))$ is the median binding function or the median bias 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 indirect inference estimator is given by:

$$\hat{\kappa}^{II}_n = b_n^{-1}(\hat{\kappa}_n). \tag{31}$$

[4]If the transition density of $X_{t+h}|X_t$ for the continuous time model is analytically available, exact simulation can be directly obtained. In this case, the Euler scheme at a finer grid is not necessary.

When a median binding function is used, the estimator is the median unbiased estimator of Andrews (1993). Typically, the binding functions cannot be computed analytically in either case. That is why II needs to calculate the binding functions via simulations. While 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 $\hat{\kappa}_n^{II}$, only monotonicity is needed for $b_n(\kappa)$ to ensure the median-unbiasedness (Phillips and Yu (2009d)).

There are several advantages in the II procedure relative to the jackknife procedure. First, indirect inference is more effective on removing the bias in parameter estimates. Phillips and Yu (2009a) provided evidence to support this superiority of indirect inference. Second, the bias reduction may be achieved often without an increase in variance. In extreme cases of root near unity, the variance of II can be even smaller than that of ML (Phillips and Yu (2009a). To see this, note that equation (31) implies:

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

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

A disadvantage in the II procedure is the high computational cost. It is expected that with the continuing explosive growth in computing power, such a drawback 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 indirect inference estimates into the pricing formulae may still yield an estimate of the price with unsatisfactory finite sample performances. This feature was illustrated in 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 (2009d) generalized the II procedure so that it is applied to the quantity of interest directly.

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

1. Given a value for the contingent-claim price p, compute $P^{-1}(p)$ (call it $\theta(p)$), where $P^{-1}(\cdot)$ is the inverse of the pricing function $P(\theta)$.

2. Let $\widetilde{\mathbf{S}}^k(p) = \{\tilde{S}_1^k, \tilde{S}_2^k, \cdots, \tilde{S}_T^k\}$ be data simulated from the time series model (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.

3. Obtain $\tilde{\phi}_n^{ML,k}(p)$, the MLE of θ, from the k'th simulated path, and calculate $\widetilde{P}_n^{ML,k}(p) = P(\tilde{\phi}_n^{ML,k}(p))$.

4. Choose p so that the average behavior of $\widetilde{P}_n^{ML,k}(p)$ is matched with $\widehat{P}_n^{ML}$ to produce a new bias corrected estimate.

The procedure can be generalized to cases where θ is an $K-$dimensional vector and where θ is obtained from cross-sectional data; see Phillips and Yu (2009d) for detailed discussions. Phillips and Yu (2009d) performed extensive Monte Carlo studies, showing that the proposed procedure works well, not only relative to ML but also relative to the jackknife procedure.

7.2 An empirical application

8 Conclusions

References

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

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

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

[4] Black, F. and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, 637-659.

[5] Duan, J.-C. 1994, Maximum likelihood estimation using price data of the derivative con- tract, *Mathematical Finance*, 4, 155-167.

[6] Duan, J.C. and A. Fulop, 2009, Estimating the Structural Credit Risk Model When Equity Prices Are Contaminated by Trading Noises, *Journal of Econometrics*, forthcoming.

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

[8] Durbin, J. and S.J. Koopman 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.

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

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

[11] Gouriéroux, C., A. Monfort, 1995, Simulation Based Econometric Methods, Oxford University Press.

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

[13] Hastings, W.K. 1970 Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57, 97–109.

[14] 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.

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

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

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

[18] Liesenfeld, R. and J.F. Richard 2003. Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics. *Journal of Empirical Finance* 10, 505–531.

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

[20] McLeish, D. 2005, *Monte Carlo Simulation and Finance*, Wiley

[21] Merton, R.C. 1974, On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance*, 29, 449-470.

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

[23] Meyer, R., and J. Yu 2000. BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal* 3, 198–215.

[24] 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.

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

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

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

[28] Pitt, M. 2002, Smooth particle filters likelihood evaluation and maximisation, Working Paper, University of Warwick.

[29] Pitt, M., and N. Shephard 1999, Filtering via simulation: Auxiliary particle filter, *The Journal of the American Statistical Association*, 94, 590–599.

[30] Richard, J.F. and W. Zhang 2006, Efficient High-Dimensional Importance, *Journal of Econometrics*, 141, 1385-1411.

[31] Sandmann, G., and Koopman, S. J. 1998, Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics*, 87, 271–301.

[32] Shephard, N. and M.K. Pitt 1997. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–667.

[33] Skaug, H. and J. Yu, 2007, Automated Likelihood Based Inference for Stochastic Volatility Models, Working Paper, Singapore Management University.

[34] Stern, S. 1997, Simulation-Based Estimation, *Journal of Economic Literature*, 35, 2006-2039

[35] Taylor, S. J. 1982, Financial returns modelled by the product of two stochastic processes — a study of the daily sugar prices 1961-75, in *Time Series Analysis: Theory and Practice, 1*, ed. Anderson, O. D., Amsterdam, North-Holland, 203–226.

[36] Zehna, P., 1966, "Invariance of Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 37, 744-744.