Journal of Economic Dynamics & Control 34 (2010) 2259–2272

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Bayesian analysis of structural credit risk models with microstructure noises

Shirley J. Huang [a], Jun Yu [b,*]

[a] Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899, Singapore

[b] School of Economics and Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Available online 15 May 2010

JEL classification:
C11
G12
G32

Keywords:
MCMC
Credit risk
Microstructure noise
Structural models
Deviance information criterion

ABSTRACT

In this paper a Markov chain Monte Carlo (MCMC) technique is developed for the Bayesian analysis of structural credit risk models with microstructure noises. The technique is based on the general Bayesian approach with posterior computations performed by Gibbs sampling. Simulations from the Markov chain, whose stationary distribution converges to the posterior distribution, enable exact finite sample inferences of model parameters. The exact inferences can easily be extended to latent state variables and any nonlinear transformation of state variables and parameters, facilitating practical credit risk applications. In addition, the comparison of alternative models can be based on deviance information criterion (DIC) which is straightforwardly obtained from the MCMC output. The method is implemented on the basic structural credit risk model with pure microstructure noises and some more general specifications using daily equity data from US and emerging markets. We find empirical evidence that microstructure noises are positively correlated with the firm values in emerging markets.

1. Introduction

Credit risk is referred to as the risk of loss when a debtor does not fulfill its debt contract and is of natural interest to practitioners in the financial industry as well as to regulators. For example, it is common practice that banks use securitization to transfer credit risk from bank's balance sheets to the market. The credit problem can well become a crisis when some of the risk lands back on banks. The turbulence in international credit markets and stock markets at the end of 2007 has largely been caused by this subprime credit problem in the US. To a certain degree, the 1997 Asian financial crisis was also caused by this credit risk problem. Not surprisingly, how to the credit risk is assessed is essential for risk management and for the supervisory evaluation of the vulnerability of lender institutions. Indeed, the Basel Committee on Banking Supervision has decided to introduce a new capital adequacy framework which encourages the active involvement of banks in measuring the likelihood of defaults. The growing need for the accurate assessment of credit risk motivates academicians and practitioners to introduce theoretical models for credit risk.

A widely used approach to credit risk modelling in practice and also in the academic arena is the so-called structural method. This method of credit risk assessment was first introduced by Black and Scholes (1973) and Merton (1974). In this approach the dynamic behavior of the value of a firm's assets is specified. If the value becomes lower than a threshold which is usually a proportion of the firm's debt value, the company is considered to be in default. For example, in Black and

* Corresponding author.
E-mail addresses: shirleyhuang@smu.edu.sg (S.J. Huang), yujun@smu.edu.sg (J. Yu).

0165-1889/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jedc.2010.05.008

Scholes (1973) and Merton (1974), a simple diffusion process is assumed for a firm's asset value, and the firm will default if its asset value is lower than its debt on the maturity date of the debt.

Since the firm's asset value is not directly observed by econometricians, the econometric estimation of structural credit risk models is nontrivial. To deal with the problem of unobservability, Duan (1994) introduces a transformed data maximum likelihood (ML) method, using observed time series data on publicly traded equity values. The idea essentially is to use the change-of-variable technique via the Jacobian, relying critically on the one-to-one correspondence between the traded equity value and the unobserved firm's asset value. Since then, this method has been applied in a number of studies; see for example, Wong and Choi (2006), Ericsson and Reneby (2004) and Duan et al. (2003). Duan et al. (2004) showed that the method is equivalent to the Moody's KMV model, a popular commercial product.

It is well known in the market microstructure literature that the presence of various market microstructure effects (such as price discreteness, infrequent trading and bid–ask bounce effects) contaminates the efficient price process with noises. There have been extensive studies on analyzing the time series properties of microstructure noises. Some earlier contributions include Roll (1984) and Hasbrouck (1993). In recent years, various specifications have been suggested for modelling microstructure noise in ultra-high frequency data in the context of measuring daily integrated volatility. Examples include the pure noise (i.e. iid) model (Zhang et al., 2005; Bandi and Russell, 2008), stationary models (Aït-Sahalia et al., 2009; Hansen and Lunde, 2006) and locally nonstationary models (Phillips and Yu, 2006, 2007). The consensus emerging from the literature is that if the microstructure noise were ignored, one would get an inconsistent estimate of the quantity of interest. This implication is also confirmed in Duan and Fulop (2009) in the context of credit risk modelling.

However, if the observed equity prices are contaminated with microstructure noises in structure credit risk models, the one-to-one correspondence between the traded equity value and the unobserved firm's asset value is broken, and hence the method developed in Duan (1994) is not applicable anymore. A fundamental difficulty is that neither the efficient prices nor microstructure noises are observable. As a result, the change-of-variable technique becomes infeasible. In an important contribution, Duan and Fulop (2009) developed a simulation-based ML method to estimate the Merton model with Gaussian iid microstructure noises. The ML method is designed to deal with nonlinear non-Gaussian state space models via particle filtering. In the credit risk model with microstructure noises, the nonlinear relationship between the contaminated traded equity value and firm's asset value is given by the option pricing model but is perturbed by microstructure noises. This gives the observation equation. The state equation specifies the dynamics of the asset value in continuous time, usually with a unit root.

The standard asymptotic theory for the ML estimator, such as asymptotic normality and asymptotic efficiency, is then called upon to make statistical inferences about the model parameters and model specifications. Most credit risk applications require the computation of nonlinear transformation of model parameters and the unobserved firm's asset value. The invariance principle is employed for obtaining the ML estimates of these quantities. The delta method is utilized to obtain the asymptotic normality and to make statistical inferences asymptotically. Duan and Fulop (2009) followed this tradition. Using simulations, Duan and Fulop checked the reliability of the standard asymptotic theory. Their results indicate that the asymptotic theory does not work well for the trading noise parameter while ML provides accurate estimates.

One reason for the departure of the finite sample distribution from the asymptotic distribution is the boundary problem. This reason has been put forward by Duan and Fulop and effectively demonstrated via Monte Carlo simulations. We believe, however, there is another reason for the departure. If the microstructure noise process is stationary, the model represents a parametric nonlinear cointegrated relationship between the observed equity value and the unobserved firm's asset value. Park and Phillips (2001) showed that in nonlinear regressions with integrated time series, the limiting distribution is nonstandard and the rate of convergence depends on the properties of nonlinear regression function. As a result, the standard asymptotic theory for ML, such as asymptotic normality, may not be valid.

The first contribution of this paper is to introduce an alternative likelihood-based inferential method for Merton's credit risk model with iid microstructure noises. The new method is based on the general Bayesian approach with posterior computations performed by Gibbs sampling, coupled with data augmentation. Simulations from the Markov chain whose stationary distribution converges to the posterior distribution enable exact finite sample inferences. We note that Jacquier et al. (1994) and Kim et al. (1998), among others, have suggested this approach in the context of a stochastic volatility model. We recently became aware that this idea has independently been discussed by Korteweg and Polson (2009) in the context of Merton's credit risk model with iid microstructure noises.[1]

There are certain advantages in the proposed 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.

[1] Our work differs from this paper in several important respects. First, while we adopt the specification of the state equation of Duan and Fulop (2009) by perturbing the log-price with an additive error, Korteweg and Polson (2009) assume a multiplication error on the state variable and require the pricing function be invertible. Second, our work goes beyond the estimation problem to encompass issues involving model comparisons. Third, we examine more flexible microstructure noise behavior based on stock prices only whereas Korteweg and Polson (2009) used the pure noise normality assumption based on multiple price relations.

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. 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 should not work well too. 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.

A disadvantage of the proposed MCMC method is that 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. Another disadvantage of the proposed MCMC method is that the model has to be fully specified whereas the MLE remains consistent even when the microstructure noise is nonparametrically specified, and in this case, MLE becomes quasi-MLE. However, other MCMC methods can be used to deal with more flexible distributions for the microstructure noise. In particular, the flexibility of the error distribution may be accommodated by using a Dirichelt process mixture (DPM) prior, leading to the so-called semiparametric Bayesian model (see Ferguson, 1973 for the detailed account of DMP, and Jensen and Maheu, 2008 for an application of DMP to volatility modelling).

The second contribution of this paper is to provide generalized models of Duan and Fulop (2009) so that we allow a more flexible behavior for microstructure noises. In particular, we consider two models. In the first specification, we model the microstructure structure noises using a Student t distribution. In the second specification, we allow the microstructure structure noises to be correlated with the shocks to the firm values. We show that it is straightforward to modify the MCMC algorithm to analyze the new models. Empirically, we find evidence of a positive correlation between the microstructure noises and the firm values in emerging markets.

The rest of the paper is organized as follows. Section 2 reviews the Merton's model and the ML method of Duan and Fulop (2009). In Section 3, we introduce the Bayesian MCMC method. Like Duan and Fulop, we put the model into the framework of nonlinear state-space methodology and describe the Bayesian approach to parameter estimation using Gibbs sampling. Section 4 discusses how the proposed method can be used for credit risk applications and for analyzing more flexible specifications for microstructure noise. We also discuss how to the model comparison using the deviance information criterion (DIC) is performed. In Section 5, we implement the Bayesian MCMC method using several datasets, including one US dataset used in Duan and Fulop (2009), and datasets from two emerging markets. Section 6 concludes.

2. Merton's model and ML 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, and 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:

$$d\ln V_t = (\mu - \sigma^2/2)\,dt + \sigma\,dW_t, \quad V_0 = c, \tag{1}$$

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

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

where $\varepsilon_t \sim N(0,1)$, 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 (i.e. $V_T < F$). 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 were 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

$$S_t \equiv S(V_t;\sigma) = V_t\Phi(d_{1t}) - Fe^{-r(T-t)}\Phi(d_{2t}), \tag{3}$$

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 $t=\tau_1,\ldots,\tau_n$. When there is no confusion, we simply write $t=1,\ldots,n$. Since the joint density of $\{V_t\}$ is specified by (2), the joint density of $\{S_t\}$ can be obtained from Eq. (3) 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 contaminates the equilibrium price process. The presence of market microstructure noises motivates Duan and Fulop (2009) to consider the following generalization to Merton's model (we call it Mod 1):

$$\ln S_t = \ln S(V_t;\sigma)+\delta v_t, \tag{4}$$

where $\{v_t\}$ is a sequence of iid standard normal variates. Eqs. (2) and (4) form the basic credit risk model with microstructure noises which was studied by Duan and Fulop (2009). Putting the model in a state-space framework, Eq. (4) is an observation equation, and Eq. (2) is a state equation. Unfortunately, the Kalman filter is not applicable here since the observation equation is nonlinear.

Let $\mathbf{X}=(\ln S_1,\ldots,\ln S_n)'$, $\mathbf{h}=(\ln V_1,\ldots,\ln V_n)'$, and $\theta=(\mu,\sigma,\delta)'$. The likelihood function of Mod 1 is given by

$$p(\mathbf{X};\theta)=\int p(\mathbf{X},\mathbf{h};\theta)\,d\mathbf{h}=\int p(\mathbf{X}|\mathbf{h};\mu)p(\mathbf{h};\theta)\,d\mathbf{h}, \tag{5}$$

where $p(\cdot)$ means the probability density function. In general, this is a high-dimensional integral which does not have a closed form expression due to the nonlinear dependence of $\ln S_t$ on $\ln V_t$.

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. The 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. As in many other simulation based methods, the 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.

Traditional particle filtering algorithms, such as the one proposed by Kitagawa (1996), sample a point $V_t^{(m)}$ when the system is advanced. To improve the efficiency, Pitt and Shephard (1999) proposed to sample a pair $(V_t^{(m)},V_{t+1}^{(m)})$. Duan and Fulop adopted this *auxiliary particle filtering* algorithm where the sequential predictive densities, and hence the likelihood function are the by-products of filtering. Unfortunately, the resulting likelihood function is not smooth with respect to the parameters. To ensure a smooth surface for the likelihood function, Duan and Fulop followed the suggestion in Pitt (2002) by using the smooth bootstrap procedure for resampling.

Since the log-likelihood function (denoted by $\ell(\theta)$) is readily available from the filtering algorithm, it is maximized numerically over the parameter space to obtain the simulation-based ML estimator (denoted by $\hat{\theta}_n$). 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. Standard asymptotic theory for ML suggests that

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{d}N(0,I^{-1}(\theta)), \tag{6}$$

where $I(\theta)$ is the limiting information matrix, and the MLE is considered optimal in the Hajék–LeCam sense, achieving the Cramér–Rao bound and having the highest possible estimation precision in the limit when $n\to\infty$. It is obvious that in this standard asymptotic theory, the rate of convergence is root-n.

Suppose $C(\theta)$ is a nonlinear function of θ and needs to be estimated. By virtue of the principle of invariance, the ML estimator of $C(\theta)$ is obtained simply by replacing θ in $C(\theta)$ with $\hat{\theta}_n$, leading to $\hat{C}_n=C(\hat{\theta}_n)$, the ML estimate of $C(\theta)$. By the standard delta method argument, the following asymptotic behavior for $\hat{C}_n$ is obtained:

$$\sqrt{n}(\hat{C}_n-C(\theta))\xrightarrow{d}N(0,V_C), \tag{7}$$

where

$$V_C=\frac{\partial C}{\partial\theta'}I^{-1}(\theta)\frac{\partial C}{\partial\theta}. \tag{8}$$

Since $\hat{C}_n$ is the ML estimator (Zehna, 1966), it retains good asymptotic properties of ML.[2] For example, it is expected to have the highest possible precision when $n\to\infty$. Not surprisingly, this plug-in estimator was suggested for credit risk applications in Duan and Fulop. Two particular examples mentioned in their paper are the credit spread of a risky corporate bond over the corresponding Treasury rate, and the default probability of a firm.

Duan and Fulop (2009) carried out Monte Carlo simulations to check the reliability of the proposed ML estimator and the standard asymptotic theory (6), based on 500 simulated samples, each with 250 daily observations. When $\delta=0.004$, it was found that both σ and μ but not δ can be accurately estimated. By examining the coverage rates, they concluded that the asymptotic distribution conforms reasonably well to the corresponding finite sample distribution for σ and μ but not

[2] However, the finite sample property of $\hat{C}_n$ may be worse than that of $\hat{\theta}_n$; see, for example, Phillips and Yu (2009).

for δ. Duan and Fulop (2009) further related the failure of the asymptotic approximation for δ to the boundary problem. In particular, for 110 out of 500 sample paths, the estimate of δ reached the lower bound. When $\delta = 0.016$, they found that the standard asymptotic distribution worked much better for δ.

In addition to the boundary problem, we believe there is another problem in the use of the standard asymptotic theory (6). While the standard asymptotic theory is well developed for stationary or weakly dependent processes, the asymptotic analysis becomes more complicated for models with integrated variables. Often the asymptotic distribution becomes nonstandard and the rate of convergence is not root-n. For example, in a simple linear process with a unit root, Phillips (1987) obtained the asymptotic distribution of the ML estimator of the autoregressive coefficient. The distribution is skewed to the left and the rate of convergence is root-n. For linear cointegration systems, Johansen (1988) showed that the ML estimator has a nonstandard limiting distribution. The asymptotic theory is even more complicated for nonlinear models with integrated time series, of which nonlinear cointegration is an important special case. Park and Phillips (2001) developed the asymptotic theory for this class of models. It was shown that the rate of convergence depends on the properties of the nonlinear regression function and can be as slow as $n^{1/4}$. The limiting distribution is nonstandard and is mixed normal with mixing variates that depend on the sojourn time of the limiting Brownian motion of the integrated process.

Clearly, the model considered in this paper is nonlinear cointegration. While both $\ln V_t$ and $\ln S_t$ are nonstationary, their nonlinear combination is stationary. The theoretical results in Park and Phillips (2001) indicate that the standard asymptotic theory may be inappropriate. However, since $\ln V_t$ is latent in our model, it would be difficult, if not impossible, to apply the theoretic results of Park and Phillips (2001) to our framework.

To examine the performance of the standard asymptotic distribution, we design a Monte Carlo study which is similar to the design in Duan and Fulop (2009). The parameter values are $\sigma = 0.3$, $\delta = 0.016$, $\mu = 0.2$. The interest rate is 5% and remains constant throughout the sample period. The initial value of V_0 is fixed at \$100 and F is fixed at \$40, both being assumed to be known. We acknowledge the fact that the specification of the initial value has important implications both for the finite sample distributions and for the asymptotic distributions because the state variable has a unit root; see, for example, Müller and Elliott (2003) for a detailed account for implications of initial conditions in unit root models. As in Duan and Fulop, 250 daily observations (1-year data) are simulated in each sample. In total, 1000 sample paths are simulated. The initial maturity is set to 10 years and, by the end of the sample period, reduces to 9 years. The filtering algorithm provided by Duan and Fulop, namely localizedfilter.dll, is implemented with 5000 particles generated to estimate the parameters. There are two differences between our design and that of Duan and Fulop. First, the initial value is fixed and assumed to be known in our design and this design represents the simplest scenario. In Duan and Fulop, the last observation is fixed and the path simulation is conducted backwards. Also, while we fix the initial value in our study, in Duan and Fulop the initial value V_0 is assumed to be the perturbed V_1^*, where V_1^* is the first period asset value obtained from the model without the microstructure noise. Second, in our design the number of simulated paths chosen is 1000 (instead of 500 as in Duan and Fulop) and the number of particles 5000 (instead of 1000 as in Duan and Fulop), with the hope that the finite sample distributions can be more accurately obtained. Bounds used for σ, δ and μ are [0.01, 20], $[10^{-7}, 1000]$ and $[-20, 20]$, respectively. Note that when $\delta = 0.016$, Duan and Fulop found little evidence of the boundary problem; see Table 6 in their paper.

Table 1 reports the mean, the median, the minimum, the maximum, the standard deviation, the skewness, the kurtosis, the Jarque-Bera (JB) test statistic for normality and its p-value, all computed from 1,000 samples. Fig. 1 plots the finite sample distributions (i.e. the histograms) and the standard asymptotic distributions. Several results emerge from the table and the figure. First, similar to what was found by Duan and Fulop, all the parameters can be accurately estimated, with the mean and the median being sufficiently close to the true value. Consistent with what was found in Merton (1980) and Phillips and Yu (2005), μ is more difficult to estimate than σ when the time span of the data is small. Second, comparing the minimum and the maximum with the bounds, we have found in all cases there is no boundary problem. Thus, the finite

Table 1
Simulation results obtained from 1000 sample paths, each with 250 daily observations.

Parameter	σ	$\delta(\times 100)$	μ
True value	0.3	1.6	0.2
Mean	0.295	1.597	0.205
Median	0.294	1.608	0.196
Minimum	0.187	0.878	−0.745
Maximum	0.398	2.143	1.042
Std. err.	0.030	0.211	0.298
Skewness	0.158	−0.231	−0.077
Kurtosis	3.108	2.967	2.780
JB stat.	4.668	8.953	2.033
p-Value	0.097	0.011	0.361

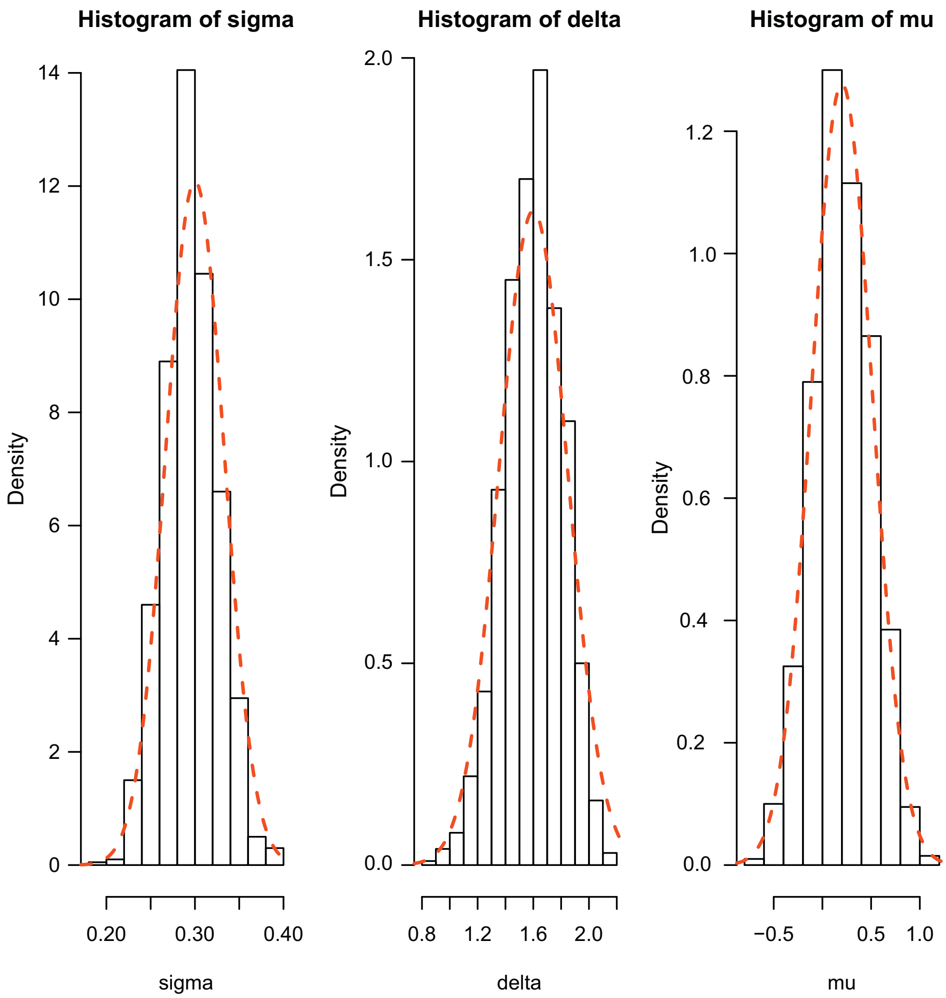

Fig. 1. Finite sample distribution (histogram) of MLE of σ, δ (multiplied by 100), μ based on the particle filtering method of Duan and Fulop (2009). The dotted line is the standard asymptotic distribution where the asymptotic variance is obtained from the Fisher information matrix.

sample distributions are not affected by the bounds. Third and most interestingly, the JB statistics suggest that the finite sample distribution is strongly nonnormal for δ and moderately nonnormal for σ and, but for μ, it conforms well to normality. In particular, the finite sample distribution for δ is skewed (-0.231). When comparing the finite sample distribution with the standard asymptotic distribution in Fig. 1, we have found that for both σ and δ, the standard asymptotic distribution is not satisfactory. Apart from the apparent skewness in the finite sample distribution of the MLEs of δ, there is strong evidence of 'peakness' in the finite sample distributions of the MLE of δ and σ, relative to the standard asymptotic distributions. For μ, the finite sample distribution conforms well to the standard asymptotic distribution. In sum, the Monte Carlo results seem to confirm our conjecture that for the model which involves nonlinear cointegration, the standard asymptotic theory may not be applicable.

3. Bayesian MCMC

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, $\mathbf{h}$, and the joint distribution of the observables, here the sequence of contaminated log-equity prices $\mathbf{X}$. The treatment of the latent state variables $\mathbf{h}$ 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

$$p(\mu,\sigma,\delta,\mathbf{h}) = p(\mu,\sigma,\delta)p(\ln V_0)\prod_{t=1}^{n} p(\ln V_t|\ln V_{t-1},\mu,\sigma). \tag{9}$$

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 equation (2). The likelihood $p(\mathbf{X}|\mu,\sigma,\delta,\mathbf{h})$ is specified by the observation equation (4) and the conditional independence assumption:

$$p(\mathbf{X}|\mu,\sigma,\delta,\mathbf{h}) = \prod_{t=1}^{n} p(\ln S_t|\ln V_t,\delta). \tag{10}$$

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

$$p(\mu,\sigma,\delta,\mathbf{h}|\mathbf{X}) \propto p(\mu)p(\sigma)p(\delta)p(\ln V_0)\prod_{t=1}^{n} p(\ln V_t|\ln V_{t-1},\mu,\sigma)\prod_{t=1}^{n} p(\ln S_t|\ln V_t,\delta). \tag{11}$$

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 $p(\theta|\mathbf{h})$ 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 $p(\theta,\mathbf{h}|\mathbf{X})$ given in (11). Note that the new likelihood function is $p(\mathbf{X}|\theta,\mathbf{h})$ 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 $\mathbf{h}$ 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 credit 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, in the paper we use the Gibbs sampler which samples each variate, one at a time, from the full conditional distributions defined by (11). 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 (2009).

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 $\mathbf{h}$.
2. Sample $\ln V_t$ from $\ln V_t|\ln V_{-t},\mathbf{X}$.
3. Sample $\sigma|\mathbf{X},\mathbf{h},\mu,\delta$.
4. Sample $\delta|\mathbf{X},\mathbf{h},\mu,\sigma$.
5. Sample $\mu|\mathbf{X},\mathbf{h},\sigma,\delta$.

Steps 2–5 form 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-called burn-in samples. The remaining samples are used to make inference.

In this paper, we make use of the all purpose Bayesian software package WinBUGS to perform the Gibbs sampling. As shown in Meyer and Yu (2000) and Yu and Meyer (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. Although more efficient MCMC algorithms, such as multi-move algorithms, can be developed for estimating credit risk models, we do not consider that possibility in the paper. One reason is that the chains generated from the single-move algorithm mix very well in the empirical applications, as we will show below.

4. Credit risk applications, flexible modelling and model comparison

4.1. Credit risk applications

One of the most compelling reasons for obtaining the estimates for the model parameters and the latent equity values is their usefulness in 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 practical important quantity is the credit spread of a risk corporate bond over the corresponding Treasure rate.

Using the notations of Duan and Fulop (2009), the credit spread is given by

$$C(V_n;\theta) = -\frac{1}{T-\tau_n}\ln\left(\frac{V_n}{F}\Phi(-d_{1n})+e^{-r(T-\tau_n)}\Phi(d_{2n})\right)-r, \tag{12}$$

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

$$P(V_n;\theta) = \Phi\left(\frac{\ln(F/V_n)-(\mu-\sigma^2/2)(T-\tau_n)}{\sigma\sqrt{T-\tau_n}}\right). \tag{13}$$

The Gibbs samplers for θ and V_n can be inserted into the formulae (12) and (13) to obtain the Markov chains for the credit spread and the default probability. Because any measurable functions of a stationary ergodic sequence is stationary and ergodic, the chains provide exact finite-sample inferences about these two quantities.

4.2. Flexible modelling of microstructure noises

Modelling the microstructure noise as an iid normal variate is a natural starting point. Duan and Fulop (2009) have convincingly shown that ignoring trading noise can lead to a significant overestimation of asset volatility and that the estimated magnitude of trading noise is in line with the prior belief. On the other hand, it is well known that the market microstructure effects are complex and take many different forms. Therefore, it is interesting to know empirically what the best way to model the microstructure noises in the context of structural credit risk models is. With this goal in mind, we introduce two more general models.

In the first model, motivated from the empirical fact that the distributions of almost all financial variables have fat tails, we assume the distribution of v_t is a Student-t with an unknown degree of freedom (call it Mod 2). That is,

$$\ln S_t = \ln S(V_t;\sigma)+\delta v_t, v_t \sim t_\kappa \tag{14}$$

and

$$\ln V_{t+1} = (\mu-\sigma^2/2)h+\ln V_t+\sigma\sqrt{h}\varepsilon_t,$$

In the second generalized model, we allow the microstructure noise to be correlated to the innovation to the equity value (call it Mod 3), that is,[3]

$$\ln S_t = \ln S(V_t;\sigma)+\delta v_t,$$

$$\ln V_{t+1} = (\mu-\sigma^2/2)h+\ln V_t+\sigma\sqrt{h}\varepsilon_t,$$

where v_t,ε_t are $N(0,1)$ and $corr(v_t,\varepsilon_t)=\rho$.

As discussed earlier, any implementation of the Gibbs sampler necessitates the specification of each of the full conditional posterior densities and of a simulation technique to sample from them. Any change in the model, such as a different prior distribution or different sampling distribution, necessarily entails changes in those full conditional densities. WinBUGS releases from the tedious task of calculating the full conditionals and chooses an effective method to sample from them. As a result, one can experiment with different types of models with very little extra programming effort. Modifications of the model are straightforward to implement by changing just one or two lines in the code. This ease of implementation appears to be in sharp contrast to the simulation-based ML method via particle filtering.

4.3. Model comparison

With alternative models being proposed, it is interesting to compare their relative performances. Duan and Fulop (2009) conducted a likelihood ratio test to compare the model with microstructure noises and the one without noises. Since their estimation method is ML with the former model nesting the latter one, the likelihood ratio test is possible. Obviously, the likelihood ratio test is not applicable in our context for two reasons. First, we have Bayesian models. Second, the two generalized models do not nest each other.

In the Bayesian context, one way of comparing the proposed models is by computing Bayes factors. Alternatively, one can use information criteria. A popular method is the Akaike information criterion (AIC; Akaike, 1973) for comparing

[3] A similar idea has been used in the context of stochastic volatility; see, for example, Yu (2005).

alternative and possibly nonnested models. AIC trades off a measure of model adequacy against a measure of complexity measured by the number of free parameters. In a nonhierarchical Bayesian model, it is easy to specify the number of free parameters. However, in a complex hierarchical model, the specification of the dimensionality of the parameter space is rather arbitrary. This is the case for all the credit risk models considered here. The reason is that when MCMC is used to estimate the models, we augment the parameter space. For example, in Mod 1, we include the n latent variables into the parameter space. As these latent variables are highly dependent with a unit root in the dynamics, they cannot be counted as n additional free parameters. Consequently, AIC is not applicable in this context (Berg et al., 2004).

Let θ denote the vector of augmented parameters. The deviance information criterion (DIC) of Spiegelhalter et al. (2002) is intended as a generalization of AIC to complex hierarchical models. Like AIC, DIC consists of two components:

$$\mathrm{DIC} = \overline{D} + p_D. \tag{15}$$

The first term, a Bayesian measure of model fit, is defined as the posterior expectation of the deviance

$$\overline{D} = E_{\theta|\mathbf{X}}[D(\theta)] = E_{\theta|\mathbf{X}}[-2\ln f(\mathbf{X}|\theta)]. \tag{16}$$

The 'better' the model fits the data, the larger the value for the likelihood. The variable $\overline{D}$, which is defined via -2 times log-likelihood, therefore attains smaller values for the 'better' models. The second component measures the complexity of the model by the *effective number of parameters*, p_D, defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean $\overline{\theta}$ of the parameters:

$$p_D = \overline{D} - D(\overline{\theta}) = E_{\theta|\mathbf{X}}[D(\theta)] - D(E_{\theta|\mathbf{X}}[\theta]) = E_{\theta|\mathbf{X}}[-2\ln f(\mathbf{X}|\theta)] + 2\ln f(\mathbf{X}|\overline{\theta}). \tag{17}$$

By defining $-2\ln f(\mathbf{X}|\theta)$ as the residual information in the data X conditional on θ, and interpreting it as a measure of uncertainty, Eq. (17) shows that p_D can be regarded as the expected excess of the true over the estimated residual information in data $\mathbf{X}$ conditional on θ. That means we can interpret p_D as the expected reduction in uncertainty due to estimation.

Spiegelhalter et al. (2002) justified DIC asymptotically when the number of observations n grows with respect to the number of parameters and the prior is nonhierarchical and completely specified. As with AIC, the model with the smallest DIC is estimated to be the one that would best predict a replicate dataset of the same structure as that observed. This focus of DIC, however, is different from the posterior-odd-based approaches, where how well the prior has predicted the observed data is addressed. Berg et al. (2004) examined the performance of DIC relative to two posterior odd approaches—one based on the harmonic mean estimate of marginal likelihood (Newton and Raftery, 1994) and the other being Chib's (1995) estimate of marginal likelihood—in the context of stochastic volatility models. They found reasonably consistent performance of these three model comparison methods. From the definition of DIC it can be seen that DIC is almost trivial to compute and particularly suited to compare Bayesian models when posterior distributions have been obtained using MCMC simulation.

5. Empirical analysis

5.1. Priors and initial values

We assume prior independence of the parameters μ,σ, and δ. We employ an uninformative prior for μ, $\mu \sim N(0.3,4)$. A conjugate inverse-gamma prior is chosen for σ, i.e. $\sigma \sim IG(3,0.0001)$. Similarly, a conjugate inverse-gamma prior is chosen for σ, i.e. $\sigma \sim IG(2.5,0.025)$. For κ and ρ, uninformative priors are used. In particular, $\kappa \sim \chi^2_{(8)}$ and $\rho \sim U(-1,1)$.

The initial values of μ,σ^2, and δ^2 are set at $\mu = 0.3$, $\delta^2 = 1.0 \times 10^{-4}$, and $\sigma^2 = 0.02$. In all cases, after a burn-in period of 10,000 iterations and a follow-up period of 100,000, we stored every 20th iteration.

5.2. US data

We implement the MCMC method using data from a company, 3M, from the Dow Jones Industrial Index. Duan and Fulop (2009) fitted Mod 1 to the same data. In addition to this basic model, we also fit the two new flexible specifications to the data. There are two purposes for using the same data as in Duan and Fulop. First, by comparing our estimates to the ML estimates obtained by Duan and Fulop, we can check whether our method can produce sensible estimates. Second, for the US data, we would like to know if the newly proposed models can perform better than Mod 1.

As explained in Duan and Fulop, the daily equity values are obtained from the CRSP database over year 2003. The initial maturity of debt is 10 years. The debt is available from the balance sheet obtained from the Compustat annual file.[4] 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. As there are 252 daily observations in the data, we set $h = \frac{1}{252}$ which is slightly different from $\frac{1}{250}$ used in Duan and Fulop. The difference is so small that the impact on the estimates should be negligible.

[4] As a referee points out, however, some care needs to be taken when the book value is used to calculate the market value of debt. While we accept this view, we use the same value of debt as in Duan and Fulop for the purpose of comparison.

Table 2 reports the estimates of posterior means and the estimates of posterior standard errors for θ in the basic credit risk model with iid normal noises. For ease of comparison, we also report the ML estimates and the asymptotic standard errors obtained in Duan and Fulop.

All the Bayesian estimates are very close to the ML counterparts. Furthermore, the two sets of standard errors are also comparable. The results show the reliability of the MCMC method for obtaining point estimates. The trace and kernel density estimates of marginal posterior distribution of model parameters are shown in Fig. 2. It can be seen that all the chains mix very well. The marginal posterior distribution is quite symmetric for both σ and μ but is slightly asymmetric for δ. All parameters pass the Heidelberger and Welch stationarity and halfwidth tests. Geweke's Z-scores for δ,μ,σ are all close to zero (0.335, -0.357, -0.224). The dependence factors from the Raftery and Lewis convergence diagnostics (estimating the 2.5 percentile up to an accuracy of ± 0.005 with probability 0.95) are 1.78, 1.03, 1.02 for δ,μ,σ respectively. All these statistics strongly suggest that the chains converge well and are indeed stationary. Fig. 3 plots the autocorrelation function for each chain. In all cases, the autocorrelation becomes negligible at a few lags, suggesting the convergence is fast.

Table 3 reports the estimates of posterior means and posterior standard errors for θ and DIC in all three specifications. In Mod 2, the posterior mean of κ, the degree of freedom parameter in the t distribution, is estimated to be 16.29. It suggests

Table 2
Bayesian estimation results and ML estimation results for the basic model using daily 3M data.

	μ		σ		$\delta \times 100$	
	Mean	Std. err.	Mean	Std. err.	Mean	Std. err.
Bayesian	0.2797	0.1273	0.1270	0.0090	0.4689	0.0649
ML	0.2798	0.1358	0.1318	0.0089	0.4044	0.0919

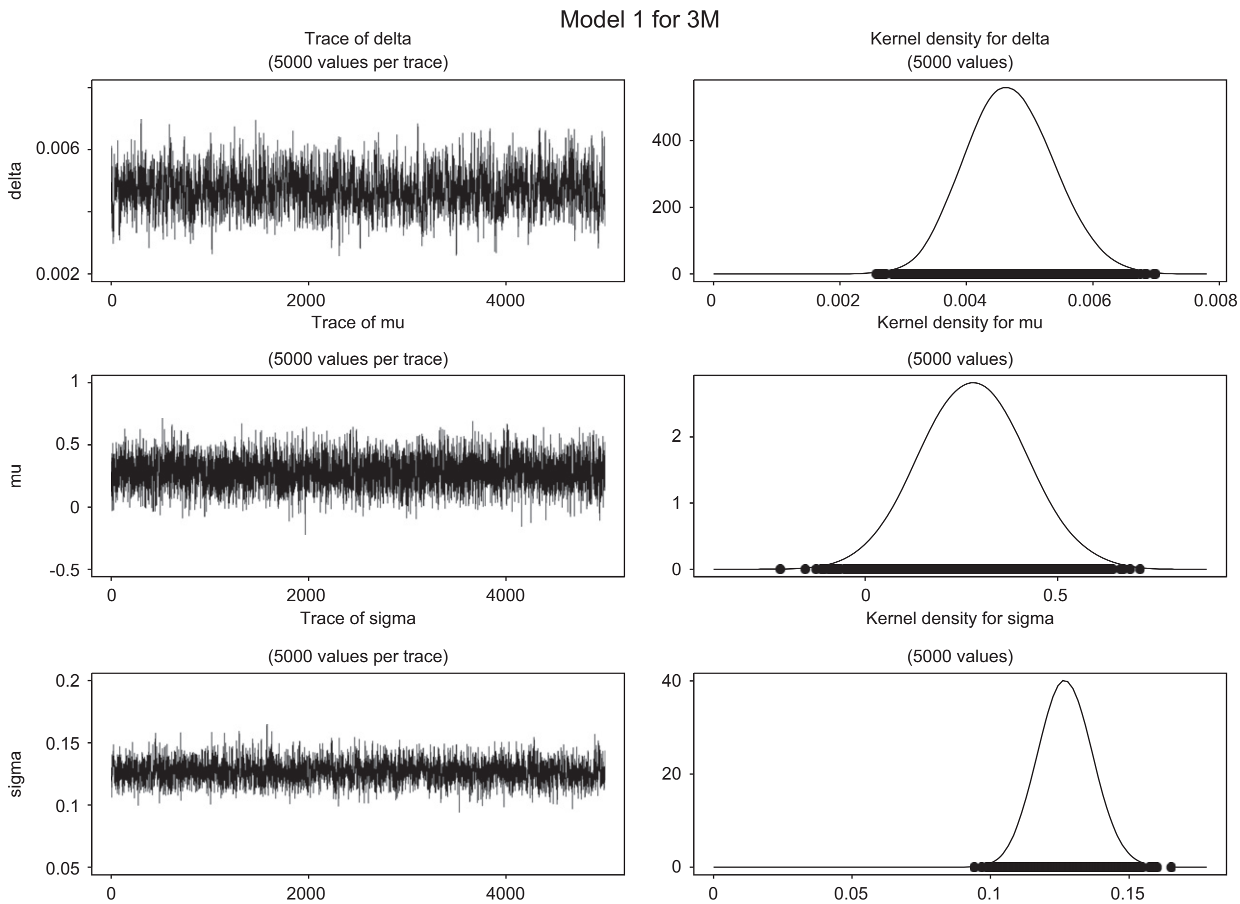

Fig. 2. Trace and kernel density estimates of the marginal posterior distribution of parameters in Mod 1.

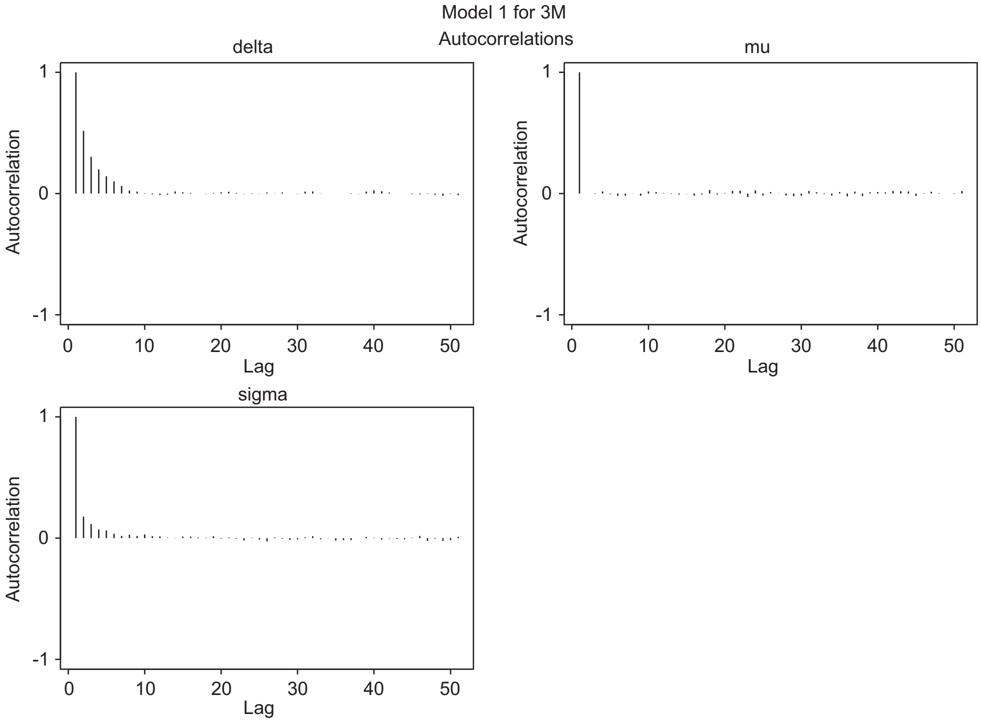

Fig. 3. Autocorrelation functions of parameters in Mod 1.

Table 3
Bayesian estimation results for alternative models using daily 3M data.

	μ		σ		$\delta \times 100$		κ or ρ		DIC
	Mean	Std. err.	Mean	Std. err.	Mean	Std. err.	Mean	Std. err.	
Mod 1	0.2797	0.1273	0.1270	0.0090	0.4689	0.0649			−1812.37
Mod 2	0.2801	0.1259	0.1256	0.0090	0.4481	0.0588	16.29	5.601	−1797.34
Mod 3	0.2803	0.1312	0.1301	0.0106	0.5479	0.0485	0.3359	0.3387	−1791.21

little evidence against normality. In Mod 3, the posterior mean of ρ is estimated to be 0.3359 with posterior standard error 0.3387. Not surprisingly, the credible interval contains zero. The estimate of κ in Mod 2 and ρ in Mod 3 seem to suggest that the two flexible models do not offer improvements to the model of Duan and Fulop. This observation is further reinforced by the DIC values for the three models. Mod 1 has the lowest DIC, followed by Mod 2, and then by Mod 3.

As explained before, with MCMC it is straightforward to obtain the smoothed estimates of the latent firm asset values and any transformation of the model parameters and the latent variables, such as the default probability. The default probability has been widely used to rate firms. In Fig. 4, we plot the observed equity values, the smoothed firm asset values and default probabilities for 3M under the preferred model, Mod 1. As can be seen, when the equity value goes up, the asset value goes up and the default probability goes down. The smoothed estimates for the default probabilities are very small and seem reasonable for 3M.

5.3. *Data from two emerging markets*

We also implement the MCMC method using datasets of two firms, both from emerging markets. The first is the Bank of East Asia listed in Hong Kong Stock Exchange while the second is DBS Bank listed in Singapore Stock Exchange. The daily closing prices over the two year period, 2003–2004, are downloaded from finance.yahoo. The balance sheets obtained from the company's website give us information about the number of outstanding shares and the total value of liabilities (debts). The initial maturity of debt is 10 years. We compound the debts for 10 years at the risk-free rate to obtain F. The risk-free rate is obtained from the Monetary Authority of Hong Kong and the Monetary Authority of Singapore, respectively. There are 496 (504) daily observations in the sample of Bank of East Asia (DBS Bank), and we set $h = \frac{1}{248}$ ($h = \frac{1}{252}$).

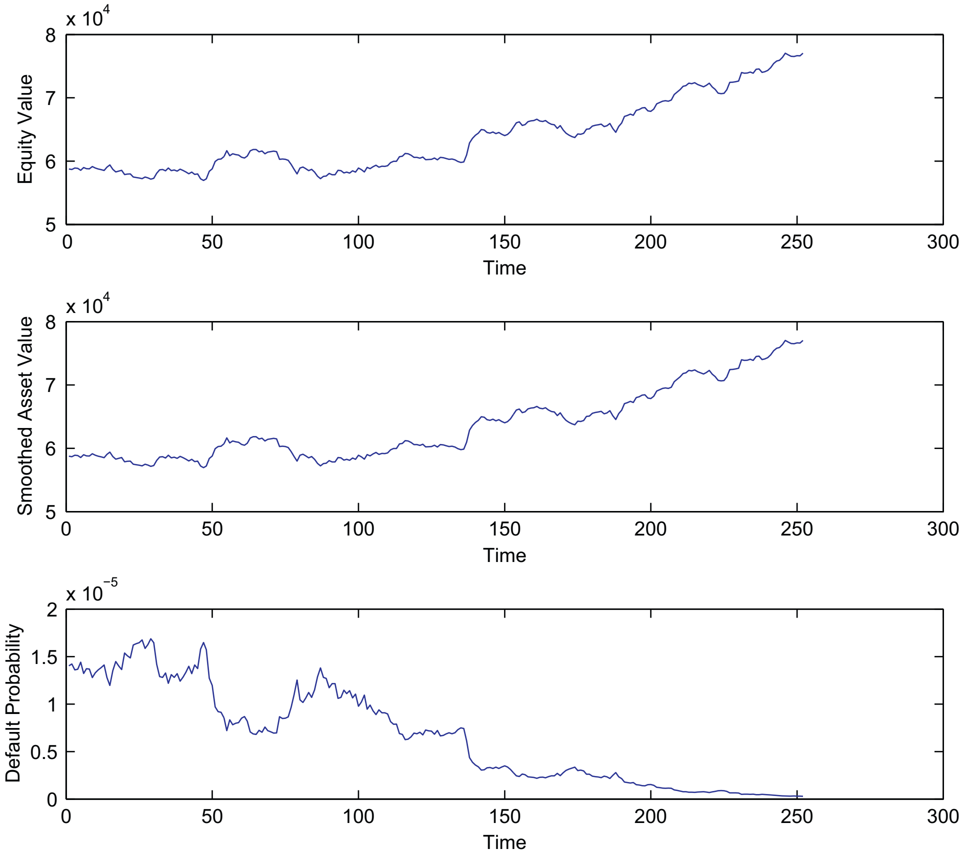

Fig. 4. The observed equity values, the smoothed firm asset values and default probabilities of 3M in Mod 1.

Table 4
Bayesian estimation results for alternative models using daily data for Bank of East Asia listed in Hong Kong Stock Exchange.

	μ		σ		$\delta \times 100$		κ or ρ		DIC
	Mean	Std. err.	Mean	Std. err.	Mean	Std. err.	Mean	Std. err.	
Mod 1	0.001116	0.1168	0.1647	0.0063	0.3245	0.03654			−3866.51
Mod 2	3.2×10^{-6}	0.1146	0.1618	0.0068	0.3225	0.03652	15.42	5.983	−3765.87
Mod 3	0.005943	0.1303	0.1856	0.0089	0.4532	0.05792	0.7566	0.1497	−3911.20

Table 5
Bayesian estimation results for alternative models using daily data for DBS Bank listed in Singapore Stock Exchange.

	μ		σ		$\delta \times 100$		κ or ρ		DIC
	Mean	Std. err.	Mean	Std. err.	Mean	Std. err.	Mean	Std. err.	
Mod 1	0.1623	0.1337	0.1893	0.00747	0.4375	0.0690			−3600.61
Mod 2	0.1624	0.1332	0.1891	0.00784	0.4209	0.0667	17.04	5.738	−3541.55
Mod 3	0.1642	0.1442	0.2039	0.01151	0.5231	0.1127	0.3804	0.2007	−3596.44

Table 4 reports the estimates of posterior means and the estimates of posterior standard errors for θ in the basic credit risk model and the two flexible models for Bank of East Asia. The estimates of δ, σ and κ for Bank of East Asia seem to have the same order of magnitude as those for 3M. For example, in Mod 2, the posterior mean of κ is estimated to be 15.42 which suggests little evidence against normality. However, in Mod 3, the posterior mean of ρ is estimated to be 0.7566 with the estimate of the posterior standard error being 0.1497. This provides strong evidence for a positive correlation between the microstructure noises and the firm values. This observation is further reinforced by the DIC values for the three models. Mod 3 has the lowest DIC, followed by Mod 1, and then by Mod 2.

Table 5 reports the estimates of posterior means and the estimates of posterior standard errors for θ in the basic credit risk model and the two flexible models for DBS Bank. Once again, the estimates of δ, σ and κ are similar to those in the previous applications. For example, in Mod 2, the posterior mean of κ is estimated to be 17.04 which suggests little evidence against normality. In Mod 3, the posterior mean of ρ is estimated to be 0.3804 with the estimate of the posterior standard error being 0.2007. The 95% credible interval includes zero while the 90% credible interval excludes zero. This provides some evidence for a positive correlation between the microstructure noises and the firm values. According to DIC, Mod 1 and Mod 3 perform similarly, followed by Mod 2 with a big gap in the DIC values.

6. Conclusion

In this paper we introduce a Bayesian method to estimate structural credit risk models with microstructure noises. We show that it is a viable alternative method to ML. The new method is applied to estimate Merton's model, augmented by various forms of microstructure noises. We have found the empirical support that microstructure noises are positively correlated with the firm values in emerging markets.

The proposed technique is very general and can be applied in other credit risk models and other forms of microstructure noises. For example, the method can be extended to a broader range of model specifications, including the Longstaff and Schwartz (1995) model with stochastic interest rates, the Collin-Dufresne and Goldstein (2001) model with a stationary leverage, and the double exponential jump diffusion model used in Huang and Huang (2003). In more complicated models, the analytic relationship between $\ln S_t$ and $\ln V_t$ may be unavailable, and hence, the Bayesian method would be computationally more involved. However, the same argument applies to alternative estimation methods, including ML.

The present paper only brings the credit risk models to daily data. With the availability of intra-day data in financial markets, one is also able to estimate the credit risk models using ultra-high frequency data, enabling more accurate estimations of model parameters, default probability, etc. It is well known from the realized volatility literature that the dynamic properties of microstructure noises critically depend on the sampling frequency (Hansen and Lunde, 2006). It is expected that a more complicated statistical model is needed for microstructure noises when ultra-high frequency data are used.

Acknowledgements

Huang gratefully acknowledges financial support from the Office of Research at Singapore Management University under Grant no. 08-C207-SMU-024. Yu gratefully acknowledges financial support from the Ministry of Education AcRF Tier 2 fund under Grant no. T206B4301-RS. We wish to thank Carl Chiarella (the Editor), an anonymous referee, Max Bruche, Jin-chuan Duan, Andras Fulop, Ser-huang Poon, and seminar participants at International Symposium on Financial Engineering and Risk Management 2008, Singapore Econometric Study Group 2009 Meeting, and Risk Management Conference 2008 for helpful discussions. We also wish to thank Jin-Chuan Duan and Andras Fulop for making their datasets and particle filtering algorithm available.

References

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2009. Ultra high-frequency volatility estimation with dependent microstructure noise. Journal of Econometrics, forthcoming.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Proceedings 2nd International Symposium Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Bandi, F.M., Russell, J., 2008. Microstructure noise, realized volatility, and optimal sampling. Review of Economic Studies 75, 339–369.

Berg, A., Meyer, R., Yu, J., 2004. Deviance information criterion for comparing stochastic volatility models. Journal of Business and Economic Statistics 22, 107–120.

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

Chib, S., 1995. Marginal likelihood from the Gibbs output. The Journal of the American Statistical Association 90, 1313–1321.

Chib, S., 2001. Markov chain Monte Carlo methods: computation and inference. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, Amsterdam, pp. 3569–3649.

Collin-Dufresne, P., Goldstein, R., 2001. Do credit spreads reflect stationary leverage ratios? Journal of Finance 56, 1929–1957

Duan, J.-C., 1994. Maximum likelihood estimation using price data of the derivative contract. Mathematical Finance 4, 155–167.

Duan, J.-C., Gauthier, G., Simonato, J.G., 2004. On the equivalence of the KMV and maximum likelihood methods for structural credit risk models. Working Paper, University of Toronto.

Duan, J.-C., Gauthier, G., Simonato, J.-G., Zaanoun, S., 2003. Estimating Merton's model by maximum likelihood with survivorship consideration. Working Paper, University of Toronto.

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.

Ericsson, J., Reneby, J., 2004. Estimating structural bond pricing models. Journal of Business 78, 707–735.

Ferguson, T., 1973. A Bayesian analysis of some nonparametric problems. Annals of Statistics 1, 209–230.

Hansen, P., Lunde, A., 2006. Realized volatility and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161.

Hasbrouck, J., 1993. Assessing the quality of a security market: a new approach to transaction-cost measurement. Review of Financial Studies 6, 191–212.

Huang, J., Huang, M., 2003. How much of the corporate-treasury yield spread is due to credit risk? Working Paper, Penn State and Stanford.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.
Jensen, M.J., Maheu, J.M., 2008. Bayesian semiparametric stochastic volatility modeling. Working Paper No. 2008-15, Federal Reserve Bank of Atlanta.
Johansen, S., 1988. Statistical analysis of cointegrated vectors. Journal of Economic Dynamic and Control 12, 231–254.
Johannes, M., Polson, N., 2009. MCMC methods for Continuous-time financial econometrics. In: Aït-Sahalia Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics 2, 1–72.
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.
Korteweg, A., Polson, N., 2009. Corporate credit spreads under parameter uncertainty. Working paper, Graduate School of Business, Stanford University.
Longstaff, F., Schwartz, E., 1995. A simple approach to valuing risky fixed and floating rate debt. Journal of Finance 50, 789–820.
Merton, R.C., 1974. On the pricing of corporate debt: the risk structure of interest rates. Journal of Finance 29, 449–470.
Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. Journal of Financial Economics 8, 323–361.
Meyer, R., Yu, J., 2000. BUGS for a Bayesian analysis of stochastic volatility models. Econometrics Journal 3, 198–215.
Müller, U.K., Elliott, G., 2003. Tests for unit roots and the initial condition. Econometrica 71, 1269–1286.
Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society, Series B 56, 3–48 (with discussion).
Park, J.S., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–167.
Phillips, P.C.B., 1987. Time series regressions with a unit root. Econometrica 55, 277–301.
Phillips, P.C.B., Yu, J., 2005. Jackknifing bond option prices. Review of Financial Studies 18, 707–742.
Phillips, P.C.B., Yu, J., 2006. Comments: realized volatility and market microstructure noise by Hansen and Lunde. Journal of Business and Economic Statistics 24, 202–208.
Phillips, P.C.B., Yu, J., 2007. Information Loss in Volatility Measurement with Flat Price Trading. Working Paper, Yale University.
Phillips, P.C.B., Yu, J., 2009. 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., 1999. Filtering via simulation: auxiliary particle filter. The Journal of the American Statistical Association 94, 590–599.
Roll, R., 1984. Simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64, 583–639 (with discussion).
Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–549.
Wong, H., Choi, T., 2006. Estimating default barriers from market information. Working Paper, Chinese University of Hong Kong.
Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.
Yu, J., Meyer, R., 2006. Multivariate stochastic volatility models: Bayesian estimation and model comparison. Econometric Reviews 25, 361–384.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.
Zehna, P., 1966. Invariance of maximum likelihood estimation. Annals of Mathematical Statistics 37, 744.