Estimation of Parameters in Mean-Reverting Stochastic Systems

Tianhai Tian, Yanli Zhou, Yonghong Wu, and Xiangyu Ge

1 School of Mathematical Sciences, Monash University, Melbourne, VIC 3800, Australia
2 School of Finance, Zhongnan University of Economics and Law, Wuhan 430073, China
3 Department of Maths & Statistics, Curtin University, Perth, WA 6845, Australia
4 Wuhan Yangtze Business University, Wuhan 430065, China

Correspondence should be addressed to Yanli Zhou; ylzou8507@gmail.com and Xiangyu Ge; xiangyu.ge@163.com

Received 8 October 2013; Revised 9 December 2013; Accepted 9 February 2014; Published 16 March 2014

Academic Editor: Xiaodong Lin

Copyright © 2014 Tianhai Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Stochastic differential equation (SDE) is a very important mathematical tool to describe complex systems in which noise plays an important role. SDE models have been widely used to study the dynamic properties of various nonlinear systems in biology, engineering, finance, and economics, as well as physical sciences. Since a SDE can generate unlimited numbers of trajectories, it is difficult to estimate model parameters based on experimental observations which may represent only one trajectory of the stochastic model. Although substantial research efforts have been made to develop effective methods, it is still a challenge to infer unknown parameters in SDE models from observations that may have large variations. Using an interest rate model as a test problem, in this work we use the Bayesian inference and Markov Chain Monte Carlo method to estimate unknown parameters in SDE models.

1. Introduction

A stochastic differential equation (SDE) can be defined as a deterministic differential equation perturbed by random disturbances that are not necessarily small. SDEs have gained popularity in recent years for their ability to model systems that are subject to fluctuations and have been widely used in a variety of disciplines including engineering, environmental science, population dynamics, biology, and medicine. In particular, SDEs are important instruments in modern finance theory and have been used to model the behavior of many key financial variables such as the instantaneous short-term interest rate, asset prices, asset returns, and their volatility. Consequently, the estimation of the parameters of SDEs from discretely-sampled data has received substantial attention in the financial econometrics literature, particularly in the last ten years [1, 2].

However, parameter estimation in nonlinear SDEs driven by Wiener processes, when only discrete observation is available, is an inherently difficult problem because theoretically an unlimited number of solutions exist for a SDE. Although parameter estimation in deterministic systems is a relatively well-studied subject (see, e.g., Beck and Arnold [3]), the estimation of the parameters of stochastic systems remains a challenge [4]. One of the reasons for this is that obtaining the solution of a set of SDEs is computationally demanding in the absence of a closed-form solution for most SDEs of practical importance. Numerical methods such as the Euler methods and the Taylor schemes combined with a Monte Carlo approach are required to calculate discrete-time trajectories of the state variables of SDEs [5]. These numerical methods require the generation of a large number of Wiener processes corresponding to different simulation trajectories, and hence accurate simulations are computationally expensive.

The methods that were developed for parameter estimation of SDEs can be classified into three different categories: maximum likelihood estimation (MLE)/simulated maximum likelihood (SML) [6–8], the methods of moments [9–11], and filtering (e.g., extended Kalman filter) [12]. Many of the methods have been developed in the context of financial modelling, where the systems of interest are characterised by long time horizons (of the order of months) and can often be sampled at regular but relatively infrequent intervals (e.g., on a daily basis). Among them, the maximum likelihood method is more reliable but has long been found to be difficult to apply to SDEs due to its computational cost. For many
SDEs, thousands of simulation trajectories or even more must be generated to ensure a low variance of the values of state variables. As a result, a large number of competing estimation procedures have been proposed in recent years.

In recent years the Bayesian inference methods have been used to estimate unknown parameters in mathematical models [13–17]. Together with the Markov-chain Monte-Carlo (MCMC) and other methods, the Bayesian inference methods have also been used to infer stochastic models in financial mathematics [18–20]. The main advantage of these methods is the ability to infer the whole probability distribution of the parameters, rather than just a single estimate. In addition, the Bayesian methods can deal with noisy data and uncertain data. Another advantage of these methods is the capability to infer parameters in either deterministic models or stochastic models. However, the potential obstacle of these methods in application is that the samples are correlated and their performances heavily depend on prior hypotheses.

A number of methods have been used to estimate the parameters in the single-factor continuous time models, including the generalized moment method [21] and Gaussian estimation methods [22]. However, our recent research work suggested that the accuracy of the estimates generated from these two methods is low, in particular, when the stepsize of observation time points is not small [23]. Thus in this work we will not test these methods again but concentrate on the proposed method that will generate accurate simulations of the stochastic model, which will lead to more accurate estimates of the model parameters. Instead, utilizing the Bayesian inference and MCMC method, we develop a numerical algorithm to estimate unknown parameters in stochastic interest rate models. The remaining part of this paper is organized as follows. Section 2 gives the stochastic models for term structure of interest rates and numerical algorithms for simulating these stochastic models. Section 3 discusses the Bayesian inference and the MCMC method. Section 4 reports the numerical results for estimating parameters in the stochastic models for the term structure of interest rates.

2. Stochastic Model and Direct Simulation Methods

We first introduce the general form of SDEs for interest rates, namely,

\[ dX = f(t, X) \, dt + g(t, X) \, dW(t), \]  

(1)

where \( f(t, X) \) is the drift term, \( g(t, X) \) is the diffusion term, and \( W(t) \) is the Wiener process whose increment follows the Gaussian distribution, namely,

\[ \Delta W_n = W(t_{n+1}) - W(t_n) \sim \mathcal{N}(0, t_{n+1} - t_n). \]  

(2)

Now we proceed to consider numerical methods for simulating the SDE. The widely used method in computational finance is the Euler-Maruyama method whose strong convergence order is just 0.5, given by

\[ X_{n+1} = X_n + hf(t_n, X_n) + g(t_n, X_n) \Delta W_n, \]  

(3)

where \( X_n \) is the numerical solution at time point \( t_n \) and \( h = t_{n+1} - t_n \). Although this method is easy to implement, its stability property is not good enough to simulate SDEs with relatively large diffusion term. In order to obtain stable simulations, a very small stepsize is required, which may lead to large computing time. To improve the stability property, the semi-implicit and fully implicit Euler method can be used to reduce the computing time [24, 25]. For example, the semi-implicit Euler method is given by

\[ X_{n+1} = X_n + hf(t_n, X_n) + g(t_n, X_n) \Delta W_n. \]  

(4)

Another approach is to use high-order methods in order to achieve more accurate simulations. The Milstein scheme uses a higher-order Taylor expansion and thus has a strong convergence order one. In the explicit Milstein method both the drift term and diffusion term are explicit, namely,

\[ X_{n+1} = X_n + hf(t_n, X_n) + g(t_n, X_n) \Delta W_n + \frac{1}{2} g(t_n, y_n) g'(t_n, y_n) (\Delta W_n)^2 - h. \]  

(5)

Similarly, the semi-implicit and fully implicit Milstein methods have been designed to improve the stability property [24, 25].

In this paper we use a SDE model of the term structure of interest rate as the test system to examine the accuracy of inference methods [1]. The first stochastic model is the CIR model (Cox, Ingersoll, and Ross), which is a linear mean reversion model and uses a diffusion process [26]. This stochastic model has been widely used to model the short interest rate [21, 27]. The CIR model states that the short rate follows a square root diffusion process, which has the following continuous-time representation:

\[ dX = \alpha(\beta - X) \, dt + \sigma \sqrt{X} \, dW(t), \]  

(6)

where \( \alpha \) is the speed of adjustment (or mean reversion), \( \beta \) represents the long run mean of the short-term interest rate, and \( \sigma \) is a constant volatility. Under this model, both the drift and volatility change with the level of the short rate.

In this work we will use the Euler-Maruyama method to generate samples of the interest rates (3). In fact, due to the linear feature of the drift term in the interest rate model, the semi-implicit method can be written in explicit form and can also be used in the Bayesian inference method. For the benchmark model, the Euler-Maruyama Scheme is

\[ X_{n+1} = X_n + \alpha(\beta - X_n) h + \sigma \sqrt{X_n} \Delta W_n, \]  

(7)

and the semi-implicit Euler scheme is given by

\[ X_{n+1} = \frac{1}{1 + \alpha h} \left( X_n + \alpha h \beta + \sigma \sqrt{X_n} \Delta W_n \right). \]  

(8)

3. Algorithm for Parameter Estimation

In this section, we establish a numerical algorithm for estimating parameters in stochastic models based on
the Bayesian inference and MCMC method. By contrast with the classical approach that the unknown parameters in a model have fixed quantity, the unknown parameters of the underlying model in the Bayesian paradigm are treated as a random variable with some prior beliefs. The heart of the Bayesian approach is the Bayes theorem which allows us to compute the conditional probability density function of model parameters \( \theta \), assuming that the model parameters are continuous random variables, given the entire data set \( y \):

\[
\mathbb{P}(\theta | y) = \frac{\mathbb{P}(y | \theta) \mathbb{P}(\theta)}{\mathbb{P}(y)}.
\] (9)

Since the probability \( \mathbb{P}(y) \) is independent of the model parameters, to maximize the joint probability density function, only the product \( \mathbb{P}(y | \theta) \mathbb{P}(\theta) \) should be considered. Thus the posterior distribution \( \mathbb{P}(\theta | y) \) can also represent updated prior belief on the parameters \( \mathbb{P}(\theta) \) updated by the current information from the data. Because we have little prior knowledge of \( \theta \), we may simply use a “noninformative” or “flat” prior.

In this work we use the Bayesian inference method derived by Joshi and Wilson [20] to infer the parameters in SDE models. It was assumed that the diffusion process \( \{X_t\} \) was observed at points \( t_0, t_1, \ldots, t_n \) and the observation vector was \( Y = \{y_0, y_1, \ldots, y_n\} \). Since the closed form of the transition densities of the diffusion processes is usually not available, the transition densities can be approximated by the densities of a numerical scheme such as the widely used Euler-Maruyama method (3). However, the observation timestep size \( \Delta t = t_{i+1} - t_i \) is normally quite large. To obtain more accurate approximation of the transition densities, a number of latent variables are introduced between every pair of consecutive observations:

\[
t_i = \tau_{0,j} < \tau_{1,j} < \cdots < \tau_{M,j} = t_{i+1}.
\] (10)

The stepsize of the latent variables \( \delta_t = \tau_{j+1,j} - \tau_{j,j} \) is small enough to ensure the accuracy and stability property of the Euler-Maruyama method. Then the transition density of the Euler scheme is

\[
P_{\text{Euler}} \{X_{j,1|1} \mid X_{j,j,\theta}\} = N(\mu_{\text{Euler}}, \sigma_{\text{Euler}}^2),
\] (11)

where \( j = 0, 1, \ldots, M - 1 \) and

\[
\mu_{\text{Euler}} = X_{j,j} + f(t_{j,j}, X_{j,j}, \theta) \delta_t,
\]

\[
\sigma_{\text{Euler}} = g(t_{j,j}, X_{j,j}, \theta) \sqrt{\delta_t}.
\] (12)

Thus we have an inference problem with unknown parameter \( \theta \) using the latent variables \( X = \{X_{j,i}\} \) for \( i = 0, 1, \ldots, n, j = 1, \ldots, M - 1 \), and the observation data \( Y \).

In the proposed Gaussian Modified Bridge Approximation with importance sampling (GaMBA-I), the computing process is given as follows.

**Algorithm 1.** Consider the following.

**Step 1.** Generate a sample of the unknown parameters \( \Theta \) using the MCMC method or other methods.

**Step 2.** Sample the solution at the latent points \( X \).

**Step 3.** Evaluate probability \( P(Y, X \mid \Theta) \).

**Step 4.** Evaluate probability \( P(X \mid Y, \Theta) \).

**Step 5.** Calculate probability

\[
P(\Theta \mid Y) \propto P(Y, X \mid \Theta) P(\Theta) P(X \mid Y, \Theta).
\] (13)

When the importance sampling technique is used, the above probability is determined by a number of samples rather than a single sample as indicated above.

**Step 6.** Accept or reject the parameter sample \( \Theta \) using the MCMC method or other methods.

The major step in this Bayesian inference method is the evaluation of the probabilities of the generated samples for the latent variables. The probability \( P(Y, X \mid \Theta) \) is

\[
P(Y, X \mid \Theta) \propto \prod_{i=1}^n P(Y_i \mid X_{M-1,j-1}, \Theta) \prod_{i=1}^n P(X_{1,j-1} \mid Y_{j-1}, \Theta) \
\times \prod_{j=1}^{M-1} \prod_{j=1}^{M-1} P(X_{j-1,j-1} \mid X_{j-1,j-1}, \Theta)
\] (14)

Here we assume that the probability for the initial observation \( y_0 \) is a constant. Each probability in the above expression
can be approximated by the transition density of the Euler method, given by

\[ P(Y, X \mid \Theta) \approx \prod_{i=1}^{n} P_{\text{Euler}}(y_i \mid X_{M-1, i-1}, \Theta) \]

\[ \times \prod_{i=1}^{n} P_{\text{Euler}}(X_{1, i-1} \mid y_{i-1}, \Theta) \]

\[ \times \prod_{i=1}^{n-1} \prod_{j=1}^{M-1} P_{\text{Euler}}(X_{j, i-1} \mid X_{j-1, i-1}, \Theta), \]  \hspace{1cm} (15)

where \( P_{\text{Euler}} \) is the Euler density in (11).

For the probability \( P(X \mid Y, \Theta) \), we need to factorise it into

\[ P(X \mid Y, \Theta) = \prod_{i=0}^{n-1} P(X^{(i)} \mid y_i, y_{i+1}, \Theta) \]

\[ = \prod_{i=0}^{n-1} P(X_{1,i}, X_{2,i}, \ldots, X_{M-1,i} \mid y_i, y_{i+1}, \Theta) \]

\[ = \prod_{i=0}^{n-1} \prod_{j=1}^{M-1} P(X_{j,i} \mid X_{j-1,i}, X_{M,j}, \Theta), \]  \hspace{1cm} (16)

where \( X_{0,i} = y_i \) and \( X_{M,i} = y_{i+1} \). Using the Modified Brownian Bridge (MBB), the density of \( P(X_j \mid X_{j-1}, X_M, \Theta) \) can be approximated by

\[ P_{\text{MBB}}(X_{j,i} \mid X_{j-1,i}, X_{M,j}, \Theta) \approx N_X(\mu_{\text{MBB}}, \sigma_{\text{MBB}}^2), \]  \hspace{1cm} (17)
where
\[
\mu_{MBB} = X_{j-1,i} + \left( \frac{X_{M,i} - X_{j-1,i}}{\tau_{M,i} - \tau_{j-1,i}} \right) \delta, \\
\sigma_{MBB} = g(X_{j-1,i}, \Theta) \sqrt{\frac{M - j}{M - j + 1}} \delta. 
\]

(18)

Thus, the solution at the latent points is sampled by using
\[
X_{j,i} = \mu_{MBB} + \sigma_{MBB}N_j, 
\]
where \(N_j\) is a sample of the standard Gaussian random variable \(N(0, 1)\).

When the importance sampling method is used, a number of samples of the latent variables are generated for \(X_k \sim P_{MBB}(X_k | Y, \Theta)\) as described in Algorithm I. Then we evaluate
\[
P_{GaMBA}(\Theta | Y) \propto \frac{1}{K} \sum_{k=1}^{K} P_{Euler}(Y, X_k | \Theta) \cdot P(\Theta). 
\]

(20)

Figure 3: Estimated model parameters using different values of the importance sampling sizes. (a) The estimated model parameters; (b) the standard deviation (std) of the estimates. The importance sampling size is \(K = 1, 10, 25, 50, 100, 200, 500\) when index \(1 \sim 7\).
the MCMC. For example, the selected initial estimate has influence on the generated sequence; in particular, it has strong influence on the initial sequence of simulations. Thus an important technique is the burn-in technique. This technique is designed to reduce the influence of initial iteration on the generated Markov Chain by discarding the first iteration sequences. Generally we discard the first half of simulated sequences and keep the remaining half of sequence to obtain the target distribution. Certainly this technique is convenient but cannot be the most efficient one because about half of computing efforts are discarded. Although more specific methods have been designed to analyze the simulation output according to the dependence of simulation on the starting values [28], we typically go with the simple burn-in approach and accept the increased Monte Carlo error involved in discarding half of simulations. To design a strategy to finish the computation, we normally monitor the convergence of all the parameters and other quantities of interest separately. Our usual approach is, for each parameter, to calculate the variance of simulations from each chain (after the first half of the chain was discarded using the burn-in technique).

Assuming we have $J$ chains from different initial estimate and the length of each chain is $G$, let $\theta_{ij}$ be the $j$th estimate in the sequence for parameter $\theta$; the variance inside the chain is

\[ W = \frac{1}{J(J-1)} \sum_{j=1}^{G} \left( \theta_{ij} - \theta_j \right)^2 \]  \hspace{1cm} (21)

and the variance between the different chains is

\[ B = \frac{1}{(J-1)G} \sum_{j=1}^{J} \left( \theta_j - \frac{1}{J} \sum_{g=1}^{G} \theta_{gj} \right)^2 \]  \hspace{1cm} (22)

Based on these values, we can calculate the value of $R$ as

\[ R = \sqrt{\frac{1}{G} \left( G - 1 + \frac{B}{W} \right)} \]  \hspace{1cm} (23)

The value of $R$ is always greater than or equal to 1. When the variance inside the chain approaches the variance between

---

**Figure 4:** Variations of the estimated parameters when the sampling of the latent variables is different. (a) The estimated model parameters; (b) the standard deviation (std) of the estimates.
the Bayesian inference and MCMC method to estimate the
In this section, we used the numerical algorithm based on

4. Numerical Results

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|}
\hline
Parameters & $\alpha(0.2)$ & $\beta(0.08)$ & $\sigma(0.05)$ \\
\hline
Estimated values & 0.1803 & 0.0792 & 0.0442 \\
Standard deviations & 0.0109 & 0.0024 & 0.0013 \\
\hline
\end{tabular}
\caption{Estimated parameters and their standard errors.}
\end{table}

the chains, the value of $R$ approaches 1. We can accept the
chain as convergent when $R < 1.2$ [29]. Another important
technique is thinning by rejecting certain part of the chain. If
$\Theta_t$ is the current candidate of the model parameters and $\Theta^*$
is the newly generated one, let

$$
\alpha = 1 \wedge \frac{P(\Theta^* | y)}{P(\Theta_t | y)}. (24)
$$

Generate a sample $r \sim U(0, 1)$, and set $\Theta_{t+1} = \Theta^*$ if $r < \alpha$.
Otherwise set $\Theta_{t+1} = \Theta_t$.

4. Numerical Results

In this section, we used the numerical algorithm based on
the Bayesian inference and MCMC method to estimate the
parameters in the CIR model (6). Figure 1 gives 5 simulations
of the CIR model with parameters $\alpha = 0.2$, $\beta = 0.08$, and $\sigma =
0.1$. When the volatility is not large, it shows that the values of
short interest rate maintain positivity, which is unlike Vasicek
model may lead to negative values of interest rate. We used
stepsize $h = 0.05$ in the numerical simulation to ensure the
accuracy and stability property of simulations.

The estimated values of the parameters in (6) for the
mean-reverting test system and their standard deviations are
given in Table 1, and more detailed simulation results of the
Bayesian inference method are presented in Figure 2. In this
test the size of the importance sampling is $K = 50$. For
each parameter, we presented the time series of the parameter
values, the cumulative means, and the histogram distribution.
Compared with the exact values ($\alpha, \beta, \sigma$) = (0.2, 0.08, 0.05),
the Bayesian inference method gives estimates with good
accuracy. For parameters $\beta$ and $\sigma$, the histogram distributions
are consistent with the cumulative means of the estimates.
However, the histogram of parameter $\alpha$ is not symmetrical
to the cumulative means.

In this study we tested the influence of the sample
size in the importance sampling on the accuracy of
the estimates. The sampling size was chosen as $K =
1, 10, 25, 50, 100, 200, 500$. Numerical results in Figure 3 show
that the sampling size is important to improve the accuracy
of the estimates but a larger sampling size does not necessarily
lead to much better accuracy, though numerical results in
Figure 3 suggest that increasing sampling size can improve
the accuracy slightly. Thus a reasonable size of the importance
sampling is sufficient to generate estimates with adequate
accuracy. This may be the reason that the sampling size is not
very large in the previous studies [20].

We have also tested the influence of different samples of
latent variables on the variation of estimates. In this test the
simulated observations $Y$ are kept unchanged. Figure 4 shows
that the difference of sampling has certain influence on the
variations of the estimates. The estimated model parameters
vary in different tests. However, the variations in both the
averaged parameter values and standard deviation are not
large, which is consistent with the numerical results using
the particle swarm optimization method to estimate model
parameters. However, the variations of estimates are smaller
than those obtained by using the genetic algorithm [30, 31].

5. Conclusions

This work presents an effective algorithm for the estimation
of parameters in SDE models. The proposed approach is
based on the implementation of the Bayesian inference and
the MCMC method. Compared with the grid method, the
MCMC based method can be used to infer stochastic models
with a large number of unknown parameters. This method
has been applied to an important stochastic model of the
term structure of interest rate, which is a fundamental issue
in the research area of financial mathematics. In addition,
the importance sampling technique was used to increase the
robustness of estimates. We have also examined the influence
of different samples of the latent variables on the variation of
estimates. Numerical results suggested that the method used
in this work is robust to such variation.

In this work, we introduced the Gaussian Modified Bridge
Approximation into the MCMC and examine the accuracy
and robustness of this approach. It is worthy to note that
the performance of the MCMC is related to a number
of important factors, such as convergence criteria, burn-in
technique, and thinning to reduce autocorrelation; thus,
other efforts are needed to discuss these issues. In addition,
it is still a challenging problem for estimating parameters
in stiff stochastic models. In the method used in this work,
the explicit Euler method was used in simulating stochastic
models. Thus, a large number of the latent variables are
needed because the stepsize of the numerical method must
be very small. However, in this case, we still have difficulties
to properly evaluate the complete likelihood $P(Y, X | \Theta)$
and conditional distribution $P(X | Y, \Theta)$ due to the product of
a large number of probabilities. Alternatively we may consider
the implicit methods or high-order methods rather than
the explicit Euler method. Thus, more effective calibration
methods should be designed for estimating parameters in stiff
SDEs.

Conflict of Interests

The authors declare that there is no conflict of interests
regarding the publication of this paper.

Acknowledgments

This research work is supported by the Australian
Research Council (ARC) (DP120104460, FT100100748,
and DP109418I) and the Chinese National Social Science
Foundation (SSF) (10BJY104).
References


Submit your manuscripts at http://www.hindawi.com