On Simulated Likelihood of Discretely Observed Diffusion Processes and Comparison to Closed-Form Approximation

Osnat Stramer Jun Yan *

June 20, 2006

Abstract

This article focuses on two methods to approximate the loglikelihood function for univariate diffusions: 1) the simulation approach using a modified Brownian bridge as the importance sampler; and 2) the recent closed-form approach. For the case of constant volatility, we give a theoretical justification of the modified Brownian bridge sampler by showing that it is exactly a Brownian bridge. We also discuss computational issues in the simulation approach such as accelerating numerical variance stabilizing transformation, computing derivatives of the simulated loglikelihood, and choosing initial values of parameter estimates. The two approaches are compared in the context of financial applications with annualized parameter values, where the diffusion model has an unknown transition density and has no analytical variance stabilizing transformation. The closed-form expansion, particularly the second-order closed-form, is found to be computationally efficient and very accurate when the observation frequency is monthly or higher. It is more accurate in the center than in the tail of the transition density. The simulation approach combined with the variance stabilizing transformation is found to be more reliable than the closed-form approach when the observation frequency is lower. Both methods performs better when the volatility level is lower, but the simulation method is more robust to the volatility nature of the diffusion model. When applied to two well known datasets of daily observations, the two methods yield similar parameter estimates in both datasets but slightly different loglikelihood in the case of higher volatility.

*Osnat Stramer is Associate Professor (E-mail: stramer@stat.uiowa.edu) and Jun Yan is Assistant Professor (E-mail: jyan@stat.uiowa.edu), Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA.

KEY WORDS: Closed-Form expansion; Continuous-time diffusion; Discrete sampling; Maximum likelihood; Simulated method; Transition density.

1. INTRODUCTION

Discrete time series are often obtained by observing a continuous-time process at a discrete sequence of observation times. It is natural to model a discrete time series by modeling the underlying process as a continuous-time process. Diffusion processes are commonly used to model the underlying process in economic and financial applications (see e.g. Sundaresan 2000). A univariate time-homogeneous diffusion process can be described as a solution to a stochastic differential equation (SDE) of the form

$$\mathrm{d}X_t = \mu(X_t;\theta)\,\mathrm{d}t + \sigma(X_t;\theta)\,\mathrm{d}W_t, \quad 0 \le t \le T, \tag{1}$$

where the drift μ and volatility σ are locally bounded Borel measurable functions, θ is a $p \times 1$ parameter vector, and W_t is a standard Brownian motion. To be specific, we assume that the process X is observed only at discrete time points $t_i = i\Delta$, $i = 0, \ldots, n$, giving observations x_i, $i = 0, \ldots, n$. The aim is to estimate θ given these observations.

This article focuses on the maximum likelihood estimate (MLE) of the parameter θ in (1). We refer to Sørensen (2004) and Aït-Sahalia (2006) for more comprehensive reviews on other methods. In order to obtain the MLE, it is necessary to have the transition density

$$p_X(\Delta, x|x_0;\theta), \tag{2}$$

that is, the conditional density of $X_{t+\Delta} = x$ given $X_t = x_0$. By the Markov property, the loglikelihood function is then

$$L_n(\theta) = \sum_{i=1}^{n} l_X(\Delta, x_i|x_{i-1};\theta), \tag{3}$$

where $l_X = \log p_X$, and the asymptotically irrelevant density of the initial observation x_0 is disregarded. The transition density is known only in rare occasions where the SDE (1) has an analytically known solution. Examples of these rare occasions are geometric Brownian motion, Ornstein-Uhlenbeck process, and Cox-Ingersoll-Ross (CIR) process (see e.g. Aït-Sahalia 1999). In general, an explicit solution to (1) is not available, so the likelihood must be approximated numerically.

An approximation of the likelihood relies on an approximation of the transition density p_X or its logarithm l_X. An approximate MLE is then obtained by maximizing the approximated likelihood.

A naive approximation of $p_X(\Delta, x|x_0;\theta)$ is the Euler (or Euler-Maruyama) scheme. The Euler approximation is a first-order approximation

$$p_X^{(1)}(\Delta, x|x_0;\theta) = \phi(x; m, v), \tag{4}$$

where $m = x_0 + \Delta\mu(x_0;\theta)$, $v = \Delta\sigma^2(x_0;\theta)$, and $\phi(\cdot; m, v)$ is the normal density with mean m and variance v. This approximation works well only when Δ, the transition interval, is sufficiently small. More sophisticated approximation methods fall into three classes (Aït-Sahalia 2002b; Sørensen 2004): 1) Data augmentation at unobserved time points (Pedersen 1995; Elerian 1999; Brandt and Santa-Clara 2002b; Durham and Gallant 2002); 2) Closed-form expansion of the transition density (Aït-Sahalia 2002a, 2005); 3) Numerical solution to the Kolmogorov forward equation implied by the transition density (Lo 1988). The data augmentation method utilizes the Euler scheme, with the transition interval between augmented data made very small. Bekos, Papaspliopoulos, Roberts, and Fearnhead (2006) recently developed an exact data augmentation approach which does not use the Euler scheme. For a general class of diffusion processes, Bekos et al. (2006) simulates unobserved data at random time skeletons from the exact distribution of the unknown transition density. This approach will not be studied here. We limit the scope of this article to the aforementioned first two classes and refer to them as the simulation approach and the closed-form approach, respectively. A brief review of these two approaches are given in the next section.

Both the simulation approach and the closed-form approach are more amenable when the volatility σ is constant. It is possible, but not always, to transform a process to a new one with constant volatility. This transformation is referred to as variance stabilization in introductory statistics. It has been used in likelihood approaches to obtain better approximation of the likelihood (Shoji and Ozaki 1998; Aït-Sahalia 2002a; Durham and Gallant 2002). It has also been used in a Bayesian context to break the dependency between missing paths and the volatility of the diffusions (Roberts and Stramer 2001). The feasibility of this transformation is therefore important. Following the definition of Aït-Sahalia (2005) in a multivariate setting, a diffusion X is reducible if there exists a one-to-one transformation $Y = h_\theta(X)$ such that the volatility of Y is the identity matrix; It is irreducible otherwise. We call a reducible diffusion X analytically reducible if h_θ is available in explicit form, and numerically reducible otherwise. Under appropriate regularity assumptions, a one-dimensional diffusion X is always reducible, at least numerically, since there exists a one-to-one

function h such that $h'_\theta(x) = 1/\sigma(x,\theta)$. By Itô's formula, the transformed process $Y_t = h_\theta(X_t)$, satisfies the SDE

$$\mathrm{d}Y_t = \tilde{\mu}(y;\theta)\,\mathrm{d}t + \mathrm{d}W_t, \tag{5}$$

where

$$\tilde{\mu}(Y,\theta) = \frac{\mu\{h_\theta^{-1}(Y);\theta\}}{\sigma\{h_\theta^{-1}(Y);\theta\}} - \frac{1}{2}\sigma'\{h_\theta^{-1}(Y);\theta\},$$

and σ' denotes the derivative of σ with respect to x, and h_θ^{-1} is the inverse function of h_θ.

The simulation approach treats the problem as a missing data problem. It augments data between two successive discretely observed values and uses importance sampling to integrate out the augmented variables; see more details in Section 2.1. Using a CIR process, which is analytically reducible, as the benchmark, Durham and Gallant (2002) find that a modified Brownian bridge sampler combined with the variance stabilizing transformation leads to much better performance than other samplers in the literature. A nice interpretation of the modified Brownian bridge sampler is given by Chib and Shephard (2002). This sampler has been successfully used for general multivariate diffusions in Golightly and Wilkinson (2005). However, a rigorous justification of the modified Brownian sampler is missing in the literature.

Aït-Sahalia (2002a) compared the results of Durham and Gallant (2002) to those of the closed-form approach in terms of speed and accuracy and concluded that for monthly observations from the CIR process, the closed-form approximation was the better in terms of the speed/accuracy trade-off. Jensen and Poulsen (2002) also give numerical comparisons of transition density approximations, including both the simulation approach and the closed-form approach. However, as with other comparisons, the diffusion processes they used were all analytically reducible, and the simulation approach could be greatly improved. To the best of our knowledge, there has been no comprehensive comparison of the two methods for univariate diffusions which are only numerically reducible.

The goal of this article is three-fold. First, we supply a theoretical justification for the modified Brownian bridge sampler of Durham and Gallant (2002) when the volatility is constant. Second, we discuss some computational issues that have not been explicitly discussed in detail previously in the numerical optimization of the simulation approach. In particular, we discuss how to accelerate numerical transformations, compute derivatives of the simulated likelihood, and choose starting values. Finally, we compare the two approaches using a diffusion model which has unknown transition

density and is only numerically reducible.

The relative advantage and disadvantage of the two approaches in practice are discussed based on results from a numerical study and real examples. The closed-form approach, both the first-order and second-order, is always faster and can be extremely accurate when the transition interval Δ is small. However, users are faced with the problem of "how small a Δ is small". It is not clear how to assess the accuracy in a given problem. There is no known way, unless compared with the simulation approach, to tell whether the approximation has been "sufficiently" accurate. On the other hand, the simulation approach can always be made as accurate as desired provided sufficient computing resource. But the issue is how much computing resource is available. The two methods give similar parameter estimates from two well known datasets of daily observations, one being the three-month treasury bill rate data and the other being the seven-day Eurodollar rate data. Nevertheless, the maximized loglikelihoods are noticeable different for the Eurodollar data, which has higher volatility than the treasury bill data. The closed-form approximation can be inaccurate at the tails in the case of high volatility and larger Δ values.

The rest of the paper is organized as follows. Section 2 reviews the simulation approximation and the closed-form approximation of the likelihood. Section 3 presents justification of the modified Brownian bridge sampler for reducible diffusion processes. Section 4 discusses computational issues. Section 5 compares the logarithm of the transition density from the two approaches. Section 6 applies the two approaches to two well-known datasets. A discussion concludes in Section 7.

2. APPROXIMATING THE LIKELIHOOD

2.1 Simulation Approach

The simulation approach to approximating the transition density $p_X(\Delta, x_i|x_{i-1}, \theta)$ is a data augmentation approach. The interval $[t_{i-1}, t_i)$ is partitioned into M subintervals $t_{i-1} = \tau_{i,0} < \tau_{i,1} < \ldots < \tau_{i,M} = t_i$, each with equal length $\delta = \Delta/M$, such that the first order Euler scheme is sufficiently accurate. The observations at $\tau_{i,1} < \ldots < \tau_{i,M-1}$ are treated as missing observations. The transition density can be expressed as

$$p_X(\Delta, x_i|x_{i-1}; \theta) = \int \prod_{m=0}^{M-1} p_X(\delta, u_{i,m+1}|u_{i,m}; \theta) d\lambda(u_{i,1}, \ldots, u_{i,M-1}), \tag{6}$$

where $u_{i,0} = x_{i-1}$, $u_{i,M} = x_i$, and λ denotes the Lebesgue measure. The integral in (6) is evaluated using importance sampling (see, for example, Robert and Casella 2005):

$$p_X(\Delta, x_i|x_{i-1};\theta) = E_q\left[\frac{\prod_{m=0}^{M-1} p_X(\delta, u_{i,m+1}|u_{i,m};\theta)}{q(u_{i,1},\ldots,u_{i,M-1})}\right], \tag{7}$$

where $q(u_{i,1},\ldots,u_{i,M-1})$, is a density function on R^{M-1} referred to as an importance sampler. Using the first-order Euler approximation (4) of each $p_X(\delta, u_{i,m+1}|u_{i,m};\theta)$, the transition density (6) is approximated by

$$p_X^{(M,K)}(\Delta, x_i|x_{i-1},\theta) = \frac{1}{K}\sum_{k=1}^{K}\frac{\prod_{m=0}^{M-1} p_X^{(1)}(\delta, u_{i,k,m+1}|u_{i,k,m};\theta)}{q(u_{i,k,1},\ldots,u_{i,k,M-1})}, \tag{8}$$

where $u_{i,k,0} = x_{i-1}$, $u_{i,k,M} = x_i$, and $\boldsymbol{u}_{i,k} = (u_{i,k,1},\ldots,u_{i,k,M-1})$ is a random sample from the importance sampler q for each $k = 1,\ldots,K$. The optimal importance sampler q is the conditional joint density of the diffusion process at times $\tau_{i,1} < \ldots < \tau_{i,M-1}$ given $X_{i-1} = x_{i-1}$ and $X_i = x_i$, denoted by $p_X(u_1,\ldots,u_{M-1}|x_{i-1},x_i)$. This choice of q is not available as it depends on the unknown transition density we are trying to approximate. Clearly, there are important rules about the sampler q for the importance sampling scheme (8) to work well. One rule is that q should provide a sufficiently good approximation to the target density $p_X(u_1,\ldots,u_{M-1}|x_{i-1},x_i)$. A second rule is that sampling $\boldsymbol{u} = (u_1,\ldots,u_{M-1})$ from q can be done efficiently.

Various proposal densities q have been proposed in the literature. Pedersen (1995) used a sampler constructed from simulating paths on each subinterval using the first-order Euler scheme, in which case,

$$p_X^{(M,K)}(\Delta, x_i|x_{i-1},\theta) = \frac{1}{K}\sum_{k=1}^{K} p_X^{(1)}(\delta, u_{i,k,M}|u_{i,k,M-1};\theta). \tag{9}$$

Although simulating from q can be done efficiently, Pedersen's sampler ignores the end point information $x_i = u_{i,M}$, and hence can be far away from the target density. Elerian, Chib, and Shephard (2001) proposed sampling $\boldsymbol{u}$ from a multivariate Gaussian or Student-t distribution, which is based on a second order Taylor expansion of the target density about its mode. This sampler provides a better approximation to the target distribution but can be computational intensive, particularly when M is large. Durham and Gallant (2002) used two additional samplers. The first one is the Brownian bridge sampler, defined as the Euler approximation of the solution H_t to the SDE

$$\mathrm{d}H_t = \frac{x_i - H_t}{t_i - t}\,\mathrm{d}t + \sigma(H_t)\,\mathrm{d}W_t, \qquad t_{i-1} \le t < t_i, \qquad H_{t_{i-1}} = x_{i-1}. \tag{10}$$

When $\sigma(\cdot)$ is constant, H_t is a Brownian bridge on $[t_{i-1}, t_i]$, starting at x_{i-1} and terminating at x_i. Choosing H_t as the underlying process is in line with the principal that an efficient proposal sampler should depend on the end point x_i. This sampler has been generalized to a multivariate setting in Chib, Pitt, and Shephard (2004). The second additional sampler from Durham and Gallant (2002) is called modified Brownian bridge sampler, defined by the recursion

$$u_{m+1} = u_m + \frac{u_M - u_m}{M - m} + \sqrt{\frac{M - m - 1}{M - m}}\delta^{1/2}\sigma(u_m, \theta)Z_m, \tag{11}$$

where $Z_m \sim N(0, 1)$. While (11) is heuristically motivated in Durham and Gallant (2002) and Chib and Shephard (2002), the relationship between the modified Brownian bridge proposal and the underlying process H_t in (10) is missing. The simulation study in Durham and Gallant (2002) using the CIR process shows that the modified Brownian bridge sampler outperforms other simulation-based methods in the literature, and that transforming the model to unit volatility first provides significant benefits. In Section 3 we show that the modified Brownian bridge sampler, when the volatility σ is a constant, is in fact exactly a Brownian bridge.

2.2 Closed-form Approach

For univariate diffusion processes, Aït-Sahalia (2002a) derived a sequence of closed-form expansions for the transition density. Under certain regularity conditions, the sequence converges to the true transition density as more correction terms are added for any fixed transition interval Δ which is smaller than a threshold $\overline{\Delta}$. This threshold $\overline{\Delta}$ depends on both the drift μ and the volatility σ. Aït-Sahalia (2005) extended the method to the multivariate setting, where a diffusion can be irreducible. The expansion for irreducible diffusions can be particularly useful in the univariate case when the variance stabilizing transformation is not analytically available. It is used in our comparison study in the sequel.

For a reducible diffusion X, the crucial step of Aït-Sahalia (2002a) is to find a closed-form transition density expansion for the transformed process $Y = h_\theta(X)$ based on a truncated Hermite series. The closed-form transition density expansion for X is then obtained by a standard variable transformation. The Hermite series approximation of the transition density of p_Y is

$$p_Y(\Delta, y|y_0; \theta) = \Delta^{-1/2}\phi\left(\Delta^{-1/2}(y - y_0)\right)\sum_{j=0}^{\infty}\eta_j(\Delta, y_0; \theta)H_j\left(\Delta^{-1/2}(y - y_0)\right), \tag{12}$$

where ϕ is the density of $N(0,1)$, H_j's are the Hermite polynomials, and Hermite coefficients

$$\eta_j(\Delta, y_0; \theta) = \frac{1}{j!} E\left[H_j\left(\Delta^{-1/2}(y - y_0)\right) | Y_t = y_0; \theta\right].$$

Aït-Sahalia (2002a) suggested approximating the conditional expectation in η_j by a Taylor expansion of order K in Δ using the infinitesimal generator of Y. With the coefficient η_j's replaced by their approximation, the terms in (12) can be gathered according to powers of Δ to obtain a closed-form expansion of the log of the transition density:

$$l_Y^{(K)}(\Delta, y|y_0; \theta) = -\frac{1}{2}\log(2\pi\Delta) + \frac{C_Y^{(-1)}(y|y_0)}{\Delta} + \sum_{k=0}^{K} C_Y^{(k)}(y|y_0; \theta)\frac{\Delta^k}{k!}, \tag{13}$$

where $C_Y^{(k)}$'s are the coefficients after the regathering. Aït-Sahalia (2005) gives an alternative approach to determine the coefficients $C_Y^{(k)}$ by solving the Kolmogorov equations using (13) as the postulated form of the solution. The partial differential equations are expressed in terms of Δ and $C_Y^{(k)}$'s on both sides. The coefficients $C_Y^{(k)}$'s are determined by matching the coefficients of Δ^k for $k = -1, 0, 1, \ldots, K$. Both approaches give the same expansion. Once (13) is available, the closed-form expansion of the log transition density of X is then of the form

$$l_X^{(K)}(\Delta, x|x_0; \theta) = -\log\sigma(x; \theta) - \frac{1}{2}\log(2\pi\Delta) + \frac{C_X^{(-1)}(x|x_0)}{\Delta} + \sum_{k=0}^{K} C_X^{(k)}(x|x_0; \theta)\frac{\Delta^k}{k!}. \tag{14}$$

For irreducible diffusions, the idea of Aït-Sahalia (2005) is to postulate a solution of form (14) and determine the coefficients $C_X^{(k)}$'s. The problem now is that these coefficients no longer have explicit solutions. Let $j_k = 2(K - k)$. Aït-Sahalia (2005) approaches the problem by deriving an explicit Taylor approximation of order j_k for $C_X^{(k)}$ in terms of $(x - x_0)$,

$$C_X^{(j_k,k)}(x|x_0) = \sum_{i=1}^{j_k} \gamma_i^k(x_0)(x - x_0)^i.$$

The expansion order of $C_X^{(k)}$ is chosen to be $j_k = 2(K - k)$ such that the approximation error due to the Taylor expansion in $(x - x_0)$ is of the same order Δ^K for each k. The coefficients $\gamma_i^{(k)}$'s in the Taylor expansion $C_X^{(j_k,k)}(x|x_0)$, $k = -1, 0, \ldots, K$, are obtained from successively solving a system of linear equations. The resulting expansion for the irreducible case is

$$\tilde{l}_X^{(K)}(\Delta, x|x_0; \theta) = -\log\sigma(x; \theta) - \frac{1}{2}\log(2\pi\Delta) + \frac{C_X^{(j_{-1},-1)}(x|x_0)}{\Delta} + \sum_{k=0}^{K} C_X^{(j_k,k)}(x|x0; \theta)\frac{\Delta^k}{k!}. \tag{15}$$

An expansion of the transition density p_X can be obtained by exponentiating $\tilde{l}_X^K$, which guarantees positivity.

3. A JUSTIFICATION OF THE MODIFIED BROWNIAN BRIDGE SAMPLER

We show in Proposition 1 that when the volatility σ is constant, the modified Brownian bridge sampler (11) is simply the exact Brownian bridge sampler. Simulating from the exact distribution of the Brownian bridge, rather than from the Euler approximation to the Brownian bridge process, may lead to better performance.

Proposition 1. *Let $q(u_1, \ldots, u_{M-1})$ be the joint density of the modified Brownian bridge sampler, defined as in* (11) *with unit volatility. Then, $q(u_1, \ldots, u_{M-1})$ is the same as the joint density of a Brownian bridge with unit volatility at time points $\tau_{i,1}, \ldots, \tau_{i,M-1}$, starting with u_0 at t_{i-1} and terminating with u_M at t_i.*

Proof. Let $f_{W|u_0,u_M}(u_1, \ldots, u_{M-1}|u_0, u_M)$ be the joint density of a Brownian bridge with unit volatility on $[t_{i-1}, t_i]$ with two endpoints u_0 and u_M at $\tau_{i,1}, \ldots, \tau_{i,M-1}$. It is sufficient to show that f is the same as q defined in (11). Rewrite $f_{W|u_0,u_M}(u_1, \ldots, u_{M-1}|u_0, u_M)$ as

$$f_{W|u_0,u_M}(u_1, \ldots, u_{M-1}|u_0, u_M) = \prod_{m=0}^{M-2} f_{W_{m+1}|W_m=u_m, W_M=u_M}(u_{m+1}|u_m, u_M). \tag{16}$$

The joint distribution of $(W_{m+1}, W_M)^T$ given $W_m = u_m$ is Gaussian $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{bmatrix} u_m \\ u_m \end{bmatrix}, \quad \text{and } \boldsymbol{\Sigma} = \begin{bmatrix} \delta & \delta \\ \delta & \delta(M-m) \end{bmatrix}.$$

Using the properties of a conditional multivariate normal distribution we obtain,

$$f_{W_{m+1}|W_m=u_m, W_M=u_M}(u_{m+1}|u_m, u_M) = \phi\left(u_{m+1}; u_m + \frac{u_M - u_m}{M-m}, \frac{M-m-1}{M-m}\delta\right).$$

Putting this back into (16) completes the proof. □

The consistency and asymptotic normality of the simulated MLE using the modified Brownian bridge importance sampler is conjectured to be true. However, a rigorous proof, which would extend the results of Pedersen (1995) and Brandt and Santa-Clara (2002b), is not straightforward and is worth further investigation.

4. COMPUTATIONAL ISSUES

In this section we discuss some computational issues in the simulation approach and provide some simple but effective suggestions. To our knowledge, these issues have not been addressed in detail

for this context in the literature. However, they are very important for practitioners applying continuous time models to real data.

4.1 Faster Numerical Transformation

Transformation to unit volatility has been shown to be important for both the simulation method and the closed-form method. When $h_\theta(x)$ is not available in closed form, one has to resort to numerically computing the forward transformation $y = h_\theta(x)$ and the backward transformation $x = h_\theta^{-1}(y)$. This can be done on a fine grid of equally spaced x values, $x_1, \ldots, x_g$, where g is the size of the grid. The corresponding value of $y_1, \ldots, y_g$ can be computed from

$$\int_c^x \frac{1}{\sigma(u,\theta)}\,\mathrm{d}u,$$

where the constant c must be chosen to be smaller than x_1. Since the grid of x is equally spaced, the forward transformation for any value of x_0 is very easy to find: locate the interval $[x_i, x_{i+1}]$ that contains x_0 and do a linear interpolation between y_i and y_{i+1}. To compute the backward transformation for any value of y_0, one ends up with a bisection search to find the interval $[y_i, y_{i+1}]$ in which y_0 falls and then interpolate between x_i and x_{i+1}. In the simulation approach, this backward transformation is needed for every augmented data point, which slows the algorithm down dramatically for large amount of augmentation.

A simple way to speed up the backward transformation is to compute a second grid for the backward transformation with equally spaced values of y this time. Consider the grid values $y'_1, \ldots, y'_g$. The corresponding values of $x'_1, \ldots, x'_g$ can be computed from the first grid efficiently, utilizing the monotonicity property of the transformation. Note that this only needs to be done once for each evaluation of the likelihood. With this second grid, the backward transformation can be computed as fast and easily as the forward transformation. The saving in computing time is dramatic for large datasets, as the number of augmented points M and the number of importance samples K increases.

4.2 Computing Derivatives of the Simulated loglikelihood

Derivatives of the loglikelihood are important for efficient maximization and evaluation of the variance-covariance matrix of the MLE. Derivatives of the transition density are also necessary in

derivative pricing in finance. For the simulated likelihood method, computing the derivatives of the loglikelihood needs to be done numerically as well. Numerical differentiation requires evaluations of the loglikelihood at the parameters and close-by values of the parameters. A naive implementation would regenerate all the sample paths in each evaluation of the loglikelihood. This can be seen by looking into an alternative expression of $p_Y^{(M,K)}$,

$$p_Y^{(M,K)}(\Delta_i, y_i|y_{i-1}, \theta) = \frac{1}{K}\sum_{k=1}^{K} G^M(\boldsymbol{u}_{i,k}, \tilde{\mu}, t_{i-1}, t_i)\phi(u_M; u_0, \Delta_i), \tag{17}$$

where

$$G^M(\boldsymbol{u}_{i,k}, \tilde{\mu}, t_{i-1}, t_i) = \exp\left\{\sum_{m=1}^{M} \tilde{\mu}(u_{i,k,m-1})(u_{,k,m} - u_{i,k,m-1}) - \frac{1}{2}\sum_{m=1}^{M} \tilde{\mu}^2(u_{i,k,m-1})\delta\right\}. \tag{18}$$

By the chain rule,

$$\frac{\partial G^M(\boldsymbol{u}(\theta_j), \tilde{\mu}, t_{i-1}, t_i)}{\partial\theta} = \sum_{m=0}^{M} \frac{\partial G^M(\boldsymbol{u}, \tilde{\mu}, t_{i-1}, t_i)}{\partial u_m}\frac{\partial u_m}{\partial\theta_j} + \frac{\partial G^M(\boldsymbol{u}, \tilde{\mu}, t_{i-1}, t_i)}{\partial\theta_j}. \tag{19}$$

Clearly, $\partial u_m/\partial\theta_j$ is complicated since u_m depends on $y_j = h_\theta(x_j)$ for $j = i-1$ and i where the transform h_θ has θ implicitly involved.

We propose a method to compute the derivatives without regenerating the sample paths. Consider a Brownian bridge $\tilde{\boldsymbol{u}} = (\tilde{u}_0, \tilde{u}_1, \ldots, \tilde{u}_M)$ with volatility one starting with 0 at time t_{i-1} and terminating with 0 at t_i. As in Roberts and Stramer (2001), this Brownian bridge $\tilde{\boldsymbol{u}}$ is connected to our sample path $\boldsymbol{u}$ through

$$u_m = u_m(\theta) = \tilde{u}_m + \frac{(M-m)y_{i-1} + my_i}{M}, \qquad m = 0, 1, \ldots, M.$$

Note that $\tilde{\boldsymbol{u}}$ is free of θ. Thus,

$$\frac{\partial u_m}{\partial\theta_j} = \frac{M-m}{M}\frac{\partial y_{i-1}}{\partial\theta_j} + \frac{m}{M}\frac{\partial y_i}{\partial\theta_j}, \quad m = 0, \ldots, M.$$

The rest of the calculation is straightforward now. For example, let X be a CIR process defined by

$$\begin{aligned} \mu(x;\theta) &= \theta_1 - \theta_2 x, \\ \sigma(x;\theta) &= \theta_3\sqrt{x}, \end{aligned} \tag{20}$$

where $\theta = (\theta_1, \theta_2, \theta_3)^T$. Then $Y_t = h_\theta(X_t) = 2\sqrt{X_t}/\theta_3$ and

$$\frac{\partial u_m}{\partial\theta_3} = \frac{M-m}{M}\left(\frac{-2\sqrt{x_{i-1}}}{\theta_3^2}\right) + \frac{m}{M}\left(\frac{-2\sqrt{x_i}}{\theta_3^2}\right).$$

For analytically reducible cases, such as the CIR example, a closed-form expression for the derivatives can be obtained given the simulated paths of $\tilde{\boldsymbol{u}}$. For processes that are only numerically reducible, $\partial y_i/\partial\theta_j$ would still need to be obtained numerically.

4.3 Choosing Starting Values of Estimates

Any iterative optimization routine needs good starting values, regardless of the methods we use to compute the likelihood. Some results reported in the literature can not be reproduced unless a set of deliberately chosen starting values are used. It is worth discussing the choice of starting values explicitly in the univariate diffusion context.

Consider the general drift and diffusion model in Aït-Sahalia (1996)

$$\begin{aligned} \mu(x;\theta) &= \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3/x, \\ \sigma(x;\theta) &= \sqrt{\beta_0 + \beta_1 x + \beta_2 x^{\beta_3}}, \end{aligned} \tag{21}$$

where $\theta = (\alpha_0, \ldots, \alpha_3, \beta_0, \ldots, \beta_3)^T$. Aït-Sahalia (1996) gave a two-stage feasible generalized least squares (FGLS) estimator from discretizing the continuous-time moments of the model. In the first stage, a linear regression for the drift equation is fitted with the crude discretizing of the data, $E(\nabla x_t|x_{t-1}) = \alpha_0+\alpha_1 x_{t-1}+\alpha_2 x_{t-1}^2+\alpha_3/x_{t-1}$, where $\nabla x_t = x_t - x_{t-1}$; The squared residuals ϵ_t^2's are used to fit the diffusion equation using non-linear least squares, $E(\epsilon_t^2|x_{t-1}) = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-1}^{\beta 3}$. In the second stage, the fitted values from the diffusion equation are used to form weights to refit the linear regression for the drift equation; The diffusion equation can also be updated based on the updated squared residuals from the drift equation. This two stage FGLS estimator has built-in discretization bias, which gets more severe for lower sampling frequencies. Nevertheless, it can be used as a reasonable starting point for the maximization of the likelihood.

There are two potential problems with this two-stage FGLS procedure. First, in some range, the best fitting diffusion model might give negative fitted values which cannot be used as weights. Second, the non-linear least square estimation itself needs initial values. We propose the following procedures to solve these issues. Instead of using the fitted values from the diffusion equation as weights to update the parameters in the drift equation, we use the fitted values from a local smoother which is readily available in most statistical software. This method provides robust weights and, therefore, robust starting estimate for the drift parameters. To avoid finding initial

values for non-linear least squares procedure, notice that for any fixed β_3, the diffusion equation is linear in the rest of the parameters. We can then work on a grid of values of β_3, using a linear regression to estimate β_0, β_1, and β_2 for each β_3. All these estimates can be used as starting values for diffusion parameters. We can also evaluate the likelihood at each set of the starting values and choose the one with the highest likelihood to start with. This method works well in the numerical application in Section 6 with daily observations.

5. COMPARING LOG TRANSITION DENSITY

We now compare the logarithm of the transition density computed from both the simulation approximation and the closed-form approximation. The first-order Euler approximation is also supplied in the comparison to illustrate situations where it performs well. The benchmark model is the general model (21) of Aït-Sahalia (1996), denoted by GEN4 in Durham (2003). This model has unknown transition density. It has non-linear drift and no closed-form variance stabilizing transformation. The simulation approximation is implemented in `C` and interfaced to `R` (R Development Core Team 2006). The closed-form expansions are obtained in Mathematica and exported into `R`.

The performance of the simulation method obviously depends on M, the number of subintervals, and K, the number of importance sampling paths. Increasing K reduces the variance of the approximation for any fixed M. Increasing M reduces the bias for any fixed K. When $M = 1$, the method reduces to the first-order Euler scheme. For the modified Brownian bridge sampler, Stramer and Yan (2005) recently established the trade-off between M and K under a given computing resource. The optimal resource allocation is shown to be $M = \sqrt{K}$ for a given computing time, and the approximation error is of order $1/M$. Four values of M are used: 2, 4, 8, and 16, each with the corresponding optimal $K = M^2$.

The performance of the simulation approach also depends on whether or not transforming the process to one with unit variance. As the variance stabilizing transformation is only numerically available, we also consider the widely used log transformation as a comprise. The log transformation turns out to give reasonably good results which are much better than those without transformation. The simulation approximation without transformation, with numerically variance stabilizing transformation, and with log transformation are all reported.

The closed-form approximations we use are of the first order and the second order, obtained from Mathematica by solving the Kolmogorov equations in Aït-Sahalia (2005). A Mathematica module is available at request. For a model as complex as model (21), the first and second order closed-form approximation generated 34 and 207 lines of Fortran code from Mathematica, respectively.

To make the comparison of practical interest, we use the estimated models from Durham (2003) for two well-known datasets: daily three-month treasury bill rate and daily seven-day Eurodollar rate; see more details about the data in Section 6. Durham (2003) illustrated that these two series behave very differently, with the Eurodollar rate exhibiting much higher volatility than the treasury bill rate. The parameters used are the estimates of the GEN4 model taken from Table 3 and Table 4 of Durham (2003), albeit this model was not found to be superior to the alternative model with constant drift (model GEN1 in Durham 2003). The parameters for the two models are presented in Table 1, labeled as low volatility and high volatility, respectively. For each model, we conduct our comparison for four levels of transition interval Δ: 1/52, 1/12, 1/4, and 1/2, corresponding to weekly, monthly, quarterly, and semi-annual observations, respectively. This is different from the existing literature which has mostly considered $\Delta = 1/12$.

[Table 1 about here.]

Two issues needs to be clarified before we present the comparison results. The first issue is, as the true density is unknown, how to assess the approximation error. This issue is approached by using the limit of the log density obtained in the simulation approximation with very large M and K as a substitute. These values are chosen to be $M = 256$ and $K = M^2$, combined with the numerical variance stabilizing transformation. Since the simulation approximation is stochastic, the limit was computed as the mean of 64 replications to ensure a good estimate of this limit. We use the term "pseudo approximate errors" to denote the approximate errors obtained using this estimated limit in place of the unknown truth. The second issue is how to choose the initial state x_0 and the range of x to do the comparison. We choose the initial state x_0 to be the median of the observed daily three-month treasure bill rate and seven-day Eurodallor rate, respectively. The range of the current state x is different under different observation frequency. We limit our attention to the interior 95% of the Δ-ahead transition distribution starting from x_0. The two endpoints of the range are chosen to be the 2.5 and 97.5 percentiles of 100,000 simulated observations from the

transition distribution. We then compare the approximations at 40 equally spaced points including the two endpoints in this range.

Figures 1–3 illustrate the comparison of pseudo approximation error of the log-transition-density for the low volatility model, where the simulation approach uses no transformation, numerical variance stabilizing transformation, and log transformation, respectively. The initial state is $x_0 = 6.076$. Each column in these figures has the same Δ. The (blue) dashed lines and the (green) long-dashed lines give respectively the pseudo approximation error of the first-order and second-order closed-form approach, which are the same within each column. Unlike the closed-form approximation approach, the approximation from the simulation approach is stochastic. We plot the pseudo approximation errors of 100 replicates in gray lines, their median in solid lines, and their 25th and 75th percentiles in dotted lines. The pseudo approximation errors of the first-order Euler scheme are plotted in (red) dot-dashed lines.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

Several observations can be made from Figures 1–3. Both approaches give better approximation than the first-order Euler scheme. By construction, the transition interval Δ is important for the performance of the closed-form approach. For this low volatility model, the second-order closed-form approximation is extremely accurate for all Δ values. The first-order closed-form approximation is very accurate for Δ as big as $1/4$ (observation frequency as low as quarterly), as the maximum pseudo error within the comparison range is only 0.05 at the right tail. At $\Delta = 1/2$ (semi-annual observation), the first-order closed-form approximation is very accurate in the center part but is not as accurate at the tail of the distribution. For the simulation approach, we see that the benefit of variance stabilizing transformation is striking by comparing the ranges of the pseudo errors in Figure 1 and 2. For a fixed configuration of Δ, M, and K, transformation results in dramatic reduction in the variation and bias of the simulated log density. With reasonably small values of M and K, the simulation approach using variance stabilizing transformation can have lower pseudo approximation error than the first-order closed-form approach, particularly at

the tails of the transition distribution and for bigger Δ values. The simulation approach with no transformation, however, can only beat the first-order closed-form approach with much larger values of M and K. For example, compare the approximation error of the log density for $\Delta = 1/2$ at these two configurations: $(M, K) = (4, 16)$ with transformation and $(M, K) = (16, 256)$ without transformation. The approximation errors under the two settings are comparable. However, the one with transformation uses much less computing resource. The simulation approximation with log transformation yields accuracy in the middle of the other two versions. For all versions of simulation approach, given Δ, one can see that, increasing M and K reduces the approximation bias and variance, particularly at the tails.

Figures 4–6 illustrate the same comparison for the high volatility model. The initial state is $x_0 = 8.111$. Similar observations can be made about the importance of transformation, transition interval, the number of auxiliary points M, and the number of sampled paths K. Some further observations are induced by the high volatility nature of the model. The difference between the two methods is visible even for $\Delta = 1/52$ (weekly observations). The second-order closed-form approximation is still very accurate for weekly data. As Δ increases (or as the observation frequency decreases), the pseudo approximation error of the closed-form approach, both first-order and second-order, gets larger relative to the simulation approach with the numerical variance stabilizing transformation. This is particularly true at the tails of the density for larger Δ. The properties of the simulation approach is, in contrast, not constrained by Δ from its construction, as long as M and K is sufficiently large. The simulation approximation is hence more robust to Δ and the volatility nature of the model.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

For both cases of low volatility and high volatility investigated here, it is interesting to note that the log transformation produces reasonably good comprises to the numerical variance stabilizing transformation. Therefore, log transformation may be used as a simpler alternative to the numerical transformation for positive processes which are only numerically reducible. It has the

advantage that the derivatives are easier to evaluate than those from the numerical transformation. These computational and numerical advantages are obtained without too much loss in accuracy, making log transformation an important tool in analyzing positive diffusions. As pointed out by an anonymous referee, this observation is not completely surprising because it is clear from the construction of the simulation approach that it is likely to perform better when the diffusion process is more slowly varying.

The two approaches differ substantially in their computational cost. On the log scales, Figure 7 presents the accuracy and speed comparison of the two approaches. Both low volatility and high volatility cases are presented. The computation is done on a Linux machine with a 2.4GHz CPU. Simulation approaches with no transformation, numerical transformation, and log transformation are all reported, from the top to the bottom. For each approach, the four points in each plot correspond to Δ values of 1/52, 1/12, 1/4, and 1/2, respectively, from the bottom up. The first-order closed-form approximation is very efficient in time, but the pseudo approximation error can be high, particularly in the high volatility case. The second-order closed-form approximation is extremely accurate in the low volatility case and for smaller Δ values. Its computational cost is less than the simulation method with $(M, K) = (4, 16)$ using the numerical transformation, but gives more accuracy in the interior 95% of the density. The high uniform absolute approximation error of the closed-form approach is essentially caused by the discrepancy at the right tail of the distribution. These observations raise the question of how to assess whether or not a "sufficiently accurate" closed-form approximation has been obtained for a given problem. At this point, there is no clear answer to this question.

[Figure 7 about here.]

We end this section with an artificial CIR example, a special case of model (21). The variance stabilizing transformation is analytically available, but we apply the two approximation approaches as if the analytical transformation is not known. True approximation errors can be computed in this case since the true density is known. This example confirms the finding that the closed-form approximation can give inaccurate result compared to the simulation approach when Δ is big. Consider the CIR model of (20) with $\theta = (0.3, 0.6, 0.5)^T$. This model specifies a high mean reverting process, with unconditional mean 0.5. We compare the approximation errors from the

two approaches in Figure 8 for Δ values at 1/12, 1/4, 1/2, and 1. The simulation approach uses numerical variance stabilizing transformation and takes (M, K) at $(8, 64)$ and $(32, 1024)$. It is interesting to note where closed-form approach stop being accurate. In the case when $\Delta = 1/2$ or 1, the closed-form approach is seen to perform not as well at the two tails of the density. The caution here is that the closed-form approximation is developed assuming that Δ is small. This raises the question of "how small is small". For financial data, daily data (probably weekly and monthly) can be considered to have small Δ, and the closed-form approach works well elegantly. But for low frequency data from other fields (Taylor, Cumberland, and Sy 1994), there is not much existing empirical studies to give guidelines. In that case, the simulation approach is more robust and can be trusted with larger M and K values. Low frequency data usually have fewer observations, which implies that the computation with large M and K can be practically affordable.

[Figure 8 about here.]

6. APPLICATIONS

We now apply the simulation approximation, the closed-form approximation, and the first-order Euler approximation to the two well known datasets in the finance literature: daily three-month treasury bill rate and daily seven-day Eurodollar rate. The treasury bill rate has much lower volatility than the Eurodollar rate. They have been analyzed by Durham (2003) using the simulated MLE. To have more control over the comparison, we use our own implementation of the simulation approach and the results are only slightly different from Durham (2003).

6.1 Daily Three-month Treasury Bill Rate

The three-month treasury bill rate data consists of 7,555 daily observations spans from January 4, 1965 to July 28, 1995. We use $\Delta = 1/252$ to correspond to daily observations. Two models are fitted: a full model (21) of Aït-Sahalia (1996) and a constrained model of (21) with $\alpha_1 = \alpha_2 = \alpha_3 = 0$. We denote them by GEN4 and GEN1, respectively, as in Durham (2003). The simulation approach uses the numerical variance stabilizing transformation, and (M, K) is chosen to be $(8, 64)$. The approximated loglikelihood function is maximized using the Nelder-Mead algorithm, which does not require gradient and is found to give the most robust results. The variance matrices of

the estimates are obtained by inverting the negative Hessian matrix of the loglikelihood function at the maximum. The results are presented in Table 2.

[Table 2 about here.]

Both methods gives similar parameter estimates and maximum loglikelihood for both GEN1 and GEN4. GEN1 is preferred to GEN4 in terms of AIC or SBC, since the increase in loglikelihood brought by GEN4 does not compensate the model complexity. As noticed in Durham (2003), the process of treasury bill rate has low volatility. Therefore, not surprisingly, the first-order Euler approximation also gives very close parameter estimates and maximum loglikelihood. The estimated diffusion function of GEN1 from the three methods are compared in Figure 9. They are very close to one another.

[Figure 9 about here.]

In our experiment, we noticed that the standard errors for parameter estimates can be difficult to obtain with smaller M and K in the simulation approach. Numerically differentiating the loglikelihood function is already a difficult problem for the simulation approach. (Aït-Sahalia 2002b). Numerically computing the variance stabilizing transformation makes the problem even harder. Closed-form approximation clearly has an advantage here.

To have a better idea on the computational cost of the two methods in practice, we summarize the time consumed in seconds for 100 evaluations of the loglikelihood for this dataset in Table 3. Maximizing the loglikelihood can involve thousands of evaluations of the loglikelihood. Given the similar estimation results in Table 2, the time consumed by the first-order closed-form approximation is much shorter than other approximations.

[Table 3 about here.]

6.2 Daily Seven-day Eurodollar Rate

The Eurodollar rate data consists of 5,505 daily observations of the seven-day Eurodollar deposit spot rate from June 1, 1973 to February 25, 1995. We did the same comparison as for the treasury bill rate data. The results are compared in Table 4.

[Table 4 about here.]

The same conclusion of Durham (2003) is obtained from the closed-form approximation approach: The GEN1 model is preferred to the GEN4 model. In contrast to the case of the treasure bill rate data, where all approximations lead to close MLEs and loglikelihood, the Eurodollar data reveals how these methods can yield noticeably different results. We first notice that the maximized loglikelihood are different across different approximation methods. For example, for the GEN1 model, the simulation approach using numerical transformation with $(M, K) = (8, 64)$ yields -941, the first-order closed-form yields -961, the second-order closed-form yields -951, and the first-order Euler scheme yields -1003.3. These differences indicate that the problem is difficult. The maximized loglikelihood from first-order and second-order closed-form approximations differ by about 10 points. The second-order closed-form result is closer to the simulation result than the first-order, but still with a difference of about 10 points. Put in the context of likelihood ratio, a 10 point difference could be a significant difference. We also note that the parameter estimates for the GEN4 model from all approximations are noticeably different in β_2 and β_3, even after considering their standard errors. For example, the point estimate of β_3 is 4.221 with standard error 0.024 from the simulation approach, but 4.011 with standard error 0.034 from the second-order closed-form. Although the difference may not be of practical significance, they are 2 standard errors apart from each other. Given all these different results, how would one choose among them? A simple but computing intensive solution is to increase M and K and see what happens. We tried $(M, K) = (16, 256)$ and the result does not differ much from that with $(M, K) = (8, 64)$. Therefore, for this more difficult problem, the simulation method has an advantage in that it gives a way to assess the accuracy against itself and against the closed-form method. This example suggests that, even for daily observations, the approximated loglikelihood can have a significant difference across different approximations for large dataset. These differences may be caused by the performance of the closed-form method at the tails of the transition density. Large dataset and heavy tailed transition density increase the frequency of extreme observations.

The estimated diffusion function in the GEN1 model from the three methods are plotted in Figure 9. They are quite close for the majority of the span of the data: The 5th and 95th sample percentiles are 3.168 and 15.589.

It is worth noting that these parameter estimates are quite different from those in Aït-Sahalia (1996), which were obtained by minimizing the distance between the parametric marginal density to a nonparametric density estimate. Aït-Sahalia (1996) imposes stationarity conditions that puts restrictions on the possible shape of the drift (see also Jones 2003, p.794). When the maximum likelihood without constraints is used, the closed-form approach of Aït-Sahalia (2005) gives close results to those from the simulation approach (Durham 2003), at least in the preferred GEN1 model.

7. DISCUSSION

In this paper we focus on the simulation method (Durham and Gallant 2002) to approximate the likelihood of discretely observed diffusion processes and compared it to the closed-form approximation (Aït-Sahalia 2005). The theoretical justification of using the modified Brownian bridge sampler proposed by Durham and Gallant (2002) in the simulated method has not been discussed before in the literature. For the case of constant volatility, we provide a rigorous justification for the advantage of using the modified Brownian bridge sampler over Euler approximation to a Brownian bridge diffusion process. For the case of non-constant volatility, further research is ongoing.

Our simulation study illustrates the importance of transforming the process of interest to one with constant volatility. Durham and Gallant (2002) has established this importance with the CIR process, where such a transformation is analytically available. In general, when this transformation cannot be done analytically, the gain in accuracy is well worth the cost of numerical transformation. We illustrate this point from a different angle in Figure 10, where the simulation methods with and without transformation are calibrated such that they use approximately the same computing time. Pseudo approximation errors of three implementations (top to bottom: no transformation, numerical transformation, and log transformation) are compared at various transition interval Δ for the low volatility model in Table 1. The configuration is similar to those illustrated in Figure 1–3. The three methods used 65, 62, and 65 seconds to evaluate the log density at 40 points for 100 replicates, respectively. The method with numerical transformation has much better accuracy. Note that the method with numerical transformation used $(M, K) = (16, 256)$ while the method without transformation used $(M, K) = (20, 400)$. Since the numerical transformation only needs

to compute the transformation grid once for each loglikelihood evaluation (see Section 4), this fixed cost takes an even smaller proportion of computing time for larger datasets. We therefore suggest that transforming to constant volatility be done whenever possible for univariate diffusions. The Brownian bridge sample, however, is general and can provide reasonably accurate results even without transformation. This is particularly important for irreducible multivariate diffusions.

[Figure 10 about here.]

Our comparison of the simulation method and the closed-form method has two distinguishing features. First, the benchmarking process (21) has unknown transition density and has no closed-form variance stabilizing transformation. Second, the simulation approach uses the modified Brownian bridge sampler which has been shown to be more efficient than other samplers. Based on the following observations from our study, the two approaches are practically complement to each other in terms of accuracy and computing speed.

The simulation approximation is stochastic. For the one-dimensional case, it should be applied with transforming the process to one with unit volatility for more accurate approximation. The simulation method is robust to the size of Δ. The approximation error can be controlled by increasing the number of subintervals M and the number of simulated sample paths K. For large values of M and K, the approximation is known to converge to the unknown truth. The issue is how much computing resources one can afford. For a given amount of computing resource, the optimal allocation of M and K is such that $K = M^2$ (Stramer and Yan 2005). The choice of M can be done by trial-and-error until an acceptable accuracy is obtained. The final approximation can be used to check the performance of closed-form approximation when the true transition density is unknown. This approach is simple to implement and is no stranger to many practitioners. It can be easily enhanced using various numerical techniques (Brandt and Santa-Clara 2002a). Compared to the closed-form approach, it is at a disadvantage in speed and accuracy for small Δ values. As illustrated in Section 5, these small Δ values covers monthly observations, and even quarterly data if the volatility is low. Another disadvantage is that it can be difficult differentiating the loglikelihood or transition density, which is important in maximizing the likelihood, getting standard errors of the estimates, and applications such as pricing derivatives. The difficulty roots in the fact that the

approximated density is not smooth. The problem gets even harder when the variance stabilizing transformation is not available in closed-form and has to be computed numerically.

The closed-form approximation works well for smaller Δ values. The first-order approximation is very fast and the second-order approximation can be extremely accurate for small transition interval Δ. It has the advantage of being non-stochastic, which makes it easy to differentiate the loglikelihood or the transition density. Expressions of the closed-forms can be obtained with software which can do symbolic calculation and export the generated code in Fortran or `C` format. A Mathematica module has been written for this purpose. One critical drawback of the closed-form approximation is that it is very difficult to assess the accuracy for a given problem. The method can only be applied to small values of Δ as it is a Taylor series expansion near $\Delta = 0$. Increasing the order of approximation does not necessarily solve this problem. The theory (Aït-Sahalia 2002a) guarantees its convergence to the true density for $\Delta \in (0, \overline{\Delta})$ under appropriate conditions. This threshold $\overline{\Delta}$ depends on both the drift and the volatility. The determination of the threshold $\overline{\Delta}$, however, can be difficult in practice. The question of "how small a Δ is small" can practically be answered with extensive comparison studies using the simulation method. Our numerical illustration shows a higher volatility model demands a smaller $\overline{\Delta}$ for the closed-form approximation to work well. From a practical point of view, we are fortunate that financial data often has monthly or higher frequency, that is $\Delta \leq 1/12$, but in other applications, for example, longitudinal AIDS data (Taylor et al. 1994), where the observation frequency is low, one should be cautious and should check the approximation against the simulation method. For a given value of transition interval Δ, our simulation results suggest that the closed-form approximation have larger errors further in the tails. The study only compares the log transition densities over the interior 95% of the distribution. As pointed out by a referee, for a dataset of 5000 observations, about 250 will occur outside of this region, which may lead to a fair amount of error in the loglikelihood. Therefore, when the tail of the transition density gets heavier, the performance of the closed-form approximation deteriorates and needs to be watched for.

The extension of the modified Brownian bridge to multivariate diffusions has been applied in Golightly and Wilkinson (2005). The closed-form approximation for general multivariate diffusions is in Aït-Sahalia (2005). A comparison study of the two different approaches in the multivariate

setting will be interesting.

REFERENCES

Aït-Sahalia, Y. (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

— (1999), "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54, 1361–1395.

— (2002a), "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach," *Econometrica*, 70, 223–262.

— (2002b), "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes: Comment," *Journal of Business & Economic Statistics*, 20, 317–321.

— (2005), "Closed-Form Likelihood Expansions for Multivariate Diffusions," Working paper, Department of Economics - Bendheim Center for Finance, Princeton University.

— (2006), "Estimating Continuous-Time Models with Discretely Sampled Data," Working paper, Department of Economics - Bendheim Center for Finance, Princeton University.

Bekos, A., Papaspliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006), "Exact and Computationally Efficient Likelihood-based Estimation for Discretely Observed Diffusion Processes (with disscusion)," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 68, 333–382.

Brandt, M. W. and Santa-Clara, P. (2002a), "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes: Comment," *Journal of Business & Economic Statistics*, 20, 321–324.

— (2002b), "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets," *Journal of Financial Economics*, 63, 161–210.

Chib, S., Pitt, M. K., and Shephard, N. (2004), "Likelihood Based Inference for Diffusion Driven Models," Working Paper 2004-W20, Economics Group, Nuffield College, University of Oxford.

Chib, S. and Shephard, N. (2002), "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes: Comment," *Journal of Business & Economic Statistics*, 20, 325–327.

Durham, G. B. (2003), "Likelihood-based Specification Analysis of Continous-time Models of the Short-term Interest Rate," *Journal of Financial Economics*, 70, 463–487.

Durham, G. B. and Gallant, A. R. (2002), "Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes," *Journal of Business & Economic Statistics*, 20, 297–338.

Elerian, O. (1999), "Simulation Estimation of Continuous-Time Models with Applications to Finance," Thesis, Nuffield College, University of Oxford.

Elerian, O., Chib, S., and Shephard, N. (2001), "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69, 959–993.

Golightly, A. and Wilkinson, D. J. (2005), "Bayesian Sequential Inference for Nonlinear Multivariate Diffusions," Working paper.

Jensen, B. and Poulsen, R. (2002), "Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques," *Journal of Derivatives*, 9, 18–32.

Jones, C. S. (2003), "Nonlinear Mean Reversion in the Short-Term Interest Rate," *Review of Financial Studies*, 16, 793–843.

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

Pedersen, A. R. (1995), "Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes," *Bernoulli*, 1, 257–279.

R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Robert, C. P. and Casella, G. (2005), *Monte Carlo Statistical Methods*, Springer-Verlag Inc, 2nd ed.

Roberts, G. O. and Stramer, O. (2001), "On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis-Hasting Algorithm," *Biometrika*, 88, 603–621.

Shoji, I. and Ozaki, T. (1998), "Estimation for Nonlinear Stochastic Differential Equations by a Local Linearization Method," *Stochastic Analysis and Applications*, 16, 733–752.

Sørensen, H. (2004), "Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey," *International Statistical Review*, 72, 337–354.

Stramer, O. and Yan, J. (2005), "Asymptotics of an Efficient Monte Carlo Estimation for the Transition Density of Diffusion Processes," Working paper, Department of Statistics and Actuarial Science, University of Iowa.

Sundaresan, S. M. (2000), "Continuous-Time Methods in Finance: A Review and an Assessment," *Journal of Finance*, 55, 1569–1622.

Taylor, J. M. G., Cumberland, W. G., and Sy, J. P. (1994), "A Stochastic Model for Analysis of Longitudinal AIDS Data," *Journal of the American Statistical Association*, 89, 727–736.

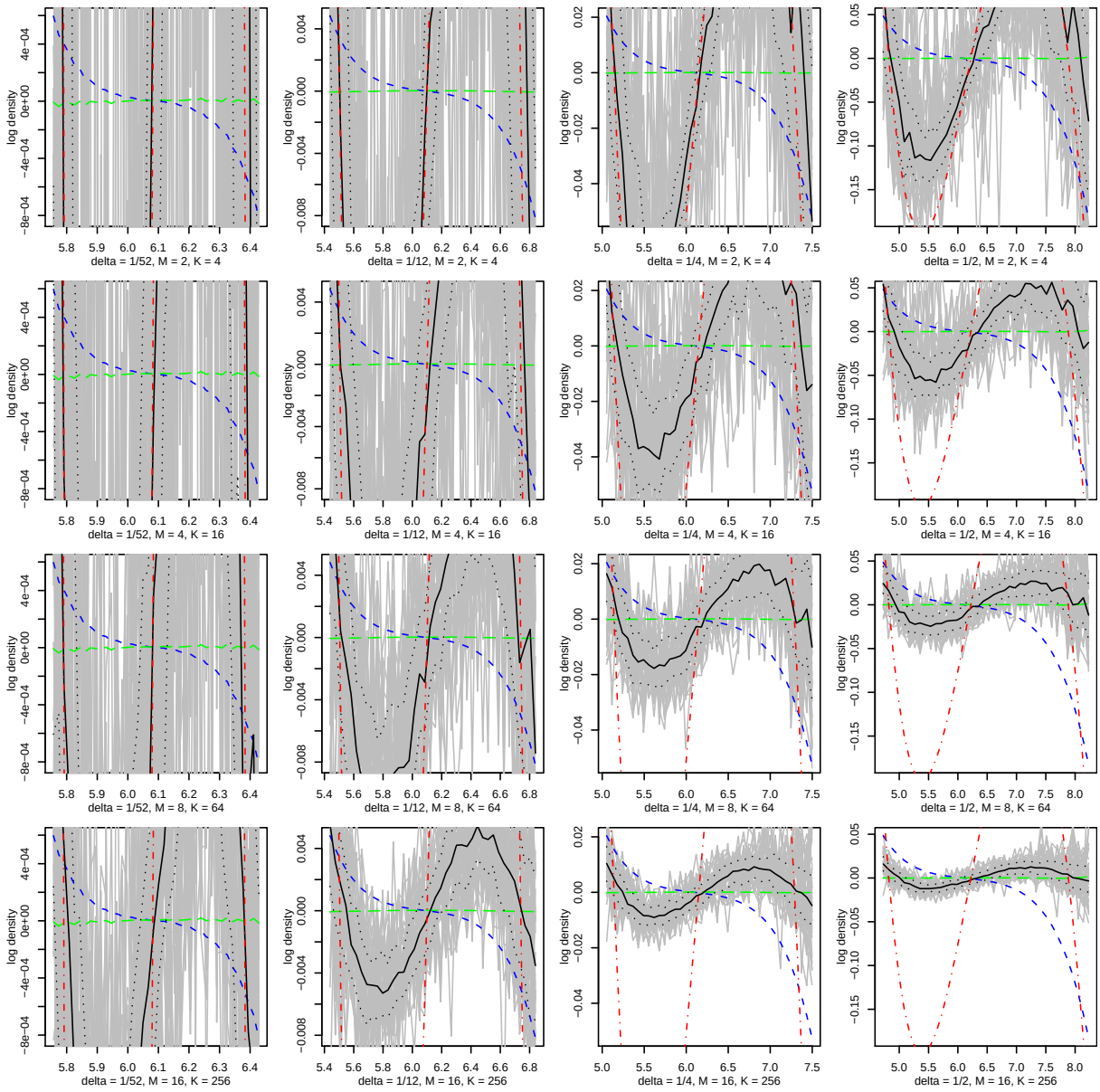

Figure 1: Pseudo approximation error comparison with the simulation approach using no transformation: low volatility case. The initial state is $x_0 = 6.076$. The (blue) dashed lines are the first-order closed-form approximation. The (green) long-dashed lines are the second-order closed-form approximation. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates. The (red) dot-dashed lines are the first-order Euler approximation.

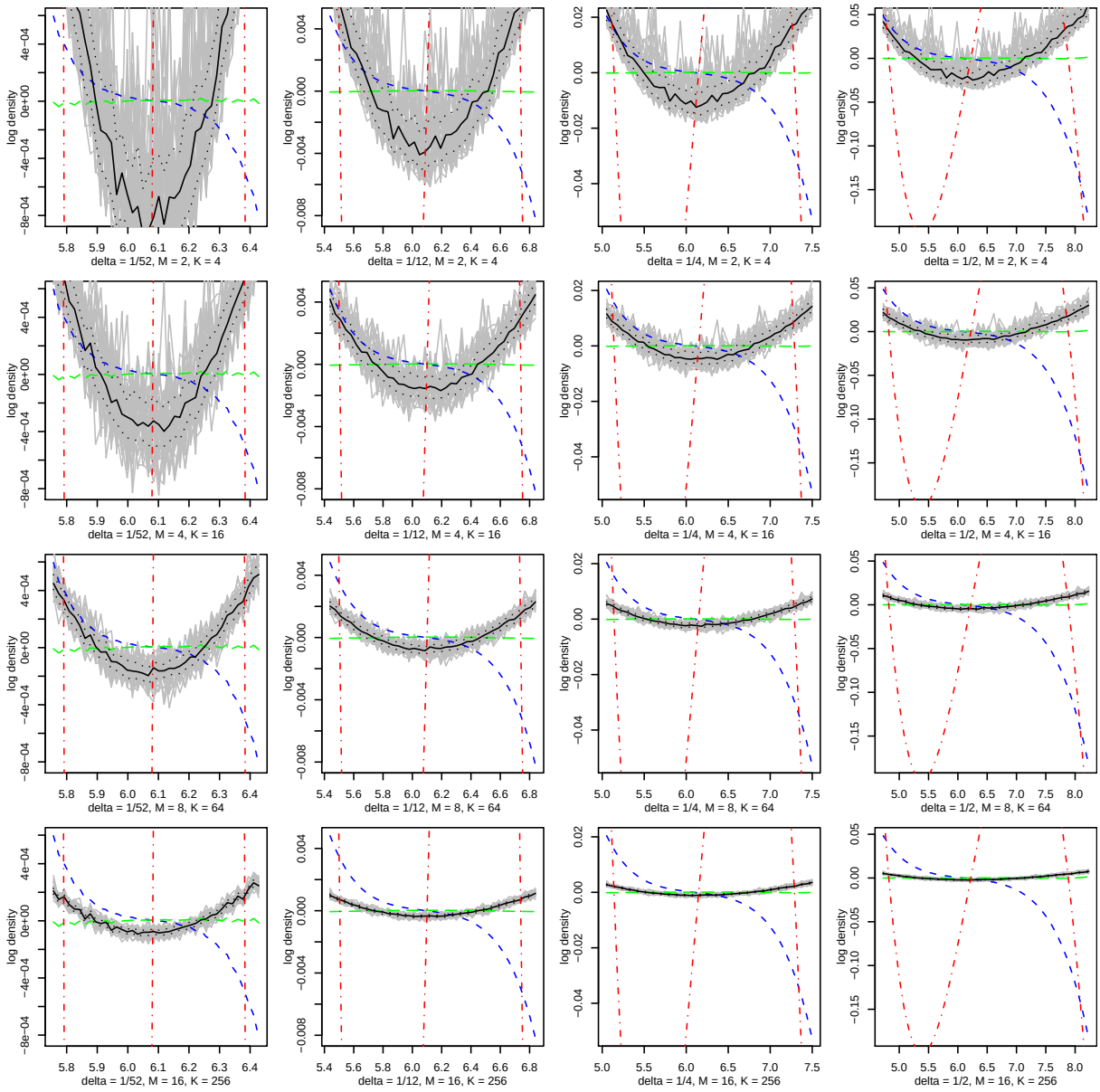

Figure 2: Pseudo approximation error comparison with the simulation approach using numerical variance stabilizing transformation: low volatility case. The initial state is $x_0 = 6.076$. The (blue) dashed lines are the first-order closed-form approximation. The (green) long-dashed lines are the second-order closed-form approximation. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates. The (red) dot-dashed lines are the first-order Euler approximation.

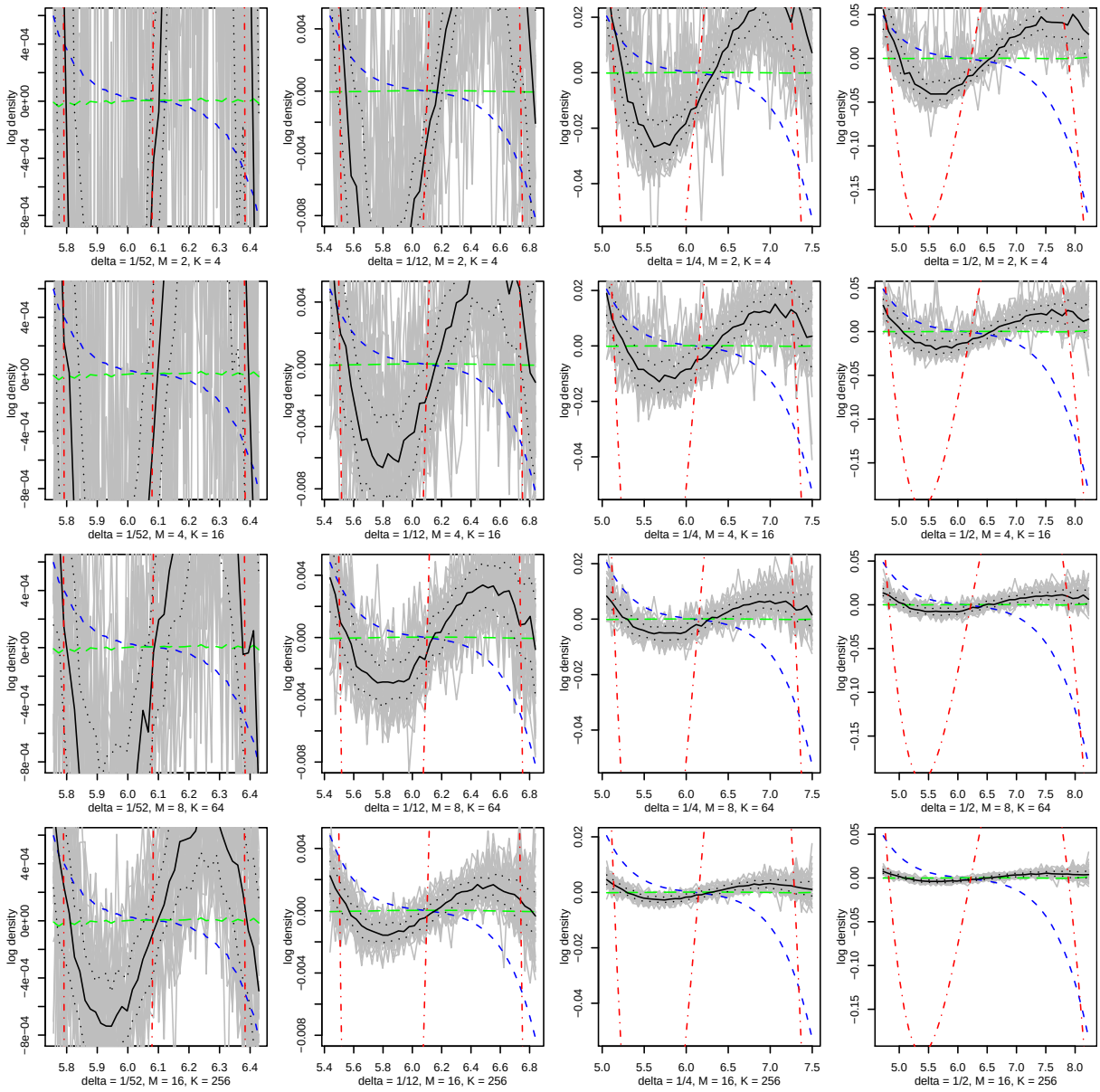

Figure 3: Pseudo approximation error comparison with the simulation approach using log transformation: low volatility case. The initial state is $x_0 = 6.076$. The (blue) dashed lines are the first-order closed-form approximation. The (green) long-dashed lines are the second-order closed-form approximation. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates. The (red) dot-dashed lines are the first-order Euler approximation.

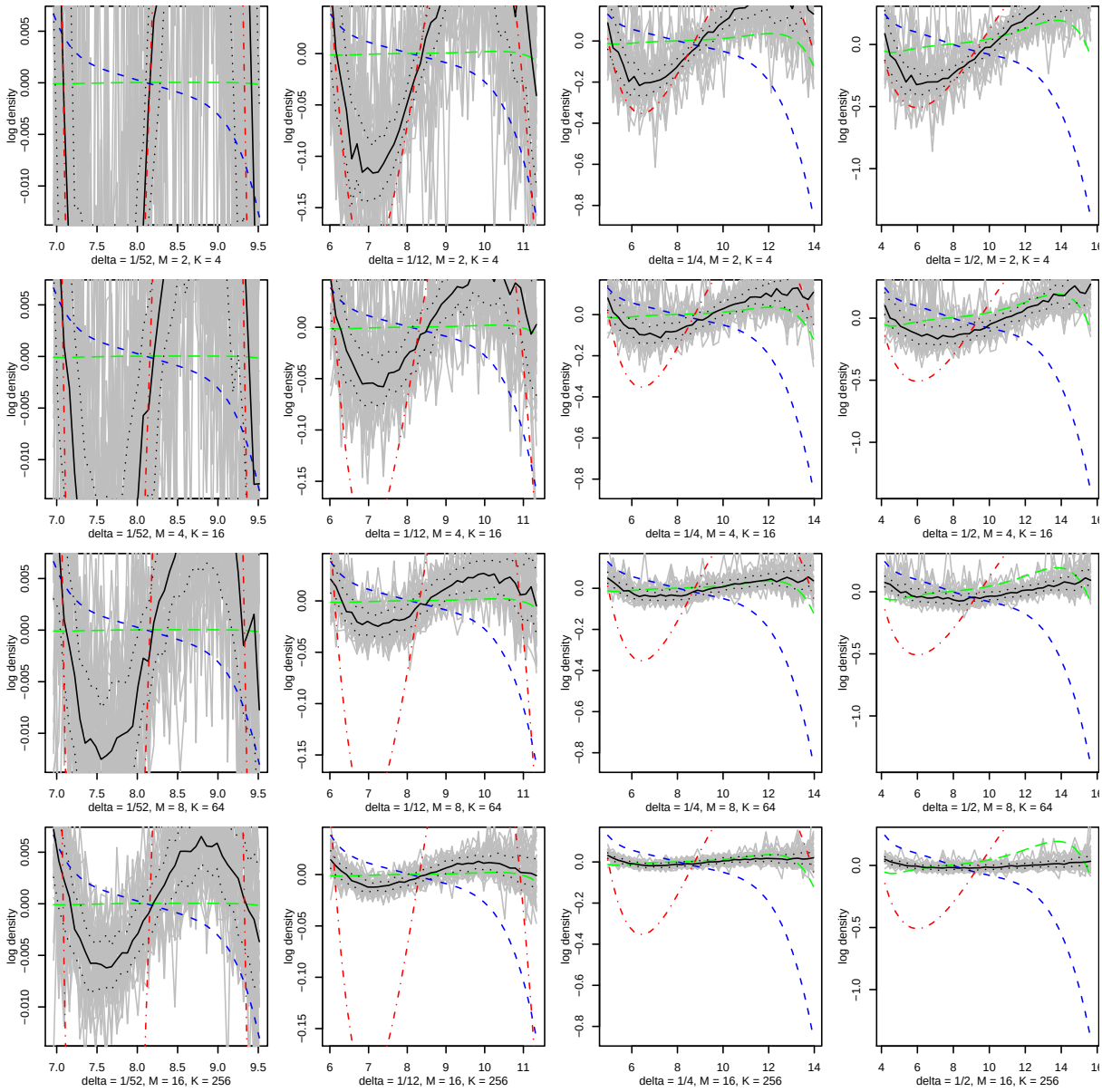

Figure 4: Pseudo approximation error comparison with the simulation approach using no transformation: high volatility case. The initial state is $x_0 = 8.111$. The (blue) dashed lines are the first-order closed-form approximation. The (green) long-dashed lines are the second-order closed-form approximation. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates. The (red) dot-dashed lines are the first-order Euler approximation.

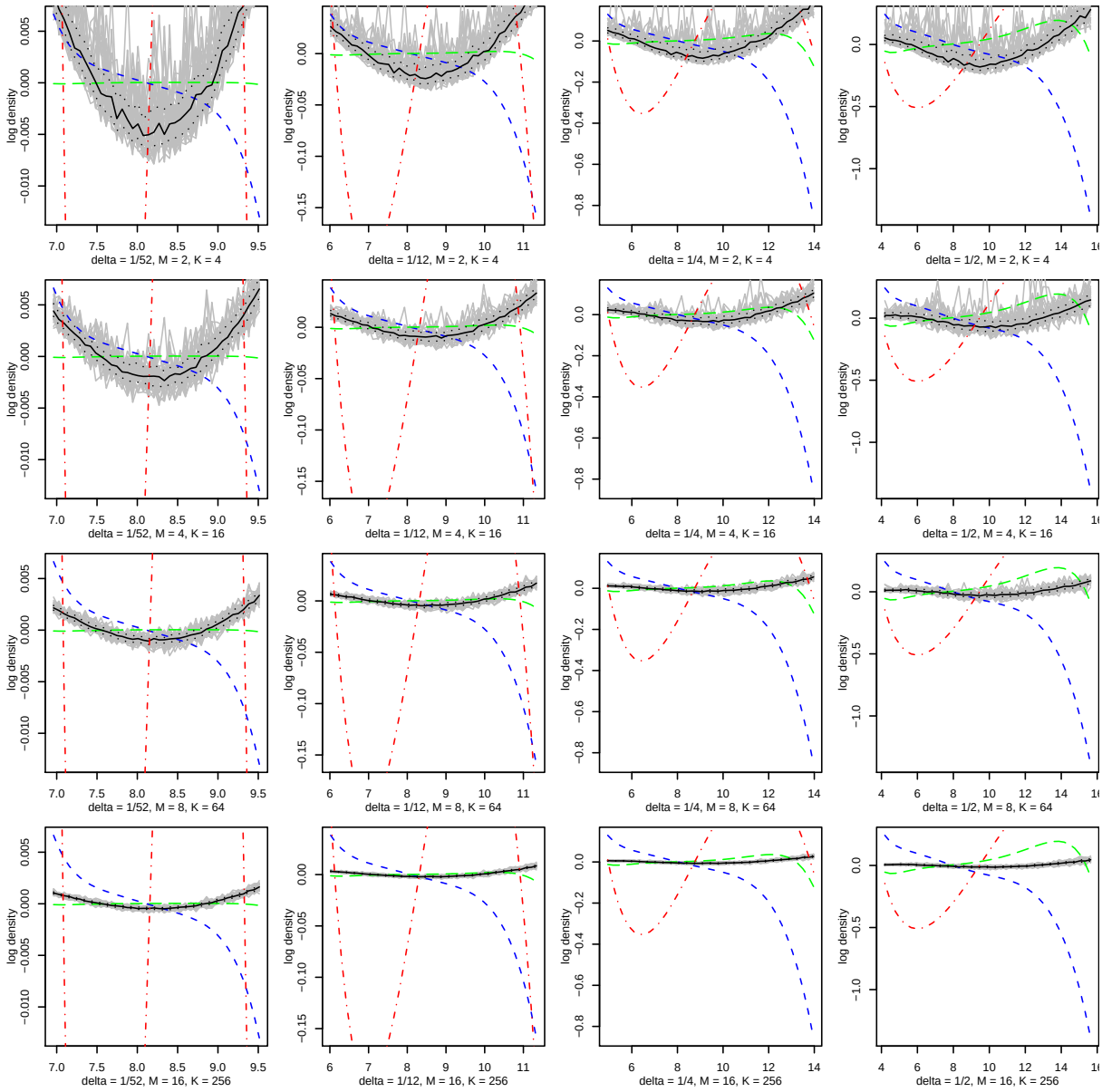

Figure 5: Pseudo approximation error comparison with the simulation approach using numerical variance stabilizing transformation: high volatility case. The initial state is $x_0 = 8.111$. The (blue) dashed lines are the first-order closed-form approximation. The (green) long-dashed lines are the second-order closed-form approximation. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates. The (red) dot-dashed lines are the first-order Euler approximation.

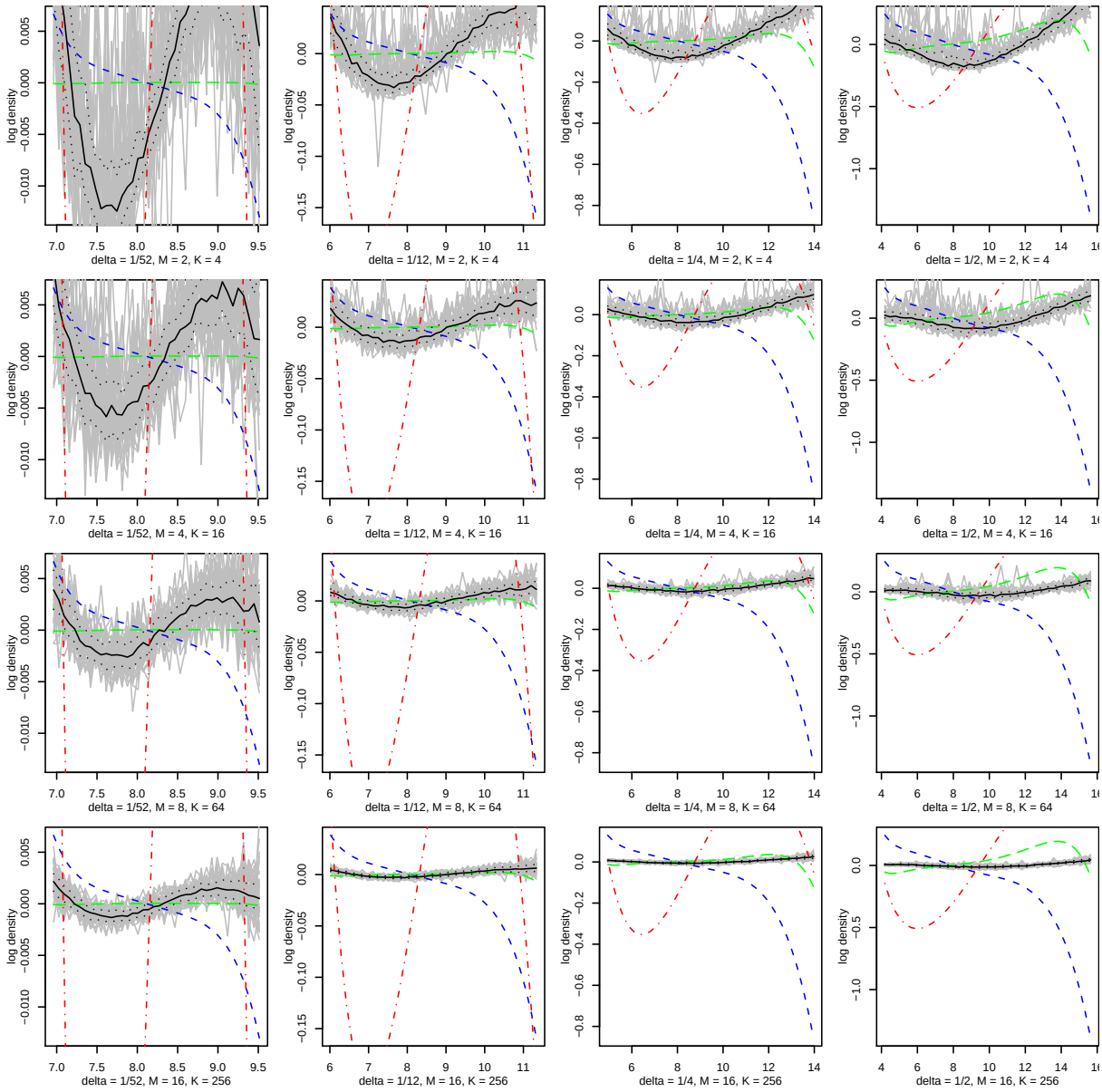

Figure 6: Pseudo approximation error comparison with the simulation approach using log transformation: high volatility case. The initial state is $x_0 = 8.111$. The (blue) dashed lines are the first-order closed-form approximation. The (green) long-dashed lines are the second-order closed-form approximation. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates. The (red) dot-dashed lines are the first-order Euler approximation.

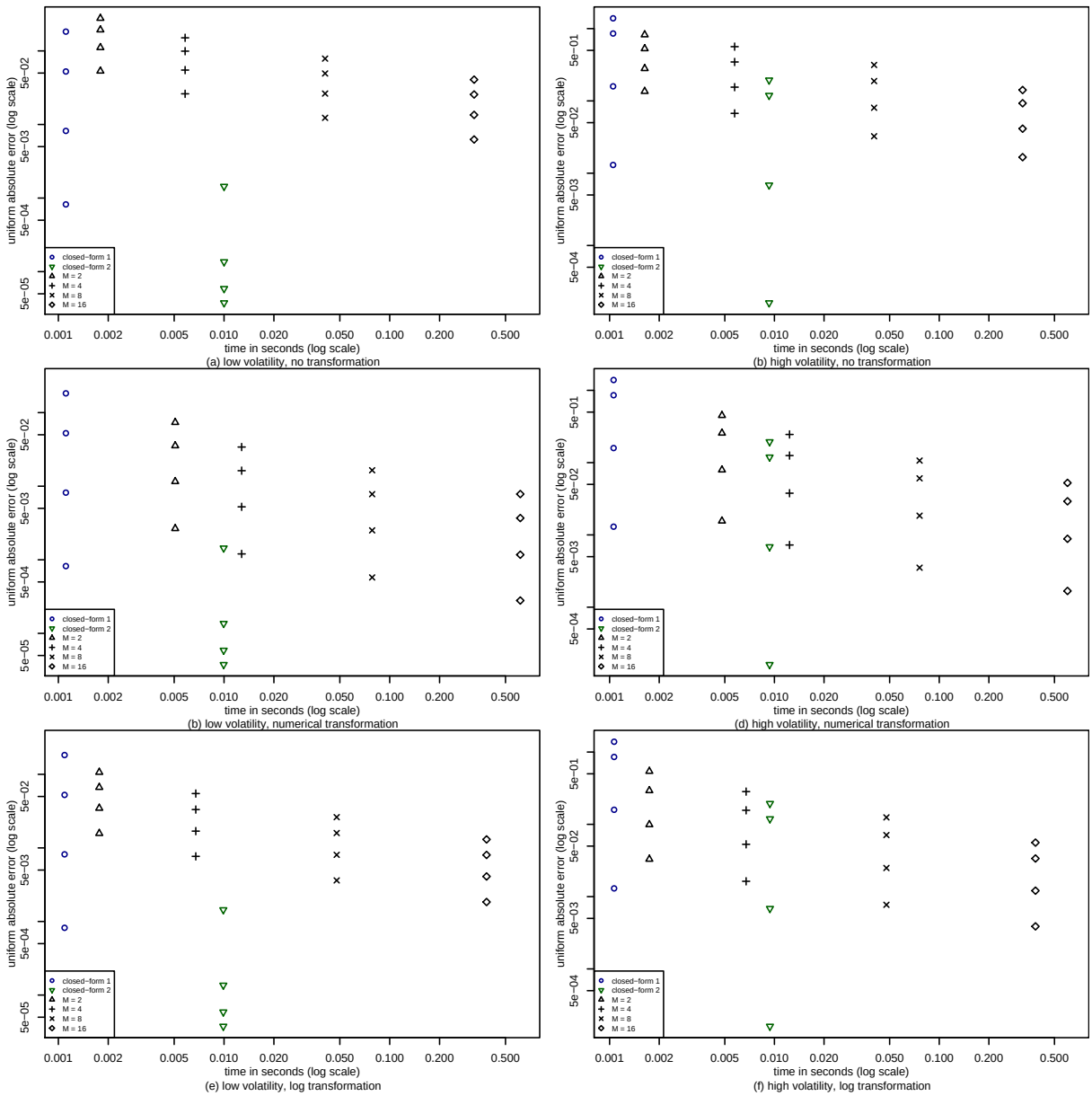

Figure 7: Accuracy and speed of different approximation methods for the log transition density. The four points for each method correspond to Δ values of 1/52, 1/12, 1/4, and 1/2, respectively, from the bottom up. The reported times are the averages of 100 replication of evaluating the log density at 40 points in the interior 95% of the Δ-ahead transition distribution starting from x_0. The reported uniform absolute errors are the average of the maximum absolute errors from the 100 replications.

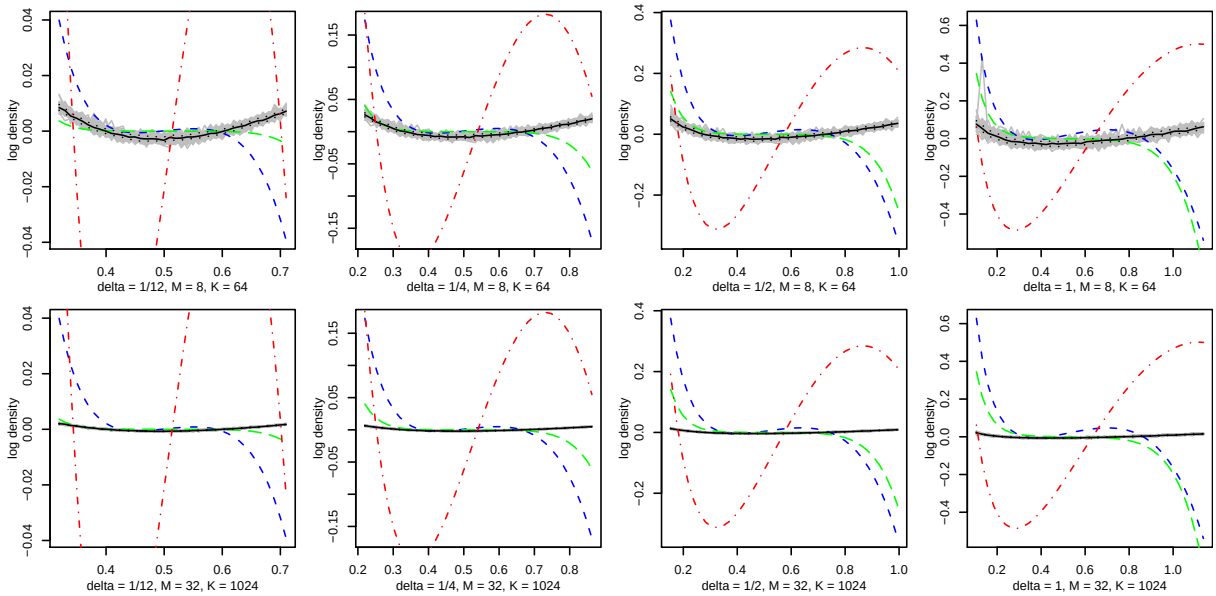

Figure 8: Approximation error comparison with the simulation approach using numerical variance stabilizing transformation: artificial CIR case. The initial state is $x_0 = 0.5$. The (blue) dashed lines are the first-order closed-form approximation. The (green) long-dashed lines are the second-order closed-form approximation. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates. The (red) dot-dashed lines are the first-order Euler approximation.

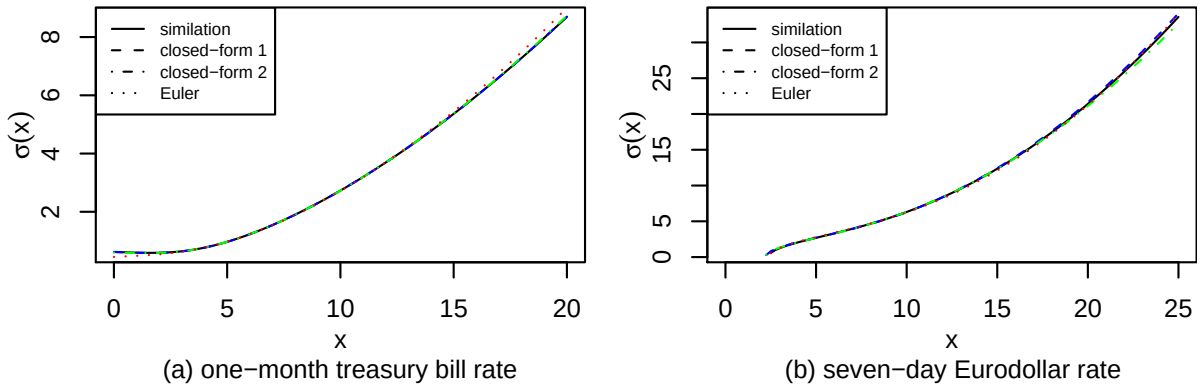

Figure 9: Comparison of the estimated diffusion function of the GEN1 model for the treasury bill rate and Eurodollar rate. The solid line is from the simulation method. The (blue) dashed line and (green) dot-dashed line are first-order and second-order closed-form approximation, respectively. The (red) dotted line is from the first-order Euler approximation.

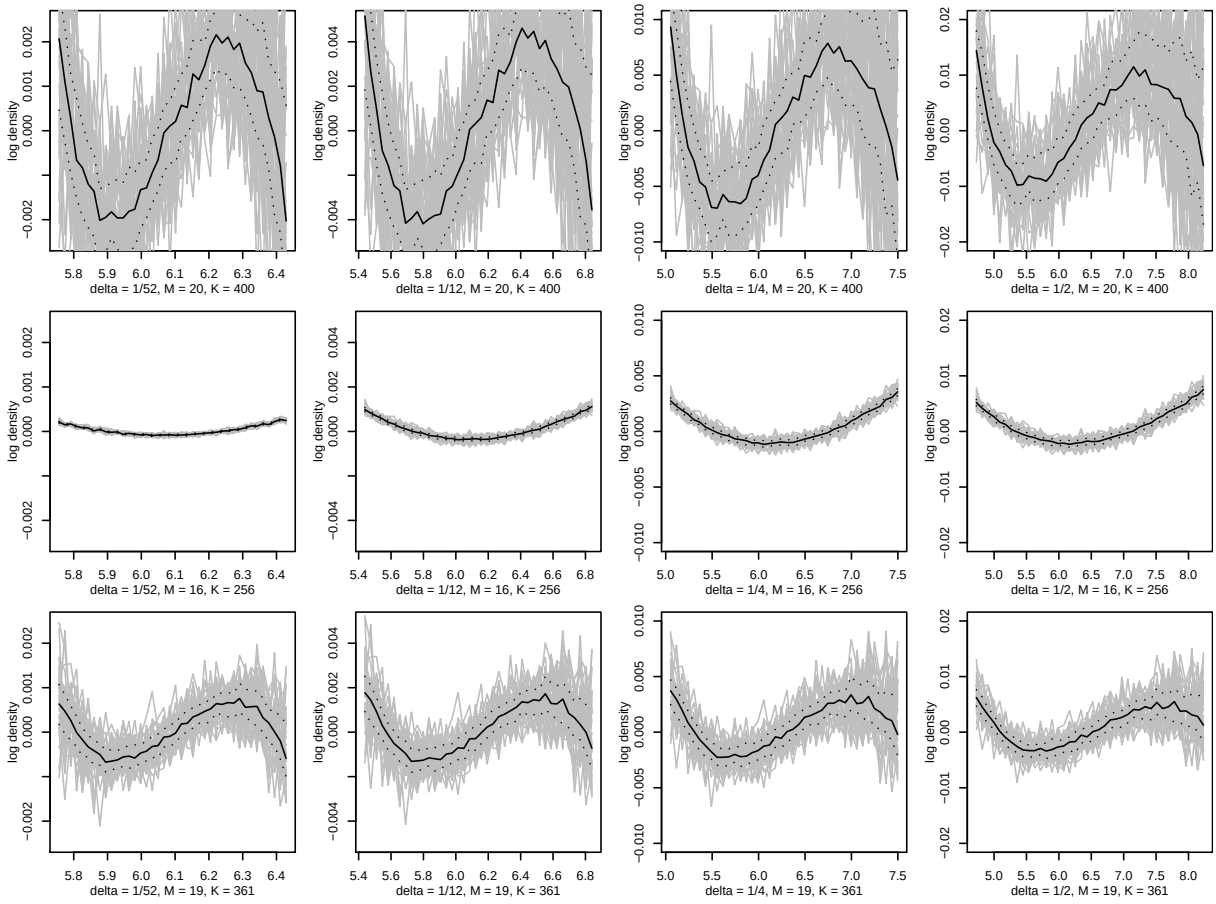

Figure 10: Pseudo approximation error comparison with given computing time. The three rows from top to bottom correspond to no transformation, numerical transformation, and log transformation, respectively. The initial state is $x_0 = 6.076$. The gray lines are 100 replicates of the simulation approximation. The solid lines are the median, and the dotted lines are the 25th and 75th percentiles of 100 replicates.

Table 1: Model parameters for the log-transition-density comparison.

Model	α_0	α_1	α_2	α_3	β_0	β_1	β_2	β_3
Low volatility	-3.3157	0.7328	-0.0503	6.2555	0.3667	-0.0275	0.0030	3.3732
High volatility	-28.8908	5.6030	-0.3119	52.4248	-5.3686	2.2718	0.0013	4.2011

Table 2: Parameter estimates and standard errors from three methods for the daily observations of the three-month treasury bill rate. The simulation method uses $(M, K) = (8, 64)$.

Method	L_n	α_0	α_1	α_2	α_3	β_0	β_1	$\beta_2 \times 10^3$	β_3
				Model GEN1					
Simulation	7813.85	0.344				0.389	-0.035	3.270	3.356
		0.203				0.112	0.032	0.184	0.024
Closed-Form 1	7813.90	0.341				0.382	-0.033	3.224	3.361
		0.203				0.112	0.032	0.180	0.024
Closed-Form 2	7813.87	0.343				0.365	-0.026	3.057	3.380
		0.203				0.112	0.032	0.165	0.023
Euler	7817.63	0.341				0.204	0.032	1.929	3.551
		0.203				0.120	0.034	0.078	0.021
				Model GEN4					
Simulation	7814.41	-3.851	0.826	-0.055	7.196	0.358	-0.023	2.999	3.387
		10.383	1.780	0.094	18.572	0.112	0.032	0.160	0.023
Closed-Form 1	7814.46	-2.980	0.673	-0.047	5.710	0.361	-0.025	3.041	3.382
		10.376	1.779	0.094	18.560	0.112	0.032	0.164	0.023
Closed-Form 2	7814.41	-3.148	0.707	-0.049	5.946	0.360	-0.024	3.035	3.383
		10.390	1.781	0.094	18.586	0.112	0.032	0.163	0.023
Euler	7817.91	-0.969	0.422	-0.040	1.305	0.332	-0.015	2.833	3.408
		10.362	1.777	0.094	18.529	0.114	0.032	0.146	0.023

Table 3: Time consumed in seconds for 100 evaluations of the loglikelihood. The simulation method uses $(M, K) = (8, 64)$. The numerical transformation uses 100 grid points.

Dataset	Nobs	Closed-form		Simulation		
		1st-order	2nd-order	No transform	Numerical transform	Log transform
Treasure-Bill	7555	7.5	103.6	728.3	1361.1	898.1
Eurodollar	5505	5.6	76.6	533.4	992.1	646.4

Table 4: Parameter estimates and standard errors from three methods for the daily observations of the seven-day Eurodollar rate. The simulation method uses $(M, K) = (8, 64)$.

Method	L_n	α_0	α_1	α_2	α_3	β_0	β_1	$\beta_2 \times 10^3$	β_3
				Model GEN1					
Simulation	−941.20	2.109				−5.506	2.319	1.276	4.238
		0.693				0.532	0.150	0.045	0.024
Closed-Form 1	−961.26	2.046				−4.911	2.202	1.252	4.255
		0.710				0.569	0.156	0.041	0.023
Closed-Form 2	−950.93	2.071				−4.625	2.078	1.908	4.098
		0.704				0.540	0.151	0.115	0.029
Euler	−1003.33	2.102				−5.438	2.339	2.259	3.990
		0.714				0.550	0.152	0.181	0.035
				Model GEN4					
Simulation	−938.80	−29.073	5.581	−0.306	53.273	−5.473	2.308	1.335	4.221
		26.313	4.046	0.186	49.806	0.538	0.151	0.051	0.024
Closed-Form 1	−958.71	−30.217	5.656	−0.301	55.896	−4.888	2.199	1.256	4.254
		26.710	4.099	0.189	50.731	0.575	0.157	0.041	0.024
Closed-Form 2	−948.45	−35.386	6.646	−0.358	63.843	−4.369	1.990	2.418	4.011
		26.293	4.033	0.185	49.951	0.552	0.155	0.188	0.034
Euler	−999.60	−16.448	3.471	−0.202	31.226	−5.944	2.512	1.269	4.204
		26.553	4.046	0.184	50.598	0.547	0.151	0.047	0.025