Bayesian Inference for Discretely Sampled Markov Processes with closed–form Likelihood Expansions*

Osnat Stramer †
Department of Statistics and Actuarial Science
University of Iowa, Iowa City, IA

Matthew Bognar ‡
Department of Statistics and Actuarial Science
University of Iowa, Iowa City, IA

Paul Schneider §
Finance Group
Warwick Business School, Coventry, UK

July 15, 2009

Abstract

This article proposes a new Bayesian MCMC methodology for estimation of a wide class of multi-dimensional Markov models. Our approach is based on the closed–form (CF) likelihood approximations of Ait-Sahalia (2002, 2008). The

*The authors are very grateful to the referees and editors for their insightful and valuable comments.

†osnat-stramer@uiowa.edu
‡matthew-bognar@uiowa.edu
§paul.schneider@wbs.ac.uk
CF likelihood approximation does not integrate to one; it is very close to one when near the MLE, but can markedly differ from one when far in the tails. We propose a MCMC algorithm that addresses the problems that arise when the CF approximation is applied in a Bayesian context. The efficacy of our approach is demonstrated in a simulation study of the Cox-Ingersoll-Ross (CIR) and Heston models, and is applied to two well known real-world datasets.

**Key Words:** Bayesian inference, MCMC, jump-diffusion process, closed-form likelihood, exchange algorithm

**Introduction**

We wish to perform inference for the parameters of a continuous-time Markov process $Y$ which is observed (possibly with noise) at discrete time points $t_i = i\Delta$ ($i = 0, \ldots, n$) yielding observations $x = (x_0, \ldots, x_n)$. We assume that the transition density $p_Y(\Delta, x|x_0, \theta)$, the conditional density of $Y_{t+\Delta} = x$ given $Y_t = x_0$, exists. By the Markov property, if all components of $Y$ at time $t_i$ $(i = 0, \ldots, n)$ are observed without noise, the likelihood function is

$$L(x|\theta) = \prod_{i=1}^{n} p_Y(\Delta, x_i|x_{i-1}, \theta).$$

A diffusion process is described as a solution to the stochastic differential equation (SDE)

$$dY_t = \mu(Y_t, \theta) dt + \sigma(Y_t, \theta) dW_t, \quad 0 \leq t \leq T,$$

where $Y_t$ takes values in $\mathbb{R}^d$, $\mu$ is a drift function of dimension $d$, $\nu = \sigma\sigma^T$ is a covariance function of dimension $d \times d$, and $W_t$ is a $k$-dimensional Brownian motion where $k \geq d$. The functions $\mu$ and $\sigma$ are known, while the parameter vector $\theta$ is unknown. A jump-diffusion processes is defined as a solution to the stochastic differential equation (SDE)

$$dY_t = \mu(Y_t, \theta) dt + \sigma(Y_t, \theta) dW_t + J_t dN_t, \quad 0 \leq t \leq T,$$

where the pure jump process $N$ has stochastic intensity $\lambda(Y_t, \theta)$ and jump size 1. The jump size $J_t$ is independent of the filtration generated by the process up to, but not
including, time $t$ and has probability density $\nu(\cdot, \theta)$. In all except a few typical cases, the transition densities are not analytically available, hindering classical and Bayesian inferential techniques. Jump-diffusion processes constitute a large and widely used class of continuous-time Markov models for financial data. For simplicity, we restrict our attention to time–homogeneous processes. Time-inhomogeneous processes can be dealt with by simply adding time as an additional state variable.

Most commonly used likelihood-based procedures for estimating the transition density $p_Y(\Delta, x_i | x_{i-1}, \theta)$ are more amenable when the covariance function $\nu$ is constant. Following the definition of Aït-Sahalia (2008) in a multivariate setting, a diffusion $Y$ is reducible if there exists a one-to-one transformation $\tilde{Y} = h(Y, \theta)$ such that the covariance function of $\tilde{Y}$ is the identity matrix; it is irreducible otherwise. One-dimensional diffusions are always reducible, while multi–dimensional typically are not. The estimation procedures fall into four main classes. 1) The discretized simulation approach (Pedersen, 1995; Elerian et al., 2001; Brandt and Santa-Clara, 2002; Durham and Gallant, 2002) is based on data augmentation schemes and estimation of the transition density using importance sampling. 2) The unbiased estimation of the diffusion transition density (Beskos et al., 2006b, 2009) builds on the exact simulation approach (Beskos and Roberts, 2005; Beskos et al., 2006b); while these methods do not have discretization error and proved to be computationally efficient, they can only be applied to reducible diffusions. 3) The closed–form (CF) expansion of the transition density (Aït-Sahalia, 2002, 2008; Egorov et al., 2003; Schaumburg, 2001; Yu, 2007). 4) The numerical solution to the Kolmogorov forward equation implied by the transition density (Lo, 1988).

The main approach in the literature for Bayesian estimation, studied independently by Jones (1999), Eraker (2001), and Elerian et al. (2001), is to consider the estimation of diffusion models with discrete observation times as a classic missing-data problem so that the complete-data likelihood can be accurately estimated. Thus, this estimation framework relies on the introduction of latent auxiliary data to complete the missing diffusion between each adjacent pair of data points. However, these algorithms can break down
due to the high dependence between the volatility coefficient and the missing data. To overcome this problem, parametrization methods are provided in Roberts and Stramer (2001), Kalogeropoulos et al. (2007), Chib et al. (2006), and Golightly and Wilkinson (2008). These Bayesian approaches are computationally intensive. A different Bayesian approach for non-noisy one-dimensional diffusions is derived in Di Pietro (2001). It is based on the analytical closed–form (CF) approximation of Aït-Sahalia (2002) for reducible diffusions.

Because the closed–form (CF) approximation is a local approximation around the maximum likelihood estimate (MLE), it cannot always be used for values of $\theta$ that are far from the MLE. As observed in Di Pietro (2001) and our simulation study (see Figure 1), the CF approximation can explode to infinity for values of $\theta$ that are far from the MLE. This is not a hindrance for practitioners simply seeking to determine the MLE, but can cause difficulties for Bayesian practitioners seeking to use MCMC techniques to sample from the posterior distribution of $\theta$. Such samplers explore the posterior distribution of $\theta$, and may require the evaluation of the likelihood far from the MLE. The MCMC sampler may get stuck in the tails of the posterior, typically when $\theta$ is far from the MLE. In fact, the approximated likelihood does not integrate to 1, its normalizer is an intractable function of the parameters $\theta$ and the state variables.

In Di Pietro (2001), the unknown normalizer is explicitly calculated for the Vasicek model or numerically approximated for the CIR model. The normalized closed–form (CF) likelihood is a density and can be used to derive the posterior. A simulation study in Di Pietro (2001) finds that, for his examples, the normalized second order ($K = 2$) CF approximation is better behaved than the un-normalized CF and outperforms the high frequency augmentation technique introduced in Elerian et al. (2001).

Aït-Sahalia (2008) provides an extension of the CF approximation for irreducible multi–dimensional diffusions. This includes a broad class of models used in the literature such as stochastic volatility models. For irreducible models, the CF approximation for the log–transition density, $\log p_Y(\Delta, x|x_0, \theta)$, is a Taylor expansion around $\Delta = 0$ and $x = x_0$. Away from $x_0$, the error is a polynomial in $(x - x_0)$. We therefore assume that
the CF approximation for \( p_Y(\Delta, x|x_0, \theta) \) is zero outside some compact set around \( x_0 \). Clearly, the normalization constant cannot be easily approximated, and even if it could, it would require tremendous computational effort. It is therefore not feasible to extend the results in Di Pietro (2001) to the CF approximation for most multivariate diffusions.

There are numerous instances in which the likelihood of interest contains a normalizer that is an intractable function of the parameters. Different (approximate) inferential approaches have been proposed in the literature; for example, see Berthelsen and Møller (2003), Heikkinen and Penttinen (1999), and Bognar (2008). A method that avoids such approximations, first proposed in Møller et al. (2006), introduces a cleverly chosen auxiliary variable into the Metropolis–Hastings algorithm so that the normalizing constants cancel in the Metropolis–Hastings ratio. A simpler and more efficient version, which inspired our work, is proposed in Murray et al. (2006).

A new asymptotically exact Bayesian approach is proposed in Atchade et al. (2008). Their algorithm requires choosing particles \( \theta^{(i)} (i = 1, \ldots, d) \) in such a way that the posterior density at each \( \theta \in \Theta \) can be well approximated by at least one of the posterior densities evaluated at \( \theta^{(i)} (i = 1, \ldots, d) \). However, the method in Atchade et al. (2008) cannot be easily applied to the CF approximation as \( \Theta \) is usually unknown.

This article proceeds as follows: Section 1 reviews the closed–form (CF) approximation of the likelihood (Aït-Sahalia, 2002, 2008; Egorov et al., 2003; Schaumburg, 2001; Yu, 2007). Section 2 presents the algorithm proposed in Murray et al. (2006) and develops a new version of the algorithm that can be used for jump-diffusion models or any other Markov process with transition density approximation that does not integrate to one; Section 2.1 presents the algorithm for non-noisy observations, Section 2.2 introduces observational noise into the model, and Section 2.3 applies it to general stochastic volatility models. A simulation study in Section 3 examines the efficacy of our approach; Section 3.1 applies our algorithm to CIR models, while Section 3.2 demonstrates its application to stochastic volatility models using a volatility proxy. Section 4 tests the algorithm on two real–world datasets, Section 5 explores some possible extensions, and Section 6 concludes and provides some discussion.
1 Closed–Form (CF) Approach

For time–homogeneous univariate diffusion processes, Aït-Sahalia (2002) derives a sequence of closed–form (CF) expansions for the transition density. The method has been generalized to time–inhomogeneous processes by Egorov et al. (2003). Aït-Sahalia (2008) extends to the multivariate setting (time–inhomogeneous diffusions can be dealt with by adding time as an additional state variable), where a diffusion is typically irreducible (i.e. cannot be transformed to a diffusion with unit covariance function). Yu (2007) establishes a framework for multivariate jump-diffusions and Schaumburg (2001) for multivariate semi–martingales with Levy–type generator. We refer to Aït-Sahalia (2006) for excellent reviews on the closed–form (CF) approach.

For a reducible diffusion \( Y \), Aït-Sahalia (2002) finds a closed–form transition density expansion for the transformed process \( \tilde{Y} = h(Y, \theta) \) (the covariance function for \( \tilde{Y} \) is the identity matrix) based on a truncated Hermite series. The closed–form transition density expansion for \( \tilde{Y} \) is then obtained by a standard variable transformation. The Hermite series approximation of the transition density of \( p_{\tilde{Y}} \) is

\[
p_{\tilde{Y}}(\Delta, \tilde{x}|\tilde{x}_0; \theta) = \Delta^{-d/2} \phi \left( \Delta^{-1/2} (\tilde{x} - \tilde{x}_0) \right) \sum_{h \in \mathbb{N}^d} \eta_h(\Delta, \tilde{x}_0; \theta) H_h \left( \Delta^{-1/2} (\tilde{x} - \tilde{x}_0) \right), \tag{2}
\]

where \( h = (h_1, \ldots, h_d)^T \in \mathbb{N}^d \), \( \phi \) is the density of the \( d \)-dimensional multivariate Normal distribution with mean zero and identity covariance matrix, \( H_h \)'s are the Hermite polynomials, and the Hermite coefficients are

\[
\eta_h(\Delta, \tilde{x}_0; \theta) = \frac{1}{h_1! \cdots h_d!} E \left[ H_h \left( \Delta^{-1/2}(\tilde{Y}_{t+\Delta} - \tilde{Y}_t) \right) | \tilde{Y}_t = \tilde{x}_0; \theta \right].
\]

Aït-Sahalia (2002, 2008) suggests approximating the conditional expectation in \( \eta_j \) by a Taylor expansion of order \( K \) in \( \Delta \) using the infinitesimal generator of \( \tilde{Y} \). With the coefficients \( \eta_j \)'s replaced by their approximations, the terms in (2) can be gathered according to powers of \( \Delta \) to obtain a closed–form expansion of the log of the transition density:

\[
l_Y^{(K)}(\Delta, \tilde{x}|\tilde{x}_0; \theta) = -\frac{d}{2} \log(2\pi\Delta) + \frac{C_{\tilde{Y}}^{(-1)}(\tilde{x}|\tilde{x}_0)}{\Delta} + \sum_{k=0}^{K} C_{\tilde{Y}}^{(k)}(\tilde{x}|\tilde{x}_0; \theta) \Delta^k k!,
\]
where the $C_{Y}^{(k)}$'s are the coefficients after the regathering. Aït-Sahalia (2008) gives an alternative approach to determine the coefficients $C_{Y}^{(k)}$ by solving the Kolmogorov equations using (1) as the postulated form of the solution. The partial differential equations are expressed in terms of $\Delta$ and $C_{Y}^{(k)}$'s on both sides. The coefficients $C_{Y}^{(k)}$'s are determined by matching the coefficients of $\Delta^k$ for $k = -1, 0, 1, \ldots, K$. Both approaches give the same expansion. Once (1) is available, the closed–form expansion of the log transition density of $Y$ is then obtained by a standard variable transformation,

$$l_{Y}^{(K)}(\Delta, x|x_0; \theta) = -\frac{1}{2} \log \det \nu(x; \theta) - \frac{d}{2} \log(2\pi \Delta)$$

$$+ \frac{C_{Y}^{(-1)}(x|x_0)}{\Delta} + \sum_{k=0}^{K} C_{Y}^{(k)}(x|x_0; \theta) \frac{\Delta^k}{k!}. \quad (3)$$

For irreducible diffusions, the idea of Aït-Sahalia (2008) is to postulate a solution of form (3) and determine the coefficients $C_{Y}^{(k)}$'s. The problem now is that these coefficients no longer have explicit solutions. Letting $j_k = 2(K+1-k)$, $k = -1, 0, \ldots, K$, Aït-Sahalia (2008) approaches the problem by deriving an explicit Taylor approximation of order $j_k$ for $C_{Y}^{(k)}$ in terms of $(x - x_0)$,

$$C_{Y}^{(j_k,k)}(x|x_0) = \sum_{i \in I_k} \beta_{i}^{(k)}(x_0)(x_1 - x_{01})^{i_1} \cdots (x_d - x_{0d})^{i_d},$$

where $I_k = \{i = (i_1, \ldots, i_d) \in \mathbb{N}^d : 0 \leq i_1 + \ldots + i_d \leq j_k\}$. The expansion order of $C_{Y}^{(k)}$ is chosen to be $j_k = 2(K+1-k)$ such that the approximation error due to the Taylor expansion in $(x - x_0)$ is of the same order, $\Delta^{K+1}$, for each $k$. The coefficients $\gamma_{i}^{(k)}$ ($i = 1, \ldots, j_k$) in the Taylor expansion $C_{Y}^{(j_k,k)}(x|x_0)$, $k = -1, 0, \ldots, K$, are obtained from successively solving a system of linear equations. The resulting expansion for the irreducible case is

$$l_{Y}^{(K)}(\Delta, x|x_0; \theta) = -\frac{1}{2} \log \det \nu(x; \theta) - \frac{d}{2} \log(2\pi \Delta)$$

$$+ \frac{C_{Y}^{(j_k,k-1)}(x|x_0)}{\Delta} + \sum_{k=0}^{K} C_{Y}^{(j_k,k)}(x|x_0; \theta) \frac{\Delta^k}{k!}. \quad (4)$$

An expansion of the transition density $p_Y$ can be obtained by exponentiating $l_{Y}^{(K)}$, which guarantees positivity.
The method can only be applied to small values of $\Delta$ as it is a Taylor series expansion near $\Delta = 0$. The theory (Aït-Sahalia, 2002, 2008) guarantees its convergence to the true density. Under certain regularity conditions, the sequence converges to the true transition density on some compact parameter space $\Theta$ as more correction terms are added. For the one-dimensional reducible case, convergence holds for any fixed transition interval $\Delta$ which is smaller than a threshold $\Delta$. This threshold $\Delta$ depends on both the drift $\mu$ and covariance $\nu$. For the multi-dimensional case, convergence holds as $\Delta \to 0$.

2 Bayesian Model Details

2.1 Data Observed Without Noise

In this section we present a MCMC algorithm that can be used to simulate from the posterior density of the parameter vector $\theta$ using the normalized closed-form (CF) approximation for the likelihood. We let

$$g^{(K)}_{CF}(\Delta, x|x_0, \theta) \equiv \exp \left( \tilde{t}^{(K)}_Y(\Delta, x|x_0; \theta) \right)$$

denote the un-normalized CF approximation of the transition density, where $\tilde{t}^{(K)}_Y$ is defined as in (4). Unless stated otherwise, we assume that $K = 2$ and notationally omit $K$. Denote the normalized CF approximation by

$$p^N_{CF}(\Delta, x|x_0, \theta) \equiv \frac{g_{CF}(\Delta, x|x_0, \theta)}{Z(x_0, \theta)}$$

where $Z(x_0, \theta) = \int g_{CF}(\Delta, x|x_0, \theta) \, dx$ is, for most models, analytically intractable. The normalized CF approximated likelihood is

$$L^N_{CF}(x|\theta) = \prod_{i=1}^n p^N_{CF}(\Delta, x_i|x_{i-1}, \theta) = \prod_{i=1}^n \frac{g_{CF}(\Delta, x_i|x_{i-1}, \theta)}{Z(x_{i-1}, \theta)}$$

and (the CF approximation of) the posterior distribution is given by

$$\pi^N_{CF}(\theta|x) \propto L^N_{CF}(x|\theta)\pi(\theta),$$

where $\pi(\theta)$ is the prior distribution on $\theta$. 

8
Standard Metropolis–Hastings algorithms for updating \( \theta \) have an acceptance ratio which involves the intractable normalizing constants. If \( \theta \) is the current value and \( \theta^* \) is the proposed value (generated from some proposal density \( q(\theta^*|\theta) \)), the Metropolis–Hastings algorithm has acceptance probability \( \min[1, R_{MH}] \) where

\[
R_{MH} = \frac{\prod_{t=1}^{n} gCF(\Delta, x_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*) \pi(\theta^*) q(\theta|\theta^*)}{\prod_{t=1}^{n} gCF(\Delta, x_i|x_{i-1}, \theta)/Z(x_{i-1}, \theta) \pi(\theta) q(\theta^*|\theta)}.
\]

It can be seen that the intractable normalizers cancel in the Metropolis–Hastings ratio and the limiting distribution yields the posterior density \( \pi(\theta|x) \). Their update procedure, which they refer to as the exchange algorithm, is executed as follows.

**Algorithm 1** (Exchange Algorithm; Murray et al. (2006)). Initialize \( \theta^{(t)} \) where \( t = 0 \), then

1. Generate \( \theta^* \) from some proposal density \( q(\theta^*|\theta^{(t)}) \).
2. Generate a sample \( w \) from \( L(w|\theta^*) = \prod_{i=1}^{n} g(w_i|\theta^*)/Z(\theta^*) \).
3. Accept \( \theta^* \) (i.e. let \( \theta^{(t+1)} = \theta^* \)) with probability \( \min[1, R_{Mur}] \) where

\[
R_{Mur} = \frac{\prod_{i=1}^{n} g(x_i|\theta^*)/Z(\theta^*) \pi(\theta^*) q(\theta^{(t)}|\theta^*) \prod_{i=1}^{n} g(w_i|\theta^{(t)})/Z(\theta^{(t)})}{\prod_{i=1}^{n} g(x_i|\theta^{(t)})/Z(\theta^{(t)}) \pi(\theta^{(t)}) q(\theta^*|\theta^{(t)}) \prod_{i=1}^{n} g(w_i|\theta^*)/Z(\theta^*)},
\]

otherwise set \( \theta^{(t+1)} = \theta^{(t)} \). Increment \( t \) and repeat. The intractable \( Z \)'s cancel, leaving only computable quantities.

The Murray et al. (2006) algorithm is not applicable in our situation since \( \theta^* \) would be accepted with probability \( \min[1, R_{Mur}] \) where

\[
R_{Mur} = \frac{\prod_{i=1}^{n} gCF(\Delta, x_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*) \pi(\theta^*) q(\theta^{(t)}|\theta^*)}{\prod_{i=1}^{n} gCF(\Delta, x_i|x_{i-1}, \theta^{(t)})/Z(x_{i-1}, \theta^{(t)}) \pi(\theta^{(t)}) q(\theta^*|\theta^{(t)})} \times \frac{\prod_{i=1}^{n} gCF(\Delta, w_i|w_{i-1}, \theta^{(t)})/Z(w_{i-1}, \theta^{(t)})}{\prod_{i=1}^{n} gCF(\Delta, w_i|w_{i-1}, \theta^*)/Z(w_{i-1}, \theta^*)}.
\]
The intractable normalizers $Z(\cdot, \cdot)$ do not cancel. Note that the exchange algorithm requires sampling $\mathbf{w}$ from the CF approximation $L_{CF}^N(\mathbf{w}|\theta)$ defined in (5) or equivalently sampling $w_i|w_{i-1}, i = 1, \ldots, n$, from $\frac{g_{CF}(\Delta, w_i|w_{i-1}, \theta^*)}{Z(w_{i-1}, \theta^*)}$. However, the normalizing constant depends on $w_{i-1}$ and thus doesn’t cancel in the acceptance probability.

To circumvent this problem we propose to draw the auxiliary variable $\mathbf{w}$ from

$$
\prod_{i=1}^n \frac{g_{CF}(\Delta, w_i|x_{i-1}, \theta^*)}{Z(x_{i-1}, \theta^*)}.
$$

This will cause the normalizing constants $Z$ to depend upon $x_{i-1}$, not $w_{i-1}$, and will consequently cancel in the acceptance probability. Two possible sampling approaches are now described.

- **Direct Approach:** Our first approach proposes a direct way of approximately simulating from $\frac{g_{CF}(\Delta, w_i|x_{i-1}, \theta^*)}{Z(x_{i-1}, \theta^*)}$. Recall that our basic assumption is that the normalized closed–form (CF) approximation is a very accurate approximation of the true transition density of the diffusion process. Thus, we propose to simulate samples from the diffusion process (1) at time $\Delta$ with a starting point $x_{i-1}$. For reducible diffusions this can be done via exact simulation (see Beskos and Roberts, 2005; Beskos et al., 2006a). For the general case, (i.e. for irreducible diffusions) this has traditionally implied the use of some of the discrete time approximation methods (Euler, Taylor expansion, etc.) which rely on small time approximate increment distributions for the diffusion (for a detailed account of these methods see Kloeden and Platen, 1992).

- **Metropolis–Hastings Approach:** A different approach is to run an inner-loop Metropolis–Hastings algorithm to simulate from $\frac{g_{CF}(\Delta, w_i|x_{i-1}, \theta^*)}{Z(x_{i-1}, \theta^*)}$. Let $w_i$ denote the current value and $w_i^*$ denote the candidate value generated from some proposal density $q(w_i^*|w_i)$. Accept $w_i^*$ (i.e. set $w_i = w_i^*$) with probability $\min[1, R_{\text{inner}}]$ where

$$
R_{\text{inner}} = \frac{g_{CF}(\Delta, w_i^*|x_{i-1}, \theta^*) q(w_i|w_i^*)}{g_{CF}(\Delta, w_i|x_{i-1}, \theta^*) q(w_i^*|w_i)}
$$

otherwise leave $w_i$ unchanged. To maximize the acceptance rate, it is important to choose a proposal density $q$ that is a good approximation of the target density $\frac{g_{CF}(\Delta, \cdot|x_{i-1}, \theta^*)}{Z(x_{i-1}, \theta^*)}$. We propose candidate-generating densities of the form $q(\cdot|w_i) =$
where \( q_x(\cdot) \) is independent of the current location \( w_i \). The following are a few possible efficient proposal density functions for \( q_x(\cdot) \):

- The measure induced by the Euler-Maruyama scheme, i.e. \( q_x(\cdot) = \phi(\cdot; x_{i-1} + \mu(x_{i-1}, \theta^*), \nu(x_{i-1}, \theta^*)) \)

- A Gaussian approximation with a higher order (than Euler) Itô-Taylor approximation for the mean and the variance (Kessler, 1997)

- A multivariate-Normal or multivariate-\( t \) approximation (with \( \nu \) degrees of freedom) with location given by the mode of \( \log g_{CF}(\Delta, \cdot|x_{i-1}, \theta^*) \), and dispersion given by the negative of the inverse Hessian evaluated at the mode. This tailoring approach was suggested by Chib and Greenberg (1994), and is used in the context of time series autoregressive-moving average models. For more details on the tailoring approach see Chib (2001).

- Langevin algorithms originally proposed by Doll et al. (1978) and popularized by Besag (1994) and Roberts and Tweedie (1996). These algorithms make use of the gradient of the target distribution \( \log g_{CF}(\Delta, \cdot|x_{i-1}, \theta^*) \) to move more often in directions in which the target density is increasing. Thus, \( q_x(\cdot) \) is a Normal or a \( t \) distribution with mean \( x_{i-1} + \frac{1}{2} \nabla \log g_{CF}(\Delta, \cdot|x_{i-1}, \theta^*) \Delta \) and variance \( \delta \). Here \( \nabla \) denotes the usual gradient differential operator. For partial implicit discretization methods to improve the performance of Langevin sampling see Beskos et al. (2008) and Casella et al. (2009).

**Algorithm 2** (Modified Exchange Algorithm). The modified version of the exchange algorithm proceeds as follows. First, choose a starting value \( \theta^{(t)} \) where \( t = 0 \), then

1. Propose a new value \( \theta^* \) from some proposal density \( q(\theta^*|\theta^{(t)}) \); the proposal density may update one randomly chosen component of \( \theta^{(t)} \) at a time, or may attempt to update multiple components of \( \theta^{(t)} \) simultaneously.

2. Generate independent draws \( w_i, i = 1, \ldots, n \), from \( g_{CF}(\Delta, w_i|x_{i-1}, \theta^*) \) using the direct or the Metropolis–Hastings approach described above.
3. Accept $\theta^*$ (i.e. set $\theta^{(t+1)} = \theta^*$) with probability $\min[1, R_d]$ where

$$R_d = \frac{\prod_{i=1}^{n} g_{CF}(\Delta, x_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*)}{\prod_{i=1}^{n} g_{CF}(\Delta, x_i|x_{i-1}, \theta^{(t)})/Z(x_{i-1}, \theta^{(t)})} \frac{\pi(\theta^*) \ q(\theta^{(t)}|\theta^*)}{\pi(\theta^{(t)}) \ q(\theta^*|\theta^{(t)})} \times \frac{\prod_{i=1}^{n} g_{CF}(\Delta, w_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*)}{\prod_{i=1}^{n} g_{CF}(\Delta, w_i|x_{i-1}, \theta^{(t)})/Z(x_{i-1}, \theta^{(t)})} \frac{\pi(\theta^*) q(\theta^{(t)}|\theta^*)}{\pi(\theta^{(t)}) q(\theta^*|\theta^{(t)})}$$

otherwise reject $\theta^*$ and set $\theta^{(t+1)} = \theta^{(t)}$. Increment $t$, and repeat.

The detailed balance equation holds for any $w = (w_1, \ldots, w_n)$ from $\prod_{i=1}^{n} g_{CF}(\Delta, w_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*)$. Summing over $w$ gives detailed balance overall. The proof follows the same steps as in Murray et al. (2006) and is therefore omitted.

### 2.2 Data Observed with Noise

Financial asset returns are often subject, especially at high frequency, to a vast array of frictions. As a result, a more realistic model might include market microstructure noise. Thus, we assume that $x_i$ are noisy observations of the process $Y$ at discrete time points $t_i = i\Delta, \ i = 0, \ldots, n$. A possible model for $X = (X_0, \ldots, X_n)$, where $X_i$ is the observed value at time $X_{t_i}$, is

$$X_i = Y_i + \epsilon_i \quad \text{where} \quad \epsilon_i \sim N(0, \Sigma), \ i = 0, \ldots, n.$$

The posterior density of parameters and unobserved data $Y = (Y_0, \ldots, Y_n)$ is

$$\pi^N_{CF}(\theta, \Sigma, y|x) \propto \pi(\theta) \pi(\Sigma) L^N_{CF}(y|\theta) \prod_{i=1}^{n} \phi(x_i; y_i, \Sigma)$$

(6)

where $L^N_{CF}$ is defined as in (5) and $\phi(\cdot; \mu, \nu)$ is the Normal density with mean $\mu$ and covariance $\nu$. Updating $\theta$ given $y$ can be done as in Section 2.1 and updating $\Sigma$ is straightforward. An additional step is required for updating the latent observations $y$. The $y_i$’s are updated one at a time, however standard Metropolis–Hastings algorithms
for updating $y$ have an acceptance ratio which involves the intractable normalizing constants. Once again we introduce an auxiliary variable into the Metropolis–Hastings algorithm so that the normalizing constants cancel in the Metropolis–Hastings ratio.

**Algorithm 3** (Updating the Latent Observations $y$). The update of $y_i$ for $i = 1, \ldots, n$ proceeds as follows:

1. Generate $y_i^*$ from some proposal density $q(y_i^*|y_i)$ using, for example, a random walk proposal.

2. Generate $w_{i+1}$ from $g_{CF}(\Delta, w_{i+1}|y_i^*, \theta)/Z(y_i^*, \theta)$ as in our modified exchange algorithm.

3. Accept $y_i^*$ with probability $\min[1, \mathcal{R}_{y_i}]$ where

$$
\mathcal{R}_{y_i} = \frac{g_{CF}(\Delta, y_i^*|y_{i-1}, \theta)/Z(y_{i-1}, \theta) g_{CF}(\Delta, y_{i+1}|y_i^*, \theta)/Z(y_i^*, \theta)}{g_{CF}(\Delta, y_i|y_{i-1}, \theta)/Z(y_{i-1}, \theta) g_{CF}(\Delta, y_{i+1}|y_i, \theta)/Z(y_i, \theta)} \times \frac{g_{CF}(\Delta, w_{i+1}|y_i^*, \theta)/Z(y_i^*, \theta) \phi(x_i; y_i^*, \Sigma) q(y_i^*|y_i)}{g_{CF}(\Delta, w_{i+1}|y_i, \theta)/Z(y_i, \theta) \phi(x_i; y_i, \Sigma) q(y_i|y_i)}.
$$

2.3 Stochastic Volatility Models

Consider models of the form $[Y_t, V_t]$ where $Y_t$ is the log-price of a stock or the short term interest with volatility $\sigma_Y(\cdot)$ which is a function of a latent diffusion $V$. We assume that $[Y_t, V_t]$ follows

$$
\begin{align*}
\text{d}Y_t &= \mu_Y(Y_t, V_t, \theta) \text{d}t + \rho \sigma_Y(Y_t, V_t, \theta) \text{d}W_t + \sqrt{1 - \rho^2} \sigma_Y(Y_t, V_t, \theta) \text{d}B_t \\
\text{d}V_t &= \mu_V(Y_t, V_t, \theta) \text{d}t + \sigma_V(Y_t, V_t, \theta) \text{d}W_t,
\end{align*}
$$

(7)

where $B$ and $W$ are two independent standard Brownian motions, and the instantaneous correlation between $\text{d}Y_t$ and $\text{d}V_t$ is controlled by $\rho$. We assume the process $Y$ is observed (possibly with noise) at discrete time points $t_i = i\Delta$ ($i = 0, \ldots, n$) yielding observations $\mathbf{x} = (x_0, \ldots, x_n)$. This class includes many interesting multi-dimensional diffusion
models. One approach often used in the literature is to infer the values of the latent process $V_t$ at times $t_i$, $i = 1, \ldots, n$, from option prices using a proxy for implied volatility (possibly observed with noise) (Aït-Sahalia and Kimmel, 2007; Johannes et al., 2008; Chernov et al., 2003; Eraker, 2004; Jones, 2003). An application of this approach within a Bayesian set-up follows directly from Section 2.2 and is performed in Section 3.2.

An alternative approach is to treat the unobserved data $V = (V_1, \ldots, V_n)$, where $V_i = V_{t_i}$, as missing data. The posterior density of the parameters and unobserved data is

$$
\pi^N_C(\theta, v|x) \propto \pi(\theta) L^N_C([x, v]|\theta),
$$

where $L^N_C$ is defined as in (5). The resulting algorithm is a series of Metropolis within Gibbs steps. It alternates between updating $\theta$ and $V$. Updating $\theta$ given $[x, v]$ can be done as in Section 2.1. Updating $V$ could lead to a very low acceptance rate, due to the discrepancy between the proposed latent process component and the true latent component. To circumvent this problem, we often split up the latent process into blocks, and cycle through each block in turn for updating. This increases the acceptance rate of each move, because there are less latent variables in each block, limiting the discrepancy between the proposed latent process component and the true latent component. See an unpublished Ph.D. thesis by O. Elerian named “Simulation estimation of continuous-time models with applications to finance”, Nuffield College, Oxford, UK, 1999. Updating $v_{i+1}, \ldots, v_{i+r}|v_i, v_{i+r+1}$ can be done via a Brownian bridge if $r$ is small. The case $r = 1$ is in Eraker (2001), where the proposal distribution is Normal with mean $\frac{1}{2}(v_{i-1} + v_{i+1})$ and variance $\frac{1}{2}\sigma^2_V(y_{i-1}, v_{i-1}, \theta)$. The more general case is developed in Durham and Gallant (2002). The last block can be updated according to the Euler scheme. For larger blocks (i.e. bigger $r$) Langevin algorithms as in Beskos et al. (2008) might prove useful.

Once again we introduce an auxiliary variable into the Metropolis–Hastings algorithm so that the normalizing constants cancel in the Metropolis–Hastings ratio. For simplicity we illustrate it with $r = 1$.

**Algorithm 4 (Updating the Unobserved Data $v$)**. The update of $v_i$ for each $i = 1, \ldots, n$ takes the following steps:
1. Generate $v_i^*$ from some proposal density $q(v_i^*|v_i)$.

2. Generate $[w_{i+1}^1, w_{i+1}^2]$ from $g_{CF}(\Delta, |\theta)/Z([x_i, v_i^*], \theta)$ as in our modified exchange algorithm.

3. Accept $v_i^*$ with probability $\min[1, R_{v_i}]$ where

   $R_{v_i} = \frac{g_{CF}(\Delta, [x_i, v_i^*][x_{i-1}, v_{i-1}], \theta)/Z([x_{i-1}, v_{i-1}], \theta)}{g_{CF}(\Delta, [x_i, v_i][x_{i-1}, v_{i-1}], \theta)/Z([x_{i-1}, v_{i-1}], \theta)} \times \frac{g_{CF}(\Delta, [x_{i+1}, v_{i+1}][x_i, v_i^*], \theta)/Z([x_i, v_i^*], \theta)}{g_{CF}(\Delta, [x_{i+1}, v_{i+1}][x_i, v_i], \theta)/Z([x_i, v_i], \theta)} q(v_i|v_i^*)

   \times \frac{g_{CF}(\Delta, [w_{i+1}^1, w_{i+1}^2][x_i, v_i^*], \theta)/Z([x_i, v_i^*], \theta)}{g_{CF}(\Delta, [w_{i+1}^1, w_{i+1}^2][x_i, v_i], \theta)/Z([x_i, v_i], \theta)} q(v_i^*|v_i)$

3. Simulation Study

3.1 CIR Model Observed With Noise

The CIR model (Cox et al., 1985) is characterized by the SDE,

$$dY_t = \beta(\alpha - Y_t)dt + \sigma \sqrt{Y_t} dW_t$$

where $\alpha$ is the mean reverting level, $\beta$ is the speed of the mean-reversion of the process, and $\sigma$ is the volatility parameter. It has a known transition density, which is a scaled non-central chi-squared distribution. We compare Bayesian analyses using

1. the exact (non-central chi-square) CIR transition density (and likelihood) in a standard Metropolis–Hastings sampler,

2. the un-normalized closed–form (CF) likelihood (see Aït-Sahalia, 1999) in a standard Metropolis–Hastings sampler, and

3. the normalized CF likelihood (i.e. the modified exchange algorithm).
For the simulation study, we produce a noisy CIR dataset as follows. First, we generate the unobserved data \( Y_t \) from the true CIR transition density with \( \Delta = 1/12, n = 500, \alpha = 0.07, \beta = 0.15, \) and \( \sigma = 0.07. \) Then, to generate the observed data \( X_t \), i.i.d. realizations of a Gaussian noise are added to \( Y_t \) where \( X_t = Y_t + N(0, \tau^2) \) and the standard deviation \( \tau = 0.005. \) The commonly analyzed FedFunds dataset yields parameter estimates close to \( \alpha = 0.07, \beta = 0.15, \) and \( \sigma = 0.07, \) so the simulated dataset mimics real-world data.

The likelihood takes the form of (6), and the prior specification is similar to Di Pietro (2001):

\[
\pi(\theta) = \pi(\alpha, \beta, \sigma, \tau) = I_{(0,1)}(\alpha)I_{(0,\infty)}(\beta)\sigma^{-1}I_{(0,\infty)}(\sigma)I_{(0,1)}(\tau)
\]

where \( I \) denotes the indicator function. The random–scan algorithm attempts a joint \((\alpha, \beta)\)–move with probability 0.5, while \(\sigma\)–moves and \(\tau\)–moves are selected with probability 0.25. The joint \((\alpha, \beta)\)–move uses a multivariate–\(t\) independence proposal with 4 degrees of freedom (we use the notation \(MV_t(\cdot)\) in the sequel), where the mean vector and covariance matrix are determined via a simple regression (obtained from the Euler approximation of the CIR model; this proposal has thicker tails than the posterior to help prevent the sampler from getting stuck in the tails). The random-walk \(\sigma\)–move generates candidate values from \(Unif(\sigma(t) - 0.01, \sigma(t) + 0.01)\) while the \(\tau\)–move uses \(Unif(\tau(t) - 0.001, \tau(t) + 0.001); \) \(\sigma(t)\) and \(\tau(t)\) are the most recent value sampled.

To determine the propensity of each algorithm (exact CIR, un-normalized CF, and normalized CF) for becoming stuck, 10 chains of length 10,000 are run (not including a 1,000 iteration burn-in for each chain) where each chain is started near the “true” value \( \theta = (\alpha, \beta, \sigma, \tau) = (0.07, 0.15, 0.07, 0.005). \) Starting the chains from a central location (as opposed to starting the chains in the tails) will allow us to witness when a chain wanders into the tails and becomes stuck (certainly, one can cause numerical instabilities in all three algorithms if the starting values are overly extreme; using central starting points eliminates this possibility).

We analyze the simulated dataset via the three algorithms under study. For each algorithm, the output from the successful (i.e. non-stuck) chains is combined; we then
compute posterior summary statistics (such as the posterior mean and quantiles), the Monte Carlo error (the Monte Carlo error, loosely speaking, provides an estimate of the standard error after accounting for the auto-correlation), and determine marginal acceptance rates (the proportion of the time a candidate was accepted) for each move-type (i.e. \((\alpha, \beta), \sigma, \text{ and } \tau\)). Moreover, to determine if a significant difference exists between the output from the exact CIR chain and the closed–form (CF) chains, we compute the \(p\)-value for the two-sample Kolmogorov-Smirnov (K-S) test. Since the K-S test assumes the data represent a random sample from each group (i.e. the data has no autocorrelation), we thin our MCMC output to every 500th iteration (yielding approximate independence) before performing the test. Hence, the K-S test is based upon (at most) a sample size of \(100,000/500 = 200\). We also compute the effective sample size (ESS); i.e. the number of independent samples that would carry the same amount of information as the available correlated samples (Kass et al., 1998). The results of the analysis are summarized in Table 1.

All code is written in the R (http://www.r-project.org) and C++ languages, where many of the computational routines are run within C++ due to its increased efficiency (we call C++ from within the main R program). The time to execute 11,000 iterations (including a 1,000 iteration burn-in) on an Intel Core 2 Duo 2.0 GHz processor was 114 seconds for the exact analysis (using the C-based non-central \(\chi^2\) function in R), 141 seconds for the normalized closed–form (CF) analysis, and 69 seconds for the un-normalized CF analysis.

To assess the likelihood of a chain getting stuck, we repeat the above analysis with 10 different datasets. Again, we generate the datasets from the true CIR transition density with \(\theta = (\alpha, \beta, \sigma, \tau) = (0.07, 0.15, 0.07, 0.005), \Delta = 1/12, \text{ and } n = 500\). We analyze each dataset with 10 different chains using central starting values. The prior specification and proposal densities remain unchanged. Hence, there are 100 individual chains for each analysis (exact, normalized CF, un-normalized CF). For the un-normalized CF analysis, 4 out of the 100 chains became stuck (see the middle entry in Table 2), while none of the exact and normalized CF chains became stuck.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>MC Error</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
<th>AR</th>
<th>KS</th>
<th>ESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>Exact</td>
<td>0.09960</td>
<td>0.00262</td>
<td>0.0296</td>
<td>0.0372</td>
<td>0.0493</td>
<td>0.0605</td>
<td>0.0856</td>
<td>0.3424</td>
<td>0.7360</td>
<td>0.5366</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>NCF</td>
<td>0.10574</td>
<td>0.00574</td>
<td>0.0292</td>
<td>0.0375</td>
<td>0.0495</td>
<td>0.0605</td>
<td>0.0858</td>
<td>0.3929</td>
<td>0.8256</td>
<td>0.3065</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>UCF</td>
<td>0.10070</td>
<td>0.00452</td>
<td>0.0294</td>
<td>0.0370</td>
<td>0.0492</td>
<td>0.0602</td>
<td>0.0837</td>
<td>0.3373</td>
<td>0.8045</td>
<td>0.5421</td>
<td>0.92</td>
</tr>
<tr>
<td>β</td>
<td>Exact</td>
<td>0.15951</td>
<td>0.00145</td>
<td>0.0066</td>
<td>0.0138</td>
<td>0.0791</td>
<td>0.1511</td>
<td>0.2274</td>
<td>0.3398</td>
<td>0.4195</td>
<td>0.5595</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>NCF</td>
<td>0.16128</td>
<td>0.00265</td>
<td>0.0049</td>
<td>0.0116</td>
<td>0.0820</td>
<td>0.1517</td>
<td>0.2299</td>
<td>0.3474</td>
<td>0.4308</td>
<td>0.3065</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>UCF</td>
<td>0.16228</td>
<td>0.00203</td>
<td>0.0055</td>
<td>0.0139</td>
<td>0.0821</td>
<td>0.1538</td>
<td>0.2304</td>
<td>0.3419</td>
<td>0.4215</td>
<td>0.5421</td>
<td>0.96</td>
</tr>
<tr>
<td>σ</td>
<td>Exact</td>
<td>0.07599</td>
<td>2.20e-4</td>
<td>0.0646</td>
<td>0.0674</td>
<td>0.0724</td>
<td>0.0759</td>
<td>0.0795</td>
<td>0.0849</td>
<td>0.0891</td>
<td>0.3753</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>NCF</td>
<td>0.07618</td>
<td>3.09e-4</td>
<td>0.0640</td>
<td>0.0672</td>
<td>0.0723</td>
<td>0.0760</td>
<td>0.0798</td>
<td>0.0858</td>
<td>0.0913</td>
<td>0.2733</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>UCF</td>
<td>0.07605</td>
<td>2.45e-4</td>
<td>0.0648</td>
<td>0.0680</td>
<td>0.0724</td>
<td>0.0759</td>
<td>0.0795</td>
<td>0.0847</td>
<td>0.0888</td>
<td>0.3750</td>
<td>0.47</td>
</tr>
<tr>
<td>τ</td>
<td>Exact</td>
<td>0.00493</td>
<td>1.03e-5</td>
<td>0.0043</td>
<td>0.0045</td>
<td>0.0047</td>
<td>0.0049</td>
<td>0.0051</td>
<td>0.0054</td>
<td>0.0056</td>
<td>0.2508</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>NCF</td>
<td>0.00491</td>
<td>1.18e-5</td>
<td>0.0042</td>
<td>0.0045</td>
<td>0.0047</td>
<td>0.0049</td>
<td>0.0051</td>
<td>0.0054</td>
<td>0.0056</td>
<td>0.2478</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>UCF</td>
<td>0.00491</td>
<td>1.09e-5</td>
<td>0.0043</td>
<td>0.0045</td>
<td>0.0047</td>
<td>0.0049</td>
<td>0.0051</td>
<td>0.0054</td>
<td>0.0056</td>
<td>0.2488</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Table 1: Marginal posterior summary statistics for the mean reverting level α speed β, volatility σ, and noise parameter τ for exact, normalized CF (NCF) and un-normalized CF (UCF) samplers. Included are the mean, Monte Carlo error (MC Error), quantiles, acceptance rates (AR), p−value for the Kolmogorov-Smirnov test (KS) comparing Exact to NCF and UCF, and effective sample size (ESS).
\[ \Delta = \frac{1}{52} \quad \Delta = \frac{1}{12} \quad \Delta = \frac{1}{4} \]

\[
\begin{array}{cccc}
MV t(2) & 0.53 & 0.96 & 0.99 \\
MV t(4) & 0 & 0.04 & 0.04 \\
MV t(40) & 0 & 0.01 & 0 \\
\end{array}
\]

Table 2: Proportion of chains that became stuck (out of 100 chains total) from the un-normalized CF sampler. Analysis compares observation frequency (weekly \( \Delta = 1/52 \), monthly \( \Delta = 1/12 \), quarterly \( \Delta = 1/4 \)) versus proposal density for a joint \((\alpha, \beta)\)-move \((MV t(2), MV t(4), MV t(40))\). None of the exact and normalized CF chains became stuck, and are thus omitted from the table.

We repeat the analysis with \( \Delta \) equal to 1/52 (with \( n = 2,000 \)), 1/12 (\( n = 500 \)), and 1/4 (\( n = 500 \)) with \( MV t(2), MV t(4), \) and \( MV t(40) \) proposals for the joint \((\alpha, \beta)\)-move. We have a total of nine analyses (three \( \Delta \)'s, three proposals), the results are summarized in Table 2. The exact and normalized CF analyses did not become stuck for any combination of proposal density and \( \Delta \). The un-normalized CF analysis became stuck quite often when using the \( MV t(2) \) proposal, but was quite robust to the \( MV t(4) \) and \( MV t(40) \) proposals. Importantly, when the un-normalized CF chains successfully mix, posterior inferences are consistent with those of the exact and normalized CF analyses. The tendency to become stuck was similar for monthly and quarterly data, but markedly decreased for weekly data.

It is interesting to examine how the normalized closed-form (CF) and un-normalized CF log-likelihood functions compare to the exact log-likelihood function. We generate a dataset \( Y_i \) with \( n = 1,000 \) from the true CIR transition density (without noise) where \( \theta = (\alpha, \beta, \sigma) = (0.07, 0.15, 0.07) \) and \( \Delta = 1/52 \). Holding \( \beta \) and \( \sigma \) fixed at 0.15 and 0.07 respectively, we compute the exact log-likelihood and un-normalized CF log-likelihood for 33 equally spaced values of \( \alpha \in (0, 1) \). The results are plotted in the left panel of Figure 1. The exact log-likelihood is depicted as a solid line and the un-normalized CF log-likelihood is depicted as a dashed line. The (approximate) normalized CF log-likelihood is shown as a dotted line (this function is approximated via numerical integration). The
Figure 1: Exact (solid line), normalized CF (dotted line), and un-normalized CF (dashed line), for $\alpha \in (0, 1)$ holding $\beta$ and $\sigma$ fixed. Data was generated from the non-noisy CIR model with $\theta = (\alpha, \beta, \sigma) = (0.07, 0.15, 0.07)$. The third order ($K = 3$) un-normalized CF log–likelihood is also shown (dot-dash-dot line); it is not plotted to the right of the $\star$ due to numerical instabilities.

The middle and right panels of Figure 1 depict the results for $n = 1,000$ with $\Delta$ equal to $1/12$ and $1/4$, respectively. The CF approximations of order $K = 2$ and $K = 3$ are extremely accurate in a neighborhood around the MLE. For values of $\alpha$ distant from the MLE, the un-normalized CF likelihood deviates markedly from the exact log–likelihood, while the normalized CF log–likelihood remains reasonably accurate. While hard to see from Figure 1 (due to scaling), the un-normalized CF log–likelihood of order $K = 3$ performs better than order $K = 2$ and is very accurate for $\alpha$ less than 0.44 when
$\Delta = 1/12$ and less than 0.20 when $\Delta = 1/4$ (these points are marked as asterisks $\star$ on the graph). However, it is still numerically unstable for larger $\alpha$ values (i.e. for values of $\alpha$ to the right of $\star$ and is therefore not plotted. Higher order of the CF approximations are discussed in further detail in Section 5.1. In addition, as can be seen from Figure 1, the reader should note that the un-normalized CF log–likelihood works best for $\Delta = 1/52$ (i.e. for values of $\alpha$ moderately far from the MLE), and progressively becomes less accurate for $\Delta = 1/12$ and $\Delta = 1/4$.

The normalized closed-form (CF) sampler was robust to the amount of noise $\tau$. In some additional simulation studies, the exact and normalized CF samplers yielded similar posterior inferences when $\tau = 0.01$ and 0.02, and the normalized CF sampler had no stability issues after increasing $\tau$.

### 3.2 Heston Model

To perform a numerical application based on the stochastic volatility model (7)–(8) we consider the Heston model where $[Y_t, V_t]$ follow

\[
\frac{dY_t}{Y_t} = \left(\mu - \frac{1}{2}V_t\right)dt + \rho \sqrt{V_t} dW_t + \sqrt{1 - \rho^2} \sqrt{V_t} dB_t
\]

\[
\frac{dV_t}{V_t} = \left(\alpha - V_t\right)dt + \sigma \sqrt{V_t} dW_t,
\]

where $B$ and $W$ are two independent standard Brownian motions, and instantaneous correlation between $dY_t$ and $dV_t$ is controlled by $\rho$. As option prices are traded assets, we need to endow the time series model (10)-(11) with risk premia for arbitrage-free pricing under the auxiliary pricing measure $Q$. To keep the simulation study simple we make the assumption of risk premia such that $W_t = W_t^Q$ and $dB_t = dB_t^Q + \frac{r - \mu}{\sqrt{1 - \rho^2} V_t} dt$ and no adjustments in the variance drift are necessary ($\alpha = \alpha^Q$ and $\beta = \beta^Q$). Prior (9) is endowed with an additional indicator function that keeps $\rho$ within $[-1, 1]$, as well as a constant for $\mu$ to express an uninformative prior.

*Instantaneous* stochastic variance is latent, even though a time series of *implied* variance is often available, for example the VIX implied volatility index published by the CBOE. To account for the stochastic nature and mean reversion of index variance
we use the fact that for short-maturity at-the-money options the Black-Scholes formula is approximately linear in volatility. Affinity of the variance $\mathbb{Q}$-drift (which is the same as the drift in (11) because we assume zero risk premia) together with Fubini’s theorem enables us to write

$$\frac{1}{\xi} E^\mathbb{Q}_t \left[ \int_t^{t+\xi} V_s \, ds \right] = A(\alpha, \beta, \xi) + B(\beta, \xi) V_t, \quad \xi > 0$$

with

$$B(\beta, \xi) = \frac{1 - e^{-\xi \beta}}{\xi \beta}, \quad A(\alpha, \beta, \xi) = \alpha(1 - B(\beta, \xi)).$$

For a nonlinear drift function the conditional expectation could be approximated using the infinitesimal generator of the variance diffusion or a simulation-based estimate. We take average expected variance as a proxy for implied variance $IV_t$:

$$IV_t \approx \frac{1}{\xi} E^\mathbb{Q}_t \left[ \int_t^{t+\xi} V_s \, ds \right], \quad (12)$$

and choose $\xi = 22/252$ as in Jones (2003). Validity of this approximation has been tested extensively in Aït-Sahalia and Kimmel (2007). Note that the approximation above works under heavy parametric assumptions only. A very general, yet data-intensive way to draw inference about spot volatility from the forward-looking VIX index is presented in Todorov and Tauchen (2009, Section 2).

To exploit available market information we consider the joint time series of log stock price and implied variance with and without observation error

$$[Y_t, IV_t] \quad \text{no observation error} \quad (13)$$

$$[Y_t, IV_t + \varepsilon_t] \quad \text{observation error,} \quad (14)$$

where $\varepsilon_t \sim N(0, \tau^2), \ t = t_0, \ldots, t_N$.

We produce ten datasets, each for $\Delta = 1/52, 1/12$, and $1/4$ with $n = 500$ observations, respectively, using the parametrization $\beta = 3, \ \alpha = 0.1, \ \mu = 0.05, \ \rho = -0.8, \ \sigma = 0.25$ (see Aït-Sahalia and Kimmel, 2007), and $\tau = 0.001$ to investigate the behavior of the MCMC samplers. MCMC samplers are constructed around the true transition
density using the expression from Lamoureux and Paseka (2005) (this expression requires only one-dimensional numeric integration instead of a two-dimensional Fourier inversion), the un-normalized CF likelihood, and the normalized CF likelihood (i.e. the modified exchange algorithms 2 and 3, respectively). Note that the model (10)–(11) is irreducible in the sense of Aït-Sahalia (2008) and the corresponding log likelihood expansions are therefore purely polynomial. The initial datapoint \[ Y_0, V_0 \] for each series is simulated from a long chain using a high frequency Euler approximation to the system (10)–(11). The datasets are then evolved from exact simulations from the Heston model employing the technique from Broadie and Kaya (2006). The time series \[ Y_t, IV_t \] is then obtained through formula (12); for \[ Y_t, IV_t + \varepsilon_t \] an observation error is added from realizations of a \( N(0, \tau^2) \) random variable.

The Markov property of the system \( Y_t, IV_t \) results in a decomposition of the log–likelihood into a sum of log-transition densities

\[
\ell(X_0, IV_0, \ldots, X_n, IV_n | \theta) = \log p(X_0, IV_0 | \theta) + \sum_{i=1}^{n} \log p(X_i, IV_i | X_{i-1}, IV_{i-1}, \theta). \tag{15}
\]

For our application we do not take into consideration the unconditional density \( p(X_0, V_0 | \theta) \) which does not exist for Heston’s model. With the linear transformation (12) and a state-independent change of variables \( V_t = \frac{IV_t - A(\alpha, \beta, \xi)}{B(\beta, \xi)} \) we can write (15) discarding the unconditional density of the first observations as

\[
\sum_{i=1}^{n} \log p(X_i, V_i | X_{i-1}, V_{i-1}, \theta) - N \log B(\beta, \xi). \tag{16}
\]

For the second case (14) the likelihood is determined through the observation equation

\[ IV_i = A(\alpha, \beta, \xi) + B(\beta, \xi)V_i + \varepsilon_t, \]

and the likelihood is

\[
\sum_{i=1}^{n} \log p(X_i, V_i | X_{i-1}, V_{i-1}, \theta) + \log p(\varepsilon_i | \theta). \tag{17}
\]

The combinations of data observed with/without noise, with/without normalizing are investigated in Table 3. We use random walk Metropolis–Hastings for the \( \mu, \sigma \) and
Table 3: Heston Model: Table displays the MCMC samplers constructed around the Heston model using the true transition density and Aït-Sahalia (2008) likelihood expansions (CF).

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Exact Observations</th>
<th>Noisy Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition density</td>
<td>CF</td>
<td>CF</td>
</tr>
<tr>
<td>MCMC</td>
<td>Plain Metropolis–Hastings</td>
<td>Plain Metropolis–Hastings</td>
</tr>
<tr>
<td>Transition density</td>
<td>CF</td>
<td>CF</td>
</tr>
<tr>
<td>MCMC</td>
<td>Algorithm 2</td>
<td>Algorithm 3</td>
</tr>
<tr>
<td>Transition density</td>
<td>Exact</td>
<td>Exact</td>
</tr>
<tr>
<td>MCMC</td>
<td>Plain Metropolis–Hastings</td>
<td>Plain Metropolis–Hastings</td>
</tr>
</tbody>
</table>

\( \rho \) parameters with \( \mu^* \sim N(\mu^{(t)}, 0.04) \), \( \sigma^* \sim N(\sigma^{(t)}, 0.05) \), \( \rho^* \sim N(\rho^{(t)}, 0.02) \), respectively. The \( \alpha, \beta \) parameters are also sampled through random walk Metropolis–Hastings for the system observed without error (13). For the system observed with observation error (14) we use proposals tailored around likelihood (17) for \( \alpha \) and \( \beta \), where \( p(X_i, V_i \mid X_{i-1}, V_{i-1}, \theta) \) is replaced by an Euler proxy. The mean of the proposal is the QML estimator and the covariance matrix is the inverse Hessian evaluated at the QML estimators scaled so that proposals are drawn from a \( \text{MV t}(4) \) distribution. For estimations using the true density, we tailor around the likelihood of the observation error only (without the Euler transition density) to increase acceptance rates. The standard deviation of the observation error \( \tau \) is sampled with a Gibbs step from an inverse gamma distribution. Computation times are very similar to Section 3.1, except for estimations using the true density, where 11,000 draws take well over 24 hours due to heavy use of adaptive numeric integration and the use of complex-valued special functions from the density representation of Lamoureux and Paseka (2005).

We collect 11,000 draws from the posterior density of the parameters for all datasets, where we discard the first 1,000 to account for the burn-in period. An investigation of the chains for the 10 datasets observed without error reveals excellent performance.
of both normalized and un-normalized closed-form (CF) for the weekly data. For the monthly data, the sampler using un-normalized CF explodes once; all chains using un-normalized CF diverge for quarterly data, while the chains using normalized CF all converge. CF estimates for quarterly data exhibit a tendency to underestimate the diffusion coefficient $\sigma$ and the speed-of-mean-reversion coefficient $\beta$: this is not surprising as the CF expansions are constructed for small time intervals.

Estimations of the systems observed with error show more extreme results: While all chains using un-normalized CF converge for weekly data, the sampler explodes for all monthly and quarterly spaced datasets. All chains using CF together with our normalization algorithms 2 and 3 converge. Parameter estimates for quarterly data exhibit similar deficiencies as for the non-noisy case.
Finally, we investigate posterior distributions of parameters for noisy monthly data using 101,000 draws, where the first 1,000 are discarded accounting for the burn in. Table 4 shows only results for the normalized CF chain and the chain using exact density. The quantiles in Table 4 reveal that there is strong agreement between the posterior distributions of the parameters. There is higher autocorrelation in the MCMC chains than in the univariate experiment from Section 3.1, however; as K-S statistics are very sensitive to autocorrelation they are not reported.

4  Real Data

For the normalizing algorithm to be useful it must also be applicable to real data. We put it to a test with FedFunds rate data observed monthly from January 1963 to December 1998 ($n = 432$) and daily log $S&P$ 500 and VIX data from March 2000 to February 2009 ($n = 2302$). We choose the square-root model for the FedFunds rate and Heston’s model for the log $S&P$ 500 and VIX data for illustrative reasons; empirical investigations have identified better models for these data sets (see Di Pietro, 2001; Chernov et al., 2003).
Figure 3: Estimated marginal posterior densities from the FedFunds analysis: exact (solid line), Euler (dashed line), normalized CF (dotted line)

4.1 CIR Model

The FedFunds rate dataset appears in Figure 2. We perform three Bayesian analyses using the exact likelihood, the Euler likelihood, and the normalized closed-form (CF) likelihood. We include an analysis based upon the Euler approximation since, in practice, this is a relatively common approximation of the exact likelihood, and since we wish to demonstrate the importance of using the modified exchange algorithm. The prior is the same as in Di Pietro (2001), \( \pi(\theta) = \pi(\alpha, \beta, \sigma) = I_{(0,1)}(\alpha)I_{(0,\infty)}(\beta)\sigma^{-1}I_{(0,\infty)}(\sigma) \), while proposal specifications are similar to Section 3.1.

We run each of the three samplers (exact, Euler, normalized CF) for 500,000 iterations proceeding a 1,000 iteration burn-in period. The estimated marginal posterior
densities appear in Figure 3. The normalized CF sampler (dotted line) yields marginal posterior density estimates that superbly mimic the exact likelihood based sampler (solid line), while the Euler-based sampler (dashed line) yields marginal posterior densities that markedly differ from the exact sampler. Trace plots for the first 50,000 iterations of the normalized CF sampler appear in Figure 4.

We also attempt an analysis based upon the un-normalized CF likelihood, but repeatedly experience the chain getting stuck at a late stage (e.g. one chain got stuck after 95,000 iterations, another after 60,000 iterations). It should be noted that the un-normalized CF sampler yields correct posterior inferences when the chains mix properly; however, executing long chains using the un-normalized CF likelihood proved elusive due to the likelihood exploding to infinity.
Table 5: Heston Model: Marginal posterior summary statistics for Heston parameters conditional on S&P 500 and VIX data from March 1, 2000 - February 27, 2009 using Ait-Sahalia (2008) likelihood expansions and the normalizing algorithm developed in this paper. Included are the mean, quantiles, acceptance rates (AR), and effective sample size (ESS).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
<th>AR</th>
<th>ESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha \beta$</td>
<td>0.080</td>
<td>0.069</td>
<td>0.073</td>
<td>0.077</td>
<td>0.080</td>
<td>0.083</td>
<td>0.087</td>
<td>0.089</td>
<td>0.25</td>
<td>819</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.091</td>
<td>1e-04</td>
<td>5e-04</td>
<td>0.002</td>
<td>0.006</td>
<td>0.012</td>
<td>0.027</td>
<td>0.049</td>
<td>0.25</td>
<td>1000</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-0.023</td>
<td>-0.084</td>
<td>-0.062</td>
<td>-0.026</td>
<td>-0.002</td>
<td>0.020</td>
<td>0.060</td>
<td>0.084</td>
<td>0.5</td>
<td>1000</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.600</td>
<td>0.581</td>
<td>0.587</td>
<td>0.595</td>
<td>0.600</td>
<td>0.605</td>
<td>0.613</td>
<td>0.618</td>
<td>0.13</td>
<td>705</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.736</td>
<td>-0.757</td>
<td>-0.750</td>
<td>-0.742</td>
<td>-0.736</td>
<td>-0.730</td>
<td>-0.720</td>
<td>-0.715</td>
<td>0.33</td>
<td>1000</td>
</tr>
</tbody>
</table>

4.2 Heston’s Model

The bivariate log S&P 500 and VIX implied volatility data from March 1, 2000 to February 27, 2009 (Figure 5) displays some very high values of implied volatility towards the end of the sample of up to 80%. This poses serious difficulties for Bayesian inference using closed-form (CF) likelihood expansions since observations deep in the tails of the transition density will invoke the polynomial nature of the expansions. As a consequence, attempting un-normalized Bayesian inference with closed-form (CF) likelihood expansions leads to the MCMC sampler getting stuck from the first iteration on. Applying the normalized algorithm (i.e. modified exchange algorithm) developed in the present paper to this difficult data provides a crucial assessment of the robustness of the algorithm.

We estimate Heston’s model using the modified exchange algorithm described in Section 2.1. We adopt prior distributions and proposal densities from Section 3.2 and produce 200,000 draws from the posterior distribution of which we record every 200th. The chains can be seen in Figure 6 and statistics of the parameter estimates are reported in Table 5. The autocorrelation in the thinned chain is virtually zero and the effective sample size is very high relative to the sample size. Our parameter estimates reflect two empirical phenomena. Firstly, in order to generate the negative risk premium on variance
risk that has been reported in the literature (Carr and Wu, 2009) the mean reversion parameter $\beta^Q$ would need to be estimated to a very high negative value, resulting in an explosive variance process. Since we impose $\beta^Q = \beta$ and constrain the variance to be non-explosive, a speed-of-mean reversion estimate close to zero is very reasonable. As a consequence the unconditional mean $\alpha$ is very difficult to locate which induces extremely high dispersion in the posterior estimates; we therefore report $\alpha \beta$ instead. Secondly, the market turmoil 2008/2009 is bound to be reflected in the volatility of variance estimate $\sigma$, hence our estimate of around 0.6 is much higher than the one from A"ıt-Sahalia and Kimmel (2007) whose sample ends in 2003. Most importantly, in the context of the present paper, the modified exchange algorithm from Section 2.1 performs well even applied to very rough financial data.
5 Future Topics

5.1 Higher Order of approximation

The accuracy of the closed–form (CF) approximation depends on the order $K$ of approximation. Our simulation study (see also Stramer and Yan, 2007) shows that for many realistic models, $K = 2$ provides a very accurate approximation for the transition density; $K = 1$ may not be sufficient, however. For example, consider the general drift and diffusion model in Aït-Sahalia (1996)

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

where $\theta = (\alpha_0, \ldots, \alpha_3, \beta_0, \ldots, \beta_3)^T$. Model (18) was applied in Stramer and Yan (2007) to two well-known datasets: daily three-month treasury bill rate and daily seven-day Eurodollar rate (see also Durham, 2003). These two series behave very differently, with the Eurodollar rate exhibiting much higher volatility than the treasury bill rate. The maximized log–likelihood from first and second order closed–form approximations differ by about 0.03 for the treasury bill rate and by 10 basis-points for the Eurodollar rate. Clearly, a higher order of approximation is needed for the latter.

However, increasing the order of approximation does not necessarily solve the problem of stability; it will improve the accuracy in the center of the distribution (i.e. near the MLE), but the higher order approximation can quickly explode to $\infty$ when far in the tails, especially for “large” $\Delta$.

Complexity and computation time of the approximation are obviously big factors when increasing the order $K$ of the approximation. The expression of the closed–form (CF) approximation can be obtained with software which can perform symbolic calculations; the expression can then be exported to Fortran or C. A Mathematica module has been written for this purpose and is available upon request. Yet, increasing the order of approximation can be a burden. For model (18), the first and second order CF approximation resulted in 34 and 207 lines of Fortran code, respectively (see Stramer and Yan,
The higher order CF approximations may dramatically increase the computational time, especially for multi-dimensional diffusions. Even for these higher order CF approximations, it appears that normalization techniques are still required for the stability of the sampler (see Figure 1).

The question is how to choose the order $K$ of the closed-form (CF) approximation. One way to increase the accuracy of the approximation is along the lines of Kim et al. (1998). The basic idea is to run the modified exchange algorithm using the order $K_1$ CF approximation, where $K_1 = 1$ or $2$. These algorithms are stable and fast (especially the first order approximation). Recall that the CF approximation is mainly problematic when exploring the posterior distribution of $\theta$ far from the MLE. We conjecture that, for most realistic models, the un-normalized CF approximation of order $K_2$, where $K_2 > K_1$, will provide an accurate approximation for all $\theta$ in the MCMC output. We therefore utilize the importance sampler approach as follows. Let $\pi_{CF_{K_1}}^N(\theta|x)$ be the approximate posterior distribution obtained from the un-normalized CF approximation of order $K_1$ (denote the draws from the posterior distribution, after accounting for the burn-in period, by $\theta^{(1)}, \ldots, \theta^{(T)}$). We can now estimate functions of the posterior of the form $E(g(\theta|x)) = \int g(\theta)\pi_{CF_{K_2}}^N(\theta|x) \, d\theta$ by re-weighting the MCMC draws as

$$\tilde{E}(g(\theta|x)) = \frac{1}{T} \sum_{t=1}^{T} g(\theta^{(t)}) r_t,$$

where $r_t = \frac{\pi_{CF_{K_2}}^N(\theta^{(t)}|x)}{\pi_{CF_{K_1}}^N(\theta^{(t)}|x)}$. If the closed-form (CF) approximation is good, we expect $r_t$ to have unit mean with small variance, i.e. the CF approximation is approximately the same for $K_1$ and $K_2$, precluding the need to consider higher order approximations. Otherwise, this procedure should be repeated with $K_3 > K_2$. The method can also be extended to stochastic volatility models.

### 5.2 Diagnostic and Model Comparison

Evaluation and comparison of Bayesian continuous time models, discussed in Elerian et al. (2001) and Di Pietro (2001), clearly requires more study. One Bayesian
approach of comparing two diffusion models is through Bayes factors, defined as the ratio of the marginal likelihood functions; Bayesians average this marginal likelihood ratio over the parameters $\theta$, given some prior density. This approach, however, requires a proper prior. The marginal likelihood can be computed using the importance scheme described in Geweke (1989), using the density ratio marginal likelihood approximation proposed in Geweke (1999), using the MCMC output as considered in Chib (1995) and Chib and Jeliazkov (2001), among other approaches. The first method may require exact evaluation of the likelihood for values of $\theta$ deep in the tail of the posterior distribution, precluding the use of the closed–form (CF) approximation. The second method can be applied to our modified exchange algorithm since the likelihood and posterior needs to be evaluated only at a single point. Conveniently, we can choose this point to be a point with high posterior density, where the un-normalized CF approximation is very accurate.

Another approach for model comparison that is closely related to Bayes factors is based on the predictive likelihood and predictive intervals. This approach allows for an improper prior. Evaluating the predictive likelihood requires evaluating the likelihood for all $\theta$ in the MCMC output. As was mentioned before, we believe that the closed–form (CF) approximation will provide an accurate approximation for most realistic models. Predictive intervals can be easily employed as they only require simulations from the transition density rather than evaluating the transition density for each $\theta$.

Another approach for model comparison that addresses the issue of model fit uses a sequence of one-step-ahead conditional predictive distribution functions. This sequence depends on the model and data only (see Tauchen, 2002; Diebold et al., 1999; Smith, 1985). For a scalar model, the cumulative distribution can be calculated explicitly when using the CF approximation. It is independently uniformly distributed on (0, 1) under the correctness of the model and can therefore be used as a non-Bayesian approach for model adequacy. The first study that incorporates parameter uncertainty in their predictive distribution is in Geweke and Amisanoy (2008). This approach can be generalized to stochastic volatility models.
We propose to study the applicability of the various model comparison procedures and diagnostics to our modified exchange algorithm, and, even more importantly, compare various models when applied to real data such as the daily S&P 500 (Geweke and Amisanoy, 2008).

6 Conclusions

The closed-form (CF) transition density is a powerful tool for the analysis of a broad class of jump-diffusion models. The intractable normalizer in the CF likelihood is close to 1 when near the MLE, but can markedly differ when far in the tails of the posterior, hindering Bayesian analysis. In this paper we have provided a Bayesian approach for inference using the CF transition density. Our algorithm is based on the exchange algorithm, first proposed in Murray et al. (2006), and avoids computation of the normalizing constant.

Our examples, based on simulation studies and the analysis of real data, demonstrate that our normalizing algorithm (i.e. the modified exchange algorithm) greatly increases the stability and mixing behavior of the MCMC sampler. Furthermore, the modified exchange algorithm is quite efficient, mixes well, and is relatively easy to implement.

We also discuss ways to choose the order of the closed-form (CF) approximation, and examine potential diagnostic and model selection procedures. These topics provide several tracks of future research, and will be assessed via their application to real-world data.
Figure 6: Heston Model: Parameter trace plots for estimation of the Heston model on joint S&P 500 and VIX data from March 1, 2000 to February 27, 2009. Bayesian estimation is performed using closed-form likelihood expansions from Aït-Sahalia (2008) using the normalizing MCMC algorithm developed in this paper.
References


