Bayesian Inference for Discretely Sampled Diffusion Processes with Closed-Form Likelihood Expansions

Osnat Stramer[1] Matthew Bognar[2] Paul Schneider[3]

April 24, 2010

[1]Department of Statistics and Actuarial Science, University of Iowa, osnat-stramer@uiowa.edu
[2]Department of Statistics and Actuarial Science, University of Iowa, matthew-bognar@uiowa.edu
[3]Finance Group, Warwick Business School, paul.schneider@wbs.ac.uk

Diffusions

- Diffusions are used to model continuous time processes and are therefore commonly used in financial models
- A diffusion process is described as a solution to the stochastic differential equation (SDE)

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

  where
  - Y_t takes values in $\Re^d$
  - μ is the drift function of dimension d
  - σ is the volatility function of dimension $d \times d$
  - θ is parameter vector
  - W_t is a $d-$dimensional Brownian motion
- The transition density

$$p_Y(\Delta, x | x_0, \theta)$$

  is the conditional density of $Y_{t+\Delta} = x$ given $Y_t = x_0$

Data, Likelihood

- Assume that the process Y is observed at discrete time points $t_i = i\Delta$, $i = 0, \ldots, n$, yielding observations

$$\mathbf{x} = (x_0, \ldots, x_n)$$

  - (Y_t may be observed with noise, $X_t = Y_t + \epsilon_t$)
- 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

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

- In most instances, the transition densities

$$p_Y(\Delta, x_i|x_{i-1}, \theta)$$

  are not analytically available
  - The likelihood is therefore not available

Cox-Ingersoll-Ross Model

- CIR model

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

  where

  - $\alpha =$ mean reverting level
  - $\beta =$ speed of the process
  - $\sigma =$ volatility parameter
  - $\theta = (\alpha, \beta, \sigma)$

- The true transition density

$$p_Y(\Delta, x|x_0, \theta)$$

  (the conditional density of $Y_{t+\Delta} = x$ given $Y_t = x_0$) is a non-central χ^2 distribution

Closed–Form Approximation-one dimensional diffusions

The analytical, closed–form (CF) approximation of the unknown transition density was introduced by Aït-Sahalia (2002) for one dimensional diffusions.

- Avoids a computationally intensive data augmentation scheme
- For univariate diffusions, the CF approximation is a non-Gaussian approximation to the transition density by means of a truncated Hermite series expansion
- Converges to the true, unknown likelihood function as the number of terms in the Taylor–like expansion increases
- The effectiveness of the CF approximation is well documented by Jensen and Poulsen (2002), Hurn, Jeisman and Lindsay (2007), and Stramer and Yan (2007).
- The CF approximation should be used with caution for very volatile models or sparse data-sets (see Stramer and Yan, 2007)

Closed Form Approximation-multivariate diffusions

- Aït-Sahalia (2008) provides an extension of the CF approximation for multidimensional diffusions. This includes a broad class of models used in the literature such as stochastic volatility 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.

Bayesian Approach Using Data Augmentation

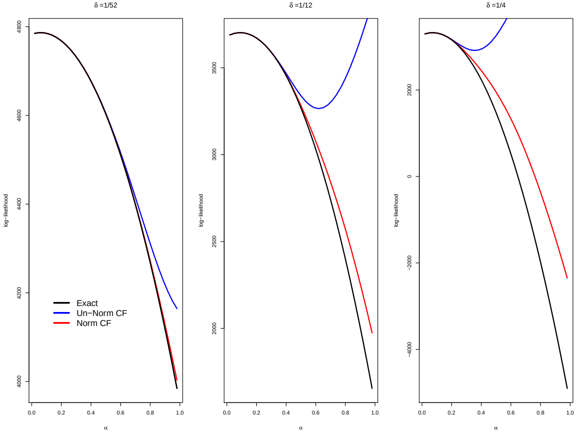
Typical approach for Bayesian estimation in discretely observed diffusion models are classical missing-data techniques

- Introduce latent auxiliary data to complete the missing diffusion between each adjacent pair of data points (computationally intensive)
- These algorithms can break down due to high dependence between the volatility coefficient and the missing data
- Re-parametrization methods can help break down the high dependency (see Roberts and Stramer (2001), Kalogeropoulos et al. (2007), Chib et al. (2006), Golightly and Wilkinson (2008))

Bayesian Approach Using the CF Approximation

- The CF likelihood does *not* integrate to 1
- The normalizer:
  - is an *intractable* function of the parameters θ
  - is very close to 1 for values of θ close to the MLE
  - can differ markedly from 1 when θ is far from MLE
  - is not a problem if we only seek to find the MLE
- Bayesian practitioners want to use MCMC techniques to sample from the posterior distribution of θ
  - May need to evaluate the likelihood far from the MLE
  - The CF approximation is least accurate in such instances
  - The MCMC sampler may get stuck in the tails of the posterior, typically when θ is far from the MLE

Comparison Of Exact, Norm CF, Un-Norm CF Likelihoods



- CIR datasets generated with $n = 500$ ($n = 1{,}000$ for weekly), $\alpha = 0.07$, $\beta = 0.15$, $\sigma = 0.07$
- β and σ held fixed; evaluate log-likelihood for $\alpha \in (0, 1)$

Addressing the Normalizer

- Univariate case: DiPietro (2001) approximated, the unknown normalizer using numerical integration
  - ▷ His simulation study, shows that the normalized CF approximation outperforms the augmentation technique introduced in Elerian et al. (2001)
  - ▷ Yet, effect of using an *estimate* of normalizing constant on the MCMC algorithm is not clear
- multidimensional diffusions
  - ▷ 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 DiPietro (2001) to the CF approximation for most multivariate diffusions.
- The first method that avoids such approximations was proposed by Møller et al. (2006)
  - ▷ Introduce a cleverly chosen auxiliary variable into the Metropolis-Hastings (M-H) algorithm so that the normalizing constants cancel in the M-H ratio
  - ▷ A simpler and more efficient version, which inspired our work, is proposed in Murray et al. (2006)

Bayesian Framework: CF Transition Density

- Denote the CF approximation of $p_Y(\Delta, x|x_0, \theta)$ by

$$g_{CF}(\Delta, x|x_0, \theta)$$

- Denote the *normalized* CF transition density by

$$p_{CF}^N(\Delta, x|x_0, \theta) \stackrel{\text{def}}{=} \frac{g_{CF}(\Delta, x|x_0, \theta)}{Z(x_0, \theta)}$$

  - ▷ $Z(x_0, \theta) = \int g_{CF}(\Delta, x|x_0, \theta)\ dx$
  - ▷ $Z(x_0, \theta)$ is <u>analytically intractable</u>

CF Likelihood & Posterior

- The CF likelihood is

$$\begin{aligned} L_{CF}^{N}(\mathbf{x}|\theta) &= \prod_{i=1}^{n} p_{CF}^{N}(\Delta, x_i|x_{i-1}, \theta) \\ &= \prod_{i=1}^{n} \frac{g_{CF}(\Delta, x_i|x_{i-1}, \theta)}{Z(x_{i-1}, \theta)} \end{aligned}$$

- Goal: use MCMC techniques to sample from the posterior distribution

$$\pi_{CF}^{N}(\theta|\mathbf{x}) \propto L_{CF}^{N}(\mathbf{x}|\theta)\pi(\theta)$$

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

Metropolis-Hastings Algorithm

Can not use a standard M-H algorithm for updating θ since the acceptance ratio involves the intractable normalizing constants

1. Let θ be the current value. Generate θ^* from some proposal density $q(\theta^*|\theta)$
2. Accept θ^* with probability $\min[1, \mathcal{R}_{MH}]$ where

$$\begin{aligned}\mathcal{R}_{MH} &= \frac{L_{CF}^N(\mathbf{x}|\theta^*)}{L_{CF}^N(\mathbf{x}|\theta)} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \\ &= \frac{\prod_{i=1}^n g_{CF}(\Delta, x_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*)}{\prod_{i=1}^n g_{CF}(\Delta, x_i|x_{i-1}, \theta)/Z(x_{i-1}, \theta)} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}\end{aligned}$$

The intractable normalizers $Z(\cdot, \cdot)$ do *not* cancel

Exchange Algorithm

- Murray et al. (2006) suggested a clever auxiliary variable algorithm to simulate from the posterior
- Assume likelihood takes the form $L(\mathbf{x}|\theta) = \prod_{i=1}^{n} g(x_i|\theta)/Z(\theta)$ where $Z(\theta)$ is an intractable normalizer
- Update procedure
  1. Generate θ^* from some proposal density $q(\theta^*|\theta)$
  2. Generate a sample $\mathbf{w}$ from

$$L(\mathbf{w}|\theta^*) = \prod_{i=1}^{n} g(w_i|\theta^*)/Z(\theta^*)$$

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

$$\mathcal{R}_{Mur} = \frac{\prod_{i=1}^{n} g(x_i|\theta^*)/Z(\theta^*)}{\prod_{i=1}^{n} g(x_i|\theta)/Z(\theta)} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \times \frac{\prod_{i=1}^{n} g(w_i|\theta)/Z(\theta)}{\prod_{i=1}^{n} g(w_i|\theta^*)/Z(\theta^*)}$$

The intractable Z's cancel

Exchange Algorithm Continued

The Murray algorithm is not applicable in our situation since

1. Generate θ^* from some proposal density $q(\theta^*|\theta)$
2. Generate a sample from

$$L_{CF}^N(\mathbf{w}|\theta^*) = \prod_{i=1}^n \frac{g_{CF}(\Delta, w_i|w_{i-1}, \theta^*)}{Z(w_{i-1}, \theta^*)}$$

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

$$\begin{aligned}\mathcal{R}_{Mur} &= \frac{\prod_{i=1}^n g_{CF}(\Delta, x_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*)}{\prod_{i=1}^n g_{CF}(\Delta, x_i|x_{i-1}, \theta)/Z(x_{i-1}, \theta)} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \\ &\times \frac{\prod_{i=1}^n g_{CF}(\Delta, w_i|w_{i-1}, \theta)/Z(w_{i-1}, \theta)}{\prod_{i=1}^n g_{CF}(\Delta, w_i|w_{i-1}, \theta^*)/Z(w_{i-1}, \theta^*)}\end{aligned}$$

The intractable normalizers $Z(\cdot, \cdot)$ do *not* cancel

Modified Exchange Algorithm

1. Propose a new value θ^* from some proposal density $q(\theta^*|\theta)$
2. Generate $\mathbf{w} = (w_1, \ldots, w_n)$ from

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

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

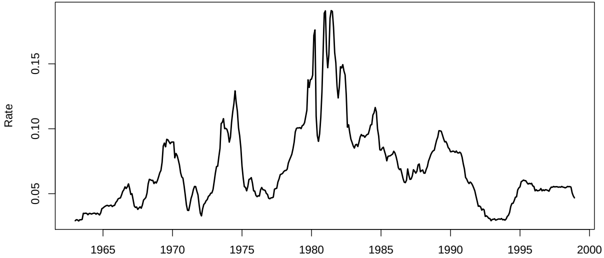
$$\begin{aligned}
\mathcal{R}_\theta &= \frac{\prod_{i=1}^{n} g_{CF}(\Delta, x_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta^*)}{\prod_{i=1}^{n} g_{CF}(\Delta, x_i|x_{i-1}, \theta)/Z(x_{i-1}, \theta)} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \\
&\quad \times \frac{\prod_{i=1}^{n} g_{CF}(\Delta, w_i|x_{i-1}, \theta^*)/Z(x_{i-1}, \theta)}{\prod_{i=1}^{n} g_{CF}(\Delta, w_i|x_{i-1}, \theta)/Z(x_{i-1}, \theta^*)} \\
&= \frac{\prod_{i=1}^{n} g_{CF}(\Delta, x_i|x_{i-1}, \theta^*)}{\prod_{i=1}^{n} g_{CF}(\Delta, x_i|x_{i-1}, \theta)} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \\
&\quad \times \frac{\prod_{i=1}^{n} g_{CF}(\Delta, w_i|x_{i-1}, \theta^*)}{\prod_{i=1}^{n} g_{CF}(\Delta, w_i|x_{i-1}, \theta)}
\end{aligned}$$

Modified Exchange Algorithm: Generating $\mathbf{w}$

The key features of the exchange algorithm in Murray et al. (2006) is that exact samples should be drawn from the likelihood. The same is needed for the modified exchange algorithm. Drawing samples from $\frac{g_{CF}(\Delta, w_i | x_{i-1}, \theta)}{Z(x_{i-1}, \theta)}$ is usually not feasible. There are two main approaches:

- For reducible diffusions this can be done via exact simulation (Beskos 2006). For the general case, (i.e. for irreducible diffusions) this has traditionally implied the use of some of the time discrete approximation methods (Euler, Taylor's expansion, etc.) which rely on small time approximate increment distributions for the diffusion
- A different approach is to run an inner-loop Metropolis-Hastings algorithm to simulate from $\frac{g_{CF}(\Delta, w_i | x_{i-1}, \theta)}{Z(x_{i-1}, \theta)}$.

Example: FedFunds Rate



- FedFunds rate observed monthly from January 1963 to December 1998 ($n = 432$)

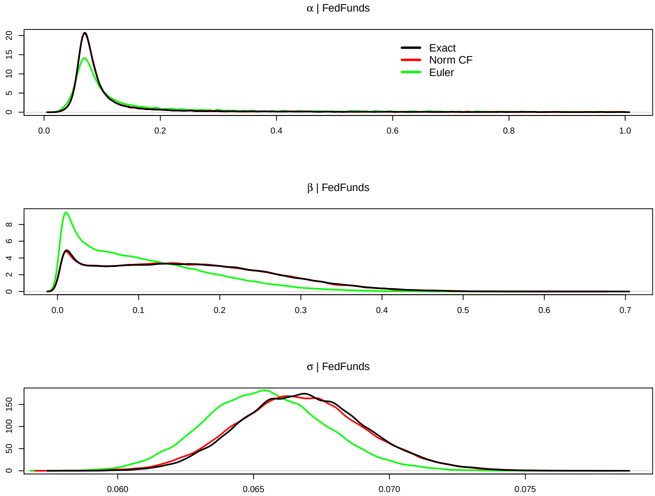
FedFunds Analysis

- Use CIR model. Compare Bayesian analyses using:
  - Exact likelihood (using standard M–H)
  - Euler likelihood (using standard M–H)
  - Un–normalized CF likelihood (using standard M–H)
  - Normalized CF likelihood (via Modified Exchange Algorithm)
- Prior: Same as in DiPietro (2001):

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

- Proposals: Joint $(\alpha, \beta)-$move uses a multivariate$-t$ proposal, random–walk proposal for σ
- Ran 500,000 iterations (after burn-in period)
  - When using the un-normalized CF likelihood, the sampler repeatedly became stuck at a late stage (e.g. one chain got stuck after 95,000 iterations, another after 60,000 iterations)

FedFunds: Estimated Marginal Posterior Densities



Example: Heston Model

- The Heston model takes the form $[Y_t, V_t]$ where
  - Y_t is the log-price for the stock S_t
  - V_t is a volatility process; V_t is latent and is estimated from the VIX implied volatility index (observed w/ or w/o noise) published by the CBOE
- We assume that $[Y_t, V_t]$ follows

$$
\begin{aligned}
dY_t &= \left(\mu - \frac{1}{2}V_t\right)dt + \rho\sqrt{V_t}\,dW_t + \sqrt{1-\rho^2}\sqrt{V_t}\,dB_t \\
dV_t &= \beta(\alpha - V_t)dt + \sigma\sqrt{V_t}\,dW_t
\end{aligned}
$$

  - B and W are independent standard Brownian motions
  - Instantaneous correlation between dY_t and dV_t is controlled by ρ
  - To keep the simulation study simple we make the assumption of risk premia such that $W_t = W_t^{\mathbb{Q}}$ and $dB_t = dB_t^{\mathbb{Q}} + \frac{r-\mu}{\sqrt{(1-\rho^2)V_t}}dt$
  - $\theta = (\alpha, \beta, \mu, \sigma, \rho)$

Heston Model Continued

- Simulation study of *non-noisy* Heston model:
  - Weekly Data: Excellent performance of both normalized and un-normalized CF
  - Monthly Data: The sampler using un-normalized CF occasionally became stuck, all normalized CF mix well
  - Quarterly Data: All chains using un-normalized CF become stuck, all normalized CF mix well. CF estimates exhibit a tendency to underestimate the diffusion coefficient σ and the speed-of-mean-reversion coefficient β
- Simulation study of *noisy* Heston model:
  - Weekly data: Excellent performance of both normalized and un-normalized CF
  - Monthly data: All un-normalized CF get stuck, all normalized CF mix properly
  - Quarterly data: All un-normalized CF get stuck, all normalized CF mix properly.

Example: S&P 500 and VIX Implied Volatility

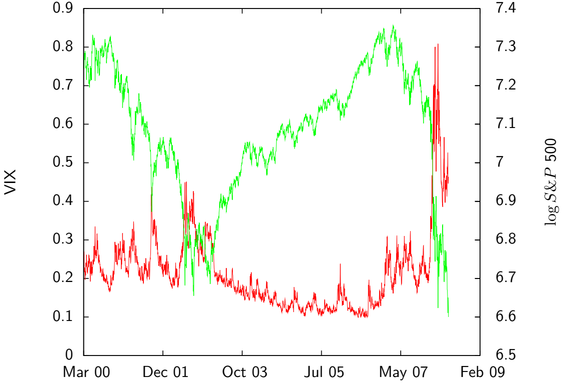
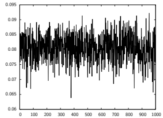
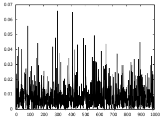


Figure: The figure depicts $\log S\&P$ 500 and VIX implied volatility data from 1 March 2000 to 27 February 2009. The scale for the VIX index can be seen on the left axis, the right axis shows the scale for $\log S\&P$ 500.

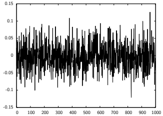
Example: S&P 500 and VIX Implied Volatility



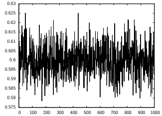
(a) $\alpha\beta$ | $S\&P500, VIX$



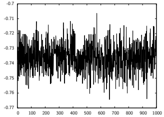
(b) β | $S\&P500, VIX$



(c) μ | $S\&P500, VIX$



(d) σ | $S\&P500, VIX$



(e) ρ |

Discussion

- The closed form (CF) transition density of Aït-Sahalia is a powerful tool for the analysis of diffusions
  - ▷ The intractable normalizer in the CF likelihood is close to 1 when near the MLE, but can markedly differ when away from the MLE
  - ▷ The unnormalized CF likelihood is least accurate when far from MLE
- The Modified Exchange Algorithm:
  - ▷ is quite efficient
  - ▷ is relatively easy to implement
  - ▷ greatly stabilizes and improves mixing behavior of the sampler

Thank you