Recursive Estimation for Continuous Time Stochastic Volatility Models using the Milstein Method

T. Koulis[a,*], A. Paseka[b,**], A. Thavaneswaran[a,]

[a] *Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada*
[b] *Department of Accounting and Finance, University of Manitoba, Winnipeg, Manitoba, Canada*

Abstract

Optimal as well as recursive parameter estimation for semimartingales had been studied in Thavaneswaran and Thompson [1, 2]. Recently, there has been a growing interest in modeling volatility of the observed process by nonlinear stochastic processes (Taylor [3]). In this paper, we study the recursive estimates for various classes of discretely sampled continuous time stochastic volatility models using the Milstein method. We provide closed form expressions for the recursive estimates for recently proposed stochastic volatility models. We also give an example of computation of the term structure of zero rates in an incomplete information environment. In this case, learning about an unobserved state variable is done jointly with the valuation procedure.

Keywords: Recursive estimation, diffusion processes, interest rate models, Milstein method

1. Introduction

In the last three decades, semimartingales have received considerable attention with the emphasis being placed on state space models. From an econometric standpoint, time-varying volatility models have been widely developed, recognizing that the volatility and the correlation of assets change over time (see for example Heston and Nandi [4]). State space models in which the conditional mean of the observed process is modeled as a stochastic process are useful in parameter estimation. For example, stochastic volatility models (Kawakatsu [5], Taylor [3]) are widely employed to estimate volatility parameters.

Thavaneswaran and Thompson [2] uses the estimating function approach for the recursive parameter estimation in models with semimartingales. Thavaneswaran and Thompson [1], Naik-Nimbalkar and Rajarshi [6] and Thompson and Thavaneswaran [7] use the estimating function method for the estimation of state space models in the Bayesian setup. Parameter estimates obtained in Thavaneswaran and Thompson [2] involve the evaluation of the stochastic integrals based on the observation of the complete path of the observed process. However, for continuous time models, it is more appropriate to study parameter estimates based on discretely observed data. In order to study the inference for diffusion processes based on discretely observed data, one has to approximate the continuous time diffusion by a discrete process. For some interest rate models (Vasicek, Cox-Ingersoll-Ross), discrete time approximation has been used to study parameter estimation (see Thavaneswaran et al. [8], Sorensen [9], and the references therein). However, the *recursive parameter estimation* based on *discrete approximation* has not been studied in the literature. **We have to define recursive par estimation. Also, is it true that people such as Ait-Sahalia have estimated**

*Corresponding author: (204) 474-8205
**Corresponding author: (204) 474-8353
Corresponding author: (204) 474-8984
Email addresses: theo.koulis@ad.umanitoba.ca (T. Koulis), paseka@cc.umanitoba.ca (A. Paseka), thavane@cc.umanitoba.ca (A. Thavaneswaran)

parameters in discretised models in the context of approximate likelihood. We need to cite his main paper I think.

In most realistic situations, the diffusion cannot be observed continuously, so discrete time approximations to stochastic integrals or a direct approach using discrete time observations is required. For extended versions of the Cox-Ingersoll-Ross (CIR) model **(reference???)**, closed form expressions for the first four conditional moments cannot be obtained easily by using Ito's formula, as was done for the non-extended CIR model**(reference???)**. Recently, Jeong and Park [10] uses the Milstein method (see Kloeden and Platen [11]) to obtain the first two conditional moments of a diffusion. For diffusion models with a finite number of parameters, Koulis and Thavaneswaran [12] use the Milstein method to obtain the first four conditional moments and to construct the optimal estimating functions for the Vasicek model of the form

$$dy_t = \mu_t dt + \sigma_t dW(t),$$

with $\mu_t = \alpha(\beta - y_t)$, $\sigma_t \equiv \sigma$, and $\alpha > 0$. One of the drawbacks of this one-factor model is that it is not in general possible to calibrate it so that it fits the presently observed term structure. For example, Kennedy [13, p. 171] points out that for the above Vasicek model, which depends on three parameters, α, β, and σ, it is not possible to choose values of those parameters so that the entire observed term structure of interest rates is fitted exactly by the model. To solve the problem, Kennedy proposes to allow time-varying parameters in the drift term of the Vasicek model.

Consider a diffusion process given by the time-homogeneous stochastic differential equation of the form

$$dy_t = a(\boldsymbol{\alpha}, y_t)dt + b(\boldsymbol{\beta}, y_t)dW(t) \tag{1.1}$$

where a and b are the drift and diffusion functions, respectively, and $W(t)$ is the standard Brownian motion. A special case of (1.1) is the Brownian motion with constant drift and diffusion coefficients:

$$dy_t = \alpha dt + \beta dW(t),$$

where $\beta > 0$. In this case, the conditional distribution of y_t given $y_0 = y$ is a normal with mean $y + \alpha t$ and variance $\beta^2 t$. If we consider the geometric Brownian motion given by

$$dy_t = \nu y_t dt + \omega y_t dW(t),$$

with $\omega > 0$, then $\log(y_t)$ becomes a Brownian motion with drift with $\alpha = \nu - \omega^2/2$ and $\beta = \omega$. In this case, the conditional distribution of $\log(y_t)$ given $\log(y_0) = \log(y)$ is also normal. The CIR process can be re-parameterized to the following form:

$$dy_t = (\alpha_1 + \alpha_2 y_t)dt + \beta\sqrt{y_t}dW(t).$$

Extended versions of the CIR process model have been proposed for modeling interest rate processes. For example, some consider the constant elasticity of variance process of the form

$$dy_t = (\alpha_1 + \alpha_2 y_t)dt + \beta_1 y_t^{\beta_2} dW(t)$$

or the nonlinear drift diffusion process (Ait-Sahalia [14]) given by

$$dy_t = (\alpha_1 + \alpha_2 y_t + \alpha_3 y_t^2 + \alpha_4 y_t^{-1})dt + \sqrt{\beta_1 + \beta_2 y_t + \beta_3 y_t^{\beta_4}}dW(t).$$

For more general extended models, the diffusion is a function of the observation y_t and hence, closed form expressions of the conditional distributions, as well as closed form expressions for the conditional moments cannot be easily obtained by solving differential equations obtained by repeated application of Itô's formula. However, the Milstein method can be used to obtain the first four conditional moments.

If we consider a discretisation in small intervals of time $t_i - t_{i-1} = h$, then the Milstein method applied to (1.1) produces

$$y_{t_i} = y_{t_{i-1}} + a(\boldsymbol{\alpha}, y_{t_{i-1}})h + b(\boldsymbol{\beta}, y_{t_{i-1}})\sqrt{h}\epsilon_{t_i} + \frac{1}{2}b(\boldsymbol{\beta}, y_{t_{i-1}})\dot{b}_y(\boldsymbol{\beta}, y_{t_{i-1}})\left(\epsilon_{t_i}^2 - 1\right)h, \tag{1.2}$$

where $\dot{b}_y = \frac{\partial b}{\partial y}$and $\epsilon_t \sim N(0,1)$, i.i.d.

Unlike the Euler method for diffusion processes, the Milstein method in (1.2) gives a non-Gaussian time series model for $y_{t_i} - y_{t_{i-1}}$. The distribution implied by the Milstein method is a mixture of a normal and chi-square distribution. Moreover, for the extended CIR model and for more general diffusion processes, Ito's approximation cannot be used to obtain closed form expressions for the first four conditional moments. In this paper, first we use the Milstein method to discretise the continuous time diffusion processes and then study the recursive estimates of latent state variables. We also show how the proposed method can be used to derive zero coupon bond prices in the incomplete information environment. In this case, the valuation exercise and the recursive estimation (learning) of the unobserved state variable are performed simultaneously by market participants.

2. State Space Models

In order to construct an optimal recursive estimate for non-normal stochastic volatility models, we start with the following discrete time example.Let the discrete-time state space model of the observed process $\{y_t\}$and the state process $\{\theta_t\}$be given by:

$$\begin{aligned} y_{t+1} &= A\theta_t + az_{t+1} + b(z_{t+1}^2 - 1) \\ \theta_{t+1} &= B\theta_t + c\eta_{t+1} + d(\eta_{t+1}^2 - 1) \end{aligned} \tag{2.1}$$

where A, B, a, b, c and d are positive constants, and possibly measurable with respect to the σ-field $\mathcal{F}_t^y$ generated by the observations of $\{y_s\}$ up to and including time t. In addition, $\{z_t\}$ and $\{\eta_t\}$ are two standard Gaussian sequences of identically distributed random variables with $\mathrm{Corr}(z_t, \eta_t) = \rho$. The following lemma will be used to prove our main Theorem.

Lemma 2.1. Assume that $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$ with $\mathrm{Corr}(Z_1, Z_2) = \rho$. Then $\mathrm{Corr}(Z_1^2, Z_2^2) = \rho^2$.

Proof: It follows from the theorem on Normal correlation that the conditional expectation and conditional variance of Z_1 given Z_2 are give by $E\left[Z_1|Z_2\right] = \rho Z_2$ and $\mathrm{Var}\left[Z_1|Z_2\right] = (1-\rho^2)$. Using the law of total expectation, we also have

$$\begin{aligned} E\left[Z_1^2 Z_2^2\right] &= E\left[Z_2^2\,\mathrm{E}\left[Z_1^2|Z_2\right]\right] = E\left[Z_2^2(1-\rho^2) + \rho^2 Z_2^4\right] \\ &= (1-\rho^2) + 3\rho^2 = 1 + 2\rho^2\ . \end{aligned}$$

Hence, the correlation between Z_1^2 and Z_2^2 is given as

$$\mathrm{Corr}(Z_1^2, Z_2^2) = \frac{E\left[Z_1^2 Z_2^2\right] - 1}{\sqrt{4}} = \rho^2.$$

The following theorem establishes the recursive estimation for the state space model (2.1).

Theorem 2.2. Given the state space model (2.1), and the class of all estimators of the form:

$$\widehat{\theta}_{t+1} = B\widehat{\theta}_t + \widehat{G}_t(y_{t+1} - A\widehat{\theta}_t)\ ,$$

the G_t, which minimizes the mean-square error, $\gamma_{t+1} = E\left[(\theta_{t+1} - \widehat{\theta}_{t+1})^2|\mathcal{F}_t^y\right]$, is given by

$$\widehat{G}_t = \frac{AB\gamma_t + \rho\,(ac + 2\rho bd)}{A^2\gamma_t + a^2 + 2b^2}.$$

Moreover, the mean-square error is given as

$$\gamma_{t+1} = \left(B - A\widehat{G}_t\right)^2\gamma_t + c^2 + 2d^2 + \widehat{G}_t^2(a^2 + 2b^2) - 2\rho\widehat{G}_t(ac + 2\rho bd).$$

Proof: The difference $\theta_{t+1} - \widehat{\theta}_{t+1}$ is given by

$$\begin{aligned}\theta_{t+1} - \widehat{\theta}_{t+1} &= B(\theta_t - \widehat{\theta}_t) + c\eta_{t+1} + d(\eta_{t+1}^2 - 1) - G_t\left(A\theta_t + az_{t+1} + b(z_{t+1}^2 - 1) - A\widehat{\theta}_t\right) \\ &= (B - AG_t)(\theta_t - \widehat{\theta}_t) + c\eta_{t+1} + d(\eta_{t+1}^2 - 1) - aG_t z_{t+1} - bG_t(z_{t+1}^2 - 1) \ .\end{aligned}$$

Squaring the above expression, taking expectations, and using the results of Lemma 2.1 it follows that the conditional mean-square error at $t+1$ is given by

$$\gamma_{t+1} = (B - AG_t)^2\,\gamma_t + c^2 + 2d^2 + G_t^2(a^2 + 2b^2) - 2\rho G_t(ac + 2\rho bd) \ .$$

Differentiating γ_{t+1} with respect to G_t and setting the first derivative to zero, we have

$$-2A(B - AG_t)\gamma_t + 2G_t(a^2 + 2b^2) - 2\rho(ac + 2\rho bd) = 0 \ .$$

Solving for G_t, we obtain

$$\widehat{G}_t = \frac{2AB\gamma_t + \rho(ac + 2\rho bd)}{2A^2\gamma_t + a^2 + 2b^2}.$$

Corollary 2.3. Let the state space model be of the form

$$\begin{aligned}y_{t+1} &= A\theta_t + z_{t+1} \\ \theta_{t+1} &= B\theta_t + \eta_{t+1}\end{aligned}$$

where $\{z_t\}$ and $\{\eta_t\}$ are two sequences of independent and identically distributed random variables having mean zero and variance σ_z^2 and σ_η^2, respectively. In the class of estimates of the form:

$$\widehat{\theta}_{t+1} = B\widehat{\theta}_{t+1} + \widehat{G}_t(y_{t+1} - A\widehat{\theta}_t) \ ,$$

the G_t which minimizes the mean-square error $\gamma_t = E\left[(\theta_t - \widehat{\theta}_t)|F_t^y\right]$ is given by

$$\widehat{G}_t = \frac{BA\gamma_t}{A^2\gamma_t + \sigma_z^2}.$$

In addition, the mean-square error is given as

$$\gamma_{t+1} = \left(B - \widehat{G}_t A\right)^2\gamma_t + \sigma_\eta^2 + \widehat{G}_t^2\sigma_z^2.$$

Proof: The result follows from Theorem 2.2 by setting $a = \sigma_z$, $b = \sigma_\eta$, $c = d = 0$, and $\rho = 0$.

3. General Model

In the continuous-time setting, consider the general state space model of the form

$$\begin{aligned}dy_t &= A(y_t)\theta_t dt + \alpha(y_t)dW_1(t), \\ d\theta_t &= B(y_t)\theta_t dt + \beta(y_t, \theta_t)dW_2(t)\end{aligned}$$

where $W_1(t)$ and $W_2(t)$ are two uncorrelated standard Brownian motions. If we consider a discretisation in small intervals of time $t_i - t_{i-1} = h$, $i = 0, 1, \ldots$, then the Milstein method gives a non-Gaussian discrete state-space model of the form:

$$\begin{aligned}y_{t_{i+1}} - y_{t_i} &= A(y_{t_i})\theta_{t_i}h + \alpha(y_{t_i})\sqrt{h}z_{t_{i+1}} + \frac{h}{2}\alpha(y_{t_i})\dot{\alpha}_y(y_{t_i})\left(z_{t_{i+1}}^2 - 1\right), \qquad (3.1) \\ \theta_{t_{i+1}} &= [1 + B(y_{t_i})h]\theta_{t_i} + \beta(y_{t_i})\sqrt{h}\eta_{t_{i+1}} + \frac{h}{2}\beta(y_{t_i}, \theta_{t_i})\dot{\beta}_\theta(y_{t_i}, \theta_{t_i})\left(\eta_{t_{i+1}}^2 - 1\right),\end{aligned}$$

where $\dot{\alpha}_y = \frac{\partial \alpha}{\partial y}$ and $\dot{\beta}_y = \frac{\partial \beta}{\partial \theta}$, and $\{z_{t_i}\}$ and $\{\eta_{t_i}\}$ are two independent standard Gaussian sequences of independent and identically distributed random variables.

We can relate the discretised model (3.1) to the discrete-time model (2.1) by letting $y_{t+1} \equiv y_{t_{i+1}} - y_{t_i}$, $\theta_{t+1} \equiv \theta_{t_{i+1}}$, $z_{t+1} \equiv z_{t_{i+1}}$, and $\eta_{t+1} \equiv \eta_{t_{i+1}}$. In addition, we have $A \equiv A(y_{t_i})$, $a \equiv \alpha(y_{t_i})\sqrt{h}$, $b \equiv \frac{h}{2}\alpha(y_{t_i})\dot{\alpha}_y(y_{t_i})$, $B \equiv [1 + B(y_{t_i})h]$, $c \equiv \beta(y_{t_i})\sqrt{h}$, $d \equiv \frac{h}{2}\beta(y_{t_i}, \theta_{t_i})\dot{\beta}_\theta(y_{t_i}, \theta_{t_i})$, and $\rho \equiv 0$. It now follows from Theorem 2.2 that the recursive estimator is of the form

$$\widehat{\theta}_{t+1} = [1 + B(y_{t_i})h]\widehat{\theta}_t + \widehat{G}_t \left(y_{t+1} - A(y_{t_i})\,\widehat{\theta}_t h\right),$$

where

$$\widehat{G}_t = \frac{A(y_{t_i})[1 + B(y_{t_i})h]\gamma_t}{A^2(y_{t_i})\gamma_t + \left(\alpha^2(y_{t_i})h + \frac{1}{2}h^2\alpha^2(y_{t_i})\dot{\alpha}_y^2(y_{t_i})\right)},$$

and the mean-square error is given as

$$\begin{aligned}\gamma_{t+1} = {} & \left(1 + B(y_{t_i})h - A(y_{t_i})\widehat{G}_t\right)^2 \gamma_t + \beta^2(y_{t_i})h + \frac{1}{2}h^2\beta^2(y_{t_i}, \widehat{\theta}_{t_i})\dot{\beta}_\theta^2(y_{t_i}, \widehat{\theta}_{t_i}) \\ & + \widehat{G}_t^2\left(\alpha^2(y_{t_i})h + \frac{1}{2}h^2\alpha^2(y_{t_i})\dot{\alpha}_y^2(y_{t_i})\right).\end{aligned}$$

Example 3.1 (Klebaner's Model)**.** Klebaner [15] considers a state space model in which the conditional mean of the observed diffusion process is modeled by the Black-Scholes process (Black and Scholes [16]) and given by:

$$\begin{aligned} dy_t &= \theta_t dt + dW_1(t), \\ d\theta_t &= \left(\mu + \frac{\sigma^2}{2}\right)\theta_t dt + \sigma\theta_t dW_2(t), \end{aligned}$$

where $W_1(t)$ and $W_2(t)$ are two independent standard Brownian motions. In this case, the Milstein method leads to

$$\begin{aligned} y_{t_{i+1}} - y_{t_i} &= \theta_{t_i} h + \sqrt{h} z_{t_{i+1}}, \\ \theta_{t_{i+1}} &= \theta_{t_i} + \left(\mu + \frac{\sigma^2}{2}\right) h\theta_{t_i} + \sigma\theta_{t_i}\sqrt{h}\eta_{t_{i+1}} + \frac{h}{2}\sigma^2\theta_{t_i}(\eta_{t_{i+1}}^2 - 1). \end{aligned} \tag{3.2}$$

We relate (3.2) to the discrete-time model (2.1) by letting $y_{t+1} \equiv y_{t_{i+1}} - y_{t_i}$, $\theta_{t+1} \equiv \theta_{t_{i+1}}$, $z_{t+1} \equiv z_{t_{i+1}}$, and $\eta_{t+1} \equiv \eta_{t_{i+1}}$. Also, we put $A \equiv h$, $a \equiv \sqrt{h}$, $b \equiv 0$, $B \equiv \left[1 + \left(\mu + \frac{\sigma^2}{2}\right) h\right]$, $c \equiv \sigma\theta_{t_i}\sqrt{h}$ and $d \equiv \frac{h}{2}\sigma^2\theta_{t_i}$. It now follows from Theorem 2.2 that the recursive estimator is of the form

$$\widehat{\theta}_{t+1} = \left[1 + \left(\mu + \frac{\sigma^2}{2}\right) h\right]\widehat{\theta}_t + \widehat{G}_t(y_{t+1} - h\widehat{\theta}_t),$$

where

$$\widehat{G}_t = \frac{\left[1 + \left(\mu + \frac{\sigma^2}{2}\right) h\right]\gamma_t}{h\left(\gamma_t + 1\right)},$$

and the mean-square error is given as

$$\gamma_{t+1} = \left(\left[1 + \left(\mu + \frac{\sigma^2}{2}\right) h\right] - h\widehat{G}_t\right)^2 \gamma_t + \sigma^2\widehat{\theta}_{t_i}^2 h + \frac{1}{2}h^2\sigma^4\widehat{\theta}_{t_i}^2 + \widehat{G}_t^2 h.$$

Example 3.2 (Hull and White Model)**.** Hull and White [17] proposed the stochastic volatility model in which the conditional variance of the observed diffusion process is modeled by a Black-Scholes process and given by:

$$\begin{aligned} dy_t &= \alpha y_t dt + \theta_t y_t dW_1(t), \\ d\theta_t^2 &= a\theta_t^2 dt + b\theta_t^2 dW_2(t). \end{aligned}$$

where $W_1(t)$ and $W_2(t)$ are two correlated standard Brownian motions with $EdW_1(t)dW_2(t) = \rho dt$. We use Ito's formula to obtain $d\theta_t$:

$$d\theta_t = \underbrace{\left(\frac{a}{2} - \frac{b^2}{8}\right)}_{\mu_\theta}\theta_t dt + \frac{b}{2}\theta_t dW_2(t).$$

To simplify Milstein approximation we treat the coefficient on $dW_1(t)$ as a function of only y_t. In this case,the Milstein method leads to

$$\begin{aligned} y_{t_{i+1}} - y_{t_i} - \alpha y_{t_i} h = \theta_{t_i} y_{t_i} \sqrt{h} z_{t_{i+1}} + \frac{1}{2}\theta_{t_i}^2 y_{t_i} h(z_{t_{i+1}}^2 - 1), \\ \theta_{t_{i+1}} = (1 + \mu_\theta h)\theta_{t_i} + \frac{b}{2}\sqrt{h}\theta_{t_i}\eta_{t_{i+1}} + \frac{b^2}{8}\theta_{t_i} h(\eta_{t_{i+1}}^2 - 1). \end{aligned}$$

We relate (3.3) to the discrete-time model (2.1) by letting $y_{t+1} \equiv y_{t_{i+1}} - y_{t_i} - \alpha y_{t_i} h$, $\theta_{t+1} \equiv \theta_{t_{i+1}}$, $z_{t+1} \equiv z_{t_{i+1}}$, $\eta_{t+1} \equiv \eta_{t_{i+1}}$. Also, we put $A \equiv 0$, $a \equiv \theta_{t_i} y_{t_i}\sqrt{h}$, $b \equiv \frac{1}{2}\theta_{t_i}^2 y_{t_i} h$, $B \equiv (1 + \mu_\theta h)$, $c \equiv \frac{b}{2}\sqrt{h}\theta_{t_i}$ and $d \equiv \frac{h}{8} b^2 \theta_{t_i}$.

It now follows from Theorem 2.2 that the recursive estimator is of the form

$$\widehat{\theta}_{t+1} = (1 + \mu_\theta h)\widehat{\theta}_t + \widehat{G}_t y_{t+1} \ ,$$

where

$$\widehat{G}_t = \frac{\rho b \left(1 + \rho b \widehat{\theta}_t \frac{h}{4}\right)}{2y_t \left(1 + \frac{1}{2}\widehat{\theta}_t^2 h\right)},$$

and the mean-square error is given as

$$\gamma_{t+1} = (1 + \mu_\theta h)^2 \gamma_t + \frac{b^2}{4} h \widehat{\theta}_t^2 + \frac{h^2}{32} b^4 \widehat{\theta}_t^2 + \widehat{\theta}_t^2 y_t^2 h \widehat{G}_t^2 \left(1 + \frac{1}{2}\widehat{\theta}_t^2 h\right) - h\rho b \widehat{\theta}_t^2 y_t \widehat{G}_t \left(1 + \rho b \widehat{\theta}_t \frac{h}{4}\right).$$

When correlation $\rho = 0$, the model simplifies to

$$\begin{aligned} \widehat{G}_t &= 0 \\ \widehat{\theta}_{t+1} &= (1 + \mu_\theta h)\widehat{\theta}_t \\ \gamma_{t+1} &= (1 + \mu_\theta h)^2 \gamma_t + \frac{b^2}{4} h \widehat{\theta}_t^2 + \frac{h^2}{32} b^4 \widehat{\theta}_t^2 \end{aligned}$$

Example 3.3 (CIR Model)**.** Consider the CIR model for observed process y_t given by

$$dy_t = k(\theta_t - y_t)dt + \sigma\sqrt{y_t} dW_1(t),$$

and the state process θ_t follows a diffusion process of the form

$$\begin{aligned} & d\theta_t = B(y_t)\theta_t dt + \beta(y_t, \theta_t) dW_2(t), \\ & EdW_1(t)dW_2(t) = 0 \end{aligned}$$

In this case, the Milstein method for y_t and θ_t leads to

$$\begin{aligned} & y_{t_{i+1}} - y_{t_i} + k y_{t_i} = k\theta_{t_i} h + \sigma\sqrt{y_{t_i} h} z_{t_{i+1}} + \frac{1}{4}\sigma^2 h(z_{t_{i+1}}^2 - 1), \qquad (3.3) \\ & \theta_{t_{i+1}} = [1 + B(y_{t_i})h]\theta_{t_i} + \beta(y_{t_i})\sqrt{h}\eta_{t_{i+1}} + \frac{h}{2}\beta(y_{t_i}, \theta_{t_i})\dot{\beta}_\theta(y_{t_i}, \theta_{t_i})\left(\eta_{t_{i+1}}^2 - 1\right), \end{aligned}$$

respectively.

We relate (3.3) to the discrete-time model (2.1) by letting $y_{t+1} \equiv y_{t_{i+1}} - y_{t_i} + ky_{t_i}$, $\theta_{t+1} \equiv \theta_{t_{i+1}}$, $z_{t+1} \equiv z_{t_{i+1}}$, $\eta_{t+1} \equiv \eta_{t_{i+1}}$, and $\rho = 0$. Also, we put $A \equiv kh$, $a \equiv \sigma\sqrt{y_{t_i}h}$, $b \equiv \frac{1}{4}\sigma^2 h$, $B \equiv [1 + B(y_{t_i})h]$, $c \equiv \beta(y_{t_i})\sqrt{h}$ and $d \equiv \frac{h}{2}\beta(y_{t_i}, \theta_{t_i})\dot{\beta}_\theta(y_{t_i}, \theta_{t_i})$.

It now follows from Theorem 2.2 that the recursive estimator is of the form

$$\widehat{\theta}_{t+1} = [1 + B(y_{t_i})h]\widehat{\theta}_t + \widehat{G}_t(y_{t+1} - kh\widehat{\theta}_t),$$

where

$$\widehat{G}_t = \frac{k[1 + B(y_t)h]\gamma_t}{k^2 h\gamma_t + \sigma^2\left(y_t + \frac{1}{8}\sigma^2\right)},$$

and the mean-square error is given as

$$\gamma_{t+1} = \left(1 + B(y_{t_i})h - kh\widehat{G}_t\right)^2 \gamma_t + \beta^2(y_t)h + \frac{h^2}{2}\beta^2(y_t, \widehat{\theta}_t)\dot{\beta}^2_\theta(y_t, \widehat{\theta}_t) + \sigma^2 h\widehat{G}^2_t\left(y_t + \frac{1}{8}\sigma^2 h\right).$$

4. Bond Valuation with Recursive Learning under Milstein Approximation

We now present the computation of a zero coupon bond price in the setting of a two-factor CIR model. In two-factor models, in general, bond yields are deterministic (and usually affine) functions of two factors. There are at least two reasons for why two-factor (or even multi-factor) models are more preferable to single-factor models. First, the empirical difficulties of fitting the shape of the term structure of zero rates and their volatilities and the variation of interest rate spreads in single-factor models are well known. Second, there are institutional restrictions on the behavior of interest rates that mandate more factors than one. Central banks tend to target certain levels (or ranges) of interest rates. These levels themselves may change over time as economic conditions change. As an example we consider a variant of the two-factor CIR model presented in [18]. The model defines the short rate as a CIR process with long-run mean (also known as central tendency) being itself a CIR process:

$$\begin{aligned} dr_t &= \kappa_r(\eta_t - r_t)dt + \sigma_r\sqrt{r_t}dz_r \\ d\eta_t &= \kappa_\eta(\theta - \eta_t)dt + \sigma_\eta\sqrt{\eta_t}dz_\eta \\ &Edz_r dz_\eta = 0 \end{aligned}$$

Milstein approximation is readily available

$$\begin{aligned} r_{t+h} - r_t &= \kappa_r(\eta_t - r_t)h + \sigma_r\sqrt{r_t h}\varepsilon_r + \frac{\sigma_r^2}{4}h\left(\varepsilon_r^2 - 1\right) \\ \eta_{t+h} - \eta_t &= \kappa_\eta(\theta - \eta_t)h + \sigma_\eta\sqrt{\eta_t h}\varepsilon_\eta + \frac{\sigma_\eta^2}{4}h\left(\varepsilon_\eta^2 - 1\right) \\ &E\varepsilon_r\varepsilon_\eta = 0 \end{aligned} \tag{4.1}$$

Note that the new state variable processes are no longer normal. Rather, they are a mixture of normal and chi-squared random variables.

Because investors do not observe η_t the task of pricing a zero coupon bond is a two-stage exercise. First, investors estimate the latent central tendency process, $\widehat{\eta}_t$. For that purpose, we assume, they use the rule described in Theorem 2.2:

$$
\begin{aligned}
\widehat{\eta}_{t+h} &= \widehat{\eta}_t + \kappa_\eta(\theta - \widehat{\eta}_t)h + \widehat{G}_t(r_{t+h} - r_t - \kappa_r(\widehat{\eta}_t - r_t)h) \\
\widehat{G}_t &= \frac{\kappa_r \kappa_\eta h \gamma_t}{\kappa_r^2 h \gamma_t + \sigma_r^2 \left(r_t + \frac{\sigma_r^2}{8} h\right)} \\
\gamma_{t+h} &= \left(1 - \kappa_\eta h - \kappa_r h \widehat{G}_t\right)^2 \gamma_t + \sigma_\eta^2 h \left(\widehat{\eta}_t + \frac{\sigma_\eta^2}{8} h\right) + \widehat{G}_t^2 \sigma_r^2 h \left(r_t + \frac{\sigma_r^2}{8} h\right) \\
&= \left((1 - \kappa_\eta h)^2 - 2\kappa_r h \widehat{G}_t + 3\widehat{G}_t \kappa_r \kappa_\eta h^2\right) \gamma_t + \sigma_\eta^2 h \left(\widehat{\eta}_t + \frac{\sigma_\eta^2}{8} h\right)
\end{aligned}
$$

Second, investors value the bond conditional on the pair $(r_t, \widehat{\eta}_t)$. Thus, investors' problem is the joint problem of estimation of the latent state process and simultaneous valuation of the bond.

The fundamental valuation principle in asset pricing states that if there is no arbitrage, then there exists a positive pricing kernel (also called stochastic discount factor (SDF)) such that the following condition is satisfied by any h-period return on any asset at any time:

$$
E_t m_{t+h} R_{t+h} = 1 \tag{4.2}
$$

In our example we are interested in an h-period return on a zero coupon default-free bond, $R_{t+h} = B_{t+h}^n / B_t^{n+h}$, where B_t^n is the time t price of a zero coupon bond with n periods remaining until maturity.

The complete information version of this model is affine, and the solution for a bond price in the complete information case is available in continuous time. Here we can start with discrete-time SDF

$$
-\ln m_{t+h} = \alpha + \beta r_t + \lambda_1 \sqrt{r_t h} \varepsilon_r + \lambda_2 h \varepsilon_r^2 \tag{4.3}
$$

Finding SDF parameter restrictions requires the knowledge of the following integral of an exponential-quadratic function of a standard normal variable, ε:

$$
E_t \exp\left(\phi_t \varepsilon + \varphi_t \varepsilon^2\right) = \frac{1}{\sqrt{1 - 2\varphi_t}} \exp\left(\frac{1}{2} \frac{\phi_t^2}{1 - 2\varphi_t}\right) \tag{4.4}
$$

with transversality condition $\varphi_t < 1/2$.

The condition that the expectation of an h-period SDF has to give us the h-period short rate allows us to find SDF coefficient restrictions:

$$
\begin{aligned}
-\ln B_t^h &= r_t h = -\ln\left(E_t \exp\left(\ln m_{t+h}\right)\right) \\
&= -\ln E_t \exp\left(-\alpha - \beta r_t - \lambda_1 \sqrt{r_t h} \varepsilon_r - \lambda_2 h \varepsilon_r^2\right)
\end{aligned}
$$

Using the fundamental pricing equation (4.2), the SDF expression (4.3), and the expression for the expectation of the exponential-quadratic function of the standard normal variable in (4.4), we have

$$
\begin{aligned}
-\ln B_t^h &= r_t h \\
&= \alpha + \beta r_t + \ln \frac{1}{\sqrt{1 + 2\lambda_2 h}} + \frac{1}{2} \frac{\lambda_1^2 r_t h}{1 + 2\lambda_2 h}
\end{aligned} \tag{4.5}
$$

For SDF (4.3) to be consistent with restriction (4.5), we must have

$$
\begin{aligned}
\alpha &= \frac{1}{2} \ln\left(1 + 2\lambda_2 h\right) \\
\beta &= h\left(1 - \frac{1}{2} \frac{\lambda_1^2}{1 + 2\lambda_2 h}\right)
\end{aligned}
$$

Inserting SDF (4.3) into the pricing equation (4.2), we obtain the following expression for the price of a zero-coupon bond maturing at time T (let $(T-t)/h = N$):

$$E_t m_{t,t+h} m_{t+h,t+2h} ... m_{T-h,T} = B_t^T$$
$$B_t^T = \exp(-\alpha N) E_t \exp\left(\sum_{n=0}^{N-1}\left(-\beta r_{t+nh} - \lambda_1 \sqrt{r_{t,t+nh}h}\,\varepsilon_{r,t+(n+1)h} - h\lambda_2 \varepsilon^2_{r,t+(n+1)h}\right)\right)$$

By definition, the yield on this bond is given by

$$y_t^T = -\frac{1}{T-t}\ln B_t^T = \frac{1}{T-t}\left(\alpha N - \ln E_t \exp\left(\sum_{n=0}^{N-1}\left(-\beta r_{t+nh} - \lambda_1 \sqrt{r_{t,t+nh}h}\,\varepsilon_{r,t+(n+1)h} - h\lambda_2 \varepsilon^2_{r,t+(n+1)h}\right)\right)\right)$$

Unfortunately, the learning implications of the model render the final bond expression non-affine in the state variables. The expectation above, however, can be easily computed using Monte Carlo integration.

When constructing the term structure of interest rates we make maturities, T, range from one year to 10 years. The discretisation time step, h, is kept constant at 1/500 of a year. As a base case for our simulations we take the following parameter values. We choose the speed of mean reversion in both the short rate and the central tendency to be $\kappa_r = \kappa_\eta = 2.0$, so that they are consistent with high persistence of the state variables. E.g., for $\kappa_r = 2.0$, the persistence of the non-Gaussian AR(1) short rate process in (4.1) is equal to $1 - \kappa_r h = 0.996$. Both κ_r and κ_η have virtually identical impact on the term structure of zero yields[1]. This influence, however, is strong as we might expect. Intuitively, larger speed of mean reversion pulls the state variables faster to the long run mean, θ. The result is that all yields are larger with the intermediate yields being affected the most, which increases the concavity of the term structure as represented in Figure 1.

The shape of the term structure strongly depends on the relative position of the current short rate with respect to the long run mean of the central tendency, θ.[2] Our model produces rich patterns of the term structure similar to non-discretised CIR models. If the short rate is below the mean, the term structure is upward-sloping, otherwise, it is inverted. For our numerical results we set the long run mean of the central tendency at 0.01 in the base case. The level of θ has a strong effect on both the levels and the curvature of the term structure, with the latter being affected the most by θ than any other parameter of the model (see Figure 2).

Our numerical simulations show that, interestingly, the instantaneous volatilities of both the short rate and the central tendency are largely irrelevant for the shape and level of the term structure. We start with the base case values of the volatilities given by $\sigma_r = \sigma_\eta = 0.01$. As an example, the yields on a 1-year and 10-year zeros in the base case are 2.01% and 3.90%, respectively. If we increase σ_r substantially to, say, 0.1, the corresponding new yields are identical to those obtained with base case parameters. Likewise, if we increase σ_η from 0.01 to 0.1, we do not see any change in any of the yields[3].

The base case risk premiums are $\lambda_1 = -0.02$ and $\lambda_2 = 0.001$. Zero yields are largely insensitive to the value of λ_1. However, the second risk premium, which is the loading on the non-Gaussian component in the SDF, has strong influence on the term structure. This non-Gaussian risk premium affects zero rates of all maturities in the same way leading to parallel shifts in the yield curve. Even though the shape of the term structure is largely not affected, the yields are very sensitive to the level of the second risk premium. E.g., a change in λ_2 from the base case level of 0.001 to 0.05 add about 980 basis points to yields of all maturities as shown in Figure 3.

[1] Due to this finding, we present simulations results only for κ_r.

[2] In our simulations we assume that both the short rate and the central tendency start at 0.01. We also assume that the posterior variance of the central tendency estimate, γ_t, starts at the level of two instantaneous standard deviations of the central tendency, η_t, i.e., $\gamma_t = 2\sigma_\eta\sqrt{\eta_t}$ per year.

[3] Only if we increase these volatilities to unrealistic levels by a factor of 1000, do the yields decline. The decline, however, is minuscule, half a basis point or less.

5. Conclusion

Recently, it has been demonstrated (McLeish [19]) that the diffusion process can be well approximated by the Milstein method rather than the Euler's method. In this paper, we study the recursive estimates for various classes of discretely sampled continuous time stochastic volatility models using the Milstein method. We also provide an example of joint valuation of a zero-coupon bond and learning about an underlying state variable under incomplete information environment.

[1] A. Thavaneswaran, M. Thompson, A criterion for filtering in semimartingale models, Stochastic Processes and their Applications 28 (2) (1988) 259–265.

[2] A. Thavaneswaran, M. E. Thompson, Optimal estimation for semimartingales, Journal of Applied Probability 23 (2) (1986) 409–417.

[3] S. Taylor, Asset Price Dynamics, Volatility, and Prediction, Princeton University Press, ISBN 9780691134796, 2011.

[4] S. L. Heston, S. Nandi, A Closed-Form GARCH Option Valuation Model, The Review of Financial Studies 13 (3) (2000) 585–625.

[5] H. Kawakatsu, Specification and estimation of discrete time quadratic stochastic volatility models, Journal of Empirical Finance 14 (3) (2007) 424–442.

[6] U. V. Naik-Nimbalkar, M. B. Rajarshi, Filtering and Smoothing Via Estimating Functions, Journal of the American Statistical Association 90 (429) (1995) 301–306.

[7] M. E. Thompson, A. Thavaneswaran, Filtering via estimating functions, Applied Mathematics Letters 12 (5) (1999) 61–67.

[8] A. Thavaneswaran, Y. Liang, N. Ravishanker, Inference for Diffusion Processes using Combined Estimating Functions, to appear in the Sri Lankan Journal of Applied Statistics, 2012.

[9] M. Sorensen, Estimating functions for diffusion-type processes, in: M. Kessler, A. Lindner, M. Sorensen (Eds.), Statistical Methods for Stochastic Differential Equations, chap. 1, Taylor & Francis, 1–108, 2012.

[10] M. Jeong, J. Y. Park, Asymptotic Theory of Maximum Likelihood Estimator for Diffusion Model, working Paper, Indiana University, 2010.

[11] P. E. Kloeden, E. Platen, Numerical solution of stochastic differential equations / Peter E. Kloeden, Eckhard Platen, Springer-Verlag, 1992.

[12] T. Koulis, A. Thavaneswaran, Estimating Functions for Diffusion Processes using Milstein's Approximation, working Paper, 2012.

[13] D. Kennedy, Stochastic Financial Models, Chapman & Hall/CRC financial mathematics series, Chapman & Hall/CRC, 2010.

[14] Y. Ait-Sahalia, Testing Continuous-Time Models Of The Spot Interest Rate, Review of Financial Studies 9 (2) (1996) 385–426.

[15] F. Klebaner, Introduction to Stochastic Calculus With Applications, Imperial College Press, 2005.

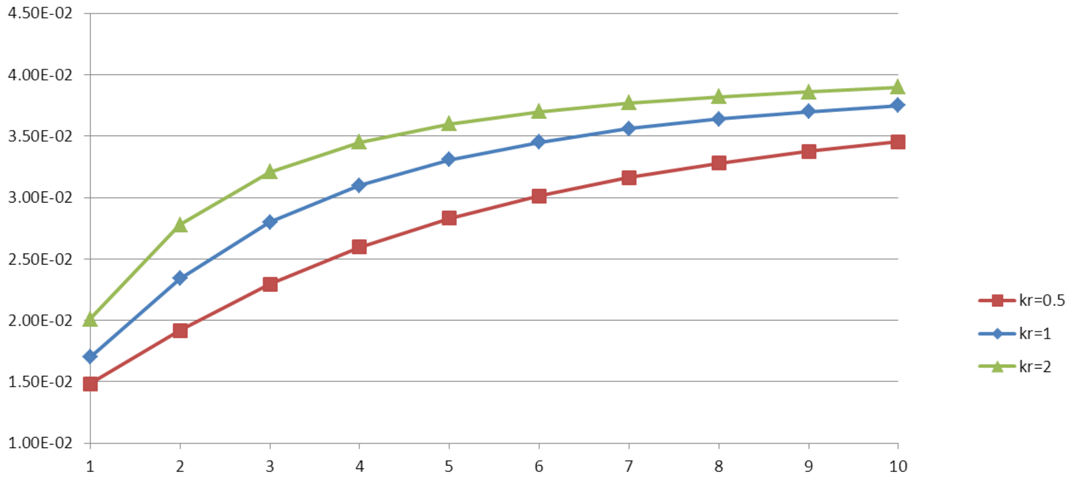
[16] F. Black, M. S. Scholes, The Pricing of Options and Corporate Liabilities, Journal of Political Economy 81 (3) (1973) 637–54.

[17] J. C. Hull, A. D. White, The Pricing of Options on Assets with Stochastic Volatilities, Journal of Finance 42 (2) (1987) 281–300.

[18] P. Balduzzi, S. R. Das, S. Foresi, The Central Tendency: A Second Factor In Bond Yields, The Review of Economics and Statistics 80 (1) (1998) 62–72, URL `http://ideas.repec.org/a/tpr/restat/v80y1998i1p62-72.html`.

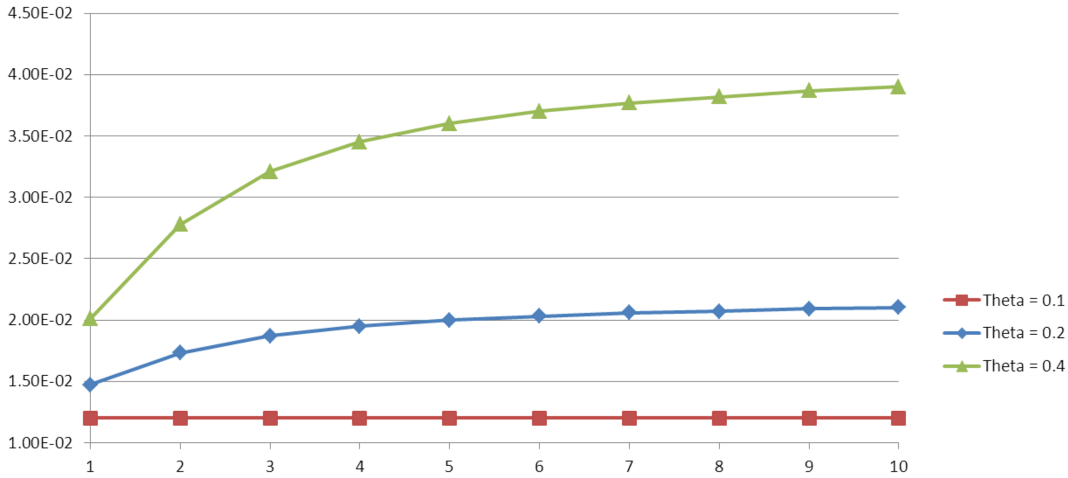
[19] D. McLeish, Monte Carlo Simulation and Finance, Wiley Finance, Wiley, 2005.

Figure 1. Term Structure as a function of the speed of mean reverison in the short rate



We use base parameters presented in the text to generate the term strucutre of zero rates. The underlying model is the discretized version of a continuous-time Cox-Ingersoll-Ross (CIR) model with the central tendency of the short rate also following a CIR process. We use Milstein discretisation scheme.
The curves represent the mean yields over 10000 Monte Carlo iterations. The time step in the Milstein scheme is 1/500 of a year. The speed of mean reversion parameter, κ_r, ranges from 0.5 to 2.
In our simulations we assume that both the short rate and the central tendency start at 0.01. We also assume that the posterior variance of the central tendency estimate, γ_t, starts at the level of two instantaneous standard deviations of the central tendency, $\gamma_t = 2\sigma_\eta\sqrt{\eta_t}$ per year.

Figure 2. Term Structure as a function of the central tendency of the short rate

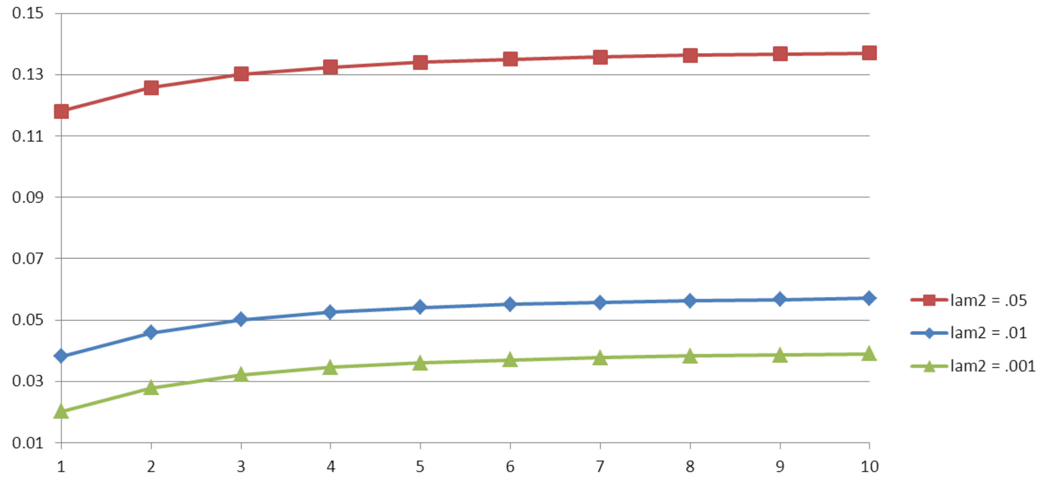


We use base parameters presented in the text to generate the term strucutre of zero rates. The underlying model is the discretized version of a continuous-time Cox-Ingersoll-Ross (CIR) model with the central tendency of the short rate also following a CIR process. We use Milstein discretisation scheme.

The curves represent the mean yields over 10000 Monte Carlo iterations. The time step in the Milstein scheme is 1/500 of a year. The long run mean of the central tendency, θ, ranges from 0.1 to 0.4.

In our simulations we assume that both the short rate and the central tendency start at 0.01. We also assume that the posterior variance of the central tendency estimate, γ_t, starts at the level of two instantaneous standard deviations of the central tendency, $\gamma_t = 2\sigma_\eta\sqrt{\eta_t}$ per year.

Figure 3. Term Structure as a function of the second rsik premium, λ_2



We use base parameters presented in the text to generate the term strucutre of zero rates. The underlying model is the discretized version of a continuous-time Cox-Ingersoll-Ross (CIR) model with the central tendency of the short rate also following a CIR process. We use Milstein discretisation scheme.
The curves represent the mean yields over 10000 Monte Carlo iterations. The time step in the Milstein scheme is 1/500 of a year. The non-Gaussian risk premium, λ_2, ranges from 0.5 to 2.
In our simulations we assume that both the short rate and the central tendency start at 0.01. We also assume that the posterior variance of the central tendency estimate, γ_t, starts at the level of two instantaneous