Estimation of the volatility component in two-factor stochastic volatility short rate models

Dmitri Danilov   and   Pranab K. Mandal

KUB, Tilburg   EURANDOM, Eindhoven

d.l.danilov@kub.nl   mandal@eurandom.tue.nl

Introduction

Continuous time models and those involving stochastic differential equations, in particular, became very popular in modern financial theory. There are several reasons for that. The continuous time setup, introduced in finance by the celebrated work of Black and Scholes (1973), is attractive from a theoretical point of view. It provides a plain and parsimonious way of representing models. A variety of techniques for pricing and hedging of derivative securities were developed in the continuous time setup, especially when underlying economic factors are described by stochastic differential equations (see Merton (1992)). Typically, the idealised continuous time setup is much simpler than discrete time considerations and therefore the derivative pricing is much simpler in continuous time, where also often analytical formulas are available.

A continuous time model for the interest rate was first proposed by Merton (1973), who introduced a Brownian motion as a candidate process. Time series of historical interest rates reveal a number of salient features such as high degree of persistence, nonnegativity and volatility clustering. It is a common practice to model these time series by stationary processes notwithstanding the fact that formal tests sometimes suggest the unit root behaviour. Merton's model, of course, captures the persistence aspect of the interest rates time series. However, it allows for negative interest rates and generates nonstationary series. Vasicek (1977) proposed to use a stochastic differential equations (SDE), namely, the Ornstein-Uhlenbeck process

$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t. \tag{1}$$

Let us highlight some of the appealing features of this model. First of all, it is clear that (1) allows for a stationary mean reverting solution. Secondly, the parameters of the model have a clear economic interpretation: μ is the "average" interest rate, κ is the persistence parameter, (small values of κ corresponds to high degree of persistence), and σ is the volatility of the

process. Note that $\kappa = 0$ corresponds to the random walk case. However, model (1) does not allow the volatility to be variable, and negative interest rates may still show up. This last drawback was corrected in the Cox, Ingersoll and Ross (CIR) (1985) model

$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t. \tag{2}$$

Their model provides not only a stationary mean reverting process, but also it does not allow the interest rate to be negative due to the so called "level" effect. Interpretation of the parameters is the same as in the case of Vasicek. The slightly more general specification

$$dr_t = \kappa(\mu - r_t)dt + \sigma r_t^\gamma dW_t, \tag{3}$$

was employed in the work of Chan, Karolyi, Longstaff, and Sanders (CKLS) (1992). The parameter γ control the "strength" of level effect and also accounts for the degree of conditional heteroscedasticity. The value $\gamma = 0$ corresponds to the homoscedastic case (Vasicek model), where the level effect is absent. The case studies in CKLS shows the estimated value of γ to be about $3/2$, in contrast to the values $1/2$ in the CIR model. The works of Longstaff and Schwartz (1992), Koedijk et al. (1994), suggest a direction for extending the CKLS model, namely, the inclusion of stochastic volatility factors. It was repeatedly mentioned in modern literature (see, e.g., Rebonatto (1996)) that one factor models fails to capture adequately the price structure of different derivative securities like yields, caps and swaptions. One of the first approaches in this direction was issued in Fong and Vasicek (FV) (1991). They proposed a model of the following form

$$\begin{cases} dr_t &= \kappa(\mu - r_t)dt &+ &\sqrt{v_t}dW_t^{(1)}, \\ dv_t &= \lambda(\nu - v_t)dt &+ &\tau\sqrt{v_t}dW_t^{(2)}. \end{cases} \tag{4}$$

As we can see, this model allows for a stationary mean reverting process whose volatility is again stationary stochastic process. Here μ is still the unconditional average of the short rate process, κ controls the degree of persistence in interest rates. In order to interpret the other parameters let us observe that the second equation in (4) is just a square root process for volatility v_t. Now we can interpret parameter ν as the unconditional average volatility. The parameter λ accounts for the degree of persistence in the volatility. Finally, the parameter τ is the unconditional infinitesimal variance of the unobserved volatility process.

As in any type of modelling, to apply the model to real life data one needs to estimate the parameters of the model. In stochastic volatility models one

further needs an efficient and reliable method for estimation of unobservable volatility component. Estimation of stochastic volatility is important in several aspects. If we know factor values in any point in time we can calculate implied term structure and therefore evaluate the adequacy of the model. Knowledge of the current value of volatility allows us to draw important economical implications: perform volatility forecast, calculate implied values of different kind of derivative securities like bond options and swaps, etc. That, in turn, can affect management decisions in many fields of economics and finance.

A number of sophisticated methods are available in order to estimate the parameters of continuous time models (e.g. GMM, EMM of Gallant and Tauchen (1996), Indirect inference method of Gourieroux, Monfort and Renault (1993) etc.). However, none of these methods provide opportunity for estimation of unobservable stochastic volatility process in a model like (4).

To emphasize the importance of estimation of stochastic volatility, suppose the short rate follows model (4). Then (see Fong and Vasicek (1991)) yields on T-maturing bonds are determined by formula

$$Y(t,T) = A(t,T) - B(t,T)r_t - C(t,T)v_t, \tag{5}$$

where functions A, B, C depend on the parameters of the short rate model and the market price of risk. As we can see, pricing formula (5) depends on v_t. Therefore, even if the model is adequate for Data Generating Process and the parameters are known, the performance of the model can be poor unless we provide a good estimator for v_t, volatility at time point t when we need to find yield or price of some other derivative security. It is not clear, however, what kind of market information should one use for estimation of v_t. This information can include only short rate time series data or yields with different maturities or even sets of option prices. In this article we work with the short rate dynamics only. The methodologies discussed in this article can be applied to many two factor stochastic volatility short rate models e.g., Fong and Vasicek (1991), Andersen and Lund (1997), etc. We have chosen to work with FV model because of its simplicity. For other models the notations will be complicated only, but would not provide any extra insight for the proposed methodology.

Note, from equation (4), that the quadratic variation of r_t is given by $<r>_t = \int_0^t v_t dt$. Therefore, if the original short rate process can be observed on any frequency then recovering v_t is trivial from $<r>_t$. Usually the best that we have is a daily series and, therefore, some indirect scheme for obtaining v_t is necessary.

The use of stochastic filtering theory is very natural here, because we want to estimate the *unobserved* volatility component from the *observed* short rates. The equation (4) as it is now, however, is not ready to receive the filtering treatment. We first discretize both the observation and the state equations to bring it in the filtering theory framework. As we shall see, the transformed equation would be nonlinear and also with non-Gaussian errors. As a naive approach we apply extended Kalman filter (see Anderson and Moore (1979)), as if the errors were Gaussian. It happens, however, that the method of extended Kalman filter (EKF) does not provide very good estimation for typical financial short rate data. We suggest a method based on Kitagawa (1987) scheme which incorporates both nonlinearity and non-Gaussianity. We also use the method of conditional moments (MCM) to estimate volatility for comparison.

The article is organized as follows. In section 1, we carry out the discretization of the FV model. In section 2, the methodologies of EKF, Kitagawa and MCM are described. A comparison of these three methods of volatility estimation on simulated data is presented in section 3. Section 4 contains the empirical analysis. Some conclusions are offered in section 5.

1 Discretization of Fong Vasicek short rate model

Recall that the short rate equation of the Fong and Vasicek model is given by (4). An application of Ito formula to the first equation yields

$$de^{\kappa t}(r_t - \mu) = e^{\kappa t}\sqrt{v_t}dW_t.$$

Integrating by parts we obtain

$$r_{t+h} = \mu + e^{-\kappa h}(r_t - \mu) + e^{-\kappa h}\int_t^{t+h} e^{\kappa(s-t)}\sqrt{v_s}dW_s.$$

Also, similarly,

$$v_{t+h} = \nu + e^{-\lambda h}(v_t - \nu) + e^{-\lambda h}\tau\int_t^{t+h} e^{\lambda(s-t)}\sqrt{v_s}dZ_s.$$

Therefore the discrete time specification of FV model has the following form,

$$\begin{aligned}
r_{t+h} &= \mu + e^{-\kappa h}(r_t - \mu) + \varepsilon_t(h), & (6)\\
v_{t+h} &= \nu + e^{-\lambda h}(v_t - \nu) + \eta_t(h), & t = 0, h, 2h, \ldots,
\end{aligned}$$

4

where h denotes the sampling interval (for example, on weekly frequency $h = 1/52$), and the innovations $\varepsilon_t(h)$ and $\eta_t(h)$ are defined as

$$\varepsilon_t(h) = e^{-\kappa h} \int_t^{t+h} e^{\kappa(s-t)} \sqrt{v_s} dW_s, \tag{7}$$

$$\eta_t(h) = \tau e^{-\lambda h} \int_t^{t+h} e^{\lambda(s-t)} \sqrt{v_s} dZ_s.$$

We approximate these innovations as

$$\varepsilon_{nh}(h) = e^{-\kappa h} \sqrt{v_{nh}} \sqrt{h} \, \varepsilon_n, \quad \eta_{nh}(h) = e^{-\lambda h} \tau \sqrt{v_{nh}} \sqrt{h} \, \eta_n,$$

where (ε_n) and (η_n) are independent standard normal random variates. Defining the transformed discrete observation to be

$$R_n = e^{\kappa h}(r_{(n+1)h} - \mu) - (r_{nh} - \mu), \quad n = 0, 1, 2, \ldots, \tag{8}$$

and denoting v_{nh} by V_n, we obtain the following discrete time state space system

$$
\begin{aligned}
R_n &= \sqrt{h} \sqrt{V_n} \, \varepsilon_n, \quad n = 0, 1, 2, \ldots, & (9)\\
V_n &= e^{-\lambda h} V_{n-1} + (1 - e^{-\lambda h})\nu + e^{-\lambda h} \tau \sqrt{h} \sqrt{V_{n-1}} \, \eta_n, \quad n = 1, 2, \ldots & (10)
\end{aligned}
$$

with initial value V_0 independent of (ε_n) and (η_n).

2 Methods of estimating stochastic volatility

2.1 Extended Kalman Filter

Standard setup of the Kalman filter is applicable to the linear state space model of the form

$$
\begin{aligned}
y_n &= Z_n\alpha_n + d_n + \varepsilon_n, & \mathsf{Var}(\varepsilon_n) &= H_n,\\
\alpha_n &= T_n\alpha_{n-1} + c_n + R_n\eta_n, & \mathsf{Var}(\eta_n) &= Q_n,
\end{aligned} \tag{11}
$$

where (ε_n) and (η_n) are independent normal random variables with zero mean. Then the conditional distribution of α_n given the observations $y_1, \ldots, y_n$ is also normal. The mean a_n and variance P_n can be calculated recursively by an application of the *one step ahead prediction equations*,

$$
\begin{aligned}
a_{n|n-1} &= T_n a_{n-1} + c_n,\\
P_{n|n-1} &= T_n P_{n-1} T_n' + R_n Q_n R_n',
\end{aligned}
$$

and *updating/filtering equations,*

$$a_n = a_{n|n-1} + P_{n|n-1}Z'_n F_n^{-1}(y_n - Z_n a_{n|n-1} - d_n),$$
$$P_n = P_{n|n-1} - P_{n|n-1}Z'_n F_n^{-1} Z_n P_{n|n-1},$$
$$F_n = Z_n P_{n|n-1}Z'_n + H_n.$$

Here $a_{n|n-1}$ and $P_{n|n-1}$ denote the conditional expectation and variance, respectively, of α_n given the observations $y_1, \ldots, y_{n-1}$.

When the state space equation is non-linear, say

$$y_n = Z_n(\alpha_n) + \varepsilon_n, \qquad \mathsf{Var}(\varepsilon_n) = H_n,$$
$$\alpha_n = T_n(\alpha_{n-1}) + R_n(\alpha_{n-1})\eta_n, \quad \mathsf{Var}(\eta_n) = Q_n, \tag{12}$$

one can use Taylor series expansion to obtain the following approximate linearised system.

$$y_n = \hat{Z}_n \alpha_n + d_n + \varepsilon_n, \qquad \mathsf{Var}(\varepsilon_n) = H_n, \tag{13}$$
$$\alpha_n = \hat{T}_n \alpha_{n-1} + c_n + \hat{R}_n \eta_n, \quad \mathsf{Var}(\eta_n) = Q_n, \tag{14}$$

where $\hat{Z}_n = \frac{d}{dx}Z_n(a_{n|n-1})$, $d_n = Z_n(a_{n|n-1}) - \hat{Z}_n a_{n|n-1}$, $\hat{T}_n = \frac{d}{dx}T_n(a_{n-1})$, $c_n = T_n(a_{n-1}) - \hat{T}_n a_{n-1}$, $\hat{R}_n = R_n(a_{n-1})$.

The Kalman filter for this approximate state-space model is then given by :

$$a_{n|n-1} = T_n(a_{n-1}),$$
$$P_{n|n-1} = \hat{T}_n P_{n-1}\hat{T}'_n + \hat{R}_n Q_n \hat{R}'_n,$$
$$F_n = \hat{Z}_n P_{n|n-1}\hat{Z}'_n + H_n,$$
$$a_n = a_{n|n-1} + P_{n|n-1}\hat{Z}'_n F_n^{-1}(y_n - Z_n(a_{n|n-1})),$$
$$P_n = P_{n|n-1} - P_{n|n-1}\hat{Z}'_n F_n^{-1}\hat{Z}_n P_{n|n-1}.$$

Smoothed estimate $a_{n|N}$ of α_n given the observations $y_1, \ldots, y_N$ is obtained by the following backward recursion :

$$a_{N|N} = a_N$$
$$a_{n-1|N} = a_{n-1} + P_{n-1}\hat{T}'_{n-1}P_{n|n-1}^{-1}(a_{n|N} - a_{n|n-1}).$$

In our setup we consider the observation y_n to be $\ln(R_n^2/h)$. From (9) we then have

$$y_n = \ln V_n + \ln \varepsilon_n^2.$$

Clearly $\ln \varepsilon_n^2$ is not Gaussian, but has the distribution of $\ln \chi_1^2$. To use EKF we replace this by a normal random variable with mean -1.270363 and variance 4.934802, the mean and variance, respectively, of a $\ln \chi_1^2$ random variable. We then apply the EKF methodology with

$$
\begin{aligned}
Z_n(x) &= \ln x - 1.270363; \quad H_n = 4.934802; \\
T_n(x) &= e^{-\lambda h} x + (1 - e^{-\lambda h}) \nu; \quad R_n(x) = \tau e^{-\lambda h} \sqrt{h} \sqrt{x}; \quad Q_n = 1.
\end{aligned}
$$

To initiate the recursion we use $V_0 = \nu$ and $P_0 = 1000$.

2.2 Kitagawa Algorithm

Extended Kalman filter method linearizes the non-linear part using Taylor series expansion. The methodology, however, depends on the Gaussian property of the error terms. When the errors are not Gaussian, which is the case of ours, Kitagawa (1987) method is more appropriate. In his paper Kitagawa treats explicitly the linear case. We present below the results for the non-linear models. The formulae are the same.

Suppose the state-space model is given by

$$
\begin{aligned}
y_n &= h(x_n, \varepsilon_n) \\
x_n &= f(x_{n-1}) + g(x_{n-1}) \eta_n
\end{aligned}
$$

where $\{\varepsilon_n\}$ and $\{\eta_n\}$ are independent white noise sequence, not necessarily Gaussian. Exploiting the Markovian property of $\{x_n\}$ and denoting the observations $(y_1, y_2, \ldots, y_n)$ by Y_n, one has the following recursive filtering scheme.

One-step-ahead prediction :

$$
f_{n|n-1}(x_n|Y_{n-1}) = \int_\infty^\infty p_{n|n-1}(x_n|x_{n-1}) f_{n-1}(x_{n-1}|Y_{n-1}) dx_{n-1}.
$$

Filtering :

$$
f_n(x_n|Y_n) = \frac{p_{y|x}(y_n|x_n) f_{n|n-1}(x_n|Y_{n-1})}{p(y_n|Y_{n-1})}.
$$

Smoothing :

$$
f_{n|N}(x_n|Y_N) = f_n(x_n|Y_n) \int_\infty^\infty \frac{f_{n+1|N}(x_{n+1}|Y_N) p_{n|n-1}(x_{n+1}|x_n)}{f_{n+1|n}(x_{n+1}|Y_n)} dx_{n+1}.
$$

7

Kitagawa method approximates all the densities by piecewise linear functions. Each density is specified by the number of segments, location of nodes and the value at each node. It is assumed that all the densities are supported on finite interval[1]. In the simplest case the nodes for all the densities are assumed same, $z_0, z_1, \ldots, z_L$, say. Then the integration in the *one-step-ahead prediction* equation is evaluated as follows.

$$
\int_\infty^\infty p_{n|n-1}(x_n|x_{n-1})f_{n-1}(x_{n-1}|Y_{n-1})dx_{n-1}
$$
$$
= \int_{z_0}^{z_L} p_{n|n-1}(x_n|x_{n-1})f_{n-1}(x_{n-1}|Y_{n-1})dx_{n-1}
$$
$$
= \sum_{i=1}^{L} \int_{z_{i-1}}^{z_i} p_{n|n-1}(x_n|x_{n-1})f_{n-1}(x_{n-1}|Y_{n-1})dx_{n-1},
$$

where using the linearity of the functions in the interval (z_{i-1}, z_i),

$$
\int_{z_{i-1}}^{z_i} p_{n|n-1}(x_n|x_{n-1})f_{n-1}(x_{n-1}|Y_{n-1})dx_{n-1}
$$
$$
\approx \left(p_{n|n-1}(x_n|z_{i-1})f_{n-1}(z_{i-1}|Y_{n-1}) + p_{n|n-1}(x_n|z_i)f_{n-1}(z_i|Y_{n-1}) \right)
$$
$$
\times \frac{(z_i - z_{i-1})}{2}.
$$

In the *filtering* equation $p(y_n|Y_{n-1})$ is evaluated as $\int_\infty^\infty p_{y|x}(y_n|x_n)f_{n|n-1}(x_n|Y_{n-1})$ and the integration is calculated as above. The integration in the *smoothing* equation is also evaluated similarly.

In our setup all the conditional distributions are Gaussian with proper mean and variance. To start the recursion we use the steady state density of v_t, a square root process, as the initial density of V_0. As for choosing the nodes for discretizing the density one should note that increasing the number of nodes will only increase the performance of the methodology. In practice one can keep on incorporating more and more nodes until the change in estimates is negligible.

2.3 Method of Conditional Moments

Recall, from equation (6) and (7), that

$$
r_{t+h} - \mu - e^{-\kappa h}(r_t - \mu) = \varepsilon_t(h) = e^{-\kappa h} \int_t^{t+h} e^{\kappa(s-t)}\sqrt{v_s}\, dW_s.
$$

---

[1]In case of infinite support, the end points of the grid are to be chosen in such a way that they cover the essential domain of the density

Hence, $E(\varepsilon_t(h)|r_t, v_t) = 0$, and

$$Var(\varepsilon_t(h)|r_t, v_t) = \int_t^{t+h} e^{2\kappa(u-t-h)} v_u du. \qquad (15)$$

Approximating the integral in the r.h.s. of (15) as $e^{-2\kappa h} v_t h$, one obtains a natural estimator, v_{ih}^*, for v_t at $t = ih$, given by

$$v_{ih}^* e^{-2\kappa h} h = \frac{1}{(2k+1)} \sum_{j=i-k+1}^{k+i} \varepsilon_{jh}^2(h),$$

that is,

$$v_{ih}^* = \frac{\sum_{j=i-k+1}^{k+i} R_j^2}{(2k+1)h}, \qquad (16)$$

where R_j's are as defined in (8). The estimator (16) is in fact an estimator of v_t by the method of conditional moments.

As we can see, the estimator (16) depends on the choice of the window size k. In our analysis to decide about the window size we have compared performances of MCM for different values of k on simulated data. The criteria of the goodness of fit used is an analog of R^2 statistic

$$R^2(k) = 1 - \frac{\sum\limits_{i=1}^{n} (V_i - V_i^*)^2}{\sum\limits_{i=1}^{n} V_i^2}. \qquad (17)$$

Based on this we have chosen $k = 10, 20$, and 50 for monthly, weekly and daily data, respectively.

3 Comparison on simulated data

We have simulated several short rate time series according to the FV model for different sets of parameter values close to the typical values. We have considered three different values for any parameter θ :

$$\theta_2 \equiv \hat{\theta} - 1.5 \times se(\hat{\theta}), \qquad \theta_3 \equiv \hat{\theta}, \qquad and \qquad \theta_4 \equiv \hat{\theta} + 1.5 \times se(\hat{\theta}),$$

where $\hat{\theta}$ is the estimate of θ obtained by applying EMM method to the real data and $se(\hat{\theta})$ is the standard error. These values are reported in section 4.2.

For each set of parameters we have generated 25 time series of length 4000 on daily frequency and of length 2000 on weekly and monthly frequencies. In all of these cases we have found that Kitagawa smoothing method

Table 1: Performances of the methods for different frequencies

$Frequency$	m_{kits}	σ_{kits}	m_{mcm}	σ_{mcm}	m_{eks}	σ_{eks}
$Monthly$	0.1628	0.0149	0.2197	0.0214	0.2168	0.0172
$Weekly$	0.0857	0.0134	0.1117	0.0208	0.1309	0.0237
$Daily$	0.0417	0.0082	0.0548	0.0110	0.0687	0.0147

m_{kits} and σ_{kits} are the average and standard deviation, respectively, of $(1 - R^2)$-values obtained by the Kitagawa smoothing method applied to 25 series simulated from FV model. The length of the series are 4000 for daily data, and 2000 for weekly or monthly data. m_{mcm}, σ_{mcm}, m_{eks}, σ_{eks} are the corresponding quantities for MCM and the extended Kalman smoothing method. Data was simulated using parameter values : $\mu = 0.0652$, $\kappa = 0.109$, $\nu = 0.000264$, $\lambda = 1.482$, and $\tau = 0.01934$.

outperforms the other methods. Here, again, we have used R^2-like quantity, given by (17), to measure goodness of fit.

To select the node points for Kitagawa method we started with a set of nodes and then if any estimate of volatility is too close to the right limit of the nodes, we increased the right limit. As for density of the nodes we compare the volatility estimates for the current set of nodes and the estimates corresponding to the nodes which has density two times the current density. If the proportional change of estimate is less than 0.1% we stop. Otherwise, we keep on doubling the number of nodes. In most of the cases we have found that the number of nodes needed are between 100 to 400.

For MCM, as mentioned in section 2.3, we have used $k = 10, 20$, and 50 for monthly, weekly and daily data, respectively.

Figure 1 on page 11 plots the $(1 - R^2)$-values obtained by applying Kitagawa smoothing, MCM, and extended Kalman Smoothing method on simulated daily, weekly and monthly data. Table 1 on this page reports the corresponding summary statistics – the average and the standard deviation of the $(1 - R^2)$-values. We see that as the frequency of data increases performances of all the methods become better with the Kitagawa smoothing method being the best in all frequencies. This can also be seen from Figure 1 on page 11.

Furthermore, we have noticed that when ν, λ and τ are fixed the goodness of fit for a method is similar for different sets of values of μ and κ. Table 2 on page 12 shows this feature when $\nu = \nu_3 = 0.000264$, $\lambda = \lambda_3 = 1.482$ and $\tau = \tau_3 = 0.01934$. Therefore, to compare the performances of these methods for different values of parameters we fix $\mu = \mu_3 = 0.0652$ and $\kappa = \kappa_3 = 0.109$

Figure 1: Performances of different methods on simulated data
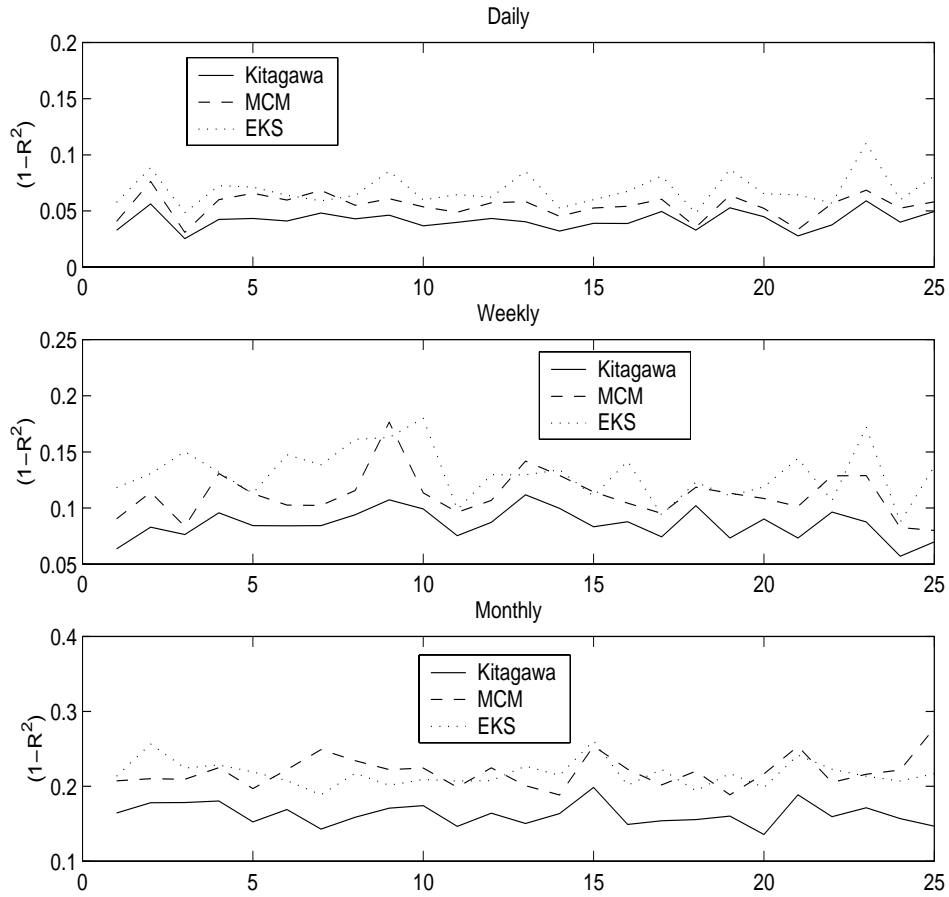
All series were simulated using parameter values : $\mu = 0.0652$, $\kappa = 0.109$, $\nu = 0.000264$, $\lambda = 1.482$, and $\tau = 0.01934$. Daily series were of length 4000 and weekly and monthly series were of length 2000.

Table 2: Performances of the methods for different μ and κ

i_μ	i_κ	m_{kits}	σ_{kits}	m_{mcm}	σ_{mcm}	m_{eks}	σ_{eks}
2	2	0.0894	0.0136	0.1166	0.0222	0.1378	0.0222
2	3	0.0884	0.0124	0.1130	0.0209	0.1346	0.0217
2	4	0.0866	0.0113	0.1101	0.0167	0.1307	0.0220
3	2	0.0856	0.0122	0.1087	0.0123	0.1365	0.0240
3	3	0.0898	0.0117	0.1158	0.0168	0.1362	0.0158
3	4	0.0838	0.0130	0.1079	0.0162	0.1337	0.0123
4	2	0.0863	0.0128	0.1099	0.0138	0.1314	0.0174
4	3	0.0912	0.0135	0.1163	0.0181	0.1365	0.0245
4	4	0.0871	0.0103	0.1107	0.0143	0.1343	0.0194

m_{kits}, σ_{kits}, m_{mcm}, σ_{mcm}, m_{eks}, σ_{eks} are as described in Table 1 but based on weekly series of length 2000. To simulate data for a row a parameter θ is set to θ_{i_θ}, where $\mu_2 = 0.0599, \mu_3 = 0.0652, \mu_4 = 0.0705$, and $\kappa_2 = 0.0577, \kappa_3 = 0.109, \kappa_4 = 0.1603$. For all entries $\nu = 0.000264, \lambda = 1.482$, and $\tau = 0.01934$.

and vary ν, λ and τ. Table 3 on page 13 presents the summary results. We see that in all cases the average $(1 - R^2)$-value for Kitagawa smoothing method is "significantly" lower than the other two methods. Another point to note is that as τ, the variance in volatility component, increases performances of all the methods decrease.

4   Empirical results

In this section we present the analysis of empirical data. Before presenting the results we describe the data and the parameter estimation of the model.

4.1   Data Description

For numerical experiments with the real data we select the yields on US Treasury Bills with maturity 3 months[2]. This maturity is short enough to believe that these yields will approximate the (unobservable) short rate sufficiently well. It is known (see e.g. Andersen and Lund (1997)) that successful estimation of multifactor stochastic volatility models require high

---

[2]Data source: H.15 Federal Reserve Statistical Release. See the web site of the Board of Governors of the Federal Reserve System *http://www.federalreserve.gov/*

Table 3: Performances of the methods for different $\nu, \lambda,$ and τ

i_ν	i_λ	i_τ	m_{kits}	σ_{kits}	m_{mcm}	σ_{mcm}	m_{eks}	σ_{eks}
2	2	2	0.0620	0.0240	0.0782	0.0286	0.1162	0.0603
2	2	3	0.0695	0.0289	0.0991	0.0526	0.1303	0.0549
2	2	4	0.0892	0.0392	0.1339	0.0670	0.1680	0.0664
2	3	2	0.0702	0.0085	0.0877	0.0128	0.1035	0.0157
2	3	3	0.0908	0.0104	0.1158	0.0146	0.1365	0.0156
2	3	4	0.1033	0.0180	0.1460	0.0275	0.1735	0.0321
2	4	2	0.0616	0.0057	0.0848	0.0108	0.0841	0.0115
2	4	3	0.1019	0.0115	0.1303	0.0141	0.1367	0.0166
2	4	4	0.1230	0.0146	0.1704	0.0220	0.1780	0.0168
3	2	2	0.0612	0.0245	0.0813	0.0258	0.1109	0.0504
3	2	3	0.0786	0.0340	0.1087	0.0508	0.1275	0.0619
3	2	4	0.1020	0.0429	0.1493	0.0780	0.1690	0.0663
3	3	2	0.0632	0.0064	0.0819	0.0108	0.0921	0.0114
3	3	3	0.0867	0.0117	0.1098	0.0138	0.1351	0.0212
3	3	4	0.1071	0.0163	0.1429	0.0261	0.1703	0.0252
3	4	2	0.0585	0.0059	0.0826	0.0089	0.0733	0.0099
3	4	3	0.0909	0.0092	0.1186	0.0126	0.1251	0.0097
3	4	4	0.1125	0.0121	0.1531	0.0163	0.1621	0.0144
4	2	2	0.0488	0.0139	0.0701	0.0153	0.0842	0.0300
4	2	3	0.0686	0.0262	0.0952	0.0480	0.1158	0.0507
4	2	4	0.0887	0.0322	0.1355	0.0705	0.1678	0.0832
4	3	2	0.0597	0.0099	0.0793	0.0133	0.0825	0.0123
4	3	3	0.0826	0.0144	0.1023	0.0156	0.1273	0.0214
4	3	4	0.0944	0.0143	0.1283	0.0195	0.1557	0.0291
4	4	2	0.0526	0.0060	0.0795	0.0110	0.0675	0.0063
4	4	3	0.0855	0.0088	0.1081	0.0144	0.1144	0.0118
4	4	4	0.1066	0.0116	0.1419	0.0176	0.1515	0.0131

m_{kits}, σ_{kits}, m_{mcm}, σ_{mcm}, m_{eks}, σ_{eks} are as in Table 1 based on weekly series of length 2000. To simulate data for a row a parameter θ is set to θ_{i_θ}, where $\nu_2 = 0.000221, \nu_3 = 0.000264, \nu_4 = 0.000307$; $\lambda_2 = 0.151, \lambda_3 = 1.482, \lambda_4 = 2.813$; and $\tau_2 = 0.01266, \tau_3 = 0.01934, \tau_4 = 0.02602$. For all entries $\mu = 0.0652$ and $\kappa = 0.109$.

Figure 2: U.S. 3-month T-Bill yield data (weekly)
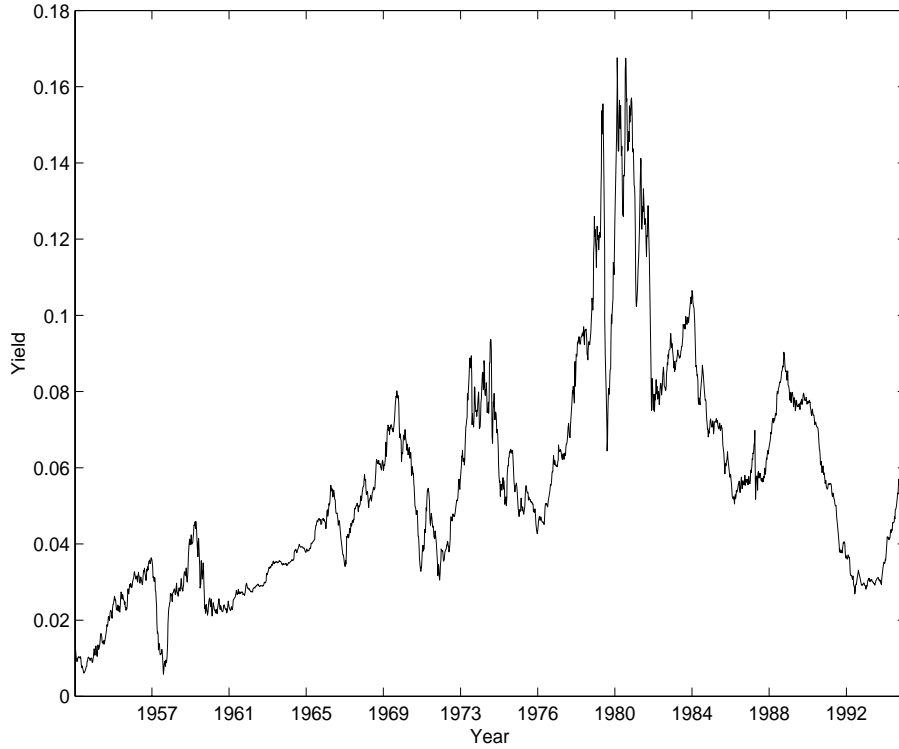
frequency data. At the same time, in order to get stable and precise estimation, we need data over long period of time. We know (see Sundaresan (1997), p. 79) that US Treasury Bills are issued each week, therefore we suggest that weekly frequency is most adequate for the short rate modelling. In our analysis we have used a dataset of 2155 weekly observations dated from January 1954 to April 1995. Figure 2 shows a plot of the data.

4.2 Choice of Parameters

Since there are many very good methods (e.g. GMM, EMM) to estimate parameters of a continuous time model, one can take advantage of those methods to estimate the parameter values. Actually, as in the case of Kalman filter, Kitagawa method also has the advantage of being able to evaluate

14

the likelihood function while performing the algorithm. However, it is to be noted that the likelihood obtained this way would only be an approximate one. Therefore, to use this for maximum likelihood estimation of parameters some special care needs to be taken to avoid numerical instability. We shall present this elsewhere once it becomes complete.

For the actual data set we have used efficient method of moments (EMM) to estimate the parameters. Below we describe the method very briefly.

4.3 Description of the EMM method

EMM is developed in a series of work by Gallant and Tauchen (1996,1997). EMM combines both efficiency and flexibility, i.e., being able to fit a sufficiently wide class of models in a routine way. By construction EMM is a Generalised Method of Moments with a specific choice of moment conditions and an estimated optimal weight matrix. The method requires an *auxiliary model* that embeds the *structural model* under consideration in a certain metric (see Tauchen (1996), Gallant and Tauchen (1997)).

EMM involves the following steps:

1. Choose an *auxiliary model* and get a maximum likelihood (ML) estimator $\tilde{\theta}_n$ of the parameters of this model.

2. Generate the 'efficient' moment conditions as:

$$m(\rho, \tilde{\theta}_n) = \int \frac{\partial \ln f(y|\tilde{\theta}_n)}{\partial \theta} p(y|\rho) dy. \qquad (18)$$

   where $f(y|\theta)$ denotes sample density according to the auxiliary model, $p(y|\rho)$ is the sample density with respect to structural model, and $\tilde{\theta}_n$ is the ML estimator of the parameters in the auxiliary model.

   Remark: In practice the right-hand side in (18) is estimated by Monte-Carlo techniques. That means that integration in (18) is replaced by averaging

$$m(\rho, \tilde{\theta}_n) = \frac{1}{N} \sum_{k=1}^{N} \frac{\partial \ln f((y_k)|\tilde{\theta}_n)}{\partial \theta} \qquad (19)$$

   by the simulated trajectory of the structural model. To simulate this trajectory the Euler approximating scheme with moderate number of intermediate steps was applied.

3. Build the chi-square estimator for ρ as:

$$\hat{\rho}_n = \underset{\rho \in \mathbb{R}^k}{argmin} \ m(\rho, \tilde{\theta}_n)' I_n^{-1} m(\rho, \tilde{\theta}_n), \qquad (20)$$

15

where I_n is some consistent estimator of $I(\theta)$, the information matrix in the auxiliary model.

In our case the *structural model* was FV model given by (4). The main requirement of the *auxiliary model* is that it should be large enough, i.e., it should "almost" nest the structural model in some sense. At the same time the auxiliary model should capture the most important features of the observed data. One of the modern methods providing a sufficiently simple and flexible framework for auxiliary model estimation is the semi-non-parametric (SNP) models (see Gallant and Tauchen (1987)). We worked with AR(L)-ARCH(M)-Hermite(K,0) model which describes density of y_t as

$$f(y_t|\theta) = C[P_K(z_t)]^2 \phi(y_t|\mu_{x_{t-1}}, \Sigma_{x_{t-1}}),\qquad(21)$$

where

$$
\begin{aligned}
C \quad &\text{is} \quad \text{the normalizing constant,}\\
P_k \quad &\text{is} \quad \text{the Hermite Polynomial of degree } K,\\
x_{t-1} \quad &\equiv \quad (y_{t-L}, \ldots, y_{t-1}) \text{ is the lag vector so that the conditional}\\
&\qquad \text{distribution of } y_t \text{ given all the past depends only on } x_{t-1},\\
\mu_{x_{t-i}} \quad &= \quad \psi_0 + \psi_1 y_{t-i} + \psi_2 y_{t-i-1} + \cdots + \psi_L y_{t-i-L+1},\\
\Sigma_{x_{t-1}} \quad &= \quad R^2_{x_{t-1}},\\
R_{x_{t-1}} \quad &= \quad \tau_0 + \tau_1|y_{t-M} - \mu_{x_{t-M-1}}| + \tau_2|y_{t-M-1} - \mu_{x_{t-M-2}}| + \\
&\qquad + \cdots + \tau_M|y_{t-1} - \mu_{x_{t-2}}|, \text{ and}\\
z_t \quad &= \quad (y_t - \mu_{x_{t-1}})/R_{x_{t-1}}.
\end{aligned}
$$

Estimation of the SNP model is done by maximum likelihood, providing consistent and asymptotically efficient estimators. A proper choice of the order of the model is made using Schwarz's Bayes information criterion (BIC) (see Schwarz (1978)) which puts a penalty for overfitting. With this criterion preferable model turns out to be AR(2)-ARCH(4)-Hermite(6,0). As for embedding the structural model, note that once discretized FV model is AR(1) with conditionally heteroscedastic innovations and therefore we can expect that AR-ARCH part of SNP will be able to incorporate this heteroscedasticity and Hermite polynomial will adjust the shape of the density of the innovations.

Moment generating conditions in (18) were estimated by Monte-Carlo, averaging the estimated scores of the AR(2)-ARCH(4)-Hermite(6,0) on a series of 200000 weekly observations generated by application of the Euler

Table 4: EMM estimates of parameters

Parameter	Estimate	t-statistic
μ	6.520	18.83
κ	0.109	3.19
ν	2.640	9.26
λ	1.482	1.67
τ	1.934	4.34

discretization scheme with 20 intervals per week to the system of SDE (4). The estimation results are reported in Table 4. For more information see Danilov and Drost (2000)[3].

4.4 Volatility Estimation

Figure 3 on page 18 shows the estimated volatilities obtained by Kitagawa smoothing, MCM and extended Kalman smoothing method. We can clearly see that all the methods under considerations reveal two periods of high volatility. The first one corresponds to years 1973-1976 approximately. The reasons for high interest rates volatility in this period are well known. The Middle East War of October 1973 when Arab countries were defeated by Israel was followed by so called "Arab oil embargo". It lead to a considerable jump in oil prices, almost quadrupled, and triggered economical crisis in US. Next few years were marked by high inflation, high interest rates and high instability of world security markets. The second period of high volatility corresponds to the monetary crisis of 1979. When the second oil price rise of 1979 happened, the United States Federal Reserve Board adopted a tight monetary policy trying to curb inflation and stem an outflow of capital. This pushed up real (and nominal) interest rates to historically high levels. A few other key developed countries followed similar contradictory policies, which triggered a worldwide recession and drove up interest rates on a world scale, see e.g. Cheru (1999). We can see that in all estimated volatility profiles at period 1979-1982 volatility is maximal.

Also, apparently, the EKS tends to 'underestimate' volatility at high volatile regions. The MCM, in turn, 'oversmoothes' volatility, especially

---

[3]These parameter estimations are obtained when the data are expressed in percentages. Since in following we use data in decimal points (divided by 100), the parameter values were renormalised appropriately.

Figure 3: Volatility estimates for weekly US 3-month T-bill yield data
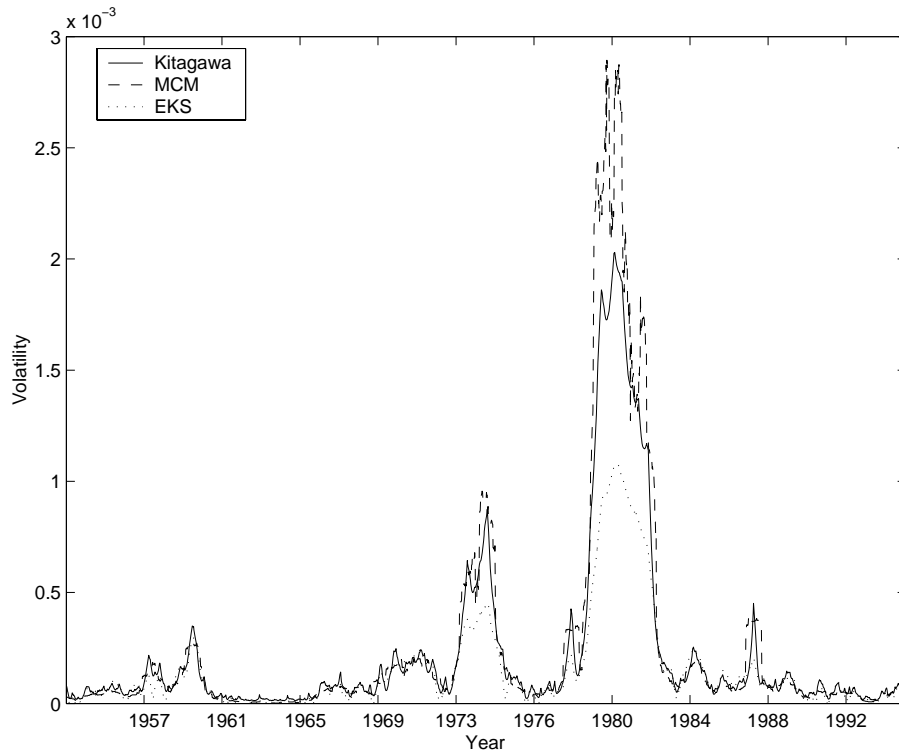
when it is low.

5 Conclusion

In this paper we have considered two factor stochastic volatility models for short term interest rates. We have employed three different methods, namely the Kitagawa (smoothing) method, method of conditional moments, and extended Kalman (smoothing) method to estimate the unobserved volatility component. Based on our analysis we find that Kitagawa method outperforms all other methods.

References

[1] B.D.O. Anderson, J.B. Moore (1979), *Optimal Filtering*, Prentice-Hall, New Jersey.

[2] T.G. Andersen, J. Lund (1997), *Estimating continuous-time stochastic volatility models of the short-term interest rate*, Journal of Econometrics 77 (1997), p.343-377.

[3] F. Black, M. Scholes (1973), *The Pricing of Options and Corporate Liabilities*, The Journal of Political Economy, Vol. 81, No. 3. (May-Jun.,1973), pp. 637-654.

[4] K.C. Chan, G.A. Karolyi, F.A. Longstaff, A.B. Sanders (1992), *An empirical comparison of alternative models of the short term interest rate*, Journal of Finance 47, 1209-1227.

[5] F. Cheru (1999), *ECONOMIC, SOCIAL AND CULTURAL RIGHTS. Effects of structural adjustment policies on the full enjoyment of human rights*, Report by the Independent Expert, Mr. Fantu Cheru, COMMISSION ON HUMAN RIGHTS Fifty-fifth session, Item 10 of the provisional agenda [4].

[6] J.C. Cox, J.E. Ingersoll, S.A. Ross (1985), *A theory of the term structure of interest rates*, Econometrica 53, 385-407.

[7] D. Danilov, F.Drost (2000), *Term Structure Models with Stochastic Volatility: Risk Premia Specifications, Exact Pricing Formulas and Empirical Evaluation of Fit*, working paper, Tilburg University.

[8] G. Fong, O. Vasicek (1991) *Fixed-income volatility management*, Journal of portfolio management, Vol. 17, N. 4 p.41.

[9] A.R. Gallant, G. Tauchen (1987), *Seminonparametric maximum likelihood estimator*, Econometrica 55, 363-390.

[10] A.R. Gallant, G. Tauchen (1996), *Which Moments to Match*, Econometric Theory 12, 657-681.

[11] A.R. Gallant, G.R. Long (1997), *Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square*, Biometrica, **84**,(1), 125-141.

---

[4]Available through the Human Rights Internet site as http://hri.ca/fortherecord1999/documentation/commission/e-cn4-1999-50.htm#C

[12] C. Gourieroux, A. Monfort, E. Renault (1993), *Indirect Inference*, Journal of Applied Econometrics 8, S85-S118.

[13] G. Kitagawa (1987), *Non-Gaussian State-Space Modeling of Nonstationary Time Series*, Journal of the American Statistical Association, Vol. 82, pp. 1032-1041.

[14] K. Koedijk, F. Nissen, P. Schotman, C. Wolff (1997) *The dynamics of Short-Term Interest Rate Volatility Reconsidered*, European Finance Review, 1 , pp. 105-130.

[15] F.A. Longstaff, E.S. Schwartz (1992), *Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model*, Journal of Finance, Vol. 47, N. 4. pp. 1259-1282.

[16] R. Merton (1973), *Theory of rational Option Pricing*, Bell Journal of Economics and Management Science 4, 141-183.

[17] R. Merton (1992), *Continuous time Finance*, Cambridge, MA: Blackwell.

[18] R. Rebonato (1996), *Interest-Rate Option Models: Understanding, Analysing and Using Models for Exotic Interest Rate Options*, Chichester, Wiley.

[19] S. Sundaresan, (1997), *Fixed Income Markets and their Derivatives*, South-Western College Publishing, 1997.

[20] G. Schwarz (1978), *Estimating the dimension of the model*, Annals of Statistics 6, 461-464.

[21] G. Tauchen (1996), *New minimum chi-square methods in empirical finance*, In D. Kreps and K. Wallis (eds.), Advances in Econometrics, Seventh World Congress. Cambridge, UK: Cambridge University Press.

[22] O. Vasicek (1977), *An equilibrium characterization of the term structure*, Journal of Financial Economics 5, 177-188.