Bayesian Estimation of CIR Model

Xiaoxia Feng[1] and Dejun Xie[2*]

[1]*Xi'an Jiaotong University and* [2]*Xi'an Jiaotong-Liverpool University*

Abstract: This article concerns the Bayesian estimation of interest rate models based on Euler-Maruyama approximation. Assume the short term interest rate follows the CIR model, an iterative method of Bayesian estimation is proposed. Markov Chain Monte Carlo simulation based on Gibbs sampler is used for the posterior estimation of the parameters. The maximum A-posteriori estimation using the genetic algorithm is employed for finding the Bayesian estimates of the parameters. The method and the algorithm are calibrated with the historical data of US Treasury bills.

Key words: Bayesian estimation, CIR model, Gibbs sampler, MAP estimation, MCMC method.

1. Introduction

The short-term interest rate is one of the key variables in economy. It is not only the main factor impacting the macro economic growth, but also an important financial instrument in the option and derivative market. For instance, bond prices and values of mortgage contracts are largely determined by the term structure of interest rates (see, for example, Hürlimann, 2011; Xie *et al.*, 2007; Duffie and Singleton, 1999). Due to its non-negativity and relative tractability in nature, Cox-Ingersoll-Ross model (Cox *et al.*, 1985; CIR hereafter) is one of the most employed interest rate models in literature. Accordingly, there exist considerable studies concerning the parameter estimation for the model. Implementation of analytical methods such as maximum likelihood estimation will be encountered with difficulties, particularly in the estimation of the drift parameters. Essentially due to the complexity of the transitional density function of CIR model, it is difficult to obtain computationally useful expressions for the unknown parameters (e.g., Kladívko, 2007). Thus previous literature has mainly focused on approximation methods. This includes the discretization of the time continuous model (e.g., Shoji and Ozaki, 1998; Yu and Phillips, 2001). However, the

---

*Corresponding author.

success of the discretization methods largely depends on the magnitude of observational interval, which is difficult to control from macroeconomic point of view. Another approach is based on nonparametric techniques to approximate the drift function, or the diffusion function, or the transitional density function itself (e.g., Aït-Sahalia, 1996; Jiang and Knight, 1997). These approximation methods, while theoretically feasible, can pose great challenges in implementation, as pointed by Robert and Stramer (2001), for instance.

Our work intends to use the Bayesian approach for the parameter estimation. To this end, we need algorithms for choosing both latent data and posteriori probability. Here we use the Gibbs sampler algorithm (Sorensen and Gianola, 2002) to create latent data for the Treasury bills yields. Gibbs sampler algorithm but not Metropolis-Hasting algorithm is embraced with the following considerations. First, it allows us to accept a candidate point without worry of the acceptance probability. Second, it generates the fully conditional posterior distribution of the parameters and the latent data. And we use the algorithm in Tanner and Wong (1987) to calculate the posteriori probability density of the three parameters appearing in the CIR model. A genetic algorithm is implemented to achieve the maximum A-posteriori (MAP hereafter) estimation of the parameters.

The paper will proceed as follows. Section 2 contains the derivation of the fully conditional posterior distribution of the parameters and the latent data. Section 3 describes the sampling algorithm and the MAP estimates of the parameters. Weekly data of 6-month US Treasury bills will be used in Section 4 to estimate the CIR process. Concluding remarks and future research directions are provided in Section 5.

2. The Derivation of the Fully Conditional Posterior Distribution

The CIR process is defined by the stochastic differential equation

$$dy(t) = \{\alpha - \beta y(t)\}dt + \sigma\sqrt{y(t)}dB(t), \qquad (2.1)$$

where $\{B(t), t \geq 0\}$ is a standard Brownian motion and $\alpha, \beta, \sigma > 0$ are the constant model parameters. Traditional methods such as maximum likelihood method attempt to estimate the parameters directly. Here we first apply the Euler-Maruyama scheme to approximate the model itself. Under the Euler-Maruyama (or Euler) approximation, the model can be rewritten as

$$y(t + \Delta^+) = y(t) + \{\alpha - \beta y(t)\}\Delta^+ + \sigma\sqrt{y(t)}\varepsilon_t \qquad (2.2)$$

where $\varepsilon_t \sim N(0, \Delta^+)$. Suppose there are T observations, and M augmented data between each pair of observations. Let $Y = (y_1, \cdots, y_T)$ denote the set of all observation data and $Y^* = (y_1^*, \cdots, y_{T-1}^*)$ the set of all augmented data, where

$y_t^* = \{y_{t,1}^*, \cdots, y_{t,M}^*\}$. According to Jones (1998), because the drift term $\alpha - \beta y(t)$ is linear in α, β and the diffusion term $\sigma\sqrt{y(t)}$ is the product of the parameter σ and a function of $y(t)$, we can partition the parameter vector $(\alpha, \beta, \sigma^2)$ as $\theta = (\psi, \sigma^2)$ where $\psi = (\alpha, \beta)$. And more importantly, we can assume that the priori probability density of the parameters satisfies $p(\alpha, \beta, \sigma) \propto 1/\sigma$. Starting with the priori probability density, we can determine, respectively, the fully conditional posterior distributions of y_t^*, ψ and σ^2.

For each $t > 0$, define $\Delta = \frac{\Delta^+}{M+1}$. Assume that $y_{t,j}^*$ is a Markov process for $j = 0, 1, \cdots, M$, we have

$$f(y_t^*|Y, \theta) = f(y_{t,1}^*, \cdots, y_{t,M}^*|Y, \theta) = \prod_{j=1}^{M} f(y_{t,j}^* \mid y_{t,j-1}^*, \theta), \qquad (2.3)$$

where $y_{t,0}^* = y_t$ and

$$y_{t,j+1}^*|y_{t,j}^*, \theta \sim N(y_{t,j}^* + (\alpha - \beta y_{t,j}^*)\Delta, \sigma^2 \triangle y_{t,j}^*), \qquad (2.4)$$

the fully conditional probability density of Y^*.

Now consider the fully conditional posterior distribution of each parameter appearing in the model. For ψ, we have

$$\begin{aligned}
p(\psi|Y, Y^*, \sigma^2) &= p(\alpha, \beta|Y, Y^*, \sigma) && (2.5)\\
&\propto p(Y^*, Y|\theta)p(\theta)\\
&\propto p(Y^*, Y|\theta)\\
&\propto \prod_{t=1}^{T-1} \prod_{j=0}^{M} f(y_{t,j+1}^*|y_{t,j}^*, \theta) && (\text{where} \quad y_{t,M+1}^* = y_{t+1})\\
&\propto \exp\{\sum_{t=1}^{T-1}\sum_{j=0}^{M} \frac{-\{y_{t,j+1}^* - [y_{t,j}^* + (\alpha - \beta y_{t,j}^*)\triangle]\}^2}{2\sigma^2 \triangle y_{t,j}^*}\}\\
&\propto \exp\{-\sum_{t=1}^{T-1}\sum_{j=0}^{M} \frac{\{[(\alpha - \beta y_{t,j}^*)\triangle + y_{t,j}^*]^2 - 2y_{t,j+1}^*(\alpha - \beta y_{t,j}^*)\triangle)\}}{2\sigma^2 \triangle y_{t,j}^*}\}\\
&\propto \exp\{-\sum_{t=1}^{T-1}\sum_{j=0}^{M} \frac{(\alpha - \beta y_{t,j}^*)^2 \triangle^2 + 2\triangle(y_{t,j}^* - y_{t,j+1}^*)(\alpha - \beta y_{t,j}^*)}{2\sigma^2 \triangle y_{t,j}^*}\}.
\end{aligned}$$

Since

$$\begin{aligned}
&(\alpha - \beta y_{t,j}^*)^2 \triangle^2 + 2\triangle(y_{t,j}^* - y_{t,j+1}^*)(\alpha - \beta y_{t,j}^*)\\
&= \triangle^2\alpha^2 + \triangle^2 y_{t,j}^{*2}\beta^2 - 2\triangle^2 y_{t,j}^*\alpha\beta + 2\triangle(y_{t,j}^* - y_{t,j+1}^*)\alpha - 2\triangle(y_{t,j}^* - y_{t,j+1}^*)y_{t,j}^*\beta,
\end{aligned}$$

we have

$$p(\psi|Y, Y^*, \sigma^2)$$
$$\propto \exp\{-\frac{\triangle A\alpha^2 + \triangle B\beta^2 - 2\triangle(T-1)(M+1)\alpha\beta - 2C\alpha - 2D\beta}{2\sigma^2}\}, \qquad (2.6)$$

where

$$A = \sum_{t=1}^{T-1}\sum_{j=0}^{M}\frac{1}{y_{t,j}^*},$$

$$B = \sum_{t=1}^{T-1}\sum_{j=0}^{M} y_{t,j}^*,$$

$$C = -\sum_{t=1}^{T-1}\sum_{j=0}^{M}\frac{y_{t,j}^* - y_{t,j+1}^*}{y_{t,j}^*},$$

$$D = \sum_{t=1}^{T-1}\sum_{j=0}^{M}(y_{t,j}^* - y_{t,j+1}^*).$$

This shows that $\psi|Y, Y^*, \sigma^2$ is distributed with some bivariate normal distribution, which hints us to write $\psi|Y, Y^*, \sigma \sim N(\mu, \Lambda^{-1})$, where $\mu = \binom{\mu_1}{\mu_2}$ and $\Lambda = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$, then

$$\begin{pmatrix} \alpha - \mu_1, \beta - \mu_2 \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}\begin{pmatrix} \alpha - \mu_1 \\ \beta - \mu_2 \end{pmatrix} = a_{11}\alpha^2 + a_{22}\beta^2 + 2a_{12}\alpha\beta$$
$$- 2(a_{11}\mu_1 + a_{12}\mu_2)\alpha - 2(a_{22}\mu_2 + a_{12}\mu_1)\beta + (2a_{12}\mu_1\mu_2 + a_{11}\mu_1^2 + a_{22}\mu_2^2). \quad (2.7)$$

Comparing formula (2.6) and formula (2.7) yields

$$\begin{cases} a_{11} = \dfrac{\triangle}{\sigma^2}A, \\ a_{22} = \dfrac{\triangle}{\sigma^2}B, \\ a_{12} = -\dfrac{\triangle}{\sigma^2}(T-1)(M+1), \end{cases} \qquad (2.8)$$

and

$$\begin{cases} a_{11}\mu_1 + a_{12}\mu_2 = C, \\ a_{22}\mu_2 + a_{12}\mu_1 = D. \end{cases} \qquad (2.9)$$

The solution of the (2.9) gives

$$\begin{cases} \mu_1 = \dfrac{a_{22}C - a_{12}D}{a_{11}a_{22} - a_{12}^2}, \\ \mu_2 = \dfrac{-a_{12}C + a_{11}D}{a_{11}a_{22} - a_{12}^2}. \end{cases} \qquad (2.10)$$

In summary of the above derivation, we now have the fully conditional posterior distribution of ψ:

$$\psi|Y, Y^*, \sigma^2 \sim N(\mu, \Lambda^{-1}) \qquad (2.11)$$

where μ_1 and μ_2 are given by (2.10) and

$$\Lambda = \begin{pmatrix} \frac{\triangle}{\sigma^2} A & -\frac{\triangle}{\sigma^2}(T-1)(M+1) \\ -\frac{\triangle}{\sigma^2}(T-1)(M+1) & \frac{\triangle}{\sigma^2} B \end{pmatrix}.$$

The fully conditional posterior distribution for σ^2 is

$$p(\sigma^2|Y, Y^*, \psi)$$
$$\propto p(Y, Y^*|\alpha, \beta, \sigma^2)p(\alpha, \beta, \sigma^2)$$
$$\propto p(Y, Y^*|\alpha, \beta, \sigma^2)p(\alpha, \beta, \sigma)(\frac{d\sigma^2}{d\sigma})^{-1}$$
$$\propto \prod_{t=1}^{T-1}\prod_{j=0}^{M} \frac{1}{\sqrt{2\pi}\sigma\sqrt{\triangle}\sqrt{y_{t,j}^*}} \exp\{\frac{-\{y_{t,j+1}^* - [y_{t,j}^* + (\alpha - \beta y_{t,j}^*)\triangle]\}^2}{2\sigma^2 \triangle y_{t,j}^*}\} \cdot \frac{1}{\sigma} \cdot \frac{1}{\sigma}$$
$$\propto (\sigma^2)^{-\frac{(T-1)(M+1)}{2}-1} \exp\{-\frac{\sum_{t=1}^{T-1}\sum_{j=0}^{M}\frac{\{y_{t,j+1}^*-[y_{t,j}^*+(\alpha-\beta y_{t,j}^*)\triangle]\}^2}{2y_{t,j}^*}}{\sigma^2}\}. \qquad (2.12)$$

Therefore,

$$\sigma|Y, Y^*, \psi \sim Inverse\text{--}Gamma(E, F), \qquad (2.13)$$

where $E = (T-1)(M+1)/2$ and

$$F = \sum_{t=1}^{T-1}\sum_{j=0}^{M}\{\frac{\{y_{t,j+1}^* - [y_{t,j}^* + (\alpha - \beta y_{t,j}^*)\triangle]\}^2}{2y_{t,j}^*}\}.$$

3. The Sampling Algorithm and the MAP Estimates of the Parameters

The major steps for the sampling process and algorithm are as follows:

Step 1: Initialize y_0, θ and generate the original values of $y_1^*, \cdots, y_{T-1}^*$ using Gibbs sampler.

Step 2: Use Gibbs sampler to

  (a) update α, β from $f(\psi|Y^*, Y, \sigma^2)$, where $y_1^*, \cdots, y_{T-1}^*$ are from the previous iteration,

  (b) update σ^2 from $f(\sigma^2|Y^*, Y, \alpha, \beta)$, where $y_1^*, \cdots, y_{T-1}^*$ are from the previous iteration and α, β are from (a).

Step 3: Update $y_1^*, \cdots, y_{T-1}^*$ from $f(y_t^*|y_t, \alpha, \beta, \sigma^2)$.

Step 4: Repeat Step 2 until the prescribed sampling size N is reached.

After the sampling of N (in this paper, $N = 1500$) values for each parameter from the fully conditional posterior distribution, using the method of Tanner and Wong (1987), we obtain the posterior density of the parameters as

$$p(\sigma^2|Y) = \frac{1}{N-500} \sum_{j=501}^{N} p(\sigma^2|Y, Y_{(j)}^*, \alpha_{(j)}, \beta_{(j)})$$

$$= \frac{1}{N-500} \sum_{j=501}^{N} \frac{\beta_{(j)}^{\alpha_{(j)}}}{\Gamma(\alpha_{(j)})} (\sigma^2)^{-\alpha_{(j)}-1} \exp(-\frac{\beta_{(j)}}{\sigma^2}), \qquad (3.1)$$

and

$$p(\alpha, \beta|Y)$$

$$= \frac{1}{N-500} \sum_{j=501}^{N} p(\sigma^2|Y, Y_{(j)}^*, \sigma_{(j)}^2)$$

$$= \frac{1}{N-500} \sum_{j=501}^{N} \frac{1}{2\pi} |\Lambda_{(j)}|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\alpha-\mu_{1(j)}, \beta-\mu_{2(j)})\Lambda_{(j)} \begin{pmatrix} \alpha - \mu_{1(j)} \\ \beta - \mu_{2(j)} \end{pmatrix}\}, \quad (3.2)$$

where the subscript j refers to the jth iteration. In order to negate the effects of initial conditions, the first 500 iterations were thrown out.

The MAP estimates of the parameters are the points where $p(\alpha, \beta|Y)$ and $p(\sigma^2|Y)$, respectively, achieve the maximum values. The genetic algorithm toolbox in MATLAB will help to realize the optimization process.

4. Estimating the CIR Process for US Treasury Bills

We present the estimating results for the CIR process using the time series of weekly observations of US 6-month Treasury bills. Because the data were observed weekly, we set the time interval as $\triangle^{\dagger} = 1/52$ (years). The number of latent points between each consecutive pair of observations is $M = 20$, which has been, to the best of our knowledge, tested by and widely used in literature. The

following Figure 1 gives the time series of historical yields of the US 6-month Treasury bills over a duration of 10 years, starting from November 17, 2000 to October 29, 2010. An example of the iteration results for parameter estimations is plotted in Figure 2.
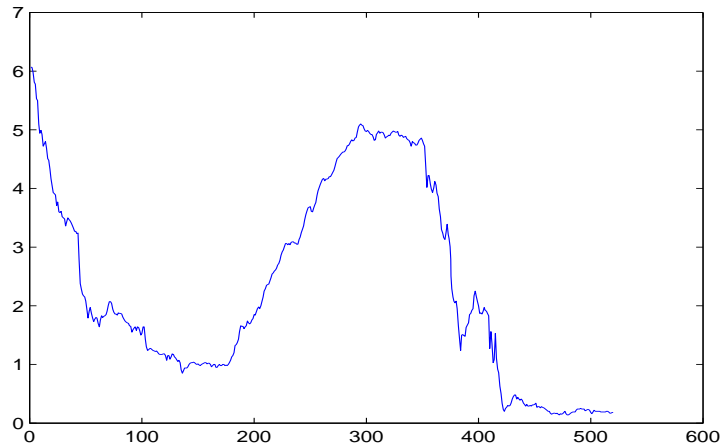

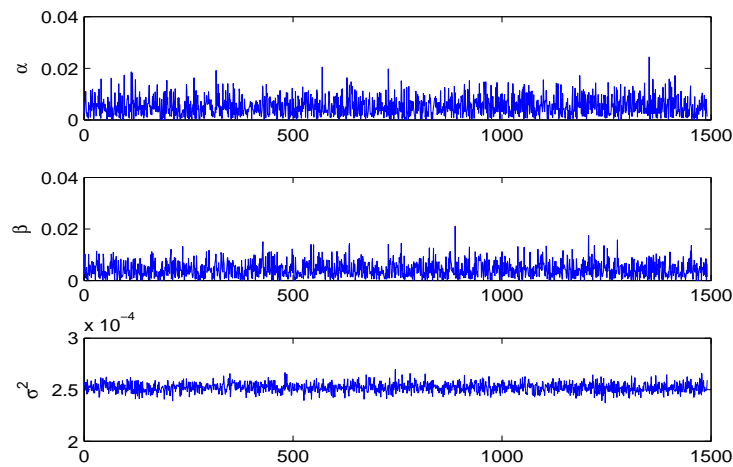Figure 1: Time series of US 6-month Treasury bills over a duration of 10 years

Figure 2: The sampled values of the parameters α, β, σ^2 from the fully conditional posterior distribution in each iteration with 6 years of observations from November 12, 2004 to October, 29, 2010. Here $M = 20, N = 1500$

In order to obtain the maximum values of (3.1) and (3.2), we take advantage of the genetic algorithm toolbox in MATLAB. The MAP estimates of the parameters using 2, 4, 6, 8, and 10 years of US Treasury bills are provided in Table 1.

Table 1: Bayesian estimates of CIR model using historical weekly data of US Treasury bills. Here the closing date is October 29, 2010. The starting dates are November 7, 2008, November 10, 2006, November 12, 2004, November 15, 2002, November 17, 2000, respectively

Year	α	β	σ^2	α/β	$\bar{y}$
2	1.586×10^{-3}	6.944×10^{-3}	0.0398	0.2284	0.2618
4	1.378×10^{-3}	8.541×10^{-4}	0.0019	1.6130	1.8207
6	1.486×10^{-3}	6.437×10^{-4}	0.0018	2.3085	2.5161
8	1.831×10^{-4}	8.633×10^{-5}	0.0019	2.1209	2.1948
10	1.641×10^{-3}	7.619×10^{-4}	0.0014	2.1538	2.3198

From Table 1, we see that the estimates of long term means with different durations are all close to the means of the historical data themselves. For example, the estimated long term mean for 2 years is 2.1209, which has a deviation only about 0.07, or 3%, from the mean of the observed data, 2.1948. And the estimates of volatility parameter for the periods with 4, 6, and 8 years are all about $\sqrt{0.002}$, almost the same with each other. However, for the parameter β, which measures the speed of adjustment for the process, the estimates with different periods are different. This also explains why it is more difficult to estimate the drift parameters than the diffusion parameters. In addition, the estimates for the model are not always consistent for different durations of observations. For instance, the long term means are different for 4 year and 6 years data sets. In the meantime, the relative bias in estimation of the drift parameter β is apparent. This suggests that there may exist regime changes in the evolution of interest rates or the efficacy of affine term structure for describing the interest evolution over large time is questionable. We would like to remark that partial empirical evidence leading to similar conjecture can be found in, say, Tang and Chen (2009).

5. Conclusion

The current work concerns the Bayesian estimation of the parameters for the CIR model. The Gibbs sampler and MCMC algorithm are adopted to simulate the fully conditional posterior distribution. The MAP estimation is implemented for finding the Bayesian estimates of the parameters. The optimal solution is sought with a genetic algorithm. The effectiveness and robustness of the method are calibrated with historical data of US 6-month Treasury bills. The experimental results tend to support the existence of regime changes in the evolution of US Treasury bills over the past 20 years. One of our future research directions would be to further investigate this hypothesis.

References

Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64**, 527-560.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica* **53**, 363-384.

Duffie, D. and Singleton, K. J. (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies* **12**, 687-720.

Hürlimann, W. (2011). Valuation of fixed and variable rate mortgages: binomial tree versus analytical approximations. *Decisions in Economics and Finance*, DOI: 10.1007/s10203-011-0117-z.

Jiang, G. J. and Knight, J. L. (1997). A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Economic theory* **13**, 615-645.

Jones, C. S. (1999). Bayesian estimation of continuous-time finance models. Working paper.

Kladívko, K. (2007). Maximum likelihood estimation of the Cox-Ingersoll-Ross process: the Matlab implementation. Technical Computing Prague.

Robert, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88**, 603-621.

Shoji, I. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications* **16**, 733-752.

Sorensen, D. and Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer, New York.

Tang, C. Y. and Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics* **149**, 65-81.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528-540.

Xie, D., Chen, X. and Chadam, J. (2007). Optimal payment of mortgages. *European Journal of Applied Mathematics* **18**, 363-388.

Yu. J. and Phillips, P. C. B. (2001). A Gaussian approach for continuous time
    models of short-term interest rate. *Econometrics Journal* **4**, 210-224.

Xiaoxia Feng
Department of Mathematics
Xi'an Jiaotong University
Xi'an 710048, China
x.x.f.09@stu.xjtu.edu.cn

Dejun Xie
Department of Mathematical Sciences
Xi'an Jiaotong-Liverpool University
Suzhou 215123, China
Dejun.Xie@xjtlu.edu.cn