Parameter Estimation in Hidden Markov Process With Kalman Filter[1]

Ping Tian and Yaozhong Hu_
Business School of Jilin Universty
Changchun, 130012 China
Department of Mathematics, University of Kansas
Lawrence, Kansas, 66045 USA}

Abstract: In this paper an asset price model described by hidden Markov process $dS(t) = \mu(t)S(t)dt + \sigma S(t)dW(t)$ is considered, where W is a standard Brownian motion and σ is an unknown constant. The mean return$\{\mu(t), 0 \le t \le T\}$ is a stochastic process not necessarily adapted to the filtration generated by the process $\{S(t), 0 \le t \le T\}$ and it contains some unknown parameters to be estimated from a continuous time observation of $S(t)$. Statistical estimators of the parameters σ and the parameters in μ based on Kalman filtering are proposed and some numerical simulations are performed for the proposed estimators.

Key Words: Hidden Markov process, Kalman Filter, parameter estimation, Stochastic Process.

1 Introdution

Let $(\Omega, \pounds, P)$ be a probability space with a complete and right continuous filtration $\{\pounds_t, 0 \le t \le T\}$ and let $W = \{W(t), 0 \le t \le T\}$ and $W' = \{W'(t), 0 \le t \le T\}$ be two independent Brownian motions adapted to the filtration $\{\pounds_t, t \le T\}$. Consider an asset whose price follows the following stochastic differential equation:

$$dS(t) = \mu(t)S(t)dt + \sigma S(t)dW(t) \tag{1.1}$$

where the drift coefficient $\mu = \{\mu(t), t \le T\}$ is mean-reverting process satisfying

$$d\mu(t) = \alpha\big(\vartheta - \mu(t)\big)dt + \beta dW'(t) \tag{1.2}$$

Here we assume that $\alpha > 0, \beta > 0, \sigma > 0$ and $\vartheta \in R$ are unknown parameters to be estimated from the observation of the process$\{S(t), 0 \le t \le T\}$. The initial value of the drift $\mu_0 = \mu(0) \in R$ is also assumed to be an unknown constant to be estimated. Since $\mu(t)$ is unobservable, we called such model (1.1) and (1.2) as hidden Markov process model. We refer to the work of Elliott, R. J. etc(1995) for a general reference. This model has wide applications in many fields such as in finance and the estimation problem has been studied earlier. Our work is motivated by the work of Frydman and Lakner(2003) where a kind of maximum likelihood type estimation method was proposed. To implement their estimators they proposed to use the EM (expectation maximization) algorithm.

For more detail information about the likelihood ratio estimator for hidden Markov chain model, we refer to the references of Frydman and Lakner(2003). See also Dembo and Zeitouni(1986) for EM algorithm application to the parameter estimation of the hidden continuous time random processes and in particular to the parameter estimation of hidden diffusions. It is necessary andvery interesting to simulate the above processes $\mu(t)$ and $S(t)$ and use the approach proposed in the work of Frydman and Lakner(2003) to estimate the parameters. However, their estimators are very complex and it is very hard for us to write computer codes to implement their algorithm.

Here in this paper we propose another method for the estimation of the parameters by using the

[1] The paper was partially supported by Humanities and social science projects of Ministry of Education 12YJCZH187, NSFC grant 11371169.

Kalman filtering technique. Since the solution to (1.1) is positive if the initial condition S(0) is positive we can make a substitute of $Y(t) = \log S(t)$. Then $Y(t)$ will satisfy a linear equation which can be considered as an observation equation for the state process $\mu(t)$. Thus we are led to a parameter estimation problem of some parameters in the state equation in an linear filtering setting. Kalman filtering technique is a natural selection in such problem. The goal of this paper is to make this possible.

This paper is organized as follows. The main idea of our work together with some basic results from the Kalman filtering theory needed in this work is briefly recalled in Section 2. Section 3 proposes some estimators for the parameters of the model by using the idea of Kalman filtering. Section 4 proposes the algorithm to compute our estimators. Section 5 presents some numerical simulation results for our estimators. First we simulate the process $\mu(t)$ and then S(t) with some given specific parameters $\alpha > 0, \beta > 0, \ \sigma > 0 \ \ \vartheta \in R$, and $\mu_0 \in R$,. And then we use our proposed estimators to estimate these parameters. Thus we can compare our estimated parameters with the true parameters. A discussion is also given there.

2 Preliminary and main idea

Kalman filtering theory have been studied extensively. Let us recall some basic results we shall use following the book of Davis. Let the state x(t) and its noise corrupted observation y(t) are given by

$$\begin{cases} dx(t) = A(t)x(t)dt + c(t)d\vartheta(t) \\ x(0) = x \end{cases} \tag{2.1}$$

and

$$\begin{cases} dy(t) = H(t)x(t)dt + G(t)dw(t) \\ y(0) = 0 \end{cases} \tag{2.2}$$

where $\vartheta(t)$ and w(t) are processes with orthogonal incrementsand the initial r.v. x is orthogonal to { $\vartheta(t)$,w(t)} and the coefficient matrices A(t),C(t),H(t),G(t) are deterministic continuous in t.

Assume that at any time t, the values of the process (y(s), $0 \le s \le t$) are observed but x(t) is unknown. We need to estimate the value of x(t). The best estimator in the mean square sense if the conditional expectation of x(t) given (y(s), $0 \le s \le t$). Namely the best estimate $\hat{x}(t) = E(x(t)|y(s), 0 \le s \le t)$. This is also the orthogonal projection of x(t) to the linear space of linear functional of (y(s), $0 \le s \le t$) when y is Gaussian process. The Kalman filtering technique gives an simple explicit way to compute $\hat{x}(t) = E(x(t)|y(s), 0 \le s \le t)$, which is described in the following theorem.

Theorem 2.1 Assume that G(t)G'(t) is strictly positive definite for all t. Then $\hat{x}(t)$ is the unique solution of the linear stochastic differential equation

$$\begin{cases} d\,\hat{x}(t) = (A - PH'(GG')^{-1}H)\,\hat{x}(t)dt + PH'(GG')^{-1}dy(t) \\ \hat{x}(0) = E[x(0)] = m_0 \end{cases} \tag{2.3}$$

where $P(t) = E[(\hat{x}(t) - x(t))(\hat{x}(t) - x(t))^T]$ is the error covariance matrix which is determined by the following matrix Riccati equation

$$\begin{cases} \dot{P} = CC' - PH'^{(GG')^{-1}}HP + AP + PA' \\ P(0) = E[(x(0) - Ex(0))(x(0) - Ex(0))^T] \end{cases} \tag{2.4}$$

The proof of this theorem is in many books (see for example the book of Davis).

Here is the main idea of our approach. First if the hidden process $\mu(t)$ given by (1.2) is observable, then we can estimate the parameters α, ϑ, μ_0 and β by $\mu(s), 0 \le s \le t$. For example, we can estimate the parameter α by he least squares estimator (or maximum likelihood estimator). See equations (3.3)-(3.9) in the next section. However, since $(\mu(s), 0 \leq s \leq t)$ is not available, we can use the Kalman filter ($\hat{\mu}(s)$,$0 \leq s \leq t$) to replace $\mu(t)$. To compute the Kalman filter $\hat{\mu}$, we need to know the true parametersα, ϑ, μ_0 and β and so on. We shall use iteration for this. First we are given an initial approximation of the parameters. We use these parameters to compute the filter $\hat{\mu}(t)$. Then we use this filter $\mu(t)$ to update the approximation of the parameters and so on. The detailed algorithm is explained in Section 5.

3 Parameter Estimators

In this section we consider the equations (1.1) and (1.2). Assume that we have the following observations $\{S(0), S(t_1), S(t_2), \cdots, S(t_n)\}$ of the process S(t) at discrete time instants $t_0, t_1, t_2, \cdots, t_n$, whert$_i = \frac{iT}{n} = ih, i = 0,1,2, \cdots, n$ (we denote $h = \frac{T}{n}$). We wish to estimate the unknown parameters $\sigma, \alpha, \beta, \vartheta$ and the initial unknown value $\mu(0)$ appeared in the equations (1.1) and (1.2). Here we assume that the interval between two consecutive observations are uniform. But other kind of observations can be treated analogously.

Firstly we make the following transformations:

$$\begin{cases} x(t) = \frac{1}{\beta}\mu(t) - \frac{1}{\beta}\mu_0 \\ y(t) = \frac{1}{\sigma}\log\left(\frac{S(t)}{S(0)}\right) + \frac{\sigma}{2}t \end{cases} \tag{3.1}$$

The equations (1.1) and (1.2) become

$$\begin{cases} dx(t) = \alpha\big(\delta - x(t)\big)dt + dw'(t) \\ dy(t) = (\lambda x(t) + \gamma)dt + dw(t) \end{cases} \tag{3.2}$$

Where x(0)=0, y(0)=0, $\delta = \frac{\vartheta - \mu_0}{\beta}$, $\lambda = \frac{\beta}{\sigma}$, and $\gamma = \frac{\mu_0}{\sigma}$.

From the second equation in (3.2) we have

$$y(t) = y(0) + \lambda \int_0^t x(s)ds + \gamma t + W(t).$$

Then we can use the so -called trajectory fitting method to estimate γ (see the work of Kutoyants(2004) and Hu, etc.(2009)):

$$\hat{\gamma} = \frac{(y(T) - y(0)) - \lambda \int_0^T x(t)dt}{T} \tag{3.3}$$

We can also use the similar technique and the first equation of (3.2) to estimate δ:.

$$\hat{\delta} = \frac{x(T) + \alpha \int_0^T x(t)dt}{\alpha T} \tag{3.4}$$

We still need to estimate λ, α and other parameters. For this reason we continue to make the following transformations:

$$\begin{cases} \xi(t) = \lambda\big(\delta - x(t)\big) = \frac{\lambda}{\beta}(\beta\delta - \mu(t) - \mu_0) \\ y'(t) = y(t) - (\lambda\delta + \gamma)t = \frac{1}{\sigma}\log\left(\frac{S(t)}{S(0)}\right) + \frac{\sigma}{2}t - \frac{\vartheta}{\sigma}t \end{cases} \tag{3.5}$$

Then we transform the equations (1.1) and (1.2) into

$$\begin{cases} d\xi(t) = -\alpha\xi(t)dt - \lambda dw'(t) \\ dy'(t) = -\xi(t)dt + dw(t) \end{cases} \tag{3.6}$$

Where $\xi(0) = \lambda\delta$, and y'(0)=y(0)=0.

The first equation of (3.6) is the well-known Ornstein-Uhlenbeck process. There are two well-known types of estimators. One is the so-called least squares estimators and the other one is the so-called the maximum likelihood estimator. These two types of estimator are the same for this simple model and have been studied since long time (see the book of Liptser, and Shiryaev, 2001). They are given by

$$\widehat{\alpha} = -\frac{\int_0^T \xi(t)d\xi(t)}{\int_0^T \xi(t)^2 dt} \tag{3.7}$$

The large deviation type results for the above estimator is studied in the paper of Bercu, and Rouault(2002) and a central limit type result was studied in the work of Hu, and Nualart(2010).

If the process $\xi(t)$ can be observed at discrete time instants $t_k = kh$, where $h = \frac{T}{n}$ and $k = 0,1,2,\cdots$, the following estimator is proposed in .

$$\widetilde{\alpha} = (\frac{1}{n}\sum_{k=0}^{n} \xi(t_k)^2)^{1/2} \tag{3.8}$$

Also in the work of Hu, and Song(2011), the central limit type and Berry-Esseen type results are obtained not only for Brownian motion but also for fractional Brownian motions.

We can use the quadratic variation method to estimate λ

$$\widehat{\lambda}^2 = \frac{\sum_{k=0}^{n-1}(\xi(t_{k+1})-\xi(t_k))^2}{T} \tag{3.9}$$

The maximum likelihood estimator for λ is also given by

$$\widehat{\lambda}^2 = \frac{\alpha}{n}\left(\xi_n - \bar{\xi_n}\right)^T M^{-1}(\xi_n - \bar{\xi_n})$$

(3.10)

Where $M = (m(i,j))_{0\le i,j\le n}$ is the covariance matrix of $\xi_n = (\xi(t_1),\cdots,\xi(t_n))^T$ with $m_{ij} = 1 - e^{-\alpha|t_i - t_j|}$ and $\bar{\xi_n} = E(\xi_n)$. Finally for the unknown parameter σ, we make the transformation $Z(t) = \log(S(t)$. Then it is well-known one estimator for σ can be

$$\widehat{\sigma}^2 = \frac{\sum_{k=0}^{n-1}(Z(t_{k+1})-Z(t_k))^2}{T} \tag{3.11}$$

The equations (3.3), (3.4), (3.7) (or (3.8)), (3.9) (or (3.10)), and (3.11) can be combined to estimate the parameters γ, δ, α, σ. However, in the above mentioned equations, except in (3.11) for the estimate of σ, we have to use x or ξ in our estimators. since we can only observe $\{S(t_0), S(t_1), S(t_2),\cdots, S(t_n)\}$ and as a consequence we can only observe $\{y(t_1), y(t_2),\cdots, y(t_n)\}$. So x(t) or $\xi(t)$ can not be used in our estimators since it is not available.

To overcome this difficulty, an EM algorithm was proposed in the work of Frydman and Lakner(2003). Here we propose a different approach, namely to use the Kalman filtering techniques (see Theorem 2.1). This approach of using Kalman filter is much simpler conceptually and is also much easy to implement.

Let $£_t^{y'} = \sigma(y'(s), 0 \le s \le t)$ be the σ-algebra generated by the observation process y' up to time instant t. Then $£_t^{y'}$ contains all the information available up to time t. The best estimation of $\xi(t)$ in the mean squares sense based on the information $£_t^{y'}$ is given by $\hat{\xi}(t) = E(\xi(t)|£_t^{y'})$, the conditional expectation of $\xi(t)$ with respect to the σ-algebra

$£_t^{y'}$. This quantity $\hat{\xi}(t)$ can be computed through the Kalman filter. In fact, from Theorem 2.1 we have

$$\begin{cases} d\hat{\xi}(t) = -(\alpha + P)\hat{\xi}(t)dt - Pdy'(t) \\ \qquad \hat{\xi}(0) = E\xi = \xi(0) = \lambda\delta \end{cases} \tag{3.12}$$

where P(t) satisfies the following linear Riccati equation

$$\begin{cases} \dot{P} = \lambda^2 - P^2 - 2\alpha P \\ P(0) = \mathrm{cov}(\xi_0) = 0 \end{cases} \tag{3.13}$$

The equation (3.13) is a common ordinary differential equation. It can be solved by a simple technique of variable separation. The equation (3.12) is a linear stochastic differential equation. It can be solved also easily. We have the following proposition for their solutions.

Proposition 3.1 The solutions to (3.12) and (3.13) are given by the following

$$\hat{\xi}(t) = e^{-\alpha t - \int_0^t P(s)ds}(\lambda\delta - \int_0^t e^{\alpha s + \int_0^s P(r)dr} P(s)dy'(s)) \tag{3.14}$$

$$P(t) = \rho_{\alpha,\lambda}\left[1 - 2\left(1 + \frac{\rho_{\alpha,\lambda}+\alpha}{\rho_{\alpha,\lambda}-\alpha}e^{2t\rho_{\alpha,\lambda}}\right)^{-1}\right] - \alpha) \tag{3.15}$$

where $\rho_{\alpha,\lambda} = \sqrt{\alpha^2 + \lambda^2}$.

4 Algorithm to find the estimators

Now we can summarize our estimators. Denote $\theta = (\alpha, \delta, \lambda, \gamma)$ the parameters in our equation. Firstly we define

$$y'(t,\theta) = \frac{1}{\sigma}\log\left(\frac{S(t)}{S(0)}\right) + \frac{\sigma}{2}t - (\lambda\delta + \gamma)t \tag{4.1}$$

$$P(t,\theta) = \rho_{\alpha,\lambda}\left[1 - 2\left(1 + \frac{\rho_{\alpha,\lambda}+\alpha}{\rho_{\alpha,\lambda}-\alpha}e^{2t\rho_{\alpha,\lambda}}\right)^{-1}\right] - \alpha \tag{4.2}$$

$$\hat{\xi}(t,\theta) = e^{-\alpha t - \int_0^t P(s,\theta)ds}(\lambda\delta - \int_0^t e^{\alpha s + \int_0^s P(r,\theta)dr} P(s, \theta)dy'(s, \theta)) \tag{4.3}$$

$$\hat{x}(t,\theta) = \delta - \frac{\hat{\xi}(t,\theta)}{\lambda} \tag{4.4}$$

where $\rho_{\alpha,\lambda} = \sqrt{\alpha^2 + \lambda^2}$.

Since we observe $\{S(t_0), S(t_1), S(t_2), \cdots, S(t_n)\}$, and y' is the function of the $S(t)$ and θ, we have $\{y'(t_1,\theta), y'(t_2,\theta), \cdots, y'(t_{n,\theta})\}$ available. Based on these values of y', we can compute $(\hat{\xi}(t,\theta),\ 0 \le t \le T)$ and $(\hat{x}(t,\theta),\ 0 \le t \le T)$. We substitute ξ and x in (3.3), (3.4), (3.7), (3.9) and use Itô-Riemann sums to approximate the stochastic integral. Thus we obtain the following estimators.

$$\hat{\alpha} = -\frac{n\sum_{i=1}^{n-1}\hat{\xi}(t_i,\theta)(\hat{\xi}(t_{i+1},\theta) - \hat{\xi}(t_i,\theta))}{T\sum_{i=1}^{n}\hat{\xi}(t_i,\theta)^2} \tag{4.5}$$

$$\hat{\lambda}^2 = \frac{\sum_{k=0}^{n-1}(\hat{\xi}(t_{k+1},\theta) - \hat{\xi}(t_k,\theta))^2}{T} \tag{4.6}$$

$$\hat{\gamma} = \frac{y'(T,\theta) - y'(0,\theta)}{T} - \frac{\hat{\lambda}\sum_{i=1}^{n}\hat{x}(t_i,\theta)}{n} \tag{4.7}$$

$$\hat{\delta} = \frac{\hat{x}(T,\theta)}{\hat{\alpha}T} - \frac{1}{n}\sum_{i=1}^{n}\hat{x}(t_i,\theta) \tag{4.8}$$

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{n-1}\left(\log\left(S(t_{i+1})\right) - \log\left(S(t_i)\right)\right)^2 \tag{4.9}$$

Use the relation between $(\alpha, \beta, \vartheta, \mu_0)$ and $(\alpha, \lambda, \gamma, \delta)$, we can have the estimators for the original parameter$\theta = (\alpha, \beta, \vartheta, \mu_0)$.We shall not write them down in this paper. Actually, we have$\beta = \lambda\sigma, \mu_0 = \gamma\sigma$ and $\vartheta = \beta\delta + \mu_0$.

5 Simulation

The solution to the equation (1.1), namely $S(t) = \mu(t)S(t)dt + \sigma S(t)dW(t)$ with the original value S(0) is given by

$$S(t) = S(0)\exp\{\sigma W(t) - \frac{\sigma^2}{2}t + \int_0^t \mu(r)dr\} \tag{5.1}$$

The hidden Markov process, namely the solution to

$$d\mu(t) = \alpha\big(\vartheta - \mu(t)\big)dt + \beta dw'(t) \tag{5.2}$$

can also be expressed explicitly as

$$\mu(t) = \mu_0 e^{-\alpha t} + \vartheta(1 - e^{-\alpha t}) + \beta e^{-\alpha t}\int_0^t e^{\alpha s} dw'(s) \tag{5.3}$$

We shall use the equations (5.1) and (5.2) to simulate a sample of the process of S(t).

Take T=10, $n = 1000$ and $h = t_i - t_{i-1} = \frac{T}{n} = \frac{1}{100}$. Given some parameters $\theta = (\alpha, \delta, \lambda, \gamma)$, we can use (5.1) and (5.2) to obtain a sample path of $S(t)$.

The estimation of the parameter σ by the formula (4.9) uses only the observed values. It can be handled separately. In our simulation we choose $\sigma = 2$. The estimated value $\hat{\sigma}$ of σ by the estimator (4.9) is 2.0335 which is very close to the original value 2.

Table 1: Estimators and the information for the unknown parameters

	Real value	Interval	Step length	Estimated value	Error
α	2	[0.2,5]	0.2	1.6560	0.3440
β	1	[0.2,5]	0.2	0.1396	0.8614
ϑ	1	[0.2,5]	0.2	2.0948	1.0948
$\mu(0)$	2	[0.2,5]	0.2	2.1308	0.1308

We use (4.5)-(4.8) to estimate the other parameters α, β, ϑ and μ_0. Because both sides of the above equations contain the estimated parameters we have in principle to solve some system of algebraic equations for the form $z = g(z)$. This can be done by using the Newton's method. If the solution to $h(z) = z - g(z)$ is between a and b (for example $h(a) < 0$ and$h(b) > 0$) and if $h(c) > 0$, where $c = \frac{a+b}{2}$, then one knows there is a solution to $z = g(z)$ between a and c. Since we are now in more than one dimension. Motivated by this algorithm we device the following scheme to find the solution to (4.5)-(4.8).We choose the values of the parameters as $\alpha = 2$, $\beta = 1$, $\sigma = 2$, $\vartheta = 1$, $\mu_0 = 2$. With these values we simulate $\mu(t)$ and then $S(t)$. To use the simulated values $S(t)$ to estimateα, β, ϑ and μ_0, we suppose that we know $\alpha \in [0.2,5]$, $\beta \in [0.2,5]$, $\vartheta \in [0.2,5]$, $\mu(0) \in [0.2,5]$. Then we uniformly partition the above domain of the parameters into small regions. For α we set $\alpha(t_1) = 0.2$, $\alpha(t_2) = 0.4, \cdots, \alpha(t_{25}) = 5$. In general, we set $\alpha(t_i) = ih$ and $h = 0.2$ For β we set $\beta(t_i) = ih_1, i = 1,2,\cdots,25, h_1 = 0.2$. For ϑ we set $\vartheta = ih$ and h =0.2. Forμ_0 we set $\mu_0(t_i) = ih$ and h =0.2. Since all the functions in our models are continues, there must have one point in the domain of parameters that is closest to the solution. Using this method we get the estimation point $\alpha = 1.6560, \beta = 0.1396$, $\vartheta = 2.0948$,$\mu_0 = 2.1307$.

References

[1] Bercu, B. and Rouault, A. Sharp large deviations for the Ornstein-Uhlenbeck process. Theory Probab. Appl. 46 (2002), 1-19.
[2] Bishwal, J. P. N. Parameter Estimation in Stochastic Differential Equations. Springer-Verlag Berlin and Heidelberg, 2000.
[3] Davis, M.H.A. Linear Estimation and Stochastic Control. London : Chapman and Hall ; New York : Wiley: distributed in the U.S.A. by Halsted Press 134-150.
[4] Dembo, A.; Zeitouni, O. Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. Stochastic Process.Appl. 23 (1986), no. 1, 91-113.
[5] Elliott, R. J.; Aggoun, L. and Moore, J. B. Hidden Markov models. Estimation and control. Springer-Verlag New York , 1995.
[6] Frydman, H. and Lakner, P. Maximum likelihood estimation of hidden Markov processes. Ann. Appl. Probab. 13 (2003), no. 4, 1296-312.
[7] Hu, Y. and Long, H. Least squares estimator for Ornstein-Uhlenbeck processes driven by stable motions. Stochastic Process. Appl. 119 (2009), 2465-2480.
[8] Hu, Y. and Nualart, D. Parameter estimation for fractional Ornstein-Uhlenbeck processes. Statist. Probab. Lett. 80 (2010), no. 11-12, 1030-1038.
[9] Hu, Y. and Song, J. Parameter estimation for fractional Ornstein-Uhlenbeck processes with discrete observations. Malliavin Calculus and Stochastic Analysis,Springer Proceedings in Mathematics & Statistics 34 (2013), pp 427-442.
[10] Kutoyants, Y. A. Statistical inference for ergodic diffusion processes. Springer-Verlag London, 2004.
[11] Liew, C. C. and Siu, T. K. A hidden Markov regime-switching model for option valuation. Insurance: Mathematics and Economics 47 (2010) 374-384.
[12] Liptser, R. S. and Shiryaev, A. N. Statistics of random processes. II. Applications. Applications of Mathematics, 6. Springer, Springer-Verlag Berlin and Heidelberg, 2001.