Drift and volatility estimation in discrete time

Article *in* Journal of Economic Dynamics and Control · February 1998

DOI: 10.1016/S0165-1889(97)00052-3

Robert J. Elliott
University of Adelaide

William C. Hunter
University of Iowa

ELSEVIER

Journal of Economic Dynamics and Control
22 (1998) 209–218

JOURNAL OF
Economic
Dynamics
& Control

Drift and volatility estimation in discrete time

Robert J. Elliott[a,*], William C. Hunter[b], Barbara M. Jamieson[a]

[a] *Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

[b] *Federal Reserve Bank, 230 South La Salle St. Chicago, IL 60604, USA*

Received 14 February 1995; accepted 7 January 1997

Abstract

In discrete time the increment of the logarithm of the price of a risky asset is supposed to involve two parameters which may be thought of as the 'drift' and 'volatility'. It is assumed these parameters take finitely many values, and that they change value like a Markov chain on this state space. Filtering and parameter estimation techniques from Hidden Markov Models are then applied to obtain recursive estimates of the 'drift' and 'volatility'. Further, all parameters in the model can be estimated. The method is illustrated by applying the results to two series of prices.

Keywords: Filtering; Hidden Markov Models; Parameter estimation; Volatility
JEL classification: C4; G0

1. Introduction

In a recent paper (Elliott and Rishel, 1994), the filtering and parameter estimation techniques of Hidden Markov Models are used to estimate the implicit interest rate of a risky asset whose price dynamics are described in continuous time by the usual log-normal stochastic differential equation. Techniques used in that paper included the Girsanov theorem and expectation maximization (EM) algorithm. The Girsanov theorem changes the drift (or interest rate), by a change of probability measure. The EM algorithm is then used to filter the logarithm of the Girsanov density, and the optimal estimate for the drift is obtained as the value which maximizes this filtered logarithm of the density. Unfortunately

* Corresponding author. Email: relliott@maths.adelaide.edu.au
Robert Elliott wishes to thank SSHRC for its support, and also the Federal Reserve Bank for its hospitality in October 1994.

PII S0165-1889(97)00052-3

this technique cannot be used in continuous time to estimate the volatility, because the probability measures corresponding to diffusions with different diffusion coefficients (volatilities) are singular.

In discrete time, however, our method will work. We consider a multiplicative model for the evolution of the price of a risky asset. The increment of the logarithm of the price involves a Gaussian noise and parameters which we suppose evolve as a finite state Markov chain. The filtering and estimation techniques for Hidden Markov Models developed in our paper (Elliott, 1994) are then applied. See also the book (Elliott et al., 1994). This enables us to obtain not only the best estimate of the state of the chain (that is, the coefficients), given the observations of the increments of the logarithm of the price, but also estimate all parameters of the model.

2. Price process

Consider a price process in discrete time, S_n, $n \in Z^+$. Suppose S_n evolves according to the dynamics $S_{n+1} = S_n \exp Y_{n+1}$ where $Y_{n+1} = g_n + \gamma_n b_{n+1}$. Here, $\{b_n\}$, $n \in Z^+$, is a sequence of i.i.d. N(0, 1) random variables.

Suppose (g_n, γ_n) takes values in a finite set $B = \{(g_i, \gamma_i),\ i = 1, \dots, N\}$. Write g for the vector $(g_1, g_2, \dots, g_N)$ and γ for $(\gamma_1, \gamma_2, \dots, \gamma_N)$. Suppose (g_n, γ_n) evolves as a Markov chain with state space B.

We can identify B with

$$\Sigma = \{e_1, e_2, \dots, e_N\},$$

where $e_i = (0, \dots, 0, 1, 0, \dots, 0)' \in R^N$. Suppose $\phi : B \to \Sigma$ gives this bijection, so for each $k, 1 \leq k \leq N,\ e_k = \phi((g_k, \gamma_k))$. Write

$$X_n = \phi(g_n, \gamma_n).$$

Then

$$g_n = \langle g, X_n \rangle$$

and

$$\gamma_n = \langle \gamma, X_n \rangle.$$

Our processes are defined on a probability space $(\Omega, \mathscr{F}, P)$ and we suppose X_n is a Markov chain with state space Σ.

(We could have g_n and γ_n each behaving as independent Markov chains on their state spaces; (g_n, γ_n) is then the (tensor) product Markov chain. See Remarks 3.1.)

Suppose X_n has transition probabilities given by

$$P(X_n = e_j \mid X_{n-1} = e_i) = a_{ji}.$$

Note $\sum_{j=1}^{N} a_{ji} = 1$, and write A for the transition matrix (a_{ji}), $1 \le i, j \le N$. We do not observe the Markov chain X; instead we observe the logarithmic increments of the price process:

$$\begin{aligned} Y_{n+1} &= \log \frac{S_{n+1}}{S_n} \\ &= g_n + \gamma_n b_{n+1} \\ &= \langle g, X_n \rangle + \langle \gamma, X_n \rangle b_{n+1}. \end{aligned}$$

Write $\{G_n\}$, $n \in Z^+$, for the complete filtration generated by the X and Y processes; $\{Y_n\}$, $n \in Z^+$, will denote the complete filtration generated by the Y process.

Write

$$\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2),$$

$$\Gamma^i(Y_n) = \phi\left(\frac{Y_n - g_i}{\gamma_i}\right) \Big/ \gamma_i \phi(Y_n),$$

$$a_i = Ae_i.$$

3. Estimation

For the price process S_n defined in Section 2, having coefficients $(g_n, \gamma_n) = (\langle g, X_n \rangle, \langle \gamma, X_n \rangle)$ which behave like a Markov chain, we can define the following filtering and estimation problem. First note that X_n can be written with dynamics

$$X_n = AX_{n-1} + M_n \tag{1}$$

where $\mathrm{E}[M_n | G_{n-1}] = 0$, so that M is a sequence of martingale increments.

The observation process Y has dynamics

$$Y_{n+1} = \langle g, X_n \rangle + \langle \gamma, X_n \rangle b_{n+1}.$$

We are, therefore, in the situation discussed in the paper of Elliott (1994) (see also the book of Elliott et al., 1994). Following Elliott (1994), recall that the analysis takes place under a probability measure $\overline{P}$ for which the $\{Y_\ell\}$ are i.i.d. N(0, 1) random variables. In fact, suppose we have a probability measure $\overline{P}$ on $(\Omega, \mathscr{F})$ such that under $\overline{P}$:

(a) X_ℓ, $\ell \in Z^+$, is a Markov chain with transition matrix A, so that

$$X_n = AX_{n-1} + M_n,$$

where $\overline{\mathrm{E}}[M_n | G_{n-1}] = 0$, and

(b) Y_ℓ, $\ell \in Z^+$, is a sequence of N(0, 1) i.i.d. random variables.
Write

$$\overline{\gamma}_\ell = \frac{\phi((Y_\ell - \langle g, X_{\ell-1}\rangle)/\langle \gamma, X_{\ell-1}\rangle)}{\langle \gamma, X_{\ell-1}\rangle \phi(Y_\ell)}$$

$$\overline{\Lambda}_n = \prod_{\ell=1}^{n} \overline{\gamma}_\ell \quad \text{for } n \geq 1$$

and define P in terms of $\overline{P}$ by putting $\mathrm{d}P/\mathrm{d}\overline{P}|_{G_n} = \overline{\Lambda}_n$.

Then under P, we still have

$$X_n = AX_{n-1} + M_n,$$

where $\mathrm{E}[M_n | G_{n-1}] = 0$ but now b_n, $n \in Z^+$, is a sequence of i.i.d N(0, 1) random variables, where

$$b_n := \frac{Y_n - \langle g, X_{n-1}\rangle}{\langle \gamma, X_{n-1}\rangle}.$$

That is, under P

$$Y_n = \langle g, X_{n-1}\rangle + \langle \gamma, X_{n-1}\rangle b_n,$$

so under P, Y is given by the 'real-world' dynamics. However, it is easier mathematically to work under $\overline{P}$.

If $\{H_\ell\}$ is any $\{G_\ell\}$ adapted sequence we write

$$\sigma_n(H_n) = \overline{\mathrm{E}}[\overline{\Lambda}_n H_n | Y_n].$$

A version of Bayes' theorem (see Elliott et al., 1994) implies that

$$E[H_\ell | Y_\ell] = \frac{\overline{\mathrm{E}}[\overline{\Lambda}_\ell H_\ell | Y_\ell]}{\overline{\mathrm{E}}[\overline{\Lambda}_\ell | Y_\ell]} = \frac{\sigma_\ell(H_\ell)}{\sigma_\ell(1)} \quad \text{say,}$$

where

$$\sigma_\ell(H_\ell) = \overline{\mathrm{E}}[\overline{\Lambda}_\ell H_\ell | Y_\ell]$$

is an unnormalized conditional expectation of H_ℓ given Y_ℓ.

Applying the results of Elliott (1994), we have the following recursive expressions. First, for the state of the Markov chain:

$$\sigma_n(X_n) = \sum_{i=1}^{N} \langle \sigma_{n-1}(X_{n-1}), e_i\rangle \Gamma^i(Y_n) a_i. \tag{2}$$

Suppose

$$N_n^{rs} = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle \, \langle X_\ell, e_s \rangle,$$

so N_n^{rs} is the number of jumps from e_r to e_j in time n. Then

$$\begin{aligned}\sigma_n(N_n^{rs} X_n) = \sum_{i=1}^{N} \langle \sigma_{n-1}(N_{n-1}^{rs} X_{n-1}), e_i \rangle \Gamma^i(Y_n) a_i \\ + \langle \sigma_{n-1}(X_{n-1}), e_r \rangle \, a_{sr} \Gamma^r(Y_n) e_s.\end{aligned} \tag{3}$$

With $J_n^r = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle$, the occupation time in e_r,

$$\begin{aligned}\sigma_n(J_n^r X_n) = \sum_{i=1}^{N} \langle \sigma_{n-1}(J_{n-1}^r X_{n-1}), e_i \rangle \Gamma^i(Y_n) a_i \\ + \Gamma^r(Y_n) \langle \sigma_{n-1}(Y_{n-1}), e_r \rangle a_r.\end{aligned} \tag{4}$$

With $f(Y)$ a function of Y and

$$G_n^r(f) = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle f(Y_\ell),$$

$$\begin{aligned}\sigma_n(G_n^r(f) X_n) = \sum_{i=1}^{n} \langle \sigma_{n-1}(G_{n-1}^r(f) X_{n-1}), e_i \rangle \Gamma^i(Y_n) a_i \\ + \Gamma^r(Y_n) \langle \sigma_{n-1}(X_{n-1}), e_r \rangle f(Y_n) a_r.\end{aligned} \tag{5}$$

For any process H_n

$$\begin{aligned}\sigma_n(H_n) &= \langle \sigma_n(H_n X_n), \mathbf{1} \rangle \\ &= \sigma_n(H_n \langle X_n, \mathbf{1} \rangle),\end{aligned}$$

where $\mathbf{1} = (1, 1, \ldots, 1)'$. As noted in Elliott (1994), we consider $\sigma_n(H_n X_n)$ for $H = N, J, G$, because, unlike $\sigma_n(H_n)$, closed-form recursions are obtained. Also, $\sigma_n(1) = \langle \sigma_n(X_n), \mathbf{1} \rangle = \overline{\mathrm{E}}[\overline{\Lambda}_n | Y_n]$.

Following Elliott (1994), the above expressions can be used to estimate the parameters of the model. The transition probabilities in the matrix A can be estimated as

$$\hat{a}_{sr}(n) = \frac{\sigma_n(N_n^{rs})}{\sigma_n(J_n^r)}. \tag{6}$$

The components of the $\hat{g}$ vector are re-estimated as

$$\hat{g}_r(n) = \frac{\sigma_n(G_n^r(Y))}{\sigma_n(J_n^r)}, \tag{7}$$

and the components of the volatility vector $\hat{\gamma}$ re-estimated as

$$\hat{\gamma}_r(n) = \frac{(\sigma_n(G_n^r(Y^2)) - 2g_r\sigma_n(G_n^r(Y)) + g_r^2\sigma_n(J_n^r))}{\sigma_n(J_n^r)}. \tag{8}$$

Remarks 3.1. If the probabilistic behaviour of the 'drift' g and 'volatility' γ are independent the model could be modified as follows. Suppose g_n takes values in a finite set $B_1 = \{g_1, g_2, \dots, g_{N(1)}\}$ and γ_n takes values in $B_2 = \{\gamma_1, \gamma_2, \dots, \gamma_{N(2)}\}$. Then there are bijections ϕ_1 (resp. ϕ_2) of B_1 (resp. B_2) with the set of unit vectors $\Sigma_1 = \{e_1, e_2, \dots, e_{N(1)}\}$ of $R^{N(\cdot)}$, (resp. the set of unit vectors $\Sigma_2 = \{f_1, f_2, \dots, f_{N(2)}\}$ of $R^{N(2)}$).

Write

$$X_n^1 = \phi_1(g_n) \in R^{N(1)}$$

and

$$X_n^2 = \phi_2(\gamma_n) \in R^{N(2)}.$$

Suppose X_n^i behaves like a Markov chain on its state space Σ_i with transition matrix A_i, so that

$$X_n^i = A_i X_{n-1}^i + M_n^i, \tag{9}$$

where M^i is a martingale increment.

If we define $X_n = X_n^1 \otimes X_n^2$, where $\otimes$ denotes the tensor, or Kronecker, product, then we can identify X_n with a unit vector in R^N, $N = N(1)N(2)$. From (9)

$$X_n = AX_{n-1} + M_n,$$

where $A = A_1 \otimes A_2$ and

$$M_n = A_1 X_{n-1}^1 \otimes M_n^2 + M_n^1 \otimes A_2 X_{n-1}^2 + M_n^1 \otimes M_n^2,$$

so $\mathrm{E}[M_n | G_{n-1}] = 0$.

Minor modifications to the algebra allow the results of the first part of this section to be applied, so obtaining recursive estimates for X_n (and so g_n and γ_n), and estimates for the parameters of the model.

4. Implementation

The proposed procedure is quite general. The only quantity not estimated here is the size of the state space of the Markov chain. Suppose we choose this to be N.

Then for any price process S_n, $n \in Z^+$, the steps are:
(1) calculate the sequence of logarithmic increments:

$$Y_{n+1} = \log \frac{S_{n+1}}{S_n},$$

(2) initially consider any set of values $\{(g_i, \gamma_i),\ i = 1, \ldots, N\}$,
(3) initially assume the elements of the transfer matrix A have any values (a_{ij}), $1 \le i,\ j \le N$, $\sum_{i=1}^{N} a_{ij} = 1, \quad a_{ij} \ge 0$,
(4) after n values of Y have been observed, calculate new estimates for $(a_{ij}), g$ and γ from Eqs. (6)–(8),
(5) use these values after further observations to re-estimate (a_{ij}), g and γ. The EM algorithm implies the estimates improve monotonically, in the sense that the expected log-likelihood increases with each re-estimation. Consequently, the model is 'self-tuning'.

5. Applications

The results of the paper were applied in two examples using a program that was written to implement the estimation procedure discussed in the paper.

In the first example, the program was run on a data set consisting of 248 monthly observations on the price of IBM stock. The sample period was from June 1975 to January 1996. The prices were processed in eight groups of 31 prices each; at the end of each pass through the data, parameter estimates were updated using the formulas given in the paper. In the analysis, the size, N, of the state space of the Markov chain was taken to be four.

Table 1 gives the initial values that were assumed for the conditional distribution of the state of the Markov chain, that is, for $\mathrm{E}[X_n, Y_n]$, the transition matrix A, and the vectors g and γ. It also gives the re-estimated values of these parameters after the sixth and seventh passes.

To assess the predictive performance of the model, predicted prices for the period February 1991–June 1993 were calculated using the formula:

$$\mathrm{E}[S_{k+1} | Y_k] = S_k \sum_{i=1}^{N} \mathrm{e}^{g_i} \mathrm{e}^{\sigma_i^2/2} \frac{\langle q_k, e_i \rangle}{\langle q_k, 1 \rangle}, \quad k = 1, 2, \ldots .$$

The formula was evaluated using estimated values of the vectors g, γ, and q_k $(k = 1, 2, \ldots)$ after the seventh pass through the data. We then regressed the actual prices for the period on the predicted prices according to the following model:

$$\text{Actual price} = \alpha + \beta^{*}\text{Predicted price} + \varepsilon.$$

Table 1
Parameter estimates, IBM stock prices

Initial values:

$E[X_n/Y_n]$ vector: $(0.25, 0.25, 0.25, 0.25)$

A matrix:

$$\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$$

g vector: $(1.0, 0.0, 1.0, 0.0)$

γ vector: $(0.25, 0.25, 0.25, 0.25)$

After the sixth pass:

$E[X_n, Y_n]$ vector: $(0.1841012, 0.3158988, 0.1841012, 0.3158988)$

A matrix:

$$\begin{bmatrix} 0.1832102 & 0.1846181 & 0.1832102 & 0.1846181 \\ 0.3167898 & 0.3153819 & 0.3167898 & 0.3153819 \\ 0.1832102 & 0.1846181 & 0.1832102 & 0.1846181 \\ 0.3167898 & 0.3153819 & 0.3167898 & 0.3153819 \end{bmatrix}$$

g vector: $(-0.0009364133, -0.0009366155, -0.0009364133, -0.0009366155)$

γ vector: $(5.883248e-07, 5.876560e-07, 5.883248e-07, 5.876560e-07)$

After the seventh pass:

$E[X_n/Y_n]$ vector: $(0.1840997, 0.3159003, 0.1840997, 0.3159003)$

A matrix:

$$\begin{bmatrix} 0.1832102 & 0.1846181 & 0.1832102 & 0.1846181 \\ 0.3167898 & 0.3153819 & 0.3167898 & 0.3153819 \\ 0.1832102 & 0.1846181 & 0.1832102 & 0.1846181 \\ 0.3167898 & 0.3153819 & 0.3167898 & 0.3153819 \end{bmatrix}$$

g vector: $(-0.02572465, -0.02572307, -0.02572465, -0.02572307)$

γ vector: $(0.007575561, 0.007575205, 0.007575561, 0.007575205)$

The regression results obtained were assessed on the basis of the 3 criteria for a good model proposed by Fama and Gibbons (1984): (1) conditional unbiasedness, that is, an intercept, α, close to zero, and a regression coefficient, β, close to one; (2) serially uncorrelated residuals; and (3) a low residual standard error. The second column of Table 2 gives the results for IBM stock. These results

Table 2
Within-sample regressions of actual prices on predicted prices

Parameter	IBM Stock	Gold
α	0.85	70.97
	(5.89)	(49.51)
β	0.99	0.80
	(0.07)	(0.14)
R-squared	0.945	0.783
Durbin–Watson	1.954	1.921
D statistic		
s	7.46	13.89

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

indicate that at the significance level of 0.05, we may conclude that the intercept, α, equals zero (although its standard error is quite large) and that the regression coefficient, β, equals one. Also, on the basis of the Durbin–Watson test, we may conclude that the residuals do not display first-order serial correlation.

In the second example, a data set of 100 monthly observations on the price of gold was analyzed. In this case, the sample period was January 1988 to April 1996. The prices were processed in four groups of 25 prices each. Again, the size, N, of the state space of the Markov chain was taken to be four.

Table 3 gives the initial and re-estimated values of the model parameters.

We again regressed actual prices for the period May 1992–March 1994 on the corresponding predicted prices, according to the model stated earlier. Predicted prices were calculated using the estimated values of the g, γ and q_k vectors after the third pass though the data. The third column of Table 2 gives the regression results for gold. They support the conclusions that the intercept is zero, the slope is one, and residuals are serially uncorrelated.

6. Conclusion

A multiplicative model for the evolution of the price of a risky asset is considered, in discrete time. The increment of the logarithm of the price involves a Gaussian noise and parameters which we suppose evolve like a finite state Markov chain. The estimation techniques of Hidden Markov models are then applied to obtain not only the best estimate of the chain (that is the coefficients), but also re-estimates of all parameters of the model.

Repetition of the estimation procedures ensures that the model and estimates improve with each iteration.

Table 3
Parameter estimates, gold prices

Initial values:

$E[X_n/Y_n]$ vector: (0.25, 0.25, 0.25, 0.25)

A matrix:

$$\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$$

g vector: (1.0, 0.0, 1.0, 0.0)

γ vector: (0.25, 0.25, 0.25, 0.25)

After the second pass:

$E[X_n, Y_n]$ vector: (0.2044194, 0.2955806, 0.2044194, 0.2955806)

A matrix:

$$\begin{bmatrix} 0.2071409 & 0.2015815 & 0.2071409 & 0.2015815 \\ 0.2928591 & 0.2984185 & 0.2928591 & 0.2984185 \\ 0.2071409 & 0.2015815 & 0.2071409 & 0.2015815 \\ 0.2928591 & 0.2984185 & 0.2928591 & 0.2984185 \end{bmatrix}$$

g vector: (−0.006556683, −0.007771985, −0.006556683, −0.007771985)

γ vector: ($5.456238e-05$, $4.970828e-05$, $5.456238e-05$, $4.970828e-05$)

After the third pass:

$E[X_n/Y_n]$ vector: (0.203848, 0.296152, 0.203848, 0.296152)

A matrix:

$$\begin{bmatrix} 0.2071409 & 0.2015815 & 0.2071409 & 0.2015815 \\ 0.2928591 & 0.2984185 & 0.2928591 & 0.2984185 \\ 0.2071409 & 0.2015815 & 0.2071409 & 0.2015815 \\ 0.2928591 & 0.2984185 & 0.2928591 & 0.2984185 \end{bmatrix}$$

g vector: (0.003540311, 0.003544279, 0.003540311, 0.003544279)

γ vector: (0.001359767, 0.001359923, 0.001359767, 0.001359923)

References

Elliott, R.J., 1994. Exact adaptive filters for Markov chains observed in Gaussian noise. Automatica 30, 1399–1408.

Elliott, R.J., Aggoun, L., Moore, J.B., 1994. Hidden Markov models: estimation and control. In: Applications of Mathematics, vol. 29, Springer, New York.

Elliott, R.J., Rishel, R.W., 1994. Estimating the implicit interest rate of a risky asset. Stochastic Processes and Applications 49, 199–206.

Fama, E.F., Gibbons, M.R., 1984. A comparison of inflation forecasts. Journal of Monetary Economics 13, 327–348.