Biometrika (1994), **81**, 1, *pp.* 11–26
Printed in Great Britain

Bayes factors for discrete observations from diffusion processes

BY NICHOLAS G. POLSON

Graduate School of Business, University of Chicago, Chicago, Illinois 60637, U.S.A.

AND GARETH O. ROBERTS

Statistical Laboratory, 16 Mill Lane, Cambridge, CB2 1SB, U.K.

SUMMARY

We present an approach to model selection for a time series of data on a fine time scale. The underlying process generating the data is modelled as a continuous time stochastic process. The underlying continuous processes are assumed to be diffusions with time varying drift and diffusion coefficient. Several approaches to modelling the diffusion coefficient are described. To perform model selection, we propose an approximation to the Bayes factor that uses only the discrete data. We illustrate our approach for several well-known processes including: Brownian motion with drift, the Ornstein–Uhlenbeck process, a mean reversion process with drift, exponential Brownian motion, and a logistic growth model. Finally, we apply our technique to data from the Standard & Poor's 500 stock index by comparing a random walk to a mean reversion model.

Some key words: Bayes factor; Brownian motion; Logistic growth; Mean reversion; Model choice; Ornstein–Uhlenbeck process; Standard & Poor's 500 stock index; Stochastic differential equation; Stochastic volatility.

1. INTRODUCTION

In this paper we provide a technique to evaluate competing continuous time models based on a Bayes factor. We discuss the use of continuous time diffusion processes to model data that arise on a fine time scale. While we focus on diffusion processes, our technique applies to other continuous semi-martingale models. Then we show how to define a Bayes factor between two continuous time processes which can be used to evaluate competing models. Several calculations for commonly-used models are provided to illustrate the Bayes factor. We construct a discrete approximation of the Bayes factor to be used for implementation. Finally, we apply the technique to choose between a random walk and a mean reversion model for daily stock data using Standard & Poor's 500 index.

While time series data are usually modelled as discrete processes, we argue that it is often easier to model the underlying continuous process when observations are relatively frequent. The use of continuous semi-martingale processes to model the underlying process has several advantages over traditional methodologies. It eliminates the need to approximate incremental distributions and so applies directly to both equally and unequally spaced observations. In addition, since semi-martingale processes are uniquely specified by the drift and diffusion coefficients only two coefficients are modelled. These techniques provide a simple and natural approach to modelling fine observations from a continuous time stochastic process.

Our approach to model selection is based on approximating the Bayes factor between competing continuous time models. We show that the Bayes factor can be calculated via

Girsanov's formula. We derive an approximation for the Bayes factor using the discrete observations on a fine time scale. The advantages of using a Bayes factor rather than a classical hypothesis test have been widely discussed; for example, see Berger & Sellke (1987). This paper demonstrates that in continuous time the advantages of our methodology are even more striking. Our Bayes factor is typically easy to compute and does not require asymptotic distributions of functionals of the process. The technique has a number of potential applications in genetics (Ethier & Kurtz, 1986, Ch. 10) and financial modelling (Cox, Ingersoll & Ross, 1985; Duffie, 1992). In this paper we provide an application in finance, choosing between a random walk model and a mean reversion process for daily data from the Standard & Poor's 500 stock index.

Section 2 discusses modelling issues involved in continuous time processes. We consider the family of continuous processes defined by the solution of the stochastic differential equation

$$dX_t = f(\theta, t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \tag{1}$$

where $f(\theta, t, X_t)$ is the drift function, $\sigma(t, X_t)$ is the diffusion coefficient and B_t is a Brownian motion. The drift function depends on the parameter θ which is modelled with a priori density $p(\theta)$. We present a simple approach for determining $\sigma(t, X_t)$ based on the quadratic variation of the fine stream of observations and discuss models that allow $\sigma(t, X_t)$ to evolve according to a stochastic differential equation. Such models are appropriate when the observation interval is too coarse.

Section 3 defines the Bayes factor between continuous time processes. First, we derive an analytical expression when the drift function is linear, $f(\theta, t, X_t) = \theta f(t, X_t) + g(t, X_t)$, with $f(.,.)$ and $g(.,.)$ known. Then an approximation to the continuous time Bayes factor is developed that uses only the discrete data. As the observation interval for the data gets finer our discrete time approximation converges to the continuous time Bayes factor.

Section 4 illustrates our approach with some well-known processes. Bayes factors, marginal beliefs and posterior distributions are derived for the following models: a Brownian motion with drift, an Ornstein–Uhlenbeck process, a mean reversion process with drift, an exponential Brownian motion, and a logistic growth model.

Section 5 applies our technique to daily data from the Standard & Poor's 500 stock index for the period 1927–1985. We use our model selection technique to evaluate a random walk model against a Markovian mean reversion model. Under a constant diffusion coefficient our results are quantitatively similar to the mean reversion studies of Fama & French (1988) and Poterba & Summers (1988). Our methodology, however, allows us to assess the probability of mean reversion and diagnose heteroscedasticities in the diffusion coefficient. In doing so, we find evidence of changing volatility in the process. When we allow for this effect the evidence for mean reversion is diminished to the extent that neither model is preferred.

2. THE MODEL AND DISCRETE APPROXIMATION

2·1. *General*

Consider a continuous time diffusion process, X_t, that evolves according to the stochastic differential equation

$$dX_t = f(\theta, t, X_t)dt + \sigma(t, X_t)dB_t, \tag{2}$$

where $f(\theta, t, X_t)$ is a drift function depending on the parameter θ, and $\sigma(t, X_t)$ is a diffusion coefficient. We model θ with a priori density $p(\theta)$. Therefore $f(\theta, t, X_t)$ represents the systematic rate of change of the process at time t and $\sigma(t, X_t)$ is an instantaneous measurement of volatility. This continuous time semi-martingale model can be viewed as an approximation to the discrete time data generating process. The martingale term $\sigma(t, X_t)\,dB_t$ provides an approximation to the 'noise' term in the discrete time model. Therefore, there is no need explicitly to model the distribution of the error term in the discrete model.

We observe data from this process on a fine time scale. Let $\{X_{t_i}, i \in I\}$ denote the discrete observation set on the process where the times are ordered at equally or unequally spaced points on the interval $[0, T]$. For the moment, suppose that we observe the full information set $X^T = \{X_t, 0 \leqslant t \leqslant T\}$. Let $P_{X^T|\theta}$ be the law of $X^T|\theta$ which evolves according to (2) and P_M be the measure induced by the martingale $dX_t = \sigma(t, X_t)\,dB_t$. The likelihood function, $L(\theta)$, is then given by Girsanov's formula (Oksendal, 1985, p. 126):

$$L(\theta) = \frac{dP_{X^T|\theta}}{dP_M}(X^T) = \exp\left[\int_0^T \frac{f(\theta, t, X_t)}{\sigma^2(t, X_t)}\,dX_t - \frac{1}{2}\int_0^T \left\{\frac{f(\theta, t, X_t)}{\sigma(t, X_t)}\right\}^2 dt\right], \tag{3}$$

$$\log L(\theta) = \int_0^T \frac{f(\theta, t, X_t)}{\sigma^2(t, X_t)}\,dX_t - \frac{1}{2}\int_0^T \left\{\frac{f(\theta, t, X_t)}{\sigma(t, X_t)}\right\}^2 dt. \tag{4}$$

For the log-likelihood to be well defined we need to impose regularity conditions on the drift and diffusion functions to ensure that, for each $T \geqslant 0$, all the $P_{X^T|\theta}$'s are absolutely continuous with respect to each other. Oksendal (1985) provides regularity conditions. For example, it is sufficient to assume that X_t is an Itô diffusion.

It is well known that, for all $t \in [0, T]$,

$$\lim_{\delta \to 0} \sum_{t_{j+1} \leqslant t} (X_{t_{j+1}} - X_{t_j})^2 = \int_0^t \sigma^2(s, X_s)\,ds \tag{5}$$

almost surely,

$$\operatorname{var}\left\{\sum_{t_{j+1} \leqslant t} (X_{t_{j+1}} - X_{t_j})^2 - \int_0^t \sigma^2(s, X_s)\,ds\right\} = O\{\delta(P^{(i)})\} \tag{6}$$

for any nonexplosive diffusion process X_t with a locally bounded continuous diffusion coefficient and sequence of partitions

$$P^{(i)} = \{0 = t_0^{(i)} < t_1^{(i)} < \ldots < t_{n_i}^{(i)} = T\},$$

where $\delta(P^{(i)}) = \sup_j(t_{j+1}^{(i)} - t_j^{(i)}) \to 0$ as $i \to \infty$.

From a modelling perspective, this suggests that the diffusion coefficient can be estimated very precisely if the time interval is 'fine' enough for (5) and (6) to hold. Thus, when $\sigma(t, X_t)$ is estimated precisely, we can use a quadrature based approximation to the log-likelihood,

$$\begin{aligned}\log L(\theta) \simeq & \sum_{i \in I, t_i < T} (X_{t_i} - X_{t_{i-1}}) \frac{f(\theta, t_{i-1}, X_{t_{i-1}})}{\sigma^2(t_{i-1}, X_{t_{i-1}})} \\ & - \frac{1}{2} \sum_{i \in I, t_i < T} (t_i - t_{i-1}) \left\{\frac{f(\theta, t_{i-1}, X_{t_{i-1}})}{\sigma(t_{i-1}, X_{t_{i-1}})}\right\}^2,\end{aligned} \tag{7}$$

where $t_0 = 0$ and X_0 is the starting point. This provides the basis for our Bayes factor approximation.

2·2. *Modelling the diffusion coefficient*

We now discuss the problem of estimating the diffusion coefficient, $\sigma(t, X_t)$. First, when the observations are on a fine time scale, we propose a simple estimate of the diffusion coefficient based on the quadratic variation process. The log-likelihood is computed by 'plugging in' this estimate into equation (7). Our analysis then proceeds conditional on this estimate. In § 3 we develop a discrete approximation to the continuous time Bayes factor with an error that converges to zero as the grid size diminishes. Secondly, when the observation interval is too coarse, we model the evolution of the diffusion coefficient explicitly. This approach is appropriate since coarse observations make the quadratic variation estimate unreliable. The log-likelihood, however, is not in closed form and we must approximate the likelihood by a discrete time nonlinear filter or by a Monte Carlo approach. Model comparison and inference can then be based on standard discrete time Bayesian methodology.

Suppose that the time scale of the observed process is 'fine' enough for the quadrature estimate (7) to reasonably approximate (4). Then (5) indicates that the information concerning the diffusion coefficient $\sigma(t, X_t)$ is contained in the process

$$Q(t) = \sum_{i \in I, t_i < t} (X_{t_i} - X_{t_{i-1}})^2.$$

This is a discretised version of the quadratic variation process, $d[X]_t$. If the diffusion coefficient is continuous in both arguments, equation (5) ensures that

$$\lim_{\varepsilon \to 0} \lim_{\delta \to 0} \frac{Q(t+\varepsilon) - Q(t-\varepsilon)}{2\varepsilon} \to \sigma^2(t, X_t) \tag{8}$$

almost surely. Therefore, in the limit as the grid size diminishes, a continuous stream of observations will identify the diffusion coefficient with probability one.

This result suggests a technique to estimate the diffusion coefficient and diagnose heteroscedasticity. For a constant diffusion, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i \in I, t_i < T} (X_{t_i} - X_{t_{i-1}})^2. \tag{9}$$

Corrections can be employed to this estimate if it is 'corrupted' by drift. For example, if the drift is $f(t, X_t)$ and the volatility is a constant, then the following estimate is more realistic

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i \in I, t_i < T} \{X_{t_i} - X_{t_{i-1}} - (t_i - t_{i-1}) f(t_{i-1}, X_{t_{i-1}})\}^2.$$

Equation (8) suggests that we can diagnose departures from the constant diffusion assumption by plotting $Q(t)$ versus t and looking for changes in the slope. Another useful diagnostic plot comes from the local estimate of a time varying diffusion coefficient $\sigma^2(t_i)$,

$$\hat{\sigma}^2(t_i) = \frac{(X_{t_i} - X_{t_{i-1}})^2}{t_i - t_{i-1}}.$$

A plot of these individual estimates versus time can identify periods of high or low volatility in the process.

The estimate (9) can be thought of as an approximate posterior mode for a discrete process with a prior on σ. Consider the discrete time model with a random diffusion coefficient where

$$X_{t_i} \mid X_{t_{i-1}}, \sigma \sim N\{X_{t_{i-1}} + hf(t_{i-1}, X_{t_{i-1}}), \sigma^2 h\}$$

with X_0 and $f(.,.)$ given. We assume that the observations are equally spaced at time intervals of size h so that $t_i = hi$ and $t_i - t_{i-1} = h$. Clearly this is a discretisation of the continuous time process $dX_t = f(t, X_t)\,dt + \sigma\,dB_t$ as B_t is an independent increments process with $B_{t+h} - B_t \sim N(0, \sigma^2 h)$. The volatility per unit time interval is modelled with an inverse gamma prior $\sigma^2 \sim \mathrm{IG}(a_0, b_0)$ with known hyperparameters. The posterior distribution for the volatility degenerates as the grid size diminishes to the estimate given by the quadratic variation process. To show this, consider the log-likelihood function with respect to Lebesgue measure,

$$\log L(\sigma) = -\frac{1}{2\sigma^2 h} \sum_{i=0}^{T/h} \{X_{t_i} - X_{t_{i-1}} - hf(t_{i-1}, X_{t_{i-1}})\}^2 - \frac{T}{h} \log \sigma.$$

The posterior distribution for σ^2 is then inverse gamma with parameters

$$a_1 = a_0 + \frac{T}{2h}, \quad b_1 = b_0 + \frac{1}{2h} \sum_i \{X_{t_i} - X_{t_{i-1}} - hf(t_{i-1}, X_{t_{i-1}})\}^2.$$

Hence,

$$E(\sigma^2 \mid \text{data}) = \frac{\sum_i \{X_{t_i} - X_{t_{i-1}} - hf(t_{i-1}, X_{t_{i-1}})\}^2 + 2hb_0}{T + 2ha_0}$$

and it is straightforward to check that $\mathrm{var}(\sigma^2 \mid \text{data}) = O(h)$. Therefore, as the grid size diminishes $h \to 0$, the posterior distribution converges to a point mass at the quadratic variation estimate $Q(T)/T$.

We now consider situations where the grid size of the observed process is too coarse for quadrature to be a reliable approximation to the likelihood. In this case, we model the stochastic evolution of the diffusion coefficient. Our approach can be viewed as modelling the experimenter's beliefs about the evolution of the diffusion coefficient within the discrete observation period.

In discrete time models, a common approach is to describe the volatilities deterministically according to a stochastic difference equation. Recent literature has focused on models for equally spaced observations where the increments of the process $y_t = \sigma_t \varepsilon_t$ have a sequence of volatilities $\{\sigma_t\}$ that evolve as a function of past volatilities and residuals. Hence, given the data sequence, the volatilities update in a deterministic manner. Statistical inference is then based on the discrete time likelihood function which can be recursively computed using a nonlinear filter. In financial applications it is natural to model the logarithm of volatility with a mean reverting component; for example, consider the volatility filter

$$\log \sigma_{t+1}^2 = \beta - \alpha \log \sigma_t^2 + \gamma \varepsilon_t - \delta |\varepsilon_t|. \tag{10}$$

Here $\omega = (\alpha, \beta, \gamma, \delta)$ is a parameter vector and σ_0 is given. Under a normal error structure, inference for ω can be performed using the posterior $p(\omega \mid y) \propto L(\omega) p(\omega)$, where

$$\log L(\omega) = -\frac{1}{2} \sum_t \left(\frac{y_t}{\sigma_t}\right)^2 - \sum_t \log \sigma_t$$

and σ_t evolves according to (10). The relationship between these classes of discrete time models and continuous time models is complex. Nelson (1990) shows how to scale such models appropriately in order to ensure a continuous time limit. Unfortunately, scaling depends on the grid size and the assessment of prior distributions for the parameters of the volatility equation is prohibitive as they have a grid size interpretation.

A natural alternative is to model the evolution of $\sigma(t, X_t)$ in continuous time as a stochastic differential equation. The stochastic terms driving the volatility can depend on the observed process via dX_t and $d[X]_t$ terms or through a new Brownian motion $d\tilde{B}_t$. For example, a diffusion model of the form

$$d\sigma_t^2 = (\beta - \alpha\sigma_t^2)\,dt + \gamma\sigma_t^2\,d\tilde{B}_t \tag{11}$$

might be an appropriate model for the evolution of volatilities in financial applications. This diffusion process prior is a continuous time counterpart to the standard conjugate prior analysis of the normal linear model as the stationary distribution for σ_t^2 is inverse gamma,

$$\sigma_t^{-2} \sim \Gamma\left(1 + \frac{2\alpha}{\gamma^2}, \frac{2\beta}{\gamma^2}\right)$$

if $2\alpha/\gamma^2 > -1$ and $\beta > 0$.

It is not possible to analyse these models using a continuous time Bayes factor due to the limiting behaviour of the quadratic variation process. Model comparison and inference has to be performed by viewing the model in discrete time as a hierarchical model where the unobserved parameter vector $\{\sigma_t\}$ has to be marginalised out to compute the likelihood function for the parameters of the volatility equation.

These two approaches can be thought of as a way of avoiding the aliasing problem. Since it is possible to have processes with different diffusion coefficients generating the same discrete sample paths of observations, model comparison is ill posed unless either we restrict ourselves to processes with the same diffusion coefficient or we explicitly model the evolution of the diffusion coefficient. Finally, we emphasize that the problem of estimating the diffusion coefficient disappears when the grid size tends to zero as demonstrated by equation (8). Our approach takes advantage of the continuous structure of the problem, hence we assume hereafter that observations are on a fine grid. Therefore, we can use the quadratic variation process to determine the diffusion coefficient. The experimenter is left to model the appropriate drift behaviour. The Bayes factor provides a technique for evaluating competing models. Drawing inferences for the drift parameter and performing model comparisons using our technique are typically straightforward for a wide class of models as we will see in the following sections. We now show how to define and implement the Bayes factor.

3. BAYES FACTOR FOR CONTINUOUS TIME MODELS

The marginal beliefs for the process, $P_{X^T}(.)$, are defined in terms of the likelihood function $P_{X^T|\theta}(.)$, as follows:

$$P_{X^T}(A) = \int P_{X^T|\theta}(A)p(\theta)\,d\theta \tag{12}$$

for all events $A \subset \mathscr{F}_T$, where $\mathscr{F}_T$ is the σ-field generated by X^T. We write $P_{X^T,\theta}(B)$ for possibly θ-dependent sets $B \subset \mathscr{F}_T \times \Theta$. We denote the marginal beliefs under the model

$\mathcal{M}_i$ by $P_{X^T}(A \mid \mathcal{M}_i)$ for $A \subset \mathcal{F}_T$. Let $p(\theta \mid \mathcal{M}_i)$ indicate the possibility of different prior distributions for the drift parameter in each model.

The Bayes factor (Jeffreys, 1961) between the models $\mathcal{M}_0$ and $\mathcal{M}_1$, denoted by $\mathrm{BF}(X^T)$, is defined by the ratio of marginal beliefs

$$\mathrm{BF}(X^T) = \frac{P_{X^T}(X^T \mid \mathcal{M}_0)}{P_{X^T}(X^T \mid \mathcal{M}_1)}. \tag{13}$$

The dominating measure for $\mathcal{M}_i$ is given by $P_{\mathcal{M}_i}$ which is induced by $dX_t = \sigma_i(t, X_t)\, dB_t$. Hence, we need identical diffusion coefficients for the Bayes factor to be well defined. A discrete time Bayes factor that compares models with different diffusion coefficients will degenerate as the grid size diminishes because the data will select the model having the 'correct' diffusion coefficient with posterior probability one.

We now compute the Bayes factor when the drift functions are of the form $\theta f_i(t, X_t) + g_i(t, X_t)$. This proves the basis for the examples in §§ 4 and 5.

THEOREM 1. *Consider the class of diffusions for model $\mathcal{M}_i$ that evolve according to*

$$dX_t = \{\theta f_i(t, X_t) + g_i(t, X_t)\}\, dt + \sigma_i(t, X_t)\, dB_t$$

and $\theta \mid \mathcal{M}_i \sim N(\mu_i, \tau_i^2)$ for known hyperparameters μ_i, τ_i. Then, the marginal beliefs $P_{X^T}(X^T \mid \mathcal{M}_i)$ are given by

$$2 \log P_{X^T}(X^T \mid \mathcal{M}_i) = \frac{1}{1 + \tau_i^2 A_i} (\tau_i^2 B_i^2 + 2\mu_i B_i - \mu_i^2 A_i) + 2C_i - \log(\tau_i^2 A_i + 1), \tag{14}$$

where

$$A_i = \int_0^T \left\{\frac{f_i(t, X_t)}{\sigma_i(t, X_t)}\right\}^2 dt, \quad B_i = \int_0^T \frac{f_i(t, X_t)}{\sigma_i^2(t, X_t)}\, dX_t - \int_0^T \frac{f_i(t, X_t) g_i(t, X_t)}{\sigma_i^2(t, X_t)}\, dt,$$

$$C_i = \int_0^T \frac{g_i(t, X_t)}{\sigma_i^2(t, X_t)}\, dX_t - \frac{1}{2} \int_0^T \left\{\frac{g_i(t, X_t)}{\sigma_i(t, X_t)}\right\}^2 dt. \tag{15}$$

Proof. The log-likelihood for model $\mathcal{M}_i$ is given by (4),

$$\log L_T(\theta \mid \mathcal{M}_i) = \int_0^T \frac{\theta f_i(t, X_t) + g_i(t, X_t)}{\sigma_i^2(t, X_t)}\, dX_t - \frac{1}{2} \int_0^T \left\{\frac{\theta f_i(t, X_t) + g_i(t, X_t)}{\sigma_i(t, X_t)}\right\}^2 dt. \tag{16}$$

The marginal beliefs $P_{X^T}(X^T \mid \mathcal{M}_i) = \int L_T(\theta \mid \mathcal{M}_i) p(\theta \mid \mathcal{M}_i)\, d\theta$ can be computed as follows: we can write

$$P_{X^T}(X^T \mid \mathcal{M}_i) = \frac{\exp(C_i)}{(2\pi)^{\frac{1}{2}} \tau_i} \int_{\mathbb{R}} \exp\left[-\frac{1}{2\tau_i^2} \{\theta^2 \tau_i^2 A_i - 2\tau_i^2 \theta B_i + (\theta - \mu_i)^2\}\right] d\theta$$

using (16), the normality of the prior, and the definitions of (A_i, B_i, C_i). After completing the square, integrating and some algebraic manipulation, we have

$$2 \log P_{X^T}(X^T \mid \mathcal{M}_i) = \frac{1}{1 + \tau_i^2 A_i} (\tau_i^2 B_i^2 + 2\mu_i B_i - \mu_i^2 A_i) + 2C_i - \log(\tau_i^2 A_i + 1), \tag{17}$$

as required. □

A special case of interest occurs when $g_i(t, X_t) = 0$ and $\mu_i = 0$. In this case, the processes evolve according to $dX_t = \theta f_i(t, X_t)\,dt + \sigma_i(t, X_t)\,dB_t$ and a priori $\theta \sim N(0, \tau_i^2)$. Theorem 1 shows that the marginal beliefs are given by

$$2 \log P_{X^T}(X^T | \mathcal{M}_i) = \frac{\tau_i^2 \tilde{B}_i^2}{1 + \tau_i^2 \tilde{A}_i} - \log(\tau_i^2 \tilde{A}_i + 1), \tag{18}$$

where $\tilde{A}_i = A_i$ and

$$\tilde{B}_i = \int_0^T \frac{f_i(t, X_t)}{\sigma_i^2(t, X_t)}\, dX_t.$$

We use a quadrature based approximation to the stochastic integrals that cannot be evaluated explicitly using Itô's formula (Oksendal, 1985, p. 32) to derive an approximation to the Bayes factor using only the discrete sequence of data. For example, if $\tilde{B}_i$ is not analytically available, we use an approximation of the form

$$\hat{B}_i \simeq \sum_{t_i} \frac{f_i(t, X_{t_{i-1}})}{\sigma_i^2(t, X_{t_{i-1}})} (X_{t_i} - X_{t_{i-1}}).$$

4. SOME SIMPLE MODELS

4·1. *General*

We now apply our methodology to several well-known continuous time Markov processes. We consider the following diffusion processes: a Brownian motion with drift, an Ornstein–Uhlenbeck process, a mean reverting process with drift, an exponential Brownian motion, and a logistic growth model. For the purpose of illustration, we assume that the diffusion coefficient is constant and $\theta \sim N(\mu, \tau^2)$ for known hyperparameters.

4·2. *Brownian motion with drift* ($\mathcal{M}_1$)

Consider the process X_t that is the solution of $dX_t = \theta\, dt + \sigma\, dB_t$ with log likelihood

$$\log L(\theta | \mathcal{M}_1) = -\frac{1}{2\sigma^2}\{\theta^2 T + \theta(X_T - X_0)\}.$$

The marginal law as given by Theorem 1 is

$$2 \log P_{X^T}(X^T | \mathcal{M}_i) = \frac{\sigma^{-2}(X_T - X_0)^2 + 2\mu\tau^{-2}(X_T - X_0) - \mu^2\tau^{-2}T}{T + \sigma^2\tau^{-2}} - \log\left(1 + \frac{\tau^2 T}{\sigma^2}\right) \tag{19}$$

under the prior $\theta \sim N(\mu, \tau^2)$. The posterior distribution can be calculated via Bayes theorem as

$$\theta | X^T \sim N\left(\frac{X_T - X_0 + \mu\sigma^2\tau^{-2}}{T + \sigma^2\tau^{-2}}, \frac{\sigma^2}{T + \sigma^2\tau^{-2}}\right). \tag{20}$$

Our procedure provides an alternative approach to a classical hypothesis test of $\mathcal{M}_0 : \theta = 0$ against $\mathcal{M}_1$, that is a process with drift versus no drift. The Bayes factor can be computed by noting that the law of $\mathcal{M}_0$, $dX_t = \sigma\, dB_t$, is the dominating measure for (19). Hence,

$$2 \log \mathrm{BF}(X^T) = \frac{\sigma^{-2}(X_T - X_0)^2 + 2\mu\tau^{-2}(X_T - X_0) - \mu^2\tau^{-2}T}{T + \sigma^2\tau^{-2}} - \log\left(1 + \frac{\tau^2 T}{\sigma^2}\right).$$

The posterior odds of $\mathcal{M}_0$ versus $\mathcal{M}_1$ is given by

$$\frac{P(\mathcal{M}_0|X^T)}{P(\mathcal{M}_1|X^T)} = (1+\tau^2\sigma^{-2}T)^{-\frac{1}{2}} \exp\left\{\frac{\sigma^{-2}(X_T-X_0)^2 + 2\mu\tau^{-2}(X_T-X_0) - \mu^2\tau^{-2}T}{2(T+\sigma^2\tau^{-2})}\right\} \frac{P(\mathcal{M}_0)}{P(\mathcal{M}_1)},$$

where $P(\mathcal{M}_0)$ is the a priori probability of model $\mathcal{M}_0$.

4·3. *Mean reversion with drift* ($\mathcal{M}_2$)

Consider a mean reversion model with drift where X_t evolves according to

$$dX_t = \{\theta - \alpha(X_t - \theta t)\}\, dt + \sigma(t, X_t)\, dB_t$$

and $\alpha > 0$. The component $-\alpha(X_t - \theta t)$ in the drift function implies that the process is 'forced' to revert to its mean. To compute the marginal beliefs, we rewrite the model as

$$dX_t = \{(1+\alpha t)\theta - \alpha X_t\}\, dt + \sigma(t, X_t)\, dB_t$$

and apply Theorem 1 with $f(t, X_t) = (1+\alpha t)$ and $g(t, X_t) = -\alpha X_t$. This yields marginal beliefs

$$2\log P_{X^T}(X^T|\mathcal{M}_2) = \frac{\tau^2 B_2^2}{1+\tau^2 A_2} + 2C_2 - \log(1+\tau^2 A_2),$$

where

$$A_2 = \int_0^T \left\{\frac{1+\alpha t}{\sigma(t, X_t)}\right\}^2 dt, \quad B_2 = \int_0^T \frac{1+\alpha t}{\sigma^2(t, X_t)}\, dX_t + \alpha \int_0^T \frac{(1+\alpha t)X_t}{\sigma^2(t, X_t)}\, dt,$$

$$C_2 = -\alpha \int_0^T \frac{X_t}{\sigma^2(t, X_t)}\, dX_t - \frac{\alpha^2}{2}\int_0^T \left\{\frac{X_t}{\sigma(t, X_t)}\right\}^2 dt.$$

The correlation structure for this process with a constant diffusion coefficient takes the form

$$\operatorname{corr}(X_s, X_t) = \exp\left\{-\left(\frac{\alpha}{\sigma}\right)|s-t|\right\}$$

for each s, t. As with all one-dimensional Markovian diffusion processes the correlations are all positive. The incremental process $Z_t^\varepsilon = X_{t+\varepsilon} - X_t$ has a negative correlation structure for $s+\varepsilon \leqslant t$, given by

$$\operatorname{corr}(Z_t^\varepsilon, Z_s^\varepsilon) = -\rho_{t-s-\varepsilon}(1-\rho_\varepsilon)^2,$$

where $\rho_{s-t} = \operatorname{corr}(X_t, X_s) = \exp\{-(\alpha/\sigma)|s-t|\}$. We use this result for the example in § 5.

4·4. *Ornstein–Uhlenbeck process* ($\mathcal{M}_3$)

The Ornstein–Uhlenbeck process is the solution of $dX_t = \theta X_t\, dt + \sigma\, dB_t$ with log-likelihood

$$\log L_T(\theta|\mathcal{M}_3) = \frac{\theta}{\sigma^2}\int_0^T X_t\, dX_t - \frac{\theta^2}{2\sigma^2}\int_0^T X_t^2\, dt,$$

where

$$\int_0^T X_t\, dX_t = \frac{1}{2}(X_T^2 - X_0^2 - \sigma^2 T).$$

The posterior distribution of the drift parameter, under a normal prior $\theta \sim N(\mu, \tau^2)$, is given by

$$\theta | X^T \sim N\left(\frac{\int_0^T X_t\, dX_t + \mu\sigma^2\tau^{-2}}{\int_0^T X_t^2\, dt + \sigma^2\tau^{-2}}, \frac{\sigma^2}{\int_0^T X_t^2\, dt + \sigma^2\tau^{-2}}\right).$$

The marginal beliefs can be derived from Theorem 1. One model for comparison is a mean reverting Brownian motion given by the solution of $dX_t = \theta \operatorname{sgn}(X_t)\, dt + \sigma\, dB_t$.

4·5. *Exponential Brownian motion* ($\mathcal{M}_4$)

Consider the process X_t that evolves to $dX_t = X_t(\theta\, dt + \sigma\, dB_t)$. The log-likelihood is given by

$$\log L(\theta | \mathcal{M}_4) = \frac{\theta}{\sigma^2} \int_0^T \frac{dX_t}{X_t} - \frac{\theta^2 T}{2\sigma^2}, \tag{21}$$

where

$$\int_0^T \frac{dX_t}{X_t} = \log X_T - \log X_0 + \frac{1}{2}\sigma^2 T.$$

Theorem 1 gives marginal beliefs

$$2 \log P_{X^T}(X^T | \mathcal{M}_4) = \frac{\tau^2}{\sigma^2 + \tau^2} \int_0^T \frac{dX_t}{X_t} - \log\left(1 + \frac{\tau^2 T}{\sigma^2}\right). \tag{22}$$

The posterior for the drift parameter is given by

$$\theta | X^T \sim N\left(\frac{\int_0^T dX_t/X_t + \mu\sigma^2\tau^{-2}}{T + \sigma^2\tau^{-2}}, \frac{\sigma^2}{T + \sigma^2\tau^{-2}}\right).$$

This model generalises to the logistic growth model considered next.

4·6. *Logistic growth model* ($\mathcal{M}_5$)

The logistic growth model, commonly used in population genetics, is the solution to $dX_t = X_t\{\theta(1 - \varepsilon X_t)\, dt + \sigma\, dB_t\}$. When $\varepsilon = 0$ this reduces to exponential Brownian motion. The marginal beliefs are given by

$$2 \log P_{X^T}(X^T | \mathcal{M}_5) = \frac{\tau^2 B_5^2}{1 + \tau^2 A_5} - \log(1 + \tau^2 A_5), \tag{23}$$

where

$$\sigma^2 A_5 = \frac{1}{\sigma^2} \int_0^T (1 - \varepsilon X_t)^2\, dt, \quad \sigma^2 B_5 = \int_0^T \frac{dX_t}{X_t} - \varepsilon(X_T - X_0).$$

Here we have taken a normal prior with mean zero. The Bayes factor between $\mathcal{M}_4$ and $\mathcal{M}_5$ can then be computed from (22) and (23).

The diffusion models considered above are all one-dimensional Markovian models. Therefore, they only allow a positive correlation structure for the data. To model a negative correlation structure a non-Markovian model is required, such as allowing the drift to depend on the full history of X_t. Our approach generalises straightforwardly to these cases.

5. Mean reversion in the Standard & Poor's 500 index

As an illustration of our technique, we choose between a Brownian motion with drift and a Markovian mean reversion process for describing daily stock returns. Our analysis employs daily data from the Standard & Poor's 500 index for the period 1927–1985. Hence, our technique can be used to assess departures from an efficient market hypothesis where the alternative is a mean reversion process. Mean reversion could result from market and fundamental values diverging and then eventually converging as they 'correct' themselves, or mean reversion might be the rational outcome of time-varying equilibria. To explore the evidence for mean reversion we calculate the Bayes factor, determine the conditional probability of mean reversion and perform diagnostics to determine heteroscedasticity in the diffusion coefficient.

We model the logarithm of the price index, $X_t = \log P_t$, as evolving according to either

$$\mathcal{M}_1 : dX_t = \theta\, dt + \sigma(t, X_t)\, dB_t$$

or

$$\mathcal{M}_2 : dX_t = \{\theta - \alpha(X_t - \theta t)\}\, dt + \sigma(t, X_t)\, dB_t.$$

Here $\alpha > 0$ and the term $-\alpha(X_t - \theta t)\, dt$ corresponds to mean reversion in the drift. To complete our model specification we assume that $\theta \sim N(0, \tau^2)$. Each of these models is a simplification of possible continuous time processes generating the data; however, our technique is straightforward to extend to mean reversion models with non-Markovian drift. The two processes have the same diffusion coefficient so our Bayes factor is well defined.

To compute the Bayes factor we need three calculations: the marginal beliefs for the two models, approximations for these beliefs based on the discrete data and an estimate of the volatility. We initially assume that $\sigma(t, X_t) = \sigma$ is constant over the period of interest. See Merton (1980) and Poterba & Summers (1986) for discussion of this assumption. Later, we relax this assumption on the basis of diagnostics indicating heteroscedasticity using techniques described in § 2.

The random walk model has marginal beliefs given by

$$2 \log P_{X^T}(X^T | \mathcal{M}_1) = \frac{(X_T - X_0)^2}{\sigma^2(T + \sigma^2 \tau^{-2})} - \log\left(1 + \frac{\tau^2 T}{\sigma^2}\right). \tag{24}$$

For the mean reversion model the marginal beliefs are given by (18) as

$$2 \log P_{X^T}(X^T | \mathcal{M}_2) = \frac{\tau^2 B_2^2}{1 + \tau^2 A_2} + 2C_2 - \log(1 + \tau^2 A_2), \tag{25}$$

where

$$\sigma^2 A_2 = T\left(1 + \alpha T + \frac{\alpha^2 T^2}{3}\right),$$

$$\sigma^2 B_2 = X_T - X_0 + \alpha \int_0^T t\, dX_t + \alpha \int_0^T (1 + \alpha t) X_t\, dt,$$

$$\sigma^2 C_2 = -\frac{\alpha}{2}(X_T^2 - X_0^2 - \sigma^2 T) - \frac{\alpha^2}{2} \int_0^T X_0^2\, dt.$$

Next, we use quadrature to approximate the stochastic integrals with only the discrete observation set $\{X_{t_i}\}$. For example, we use the approximations:

$$\int_0^T (1+\alpha t)\,dX_t \simeq \sum_{t_i} (1+\alpha t_{i-1})(X_{t_i} - X_{t_{i-1}}),$$

$$\int_0^T X_t^2\,dt \simeq \sum_{t_i} X_{t_{i-1}}^2 (t_i - t_{i-1}).$$

Then, we estimate the diffusion coefficient, σ, using

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{i \in I, t_i < T} (X_{t_i} - X_{t_{i-1}})^2.$$

We diagnose departures from the assumption of constant volatility by plotting the quadratic variation process and looking for changes in the slope of this process. Finally, we take $\tau^2 = 1$ throughout our analysis. This seems reasonable a priori and our results are very insensitive to changes in this parameter due to the amount of data.

Using these approximations, we compute the Bayes factor, $\mathrm{BF}(X^T)$, between the competing models. If the Bayes factor is greater than one then there is evidence in the data against the model $\mathcal{M}_1$ in favour of the model $\mathcal{M}_2$. See Jeffreys (1961, Appendix B) for further discussion of a scale on which to grade the Bayes factor. We can also assess a probability of mean reversion. Suppose that we specify a priori odds $O = P(\mathcal{M}_2)/P(\mathcal{M}_1)$ of mean reversion versus a random walk model. Typically $O = 1$. In light of evidence in the data, these odds are updated by the Bayes factor to posterior odds. An assessment of the probability of mean reversion is given by

$$P(\mathcal{M}_2 \mid \text{data}) = \frac{\mathrm{BF}}{O + \mathrm{BF}}. \tag{26}$$

Now we have a simple interpretation of the Bayes factor on a probability scale.

Figure 1 plots the daily Standard & Poor's 500 index for the period 1927–1985. Figure 2 plots the logarithm of prices. The Bayes factor and conditional probability of mean reversion are computed using the approximations described above. Figure 3 plots the probability of mean reversion $P(\mathcal{M}_2 \mid \text{data})$ versus α with a priori odds $O = 1$. We see that there is evidence in the data for a mean reversion parameter in the range $\alpha \in (0, 0{\cdot}0005)$. For the purpose of illustration, suppose that we were to a priori compare the model $\mathcal{M}_2 : \alpha = 0{\cdot}0002$ to a random walk model $\mathcal{M}_1 : \alpha = 0$. In this case, $\mathrm{BF} = 5{\cdot}2$ and the probability of mean reversion is 0·84.

Mean reversion implies that the autocorrelations of long horizon returns are negative. For $\alpha = 0{\cdot}0002$ and $\sigma = 0{\cdot}0114$, the logarithm of returns $Z_t = \log (P_{t_i}/P_{t_{i-1}})$ has correlation structure

$$\operatorname{corr}(Z_s, Z_t) = -\frac{1}{2}(1 - \rho_{s-t})^2,$$

where $\rho_{s-t} = \exp\{-(\alpha/\sigma)|s-t|\}$. Under these parameter values, one and two year returns have a negative correlation of $-0{\cdot}006$ and $-0{\cdot}020$ which compare with sample correlations of $-0{\cdot}011$ and $-0{\cdot}018$. Hence, this sample Markovian mean reversion model provides a plausible explanation of the negative sample correlations of long horizon returns.

A purely subjectivist analysis would model the mean reversion parameter α with a prior

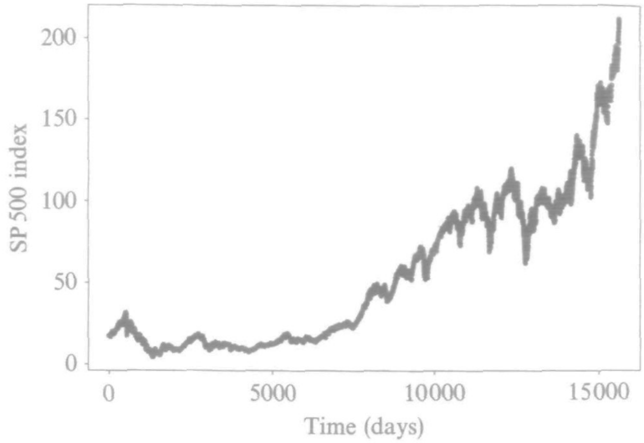

Fig. 1. Standard & Poor's 500 index 1927–1985.

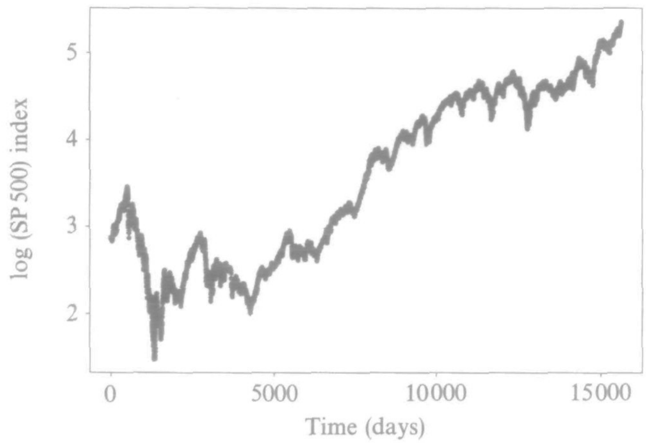

Fig. 2. Logarithm of Standard & Poor's 500 index for 1927–1985.

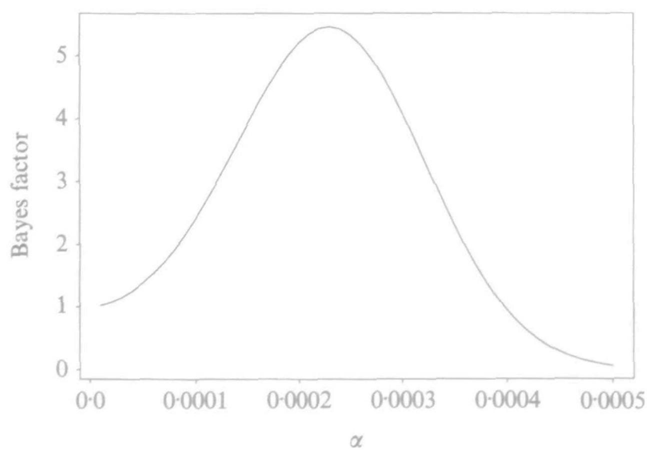

Fig. 3. Bayes factor for mean reversion.

measure $p(\alpha)$. The Bayes factor, as a function of α, is a marginal likelihood and can be used to re-weight the prior and give a posterior density $p(\alpha|\text{data})$. Rather than using a one-number summary, namely the Bayes factor, the experimenter can use the posterior to draw inferences about the magnitude of the mean reversion. In this framework, evidence for mean reversion will generally be diminished mainly due to the marginalisation over α in computing the Bayes factor. We do not explore this direction further. Instead, we demonstrate the presence of heteroscedasticity and if we model this, the evidence for mean reversion is greatly diminished.

Figure 4 plots the quadratic variation process for the period. Therefore, it is reasonable to assume heteroscedasticity in the diffusion coefficient as there is a change in the slope of the process. We consider a model where a period of high volatility is followed by a lower one with a changepoint around 1951. This is plausible from the quadratic variation plot and from the data in Fig. 2. Using piecewise estimates from the quadratic variation process, we obtain $\hat{\sigma}_1 = 0{\cdot}0144$ for the period 1927–1951 and $\hat{\sigma}_2 = 0{\cdot}0077$ for the period 1951–1985. Quantitatively, our results are robust to the changepoint in the window 1947–1955.

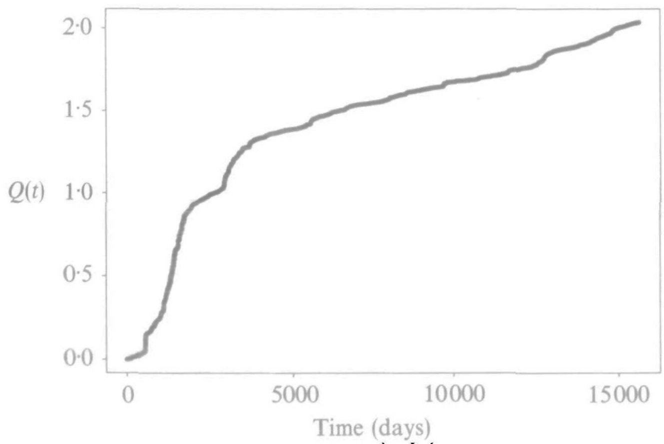

Fig. 4. Quadratic variation process, $Q(t)$, for 1927–1985.

It is straightforward to extend (24) and (25) to compute the Bayes factor for a diffusion coefficient with a changepoint. Figure 5 plots the Bayes factor versus α. The evidence for mean reversion is diminished. The Bayes factor is only slightly greater than one for small α values, the plausible range for the mean reversion parameter is now $\alpha \in (0, 0{\cdot}00002)$. This indicates that there is virtually no evidence in the data when comparing a Markovian mean reversion model against the random walk model if we relax the assumption of constant volatility.

Fama & French (1988) and Poterba & Summers (1988) find that long horizon returns have negative autocorrelation and suggest that this is due to a mean reversion in prices. Fama & French (1988) find evidence for mean reversion by regression multiperiod returns on lagged multiperiod returns. This suggests that a non-Markovian model might be more appropriate for investigating mean reversion. Poterba & Summers (1988) find weak evidence in the data for a mean reverting component although they find that a random walk model cannot be rejected at conventional statistical levels. They also note that classical tests have low power for alternatives against an efficient random walk behaviour for prices. Another approach to explore nonrandom walk behaviour is to compare the relative varia-

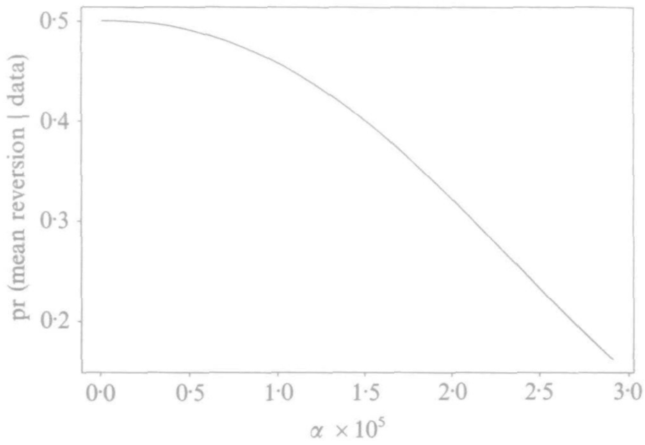

Fig. 5. Probability of mean reversion.

bility of returns over different horizons using variance-ratio tests. Short horizon returns, however, tend to have a small positive autocorrelation contrary to a mean reversion model. The daily autocorrelation of the returns on the Standard & Poor's 500 index is 0·064. Gallant, Rossi & Tauchen (1992) model the series in discrete time allowing for volume effects and conditional heteroscedasticity. In discrete time there are many other effects one might wish to control for, in particular, weekend and monthly trading effects.

Our technique is flexible enough to allow non-Markovian models and can provide assessments of the probability of mean reversion. It is also straightforward to extend the model to including explanatory variables for the parameters (θ, α) in a hierarchical fashion and to model these parameters with a priori distributions. For predictive purposes, we might allow for a two-point mixture between the random walk and mean reversion model. The weights on the models will adapt as we observe the process. This approach is particularly appropriate when there is weak evidence in the data concerning model selection.

6. Discussion

For data arising as fine observations from a continuous time stochastic process, our approach has several advantages over a traditional approach to inference. There is no need to compute asymptotic distributions of functionals of the process or to compute conditional distributions given ancillary statistics. Prediction becomes more natural within our approach and can be performed for arbitrary time points.

There are many interesting and important issues still to be resolved. The suitability of our approach depends on the validity of the approximations used. First, more sophisticated smoothing techniques for estimating the diffusion coefficient are clearly appropriate in many applications. Plotting the quadratic variation process and comparing different volatility estimates can provide useful diagnostic information. Secondly, a limiting feature of using Markov models in one dimension is the positivity constraint that this imposes on the correlation structure. Our approach, however, extends naturally to a non-Markovian framework where negative correlations can be modelled. The multidimensional generalisations, in both state and parameter space, are also conceptually straightforward.

ACKNOWLEDGEMENT

We would like to thank the referee and the editor for many helpful comments.

REFERENCES

BERGER, J. O. & SELLKE, T. (1987). Testing a point null hypothesis: The irreconcilability of P values and evidence (with discussion). *J. Am. Statist. Assoc.* **82**, 112–39.

COX, J., INGERSOLL, J. & ROSS, S. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–407.

DUFFIE, D. (1992). *Dynamic Asset Pricing Theory.* Princeton: Prentice-Hall.

ETHIER, S. N. & KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* New York: Wiley.

FAMA, E. F. & FRENCH, K. R. (1988). Permanent and temporary components of stock prices. *J. Polit. Economy* **96**, 246–73.

GALLANT, A. R., ROSSI, P. E. & TAUCHEN, G. (1992). Stock prices and volume. *Rev. Financ. Studies* **5**, 199–242.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Oxford University Press.

MERTON, R. C. (1980). On estimating the expected return on the market. *J. Financ. Econ.* **8**, 323–61.

NELSON, D. B. (1990). ARCH models as diffusion approximations. *J. Economet.* **45**, 7–38.

OKSENDAL, B. (1985). *Stochastic Differential Equations*, 3rd ed. Berlin-Heidelberg-New York: Springer-Verlag.

POTERBA, J. M. & SUMMERS, L. H. (1986). The persistence of volatility and stock market fluctuations. *Am. Econ. Rev.* **78**, 1142–51.

POTERBA, J. M. & SUMMERS, L. H. (1988). Mean reversion in stock prices: evidence and implications. *J. Financ. Econ.* **22**, 27–59.

[*Received February* 1991. *Revised July* 1993]