The Annals of Applied Probability
2003, Vol. 13, No. 4, 1296–1312

MAXIMUM LIKELIHOOD ESTIMATION OF HIDDEN MARKOV PROCESSES

BY HALINA FRYDMAN AND PETER LAKNER

New York University

We consider the process $dY_t = u_t\,dt + dW_t$, where u is a process not necessarily adapted to $\mathcal{F}^Y$ (the filtration generated by the process Y) and W is a Brownian motion. We obtain a general representation for the likelihood ratio of the law of the Y process relative to Brownian measure. This representation involves only one basic filter (expectation of u conditional on observed process Y). This generalizes the result of Kailath and Zakai [*Ann. Math. Statist.* **42** (1971) 130–140] where it is assumed that the process u is adapted to $\mathcal{F}^Y$. In particular, we consider the model in which u is a functional of Y and of a random element X which is independent of the Brownian motion W. For example, X could be a diffusion or a Markov chain. This result can be applied to the estimation of an unknown multidimensional parameter θ appearing in the dynamics of the process u based on continuous observation of Y on the time interval $[0, T]$. For a specific hidden diffusion financial model in which u is an unobserved mean-reverting diffusion, we give an explicit form for the likelihood function of θ. For this model we also develop a computationally explicit E–M algorithm for the estimation of θ. In contrast to the likelihood ratio, the algorithm involves evaluation of a number of filtered integrals in addition to the basic filter.

1. Introduction.

Let $(\Omega, \mathcal{F}, P)$, $\{\mathcal{F}_t, t \le T\}$ be a filtered probability space and $W = \{W_t, t \le T\}$ a standard Brownian motion. We assume that filtration $\{\mathcal{F}_t, t \le T\}$ is complete and right continuous. We consider process Y with the decomposition

$$dY_t = u_t(\theta)\,dt + dW_t, \tag{1}$$

where u is a measurable, adapted process, and θ is an unknown, vector-valued parameter. We assume that Y is observed continuously on the interval $[0, T]$, but that neither $u = \{u_t, t \le T\}$ nor W are observed. The purpose of this work is to develop methods for the maximum likelihood estimation of θ. Our interest in this work was motivated by the following model from finance. Suppose that we observe continuously a single security with the price process $\{S_t; t \le T\}$ which follows the equation

$$dS_t = \mu_t S_t\,dt + \sigma S_t\,dW_t, \tag{2}$$

Received February 2001; revised July 2002.
AMS 2000 subject classifications. 62M05, 60J60, 60J25.
Key words and phrases. Hidden diffusion financial models, likelihood ratio, maximum likelihood estimation, E–M algorithm, filtered integrals.

where $\sigma > 0$ is a known constant, but the drift coefficient $\mu = \{\mu_t, t \leq T\}$ is an unobserved mean-reverting process with the dynamics

$$d\mu_t = \alpha(\upsilon - \mu_t)\,dt + \beta\,dW_t^1. \tag{3}$$

Here $\alpha > 0$, $\beta > 0$ and υ are unknown parameters, and W^1 is a standard Brownian motion independent of W. The initial value of the security S_0 is an observable constant whereas the initial value of the drift, μ_0, is an unknown constant. The model in (2) and (3) has been recently found to be very flexible in modeling of security prices. Its versions have been used extensively in the area of dynamic asset management [see, e.g., Bielecki and Pliska (1999), Lakner (1998), Kim and Omberg (1996) and Haugh and Lo (2001)]. We now transform the model defined by (2) and (3) to the form in (1) by introducing the following changes of variables. The reason for the transformation is that we want the law of $\{Y_t, t \leq T\}$ to be equivalent to that of a Brownian motion. Let

$$X_t = \frac{1}{\beta}\mu_t - \frac{1}{\beta}\mu_0,$$

$$Y_t = \frac{1}{\sigma}\log\left(\frac{S_t}{S_0}\right) + \frac{\sigma}{2}t.$$

Then the hidden process is

$$dX_t = \alpha(\delta - X_t)\,dt + dW_t^1, \tag{4}$$

where $\delta = \upsilon/\beta$, and the observed process is

$$dY_t = (\lambda X_t + \nu)\,dt + dW_t, \tag{5}$$

where $\lambda = \beta/\sigma$ and $\nu = \mu_0/\sigma$. Let $\theta = (\alpha, \delta, \lambda, \nu)$. We note that $X_0 = 0$ and $Y_0 = 0$, and that (5) is in the form (1) with $u_t(\theta) = \lambda X_t + \nu$.

We summarize our results. Our main result is a new formula for the likelihood ratio of the law of the process Y, specified in (2), with respect to the Brownian measure. Naturally, the likelihood ratio is a function of θ. This result generalizes a result in Kailath and Zakai (1971) that gives a formula for the likelihood ratio in the case when u is adapted to $\mathcal{F}^Y = \{\mathcal{F}_t^Y, t \leq T\}$, the filtration generated by the observed process Y. Here we extend this result to the case when u is not adapted to $\mathcal{F}^Y$. We note that in our financial model example (5) u is not adapted to $\mathcal{F}^Y$. We establish the new formula under a stronger (Proposition 1) and a weaker (Theorem 2) set of sufficient conditions. These conditions allow u to be a functional of Y, and of a random element X which is independent of a Brownian motion W. For example, X could be a diffusion or a Markov chain. The likelihood ratio involves only one basic filter, that is, the expectation of u conditional on the observed data. The new formula makes it possible to directly maximize the likelihood function with respect to the parameters. It may also be useful in developing the likelihood ratio tests of various hypotheses about the parameters of hidden Markov processes. For the financial model in (4) and (5) we give an explicit form of the likelihood function.

We also consider the Expectation–Maximization (E–M) algorithm approach to the maximization of the loglikelihood function for the model in (4) and (5). Dembo and Zeitouni (1986) extended the E–M algorithm to the parameter estimation of the hidden continuous time random processes and in particular to the parameter estimation of hidden diffusions. The financial model in (4) and (5) generalizes an example in Dembo and Zeitouni (1986). For this model we develop a computationally explicit E–M algorithm for the maximum likelihood estimation of θ. The expectation step of this algorithm requires evaluation of filtered integrals, that is, evaluation of the expectations of stochastic integrals conditional on observed data. As remarked by Elliott, Aggoun and Moore [(1997), page 203], Dembo and Zeitouni (1986) express the filtered integrals in the form of nonadapted stochastic integrals which are not well defined. In our development we express a required filtered integral in terms of adapted stochastic integrals (Proposition 4). In contrast to the likelihood function, the E–M algorithm involves a number of filters in addition to the basic filter.

The asymptotic properties of the maximum likelihood estimator of θ when u is adapted to $\mathcal{F}^Y$ have been studied by Feygin (1976) and Kutoyants (1984). The asymptotic inference for θ when u is not adapted to $\mathcal{F}^Y$ has been considered by Kutoyants (1984) and Kallianpur and Selukar (1991) in a special case of our model obtained by setting $\delta = \nu = 0$. (This special model is known as the Kalman filter.) The extension of these results to a more general linear filtering model, in particular, to our model in which hidden process has a mean-reverting drift, is an interesting topic for the future investigation. Whereas Kutoyants (1984) and Kallianpur and Selukar (1991) were concerned with the asymptotic properties of the estimators in a similar model to ours, the objective of this paper is to derive an explicit form of the likelihood function and a computationally explicit E–M algorithm for the maximum likelihood estimation of the model parameters.

This paper is organized as follows. The new formula for the likelihood ratio is established in Section 2. Section 3 presents the extended E–M algorithm and some of its properties. In Sections 4 and 5 we apply the E–M algorithm to the estimation of θ in model in (4) and (5). The Appendix contains proofs of two propositions.

2. The likelihood ratio. In this section we generalize a result of Kailath and Zakai (1971), which also can be found in Kallianpur [(1980), Theorems 7.3.1 and 7.3.2]. We consider the process in (1), but to simplify notation, we suppress the dependence of u on θ, so that

$$dY_t = u_t\,dt + dW_t. \tag{6}$$

Our standing assumption is

$$\int_0^T u_t^2\,dt < \infty. \tag{7}$$

Let $\mathcal{P}_y$ be the measure induced by Y on $\mathcal{B}(C[0,T])$ and $\mathcal{P}_w$ be the Brownian measure on the same class of Borel sets. We know from Kailath and Zakai (1971) that (7) implies $\mathcal{P}_y \ll \mathcal{P}_w$. The same authors give a formula for the likelihood ratio $\frac{d\mathcal{P}_y}{d\mathcal{P}_w}$ for the case when u is adapted to $\mathcal{F}^Y$. Here we are going to extend this result to the case when u is not adapted to $\mathcal{F}^Y$.

Let $\{\upsilon_t,\ t \le T\}$ be the $\mathcal{F}^Y$-optional projection of u, thus

$$E\big[u_t|\mathcal{F}_t^Y\big] = \upsilon_t \qquad \text{a.s.} \tag{8}$$

whenever

$$E|u_t| < \infty.$$

The following proposition is closely related to the result stated already in Lakner (1998). For the sake of the exposition we restate it here in the form most suitable for our purposes and include its proof in the Appendix.

PROPOSITION 1. *Additionally to assumption* (7), *we assume that*

$$E|u_t| < \infty \qquad \textit{for all } t \in [0,T], \tag{9}$$

$$\int_0^T \upsilon_s^2\, ds < \infty \qquad \textit{a.s.} \tag{10}$$

and

$$E\left[\exp\left\{-\int_0^T u_s\, dW_s - \tfrac{1}{2}\int_0^T u_s^2\, ds\right\}\right] = 1. \tag{11}$$

Then $\mathcal{P}_y$ is equivalent to $\mathcal{P}_w$ and

$$\frac{d\mathcal{P}_y}{d\mathcal{P}_w} = \exp\left\{\int_0^T \upsilon_s\, dY_s - \frac{1}{2}\int_0^T \upsilon_s^2\, ds\right\}. \tag{12}$$

REMARK. The formula on the right-hand side is an $\mathcal{F}_T^Y$-measurable random variable, thus it is also a path-functional of $\{Y_t,\ t \le T\}$. Thus this expression represents a random variable and a functional on $C[0,T]$ simultaneously, and in (12) it appears in its capacity as a functional on the space of continuous functions.

In the rest of this section we are going to strengthen Proposition 1 by eliminating condition (11). This will be a generalization again of a result by Kailath and Zakai (1971), where u is assumed to be $\mathcal{F}^Y$-adapted. Our objective here is to have a result which applies to a situation when u depends on Y and possibly on another process which is independent of W. This other process may be a Brownian motion, or possibly a Markov chain. In order to achieve this generality we assume that $\mathcal{X}$ is a complete separable metric space and $X : \Omega \mapsto \mathcal{X}$ is an $\mathcal{F} \mapsto \mathcal{B}(\mathcal{X})$-measurable random element, independent of the Brownian motion W. We denote by $\mathcal{P}_X$ the measure induced by X on $\mathcal{B}(\mathcal{X})$. We assume that

$$u_s = \Phi(Y_0^s, s, X), \qquad s \in [0, T], \tag{13}$$

where Y_0^s represents the path of Y up to time s. Before spelling out the exact conditions on Φ we note that X may become a Brownian motion if $\mathcal{X} = C[0, T]$, or a Markov chain if $\mathcal{X} = D[0, T]$ (the space of right-continuous functions with left limits) with the Skorohod topology. Other examples are also available.

The conditions for Φ are the following. First, we give a notation that for every $f \in C([0, T])$ and $s \in [0, T]$ we denote by $f^s \in C[0, s]$ the restriction of f to $[0, s]$. We assume the following:

For every $s \in [0, T]$ fixed, $\Phi(\cdot, s, \cdot)$ is a mapping from $C[0, s] \times \mathcal{X}$ to $\Re$ such that:

(i) The $C[0, T] \times [0, T] \times \mathcal{X} \mapsto \Re$ mapping given by $(f, s, x) \mapsto \Phi(f^s, s, x)$ is measurable with respect to the appropriate Borel-classes.

(ii) For every $s \in [0, T]$ and $f \in C[0, s]$ the random variable $\Phi(f, s, X)$ is $\mathcal{F}_s$-measurable.

(iii) The following "functional Lipschitz" condition [Protter (1990)] holds: for every $f, g \in C([0, T])$ and $\mathcal{P}_X$-almost every $x \in \mathcal{X}$ there exists an increasing (finite) process $K_t(x)$ such that

$$\left|\Phi(f^t, t, x) - \Phi(g^t, t, x)\right| \le K_t(x) \sup_{u \le t} |f(u) - g(u)|$$

for all $t \in [0, T]$.

Our basic setup now is Y with a decomposition (6), u of the form (13), where X is a random element independent of W and Φ a functional satisfying (i)–(iii).

THEOREM 2. *Suppose that* (7) *holds, that is, in our case*

$$\int_0^T \Phi^2(Y_0^s, s, X)\, ds < \infty \tag{14}$$

and also

$$\int_0^T \Phi^2(W_0^s, s, X)\, ds < \infty. \tag{15}$$

Then

$$E\left[\exp\left\{-\int_0^T u_s\, dW_s - \frac{1}{2}\int_0^T u_s^2\, ds\right\}\right] = 1. \tag{16}$$

REMARK. It is interesting to observe that if Φ does not depend on Y, that is, u is independent of W then (16) follows from (7) only.

REMARK. Conditions (14) and (15) are obviously satisfied if the mapping

$$s \mapsto \Phi(f^s, s, x)$$

is continuous for every $(f, x) \in C[0, T] \times \mathcal{X}$.

PROOF OF THEOREM 2. The main idea of the proof is to derive our theorem from the results of Kailath and Zakai by conditioning on $X = x$. Let $Q(x; \cdot)$, $x \in \mathcal{X}$, be a regular conditional probability measure on $\mathcal{F}$ given X. The existence of such regular version is well known in this situation. It is shown in Proposition A.1 in the Appendix that the independence of the Brownian motion W and X implies that $\{(W_t, \mathcal{F}_t^W),\ t \le T\}$ remains a Brownian motion under $Q(x; \cdot)$ for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$. We also show in the Appendix that if A is a zero probability event under P then for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$ we have $Q(x; A) = 0$. [The "$\mathcal{P}_X$-almost every" is important here; obviously $Q(x; \cdot)$ is not absolutely continuous with respect to P.] Accordingly, for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$

$$\int_0^T \Phi^2(Y_0^s, s, x)\, ds < \infty, \qquad Q(x; \cdot)\text{-a.s.}$$

and

$$\int_0^T \Phi^2(W_0^s, s, x)\, ds < \infty, \qquad Q(x; \cdot)\text{-a.s.}$$

Since $\mathcal{X}$ is assumed to be a complete measurable metric space, we have

$$Q\bigl(x; \{\omega \in \Omega : X(\omega) = x\}\bigr) = 1$$

for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$ [Karatzas and Shreve (1988), Theorem 5.3.19]. What follows is that under $Q(x; \cdot)$ our equations (6) and (13) become

$$dY_t = \Phi(Y_0^t, t, x)\, dt + dW_t.$$

It follows from Theorem V.3.7 in Protter (1990) and our condition (iii) that the above equation has a unique (strong) solution. Thus the process Y must be adapted to the $Q(x; \cdot)$—augmentation of $\mathcal{F}^W$. However, now we are back in the situation studied by Kailath and Zakai (1971). For every $x \in \mathcal{X}$ let $\mathcal{P}_Y(x; \cdot)$ be the measure induced by Y on $\mathcal{B}(C[0, T])$ under the conditional probability measure $Q(x; \cdot)$. Kailath and Zakai, Theorems 2 and 3 imply that for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$ we have $\mathcal{P}_Y(x; \cdot) \sim \mathcal{P}_w$ and

$$\frac{d\mathcal{P}_w}{d\mathcal{P}_Y(x; \cdot)} = \exp\biggl\{-\int_0^T \Phi(Y_0^t, t, x)\, dY_t + \frac{1}{2}\int_0^T \Phi^2(Y_0^t, t, x)\, dt\biggr\}.$$

This implies that the integral of the right-hand side with respect to the measure $\mathcal{P}_Y(x; \cdot)$ is 1, which can be expressed as

$$E\biggl[\exp\biggl\{-\int_0^T u_t\, dY_t + \tfrac{1}{2}\int_0^T u^2\, dt\biggr\}\bigg| X = x\biggr] = 1$$

for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$. Formula (16) now follows. $\square$

We summarize the results of Proposition 1 and Theorem 2 in the following corollary.

COROLLARY 3. *Let $X : \Omega \mapsto \mathcal{X}$ be a measurable mapping taking values in the complete separable metric space $\mathcal{X}$. Assume that X is independent of the Brownian motion W, Y has the form* (6) *and u is as specified in* (13) *where Φ satisfies conditions* (i)–(iii). *Additionally we assume*

$$E[\Phi(Y_0^s, s, X)] < \infty,$$

$$\int_0^T v_s^2 \, ds < \infty \qquad a.s.,$$

$$\int_0^T \Phi^2(Y_0^s, s, X) \, ds < \infty \qquad a.s.,$$

$$\int_0^T \Phi^2(W_0^s, s, X) \, ds < \infty \qquad a.s.,$$

where v is the optional projection of u to $\mathcal{F}^Y$. Then $\mathcal{P}_y \sim \mathcal{P}_w$ and

$$\frac{d\mathcal{P}_y}{d\mathcal{P}_w} = \exp\left\{ \int_0^T v_s \, dY_s - \frac{1}{2} \int_0^T v_s^2 \, ds \right\}.$$

Furthermore,

$$\begin{aligned} &E\left[\exp\left\{ -\int_0^T u_s \, dW_s - \tfrac{1}{2} \int_0^T u_s^2 \, ds \right\}\right] \\ &\qquad = E\left[\exp\left\{ -\int_0^T v_s \, dW_s - \tfrac{1}{2} \int_0^T v_s^2 \, ds \right\}\right] = 1. \end{aligned}$$

REMARK. For the security price model in (4) and (5) the required optional projection is of the form $v_t = \lambda E(X_t | \mathcal{F}_t^Y) + \nu$. We compute $E(X_t | \mathcal{F}_t^Y)$ explicitly in Section 5.

3. E–M algorithm. We briefly describe the extended E–M algorithm presented in Dembo and Zeitouni (1986). In the next section we apply this algorithm to the estimation of θ in the model defined in (4) and (5). Let $\mathcal{F}_T^X$ be the filtration generated by $\{X_t, 0 \le t \le T\}$, $\mathcal{F}_T^Y$ the filtration generated by $\{Y_t, 0 \le t \le T\}$, and $\mathcal{F}_T^{X,Y}$ the filtration generated by $\{X_t, 0 \le t \le T\}$ and $\{Y_t, 0 \le t \le T\}$. We identify Ω as $C[0, T] \times C[0, T]$ and $\mathcal{F}$ as the Borel sets of Ω. Let X be the first coordinate mapping process and Y be the second coordinate mapping process, that is, for $(\omega_1, \omega_2) \in C[0, T] \times C[0, T]$ we have $X_t(\omega_1, \omega_2) = \omega_1(t)$ and $Y_t(\omega_1, \omega_2) = \omega_2(t)$. Let P_θ be a family of probability measures such that $P_\theta \sim P_\phi$

for all $\theta, \phi \in \Theta$, where Θ is a parameter space. Also let E_θ denote the expectation under the measure P_θ. We denote by P_θ^Y the restriction of P_θ to $\mathcal{F}_T^Y$. Then obviously

$$\frac{dP_\theta^Y}{dP_\phi^Y} = E_\phi\left(\frac{dP_\theta}{dP_\phi}\Big|\mathcal{F}_T^Y\right). \tag{17}$$

The maximum likelihood estimator $\widehat{\theta}$ of θ is defined as

$$\widehat{\theta} = \arg\max_\theta \left(\frac{dP_\theta^Y}{dP_\phi^Y}\right) \qquad \text{for some constant } \phi \in \Theta.$$

It should be clear that this definition does not depend on the choice of ϕ. The maximization of dP_θ^Y/dP_ϕ^Y is equivalent to the maximization of the log-likelihood function

$$L_\phi(\theta) \triangleq \log \frac{dP_\theta^Y}{dP_\phi^Y}. \tag{18}$$

Now by (17) and (18),

$$L_\phi(\theta') - L_\phi(\theta) = \log \frac{dP_{\theta'}^Y}{dP_\theta^Y} = \log E_\theta\left(\frac{dP_{\theta'}}{dP_\theta}\Big|\mathcal{F}_T^Y\right),$$

and from Jensen's inequality,

$$L_\phi(\theta') - L_\phi(\theta) \geq E_\theta\left(\log \frac{dP_{\theta'}}{dP_\theta}\Big|\mathcal{F}_T^Y\right), \tag{19}$$

with equality if and only if $\theta' = \theta$. The E–M algorithm is based on (19) and proceeds as follows. In the first step we select an initial value θ_0 for θ. In the $(n+1)$st iteration, $n \geq 0$, the E–M algorithm is applied to produce θ_{n+1}. Each iteration consists of two steps: the expectation, or E, step and the maximization, or M, step. In the E-step of the $(n+1)$st iteration the following quantity is computed

$$Q(\theta, \theta_n) \triangleq E_{\theta_n}\left(\log \frac{dP_\theta}{dP_{\theta_n}}\Big|\mathcal{F}_T^Y\right).$$

In the M-step $Q(\theta, \theta_n)$ is maximized with respect to θ to obtain θ_{n+1}. Iterations are repeated until the sequence $\{\theta_n\}$ converges. It is clear from (19) that

$$L_\phi(\theta_{n+1}) \geq L_\phi(\theta_n), \tag{20}$$

with equality if and only if $\theta_{n+1} = \theta_n$. Thus at each iteration of the E–M algorithm the likelihood function increases. In the M-step the maximizer of θ is obtained as the solution of

$$\frac{\partial}{\partial\theta} Q(\theta, \theta_n) = \frac{\partial}{\partial\theta} E_{\theta_n}\left(\log \frac{dP_\theta}{dP_{\theta_n}}\Big|\mathcal{F}_T^Y\right) = 0, \tag{21}$$

or, if we can interchange the differentiation and expectation operations, the maximizer of θ is the solution of

$$E_{\theta_n}\left(\frac{\partial}{\partial\theta}\log\frac{dP_\theta}{dP_{\theta_n}}\Big|\mathcal{F}_T^Y\right)=0.$$

A corollary of (20) is that the maximum likelihood estimator is a fixed point of the E–M algorithm. It is stated in Dembo and Zeitouni (1986) that under some additional technical conditions all limit points of the sequence $(\theta_n)_{n\geq 0}$ are stationary points of the log-likelihood function. In particular, if the log-likelihood function is concave in the parameter then it follows that $(\theta_n)_{n\geq 0}$ converges to the unique maximum likelihood estimator. For a more detailed discussion of the convergence properties of the E–M algorithm and additional references we refer the reader to Dembo and Zeitouni (1986).

4. Estimation of the drift of the security price process with the E–M algorithm. We develop the E–M algorithm for the estimation of $\theta=(\alpha,\delta,\lambda,\nu)$ in the security price model in (4) and (5). By Girsanov's theorem, if also X process is completely observed, the log-likelihood function of θ is given by

$$\begin{aligned}\frac{dP_\theta}{dP_{\overline{W}}}&=\exp\left\{\int_0^T\alpha(\delta-X_s)\,dX_s-\frac{1}{2}\int_0^T\alpha^2(\delta-X_s)^2\,ds\right\}\\&\quad\times\exp\left\{\int_0^T(\lambda X_s+\nu)\,dY_s-\frac{1}{2}\int_0^T(\lambda X_s+\nu)^2\,ds\right\},\end{aligned}$$

where $\overline{W}=(W^1,W)$ is a two-dimensional Brownian motion and $P_{\overline{W}}$ is the associated Brownian measure on $ß(C[0,T]\times C[0,T])$. Hence

$$\begin{aligned}\log\frac{dP_\theta}{dP_{\theta_n}}&=\left\{\int_0^T\alpha(\delta-X_s)\,dX_s-\frac{1}{2}\int_0^T\alpha^2(\delta-X_s)^2\,ds\right\}\\&\quad+\left\{\int_0^T(\lambda X_s+\nu)\,dY_s-\frac{1}{2}\int_0^T(\lambda X_s+\nu)^2\,ds\right\}\\&\quad+f(\theta_n),\end{aligned}$$

where $f(\theta_n)$ is a term which does not depend on θ. Since, for our problem, we can interchange the operations of differentiation and integration, the equations in (21) take the form

$$\begin{aligned}&E_{\theta_n}\left[\frac{\partial}{\partial\alpha}\log\frac{dP_\theta}{dP_{\theta_n}}\Big|\mathcal{F}_T^Y\right]\\&\quad=E_{\theta_n}\left[\int_0^T(\delta-X_s)\,dX_s-\alpha\int_0^T(\delta-X_s)^2\,ds\,\Big|\mathcal{F}_T^Y\right]=0,\end{aligned}\tag{22}$$

$$
\begin{aligned}
&E_{\theta_n}\left[\frac{\partial}{\partial \delta}\log\frac{dP_\theta}{dP_{\theta_n}}\Big|\mathcal{F}_T^Y\right] \\
&\quad = E_{\theta_n}\left[\alpha X_T - \int_0^T \alpha^2(\delta - X_s)\,ds\,\Big|\mathcal{F}_T^Y\right] = 0,
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
&E_{\theta_n}\left[\frac{\partial}{\partial \lambda}\log\frac{dP_\theta}{dP_{\theta_n}}\Big|\mathcal{F}_T^Y\right] \\
&\quad = E_{\theta_n}\left[\int_0^T X_s\,dY_s - \int_0^T (\lambda X_s + \nu)X_s\,ds\,\Big|\mathcal{F}_T^Y\right] = 0,
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
&E_{\theta_n}\left[\frac{\partial}{\partial \upsilon}\log\frac{dP_\theta}{dP_{\theta_n}}\Big|\mathcal{F}_T^Y\right] \\
&\quad = E_{\theta_n}\left[Y_T - \int_0^T (\lambda X_s + \nu)\,ds\,\Big|\mathcal{F}_T^Y\right] = 0.
\end{aligned}
\tag{25}
$$

Let

$$
\begin{aligned}
m(t, T, \theta_n) &= E_{\theta_n}(X_t|\mathcal{F}_T^Y), \\
n(t, T, \theta_n) &= E_{\theta_n}(X_t^2|\mathcal{F}_T^Y).
\end{aligned}
$$

Then, recalling that $\int_0^T X_s\,dX_s = \frac{1}{2}X_T^2 - \frac{T}{2}$, (22) becomes

$$
\begin{aligned}
&\delta m(T, T, \theta_n) - \frac{1}{2}n(T, T, \theta_n) + \frac{T}{2} - \alpha\delta^2 T \\
&\qquad + 2\alpha\delta\int_0^T m(s, T, \theta_n)\,ds - \alpha\int_0^T n(s, T, \theta_n)\,ds = 0.
\end{aligned}
\tag{26}
$$

Equation (23) takes the form

$$
m(T, T, \theta_n) - \alpha\delta T + \alpha\int_0^T m(s, T, \theta_n)\,ds = 0. \tag{27}
$$

Multiplying (27) by δ and subtracting the result from (26) gives

$$
-\frac{1}{2}n(T, T, \theta_n) + \frac{T}{2} + \alpha\delta\int_0^T m(s, T, \theta_n)\,ds - \alpha\int_0^T n(s, T, \theta_n)\,ds = 0. \tag{28}
$$

From equations (27) and (28) we obtain the following updated values of α and δ:

$$
\alpha_{n+1} = \frac{-\frac{1}{2}Tn(T, T, \theta_n) + T^2/2 + m(T, T, \theta_n)\int_0^T m(s, T, \theta_n)\,ds}{T\int_0^T n(s, T, \theta_n)\,ds - [\int_0^T m(s, T, \theta_n)\,ds]^2}, \tag{29}
$$

$$
\delta_{n+1} = \frac{m(T, T, \theta_n)}{T\alpha_{n+1}} + \frac{\int_0^T m(s, T, \theta_n)\,ds}{T}. \tag{30}
$$

Similarly we write (25) as

$$Y_T - \lambda \int_0^T m(s,T,\theta_n)\,ds - \nu T = 0 \tag{31}$$

and (24) as

$$E_{\theta_n}\left[\int_0^T X_s\,dY_s \,\middle|\, \mathcal{F}_T^Y\right] - \lambda \int_0^T n(s,T,\theta_n)\,ds - \nu \int_0^T m(s,T,\theta_n)\,ds = 0. \tag{32}$$

We compute the first term of (32) in the following proposition. Let

$$m_s(\theta_n) = E_{\theta_n}(X_s|\mathcal{F}_s^Y),$$

and to simplify notation we will write

$$m(t,T) = m(t,T,\theta),$$
$$n(t,T) = n(t,T,\theta),$$
$$m_s = m_s(\theta).$$

PROPOSITION 4.

$$\begin{aligned} E_\theta\left[\int_0^T X_s\,dY_s \,\middle|\, \mathcal{F}_T^Y\right] &= \int_0^T m_s\,dY_s - \int_0^T m_s(\nu + \lambda m_s)\,ds \\ &\quad + \nu \int_0^T m(s,T)\,ds + \lambda \int_0^T m(s,T) m_s\,ds \end{aligned} \tag{33}$$

PROOF. Let

$$\widetilde{W}_t = Y_t - \nu t - \lambda \int_0^T m_s\,ds. \tag{34}$$

It is known that the process $(\widetilde{W}_t, \mathcal{F}_t^Y)$, $0 \le t \le T$, is a Brownian motion [see, e.g., Lemma 11.3 in Lipster and Shiryayev (1978)]. By Theorem 12.1 of Lipster and Shiryayev (1978), m_t satisfies the following equation:

$$dm_t = \alpha(\delta - m_t)\,dt - \lambda\gamma_t(\nu + \lambda m_t)\,dt + \lambda\gamma_t\,dY_t, \qquad m_0 = 0, \tag{35}$$

where $\gamma_t = V(X_t|\mathcal{F}_t^Y)$ is a deterministic function satisfying

$$\frac{d\gamma_t}{dt} = -\lambda^2\gamma_t^2 - 2\alpha\gamma_t + 1, \qquad \gamma_0 = 0. \tag{36}$$

From (34)

$$d\widetilde{W}_t = dY_t - (\nu + \lambda m_t)\,dt, \tag{37}$$

and thus

$$dm_t = \alpha(\delta - m_t)\,dt + \lambda\gamma_t\,d\widetilde{W}_t. \tag{38}$$

We see from (38) that m_t is $\mathcal{F}_t^{\widetilde{W}}$-adapted. Then Y_t is $\mathcal{F}_t^{\widetilde{W}}$-adapted by (34). Hence,

$$
\begin{aligned}
&E_\theta\left[\int_0^T X_s\,dY_s \,\middle|\, \mathcal{F}_T^Y\right] \\
&\quad = E_\theta\left[\int_0^T X_s\,dY_s \,\middle|\, \mathcal{F}_T^{\widetilde{W}}\right] \\
&\quad = E_\theta\left[\int_0^T X_s\,d\widetilde{W}_s + \int_0^T X_s(\nu+\lambda m_s)\,ds \,\middle|\, \mathcal{F}_T^{\widetilde{W}}\right] \\
&\quad = E_\theta\left[\int_0^T X_s\,d\widetilde{W}_s \,\middle|\, \mathcal{F}_T^{\widetilde{W}}\right] + \nu\int_0^T m(s,T)\,ds + \lambda\int_0^T m(s,T)m_s\,ds.
\end{aligned}
\tag{39}
$$

Now, since the assumptions of Theorem 5.14 in Lipster and Shiryayev [(1977), page 185] hold in our case, it follows by this theorem that

$$
E_\theta\left[\int_0^T X_s\,d\widetilde{W}_s \,\middle|\, \mathcal{F}_T^{\widetilde{W}}\right] = \int_0^T m_s\,d\widetilde{W}_s. \tag{40}
$$

Thus by (40), (39) and (37),

$$
\begin{aligned}
&E_\theta\left[\int_0^T X_s\,dY_s \,\middle|\, \mathcal{F}_T^Y\right] \\
&\quad = \int_0^T m_s\,d\widetilde{W}_s + \nu\int_0^T m(s,T)\,ds + \lambda\int_0^T m(s,T)m_s\,ds \\
&\quad = \int_0^T m_s\,dY_s - \int_0^T m_s(\nu+\lambda m_s)\,ds + \nu\int_0^T m(s,T)\,ds \\
&\qquad + \lambda\int_0^T m(s,T)m_s\,ds. \qquad \square
\end{aligned}
$$

By substituting (33) into (32) and using (31), we obtain the updated values for λ and ν:

$$
\begin{aligned}
\lambda_{n+1} = {} & T\int_0^T m_s(\theta_n)\,dY_s - Y_T\int_0^T m_s(\theta_n)\,ds \\
& \times\Bigg(T\int_0^T \left[m_s^2(\theta_n) + n(s,T,\theta_n) - m(s,T,\theta_n)m_s(\theta_n)\right]ds \\
& \qquad - \int_0^T m(s,T,\theta_n)\,ds\int_0^T m_s(\theta_n)\,ds\Bigg)^{-1}
\end{aligned}
\tag{41}
$$

and

$$
\nu_{n+1} = \frac{Y_T - \lambda_{n+1}\int_0^T m(s,T,\theta_n)\,ds}{T}. \tag{42}
$$

To implement the E–M algorithm using equations (29), (30), (41) and (42) we need the expressions for m_s, $m(s,T)$ and $n(s,T)$. These expressions are obtained in the next section.

5. Expressions for filters. We obtain m_t as the solution of (35). Letting

$$\Phi_t = \exp\left(-\alpha t - \lambda^2 \int_0^t \gamma_s \, ds\right), \tag{43}$$

the solution is

$$m_t = \Phi_t \left[\lambda \int_0^t \gamma_s \Phi_s^{-1} (dY_s - \nu \, ds) + \alpha\delta \int_0^t \Phi_s^{-1} \, ds\right]. \tag{44}$$

The function γ_t is obtained as the solution of (36) and is given by

$$\gamma_t = \frac{b}{\lambda^2} \frac{\exp(2at) - d}{\exp(2at) + d} - \frac{\alpha}{\lambda^2}, \tag{45}$$

where $b = \sqrt{\alpha^2 + \lambda^2}$, $d = (b - a)/(b + a)$.

For evaluation of m_t it is useful to note the following expressions. Substituting (45) into (43), after some algebra, gives $\Phi_t = 2b/[(b + \alpha)\exp(bt) + (b - \alpha) \times \exp(-bt)]$, and also $\gamma_s \Phi_s^{-1} = \sinh(bs)/b$.

Application of Lemma 12.3 and Theorem 12.3 in Lipster and Shiryayev [(1978), pages 36 and 37] to our model shows that

$$\begin{aligned} m(s, t) = (\varphi_s^t)^{-1} m_t - \alpha\delta \int_s^t (\varphi_s^u)^{-1} \, du \\ - \lambda \int_s^t (\varphi_s^u)^{-1} \gamma(u, s)(dY_u - \nu \, du), \qquad s \le t \end{aligned}$$

and

$$\gamma(s, t) = \left[1 + \lambda^2 \gamma_s \int_s^t (\varphi_s^u)^2 \, du\right]^{-1} \gamma_s, \qquad s \le t, \tag{46}$$

respectively, where φ_s^t, $t \ge s$, is the solution of

$$\frac{d\varphi_s^t}{dt} = \big(-\alpha - \lambda^2 \gamma(t, s)\big)\varphi_s^t, \ \varphi_s^s = 1, \tag{47}$$

and for $t > s$, $\gamma(t, s) = V(X_t | \mathcal{F}_t^Y, X_s)$ satisfies

$$\frac{d\gamma(t, s)}{dt} = -2\alpha\gamma(t, s) + 1 - \lambda^2 \gamma(t, s)^2, \qquad \gamma(s, s) = 0. \tag{48}$$

It can be easily verified that the solution to (48) is

$$\gamma(t, s) = \frac{b}{\lambda^2} \frac{\exp[2b(t - s)] - d}{\exp[2b(t - s)] + d} - \frac{\alpha}{\lambda^2}, \qquad t \ge s. \tag{49}$$

Using (49), the solution to (47) is

$$\varphi_s^t = \exp\left[-b \int_s^t \frac{\exp[2b(u - s)] - d}{\exp[2b(u - s)] + d} \, du\right] = (1 + d) \frac{\exp[b(t - s)]}{\exp[2b(t - s)] + d}.$$

For evaluation of $m(s,t)$ it is useful to note that $(\varphi_s^t)^{-1}\gamma(t,s) = \sinh(b(t-s))/b$, and to evaluate $\gamma(s,t)$ we compute the integral

$$\int_s^t (\varphi_s^u)^2\, du = \frac{\exp[2b(t-s)]-1}{(b+\alpha)\exp[2b(t-s)]+b-\alpha}.$$

Clearly $n(s,t) = \gamma(s,t) + m(s,t)^2$, $0 \le s \le t$.

APPENDIX

PROOF OF PROPOSITION 1. We define the process

$$Z_t = \exp\left\{-\int_0^t u_s\, dW_s - \tfrac{1}{2}\int_0^t u_s^2\, ds\right\}, \qquad t \in [0,T],$$

and the probability measure P_0 by

$$\frac{dP_0}{dP} = Z_T,$$

which is a bona fide probability measure by condition (11). We are going to repeatedly use the following form of “Baye’s theorem”: for every $\mathcal{F}_t$-measurable random variable V we have

$$E_0[V|\mathcal{F}_t^Y] = \frac{1}{E[Z_t|\mathcal{F}_t^Y]} E[Z_t V|\mathcal{F}_t^Y], \qquad t \in [0,T], \tag{50}$$

where E_0 is the expectation operator corresponding to the probability measure P_0. According to Girsanov’s theorem Y is a Brownian motion under P_0. Also, one can see easily that

$$\frac{d\mathcal{P}_y}{d\mathcal{P}_w} = E_0\left[\frac{1}{Z_T}\Big|\mathcal{F}_T^Y\right] = \left(E[Z_T|\mathcal{F}_T^Y]\right)^{-1}.$$

The equivalence of $\mathcal{P}_y$ and $\mathcal{P}_w$ already follows, since $Z_T > 0$ (under P and P_0) thus also

$$E_0\left[\frac{1}{Z_T}\Big|\mathcal{F}_T^Y\right] > 0,$$

hence the same expression, as a path-functional of $\{Y_t,\ t \le T\}$ is $\mathcal{P}_y$-almost everywhere positive. We still need to show (12). We start with the identity

$$\frac{1}{Z_t} = 1 + \int_0^t \frac{1}{Z_s} u_s\, dY_s.$$

Theorem 5.14 in Lipster and Shiryayev (1977) implies that

$$E_0\left[\frac{1}{Z_t}\Big|\mathcal{F}_t^Y\right] = 1 + \int_0^t E_0\left[\frac{1}{Z_s}u_s\Big|\mathcal{F}_s^Y\right] dY_s, \tag{51}$$

as long as the following two conditions hold:

$$E_0\left[\frac{1}{Z_s}|u_s|\right] < \infty, \tag{52}$$

$$\int_0^T \left(E_0\left[\frac{1}{Z_s}u_s\middle|\mathcal{F}_s^Y\right]\right)^2 ds < \infty \qquad \text{a.s.} \tag{53}$$

However, (52) follows from condition (9). In order to show (53) we use the rule (50) and compute

$$\begin{aligned}\int_0^T \left(E_0\left[\frac{1}{Z_s}u_s\middle|\mathcal{F}_s^Y\right]\right)^2 ds &= \int_0^T (E[Z_s|\mathcal{F}_s^Y])^{-2} v_s^2\, ds \\ &= \int_0^T \left(E_0\left[\frac{1}{Z_s}\middle|\mathcal{F}_s^Y\right]\right)^2 v_s^2\, ds.\end{aligned}$$

Now we observe that by (51), $(s,\omega) \mapsto E_0[\frac{1}{Z_s}|\mathcal{F}_s^Y]$ is a P_0-local martingale (and also a martingale but this is not important at this point), adapted to the filtration generated by the P_0-Brownian motion Y, thus it must be continuous [Karatzas and Shreve (1988), Problem 3.4.16]. This fact and condition (10) imply (53). Now using (51) and (50), we get

$$E_0\left[\frac{1}{Z_t}\middle|\mathcal{F}_t^Y\right] = 1 + \int_0^t \frac{1}{E[Z_s|\mathcal{F}_s^Y]} v_s\, dY_s = 1 + \int_0^t E_0\left[\frac{1}{Z_s}\middle|\mathcal{F}_s^Y\right] v_s\, dY_s,$$

which implies (12). □

PROPOSITION A.1. *Suppose that $\{(W_t, \mathcal{F}_t), t \le T\}$ is a Brownian motion, $\mathcal{X}$ is a complete separable metric space, $X : \Omega \to \mathcal{X}$ is a measurable mapping, and $\{Q(x; \cdot), x \in \mathcal{X}\}$ is a family of regular conditional probability measures for $\mathcal{F}$ given X. Let $\mathcal{P}_X$ be the measure induced by X on $\mathcal{B}(\mathcal{X})$. Then*:

(a) *for every $A \in \mathcal{F}$ such that $P(A) = 0$ we have $Q(x; A) = 0$ for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$,*

(b) *for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$, the process $\{(W_t, \mathcal{F}_t^W), t \le T\}$ is a Brownian motion under the probability measure $Q(x; \cdot)$.*

PROOF. Statement (a) is an obvious consequence of the definition of regular conditional probability measure. Indeed, for every $A \in \mathcal{F}$ we have

$$Q(x; A) = P(A|X = x), \qquad \mathcal{P}_X\text{-a.e.}, x \in \mathcal{X}$$

and thus

$$Q(X; A) = P(A|X), \qquad P\text{-a.s.}$$

This implies that if A is a P-zero event then $E[Q(X; A)] = P(A) = 0$ and (a) now follows. Next we are going to prove (b). It follows from (a) that the process W

has $Q(x;\cdot)$-almost surely continuous paths for $\mathcal{P}_X$-almost every $x \in \mathcal{X}$. We need to show that the finite dimensional distributions of W under $Q(x;\cdot)$ coincide with those under the probability measure P. Let us fix a positive integer n and time points $0 \le t_1 \le \cdots \le t_n \le T$. We denote by $\mathcal{Q}$ the set of rational numbers. For every $q = (q_1, \dots, q_n) \in \mathcal{Q}^n$ there exists a set $B_q \in \mathcal{B}(\mathcal{X})$, $\mathcal{P}_X(B_q) = 1$ such that for all $x \in B_q$,

$$\begin{aligned} &Q\big(x; W(t_1) \le q_1, \dots, W(t_n) \le q_n\big) \\ &\quad = P\big(W(t_1) \le q_1, \dots, W(t_n) \le q_n | X = x\big) \\ &\quad = P\big(W(t_1) \le q_1, \dots, W(t_n) \le q_n\big), \end{aligned}$$

where the last step follows from independence of X and W under P. Now we define

$$B = \bigcap_{q \in \mathcal{Q}^n} B_q,$$

and note that $\mathcal{P}_X(B) = 1$ since $\mathcal{Q}^n$ is countable. It follows that the equation above holds for every $x \in B$ which completes the proof. $\square$

Acknowledgments. Halina Frydman completed part of this work while visiting the Department of Biostatistics at the University of Copenhagen. She thanks the department for the hospitality. We also thank the anonymous referee for the comments which substantially improved the paper.

REFERENCES

BIELECKI, T. R. and PLISKA, S. R. (1999). Risk sensitive dynamic asset management. *J. Appl. Math. Optim.* **39** 337–360.

DEMBO, A. and ZEITOUNI, O. (1986). Parameter estimation of partially observed continuous time stochastic processes. *Stochastic Process. Appl.* **23** 91–113.

ELLIOTT, R. J., AGGOUN, L. P. and MOORE, J. B. (1997). *Hidden Markov Models*. Springer, Berlin.

FEYGIN, P. D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Adv. in Appl. Probab.* **8** 712–736.

HAUGH, M. B. and LO, A. W. (2001). Asset allocation and derivatives. *Quant. Finance* **1** 45–72.

KAILATH, T. and ZAKAI, M. (1971). Absolute continuity and Radon–Nikodym derivatives for certain measures relative to Wiener measure. *Ann. Math. Statist.* **42** 130–140.

KALLIANPUR, G. (1980). *Stochastic Filtering Theory.* Springer, New York.

KALLIANPUR, G. and SELUKAR, R. S. (1991). Parameter estimation in linear filtering. *J. Multivariate Anal.* **39** 284–304.

KARATZAS, I. and SHREVE, S. E. (1988). *Brownian Motion and Stochastic Calculus*. Springer, New York.

KIM, T. and OMBERG, E. (1996). Dynamic nonmyopic portfolio behavior. *Rev. Financial Studies* **9** 141–161.

KUTOYANTS, Y. A. (1984). *Parameter Estimation for Stochastic Processes*. Helderman, Berlin.

LAKNER, P. (1998). Optimal trading strategy for an investor: The case of partial information. *Stochastic Process. Appl.* **76** 77–97.

LIPSTER, R. S. and SHIRYAYEV, A. N. (1977). *Statistics of Random Processes* **1**. Springer, New York.

LIPSTER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes* **2**. Springer, New York.

PROTTER, P. (1990). *Stochastic Integration and Differential Equations*. Springer, New York.

DEPARTMENT OF STATISTICS
AND OPERATIONS RESEARCH
STERN SCHOOL OF BUSINESS
NEW YORK UNIVERSITY
44 WEST 4TH STREET
NEW YORK, NEW YORK 10012
E-MAIL: hfrydman@stern.nyu.edu
plakner@stern.nyu.edu