Recursive Parameter Estimation for Semimartingales

by

A. Thavaneswaran

and

Muhammad K. Habib

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1819

March 1987

Recursive Parameter Estimation for Semimartingales

A. Thavaneswaran and M.K. Habib

Department of Biostatistics
University of North Carolina
Chapel Hill, NC 27514 USA

A recursive estimation algorithm is presented for a semimartingale model based on the theory of optimal estimating functions. Strong consistency and asymptotic normality of the recursive estimate generated by the algorithm are established. This recursive algorithm may be used to handle the problems of missing observations and censored observations.

AMS 1980 Subject Classification: 60G 60K99,62F99

KEYWORDS AND PHRASES: Asymptotic normality, consistency, optimal estimation, recursive estimation, semimartingales

Research supported by the Office of Naval Research under contract N00014-83-K-0387.

1. Introduction

This paper deals with the problem of recursively estimating a p-dimensional parameter, θ, that occurs in the predictable part of a semimartingale model. Semimartingales have been shown to serve as models of observable stochastic processes occurring in many applied areas such as economics and finance (Aase, 1985) and neurophysiology (Habib and Thavaneswaran, 1986). Off-line estimation procedures using the method of optimal estimation have been analyzed in Thavaneswaran and Thompson (1986), quasi-likelihood has been treated in Hutton and Nelson (1986) and quasi-least squares has been investigated in Christopeit (1986). Aase (1981, 1982, 1983) has proposed and studied the recursive estimates for several regression type models in the i.i.d. and non i.i.d. setup. Following Aase, (1982) Thavaneswaran (1986) proposed a recursive estimate for a nonlinear counting process model and studied its asymptotic properties. In this paper a recursive estimate of θ based on the theory of optimal estimating functions is derived and its asymptotic properties are studied for multivariate semimartingales with multidimensional parameters. This approach enables us to study the recursive estimation for Itô-Markov models as a special case and to our knowledge the recursive estimation for Itô-Markov models has not been treated earlier. The resulting algorithms are related to the Kalman-Bucy type algorithm in the special cases. Advantages for using the recursive estimate over its off-line version are also included.

A semimartingale is a stochastic process which can be represented as a sum of a process of bounded variation and a local martingale. In the case of continuous time processes, a typical example of such a process is a process $(X_t, t\geq 0)$
with independent increments for which $E|X_t|$ is finite and a function of locally bounded variation. The class of semimartingales includes point processes (counting processes, Poisson processes, extended gamma processes, branching processes), Itô processes, Itô-Markov processes, diffusion processes, etc. In Section 2, the problem is formulated and in Section 3, Godambe's optimality Theorem is proved. Sections 4 and 5 deal with recursive estimates and their asymptotic properties.

2. Notations and Conventions

We assume that all the semimartingales that will appear in the sequel are defined on some fixed complete probability space (Ω, F, P) for each P in a family $\{P\}$ of probabilty measures. We also assume a family $F=\{F_t, t\geq 0\}$ of σ-algebras satisfying the standard conditions ($F_s \leq F_t \leq F$ for $s \leq t$, F_0 is augmented by sets of measure zero of F, and $F_t = F_{t+}$, where $F_{t+} = \bigcap_{s\geq t} F_s$). We denote by D the space of right continuous functions $X = (X_t, t\geq 0)$ having limits on the left and use $X = (X_t, F_t)$ to denote an $\{F_t\}$ adapted random process (X_t) with trajectories in the space D. Assume that the process $X=(X_t, F_t)$ is a semimartingale for each P, that is, for each P it can be represented in the form

$$X_t = V_t + H_t \tag{2.1}$$

where $V=(V_t, F_t)$ is a locally bounded variation process and $H=(H_t, F_t)$ locally square-integrable martingale.

We assume that V_t is an absolutely continuous process of the form

$$V_t = \int_0^t f_{s,\theta} \, d\lambda_s$$

where $\theta \in R_+^p$ and f is some other observed process which is assumed to be predictable. The nonrandom parameter θ is assumed to be unknown and it will be estimated on the basis of the observations X_t and f_t. $\lambda = (\lambda_t, F_t)$ is a real monotonically increasing right continuous process $\lambda_0 = 0$. Hence, the model (2.1) takes the form

$$X_t = \int_0^t f_{s,\theta} \, d\lambda_s + H_{t,\theta}$$

We further assume that $< H >_{t,\theta} = \int_0^t b_{s,\theta} \, d\lambda_s$ where $\{\lambda_t\}$, $\{b_{t,\theta}\}$ and $\{f_t\}$ are predictable processes with respect to $\{F_t\}$. This is analogous to a regression model where only the form of the dependence structure among the error terms are specified.

3. Godambe's Optimality Criteria

Following Godambe (1985), consider a parameter θ to be a function of $p \in P$ (where P is a family of probability measures). Let $G(X,\theta) = (G_t(X,\theta), F_t)$ represent a family of processes indexed by θ such that $E_P\, G_t(V,\theta) = 0$. This corresponds to the unbiasedness property of Godambe's (1960) optimality criteria which, adapted to this situation, says that G° is optimal in L - the class of unbiased estimating forms if $Q = A_h - A_{h^{\circ}}$ is nonnegative definite for all $G \in L$ and for all P, where

$$h(X) = E\left\{ \left[\frac{\partial G}{\partial \theta} \right]^{-1} \right\} G(X,\theta),$$

$$h^{\circ}(X) = E\left\{ \left[\frac{\partial G^{\circ}}{\partial \theta} \right]^{-1} \right\} G^{\circ}(X,\theta),$$

and A_h is the variance - covariance matrix for h under θ_0.

The following lemma states a sufficient condition (due to M.E. Thompson) for optimality. (c.f. Thavaneswaran (1985), p.57)

Lemma: - G° is optimal in L if

$$E\left[\frac{\partial G}{\partial \theta} \right] = K\, E\left[G G^{\circ T} \right] \text{ for all } G \in L.$$

where T denotes the transpose and K is a constant matrix.

Let $X_t = \int_0^t f_{s,\theta}\, d\lambda_s + H_{t,\theta}$ and $\theta \in R^p$. Then it is natural to look for p - dimensional estimating functions of the form

$$G_{t,\theta} = \int_0^t a_{s,\theta}\, dH_{s,\theta}\ ,$$

generated by $\{a_{s,\theta}\}$.

Then

$$E\left[G G^{\circ T} \right] = E \int_0^t a_{s,\theta}\, d\langle H \rangle_{s,\theta}\, a^{\circ T}_{s,\theta}$$

where T denotes the transpose, and

$$G^{o}_{t,\theta} = \int_0^t a^{o}_{s,\theta} \, dH_{s,\theta}$$

Hence

$$E\left[GG^{oT} \right] = E \int_0^t a_{s,\theta} \, b_{s,\theta} \, a^{oT}_{s,\theta} \, d\lambda_s$$

Moreover,

$$E\left[\frac{\partial G_{t,\theta}}{\partial \theta} \right] = E \int_0^t a_{s,\theta} \, \partial/\partial\theta \, (dH_{s,\theta})$$

$$= - E \int_0^t a_{s,\theta} \, f^{T}_{s,\theta} \, d\lambda_s$$

where • denotes the derivative with respect to θ. Hence the optimal estimating function is given by

$$G^{o}_{t,\theta} = \int_0^t \dot{f}^{T}_{s,\theta} \, b^{+}_{s,\theta} \, dH_{s,\theta}$$

provided that $b^{+}_{s,\theta}$; the inverse of $b_{s,\theta}$ exists.

4. Recursive Estimation

Optimal estimator $\hat{\theta}$ satisfies the equation

$$G^{o}_{t,\hat{\theta}} = \int_0^t a^{o}_{s,\hat{\theta}_s} \, (dX_s - d_{s,\hat{\theta}_s} \, \lambda X_s) = 0$$

If G is a smooth function of both θ and t, it follows from the implicit function theorem that $\hat{\theta}_t$ satisfies the equation

$$d\hat{\theta}_t = - \left[\hat{G}^{o}_t \, (\hat{\theta}_t) \right]^{-1} \partial/\partial t \, G^{o}_{t,\hat{\theta}_t}$$

we have

$$d\hat{\theta}_t = k_t \, a^o_{t.\hat{\theta}_t} \left[dX_t - f_{t,\hat{\theta}_t} \, d\lambda_t \right]$$

where

$$k_t^{-1} = \hat{G}^o_t \, (\hat{\theta}_t) = \int_0^t a^o_{s,\hat{\theta}} \, f^T_{s,\hat{\theta}} \, d\lambda_s$$

provided that $a^o_{s,\theta}$ is independent of θ.

Note : In general $a^o_{s,\theta}$ is independent of θ for models having linear intensity and having the variance process of the martingale independent of θ. When $a^o_{s,\theta}$ depends on θ, this corresponds to the approximation obtained by equating the martingale term

$$\int_0^t \left[\frac{\partial a^o_{s,\theta}}{\partial \theta} \right]_{\hat{\theta}} dH_{s,\hat{\theta}} = 0.$$

(c.f. Spreij (1986), p.284). Furthermore,

$$a^o_{s,\theta} = f^T_{s,\theta} \, b^+_{s,\theta}$$

implies that

$$k_t^{-1} = \int_0^t a^o_{s,\hat{\theta}} \, b_{s,\hat{\theta}} \, a^{oT}_{s,\hat{\theta}} \, d\lambda_s$$

$$dk_t = -k_t \left(a^o_{t,\hat{\theta}} \, b_{t,\hat{\theta}} \, a^{oT}_{t,\hat{\theta}} \right) k_t \, d\lambda_t \, .$$

Thus we have the following recursive algorithm for the multivalued semimartingales and multiparameter case:

$$d\hat{\theta}_t = k_t \, a^o_{t,\hat{\theta}} \left[dX_t - f_{t,\hat{\theta}} \, d\lambda_t \right]$$

and

$$dk_t = -k_t \left(a^o_{t,\hat{\theta}} \, b_{t,\hat{\theta}} \, a^{oT}_{t,\hat{\theta}} \right) k_t \, d\lambda_t$$

Example 4.1: Now we consider the following model as in Aase (1982).

$$dX_t = \left[G(t,F_t^X) + H(t,F_t^X)\ \theta \right] dt + \sigma(t,F_t^X)\ dW_t$$

this corresponds to the case where

$$f_{s,\theta} = G + H\theta,$$

$$\lambda_s = s \text{ and } H_{t,\theta} = \int_0^t \sigma(s,F_s^X)\ dW_s.$$

$$b_{t,\theta} = \sigma(t,F_t^X)\ \sigma^T(t,F_t^X)$$

Hence $a^o_{t,\theta} = H_t^T(\sigma\sigma^T)^{-1}$ which is independent of θ and

$$a^o_{t,\theta}\ b_{t,\theta}\ a^o_{t,\theta} = H_t^T\left(\sigma\sigma^T \right)^{-1} H_t$$

The recursive estimates are given by

$$d\hat{\theta}_t = k_t\ H^T \left(\sigma\sigma^T \right)^{-1} [dX_t - (G+H\hat{\theta}_t)\ dt]$$

$$\dot{k}_t = -k_t\ H_t^T \left(\sigma\sigma^T \right)^{-1} H_t k_t$$

which leads to the same scheme as in Aase (1982) motivated somewhat differently using filtering theory. More interestingly if σ depends on θ then the likelihood cannot be defined, as the measures induced by X_t and W_t become singular. However the same algorithm can be used to obtain the recursive optimal estimate.

If we take θ_t, the signal process in the filtering setup as $\theta_t = \theta$, and assume that θ has multivariate normal distribution then the algorithm obtained here is the same as the nonlinear version of the Kalman filter. (Liptser and Shiryayev, 1978)

Example 4.2: Now we consider the recursive estimation in the following counting process setup as in Spreij (1986, p. 281).

$$dX_t = f^T(X_t)\,\theta dt + dm_t, \quad X_0 = 0$$

then this corresponds to the case with r=1, $\lambda_t = t$, $f_{t,\theta} = f^T(X_t)\theta$, is predictable $H_t = M_t$, purely discontinuous martingale and $b_{t,\theta} = f^T(X_{t-})\,\theta$.

Hence $a^o_{t,\theta} = \dfrac{F(X_t)}{F^T(X_t)\theta}$ and the recursive algorithm turns out to be

$$d\hat{\theta}_t = K_t \frac{f^T(X_t)}{f^T(X_t)\,\hat{\theta}_{t-}} (dX_t - f^T(X_{t-})\,\hat{\theta}_t dt)$$

$$dk_t = \frac{-k_t f(X_t)\, f^T(X_t)\, k_t dt}{f^T(X_t)\,\hat{\theta}_{t-}}$$

which lead to the same algorithm motivated somewhat differently in Spreij (1986). It is of interest to note that the recursive estimate obtained by the method of least squares is not efficient while the one obtained by the optimal estimation technique turns out to be the approximate maximum likelihood estimate and hence is efficient. Furthermore, the same algorithm may be used as long as $< M >_t = f^T(X_t)\theta$ and we do not need to impose any regularity conditions for absolute continuity of measures induced by X_t and the standard Poisson process.

Application of this algorithm in survival analysis with censored observations and to a random periodic intensity model together with the connection to Poisson-Gamma filter see Thavaneswaran (1986).

Example 4.3: Next we consider the Ito-Markov model of the form

$$X_t = \int_0^t f(X_s)\theta\, ds + \int_0^t \sigma(X_s)dW_s + \int_0^t\int_Z C(X_s)\, q(ds,dz)$$

where f,σ,c are predictable processes W is a standard Wiener process, q is the martingale measure corresponding to a Poisson process P(ds,dz) given by $q(ds,dz) = P(ds,dz) - ds \otimes \alpha(dz)$ with $\alpha(dz)$ as the Levy measure of P.

This corresponds to our setup with $\lambda_t = t$, $f_{s,\theta} = f(X_s)\theta$,

$$H_{t,\theta} = H_t = \int_0^t \sigma(X_s)dW_s + \int_0^t\int_Z C(X_s)\, q(ds,dz)$$

Hence $a^o_s = f^T(X_s)b^+_s$ and the recrusive algorithm is given by

$$d\hat{\theta}_t = k_t\, f^T(X_t)b^+_t \left[dX_t - f(X_t)\, \hat{\theta}_t dt \right]$$

$$dk_t = -k_t\, (f^T(X_t)\, b^+_t\, f(X_t))\, k_t dt$$

Since $a^o_{s,\theta}$ is independent of θ, this gives an exact algorithm. The fact that the estimator depends on σ and C is not a serious restriction because of

Levy's result for Brownian motion and its extension the conditional variance b_t can be estimated nonparametrically with variance as small as we please for a given data stretch in continuous time (see Brown & Hewit (1975)) so it can be assumed to be known.

5. Asymptotic Properties

In this section we establish the strong consistency and asymptotic normality of the recursive estimate for a semimartingale model with linear intensity.

Strong Consistency. Suppose we have

$$dX_t = f(X,t)\theta d\lambda_t + dH_{t,\theta}$$

Then the estimate $\hat{\theta}_t$, k_t the gain matrix satisfying the recessive form

$$d\hat{\theta}_t = k_t\, a^o_{t,\hat{\theta}} \left[dX_t - f(X_t)\, \hat{\theta}_t d\lambda_t \right]$$

and

$$dk_t = -k_t \left(a^o_{t,\hat{\theta}}\, b_{t,\hat{\theta}}\, a^{oT}_{t,\hat{\theta}} \right) k_t d\lambda_t$$

can be written as

$$\hat{\theta}_t = k_t \left[\theta_0 k_0^{-1} + \int_0^t a^o_{s,\theta} \left(dX_s - f(X_s)\hat{\theta} d_s\lambda \right) \right]$$

with

$$k_t = \left[I + k_0 \int_0^t \left(a^o_{s,\hat{\theta}} b_{s,\hat{\theta}} a^{oT}_{s,\theta} \right) d\lambda_s \right]^{-1}$$

or equivalently,

$$\hat{\theta}_t = \theta + \left[I + k_0 \int_0^t \left(a^o_{s,\theta} b_{s,\hat{\theta}} a^{oT}_{s,\hat{\theta}} \right) d\lambda_s \right]^{-1}$$

$$\cdot \left[(\theta_0 - \theta) + k_0 \int_0^t a_{s,\hat{\theta}} dH_{s,\theta} \right]$$

Let $M_t = \int_0^t a^o_{s,\hat{\theta}} dH_{s,\theta}$ then, $< M >_t = \int_0^t a^o_{s,\hat{\theta}} b^o_{s,\hat{\theta}} a^{oT}_{s,\hat{\theta}} d\lambda_s$,

In general $b_{s,\theta}$ is independent of θ and $< M >_t$ does not depend on θ through b. Hence for any increasing function T(t) of t, we have

$$\hat{\theta}_t = \theta + \left[\frac{I}{T(t)} + k_0 \frac{< M >_t}{T(t)} \right]^{-1} \left[\frac{\theta_0 - \theta}{T(t)} + k_0 \frac{< M >_t}{T(t)} < M >_t^{-1} M_t \right] \tag{5.1}$$

<u>Theorem</u> <u>5.2</u> - Let $\theta \in$ open subset of $\mathbb{R}^p$. Let $T(t) \to \infty$ and

$$\liminf_{t \to \infty} \frac{< M >_t}{T(t)} = \sum \text{, a symmetric positive definite matrix.}$$

The $\lim_{t\to\infty} \hat{\theta}_t = \theta$ a.s. (i.e.) $\hat{\theta}_t$ is strongly consistent estimate of θ.

<u>Proof</u>. By the hypothesis of the Theorem the denominator of the second term in (5.1) $\to \sum$ a.s. and by the martingale strong law of large numbers the numerator $\to 0$ a.s.. Hence the result follows.

<u>Asymptotic</u> <u>Normality</u>. In this section we prove the asymptotic normality of the recursive estimate using a martingale central limit theorem.

The recursive estimate of $\hat{\theta}_t$ may be written as

$$\hat{\theta}_t = \theta + \left[k_0^{-1} I + \langle M \rangle_t \right]^{-1} \left[k_0^{-1} (\theta_0 - \theta) + M_t \right] \quad (5.2)$$

where $M_t = \int_0^t a_s^o \, dH_s'$ and $\langle M \rangle_t = \int_0^t a_s^o \, b_s \, a_s^{oT} \, d\lambda_s$.

Theorem 5.3: - Let M be a locally square integrable martingale defined above. Assume that there exists a function $T:[0,\infty) \to [0,\infty)$ with $T(t) \to \infty$ s.t.

(i) $\lim_{t\to\infty} T^{-1}(t) \langle M \rangle_t = \Sigma$ in probability,

$\Sigma \in \mathbb{R}^{p \times p}$ is a symmetric positive definite nonrandom matrix,

(ii) $\lim T^{-1}(t) \int_0^t a_s^o \, I \left\{ |a^o_{s(ij)}| > \epsilon \, T(t) \right\} b_s \, a_s^{oT} \, d\lambda_s = 0$ a.s.

Then,

$$\text{(a)} \quad \langle M \rangle_t^{-1/2} M_t \xrightarrow{L} N(0,I)$$

and

$$\text{(b)} \quad \Sigma^{-1/2} (\hat{\theta}_t - \theta) \xrightarrow{L} N(0,I)$$

Proof : In the one dimensional case when M_t has continuous trajectories as in the case of diffusion process model, (a) is a restatement of Kunita and Watanabe (1967), and when M_t has purely discontinous trajectories with unit jumps this corresponds to the Metivier's ((1982), p. 200) result. In general (a) follows from Linkov (1982). Applying (i), (ii) & (a) in (5.2) gives (b).

Example 5.4: - Now we consider a periodic intensity Itô-Markov model of the form

$$dX_t = (3 \quad 3+2\sin t) \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} dt + dW_t + (5+2\sin t)^{1/2} dM_t^d$$

Hence $H_t = \int_0^t dW_s + \int_0^t (5+2\sin s)^{1/2} dM_t^d$

and

$$< M >_t = \int_0^t \begin{bmatrix} 3 \\ 3+2\sin s \end{bmatrix} [3 \quad 3+2\sin s] \frac{ds}{6+2\sin s}$$

$$\frac{1}{t} < M >_t = \frac{1}{t} \int_0^t \begin{pmatrix} \frac{9}{6+2\sin s} & \frac{9+6\sin s}{6+2\sin s} \\ \frac{9+6\sin s}{6+2\sin s} & \frac{9+12\sin s + 4\sin^2 s}{6+2\sin s} \end{pmatrix} ds$$

Using the fact that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{dx}{a+b\sin x} = \frac{1}{\sqrt{a^2 - b^2}} \quad \text{for all} \quad a>b\geq 0.$$

We have

$$\text{as } t\to\infty, \quad \frac{1}{t} < M >_t \longrightarrow \begin{pmatrix} \frac{9}{\sqrt{32}} & 3 - \frac{9}{\sqrt{32}} \\ 3 - \frac{9}{\sqrt{32}} & \frac{9}{\sqrt{32}} \end{pmatrix}$$

a symmetric positive definite matrix. Hence assumption (i) of Theorem 5.3 is satisfied. To establish the assumption (ii) of Theorem 5.3 it is sufficient to remark that the periodic functions in the integrand of $< M >_t$ are bounded. Hence the normality and strong consistency of the estimate follows.

REFERENCES

Aase K.K. (1985) R & D Projects analyzed by semimartingale methods. J. Appl. Prob. 22, 288-299.

Aase K.K. (1983) Recursive estimation in nonlinear time series models of autoregressive type. J. R. Statist. Soc.B. 45, 228-237.

Aase K.K. (1982) Stochastic continuous-time model reference adaptive system with decreasing gain. Adv. in Prob. 14 , 763-788.

Aase K.K. (1981) Model reference adaptive systems applied to regression theory. Statist. Neerlandica 3, 123-155.

Christopeit N. (1986) Quasi-least-squares estimation in semimartingale regression models. Stochastics 16, 255-278.

Elliot R.J. (1982) Stochastic Calculus and Applications. Springer-Verlag, New York.

Godambe V.P. (1960) An optimum property of regular maximum likelihood equation. Ann. Math. Statist. 31, 1208-11.

Godambe V. P. and Thompson M.E. (1985) Logic of least square revisited. Preprint.

Godambe V. P. (1985) The foundations of finite sample estimation in stochastic processes. Biometrika 12, 419-428.

Habib M.K. and Thavaneswaran A. (1986) Optimal estimation for semimartingales neuronal models. Preprint.

Hutton J.E. and Nelson P.I. (1986) Quasi-likelihood estimation for semimartingale. Stoch. Proc. Appl., 22 , 245-257.

Kunita H. ane Watanabe S. (1967) On square integrable martingales. Nagoya Math. J. 30, 209-245.

Lindsay B. (1985) Using emperical partially Bayes inference for increased efficiency. Ann. Statist. 13, 914-931.

Linkov Yu.N. (1982) Asymptotic properties of statistica estimators and tests for Markov processes. Theor. Prob. and Math Statist. 25, 83-89.

Liptser R.S. and Shiryaryev (1980) A strong law of large numbers for local martingales. Stochastics 3, 217-228.

Liptser R.S. and Shiryayev A.N. (1978) Statistics of Random Processes. 2. Springer Verlag, New York.

Metivier M. (1982) Semimartingales. Walter de Cruyter, New York.

Spreij P. (1986) Recursive parameter estimation for counting processes with linear intensity. Stochastics, 18, 277-312.

Thavaneswaran A. and Thompson M.E. (1986) Optimal estimation for semimartingales. J. Appl. Prob, 23, 409-417.

Thavaneswaran A. (1986) Model reference adaptive system estimates for counting processes. Statistica Neerlandica, 40,. 65-72.