Journal of the Korean Statistical Society (1998), **27**: *2, pp 221–235*

A Quasi-Likelihood Approach to Nonlinear Filtering Problems †

Yoon Tae Kim[1]

ABSTRACT

Suppose that an observed process can be written as the additive model of the signal process and the noise process with unknown parameters. In practice the signal process is not directly observed. We consider the problem of estimating parameter from the observation process using the quasi-likelihood method.

Keywords: Quasi-likelihood method; Nonlinear filtering model; Innovation process; Predictable compensator; Dual predictable projection

1. INTRODUCTION

The standard method of estimation for parameters in diffusion models involves the maximum likelihood method. From Girsanov's theorem the Radon-Nikodym derivative of the measures, induced by the observed processes corresponding to the parameters, can be calculated (e.g. see Liptser and Shiryaev (1977)). However, if the Radon-Nikodym derivative does not exist, is not known, or is difficult to calculate, then the likelihood method is not available. To deal with these situations, a new approach of the Quasi-Likelihood (Q.L.) method enables us to obtain estimators under very general conditions. But the Quasi Likelihood Estimators (Q.L.E's) coincide with the maximum likelihood estimators in a context where the Radon-Nikodym derivative is available. A general framework for this new approach has been constructed in connection with conventional statistical problems by Godambe and Heyde (1987).

In the estimation of parameters in nonlinear filtering models, the method of maximum likelihood is frequently difficult and sometimes impossible to apply. The latter case occurs when probability measures corresponding to two different parameters values are mutually singular or not absolutely continuous. In Section

†This research was supported by RESEARCH FUND, Hallym University.
[1]Department of Statistics, Hallym University, Chunchon 200-702, Korea.

2, we refer, in paticular, to the work of Hutton and Nelson (1986) and Heyde (1994) among others which are closely related with the application of the Q.L. method to the nonlinear filtering problem. In Section 3 and 4, we present the Q.L. method as applied to the nonlinear filtering problems and prove the theorem concerning the derivation of the Quasi Score(Q.S.) estimating function. Also we give several examples of parameter estimation illustrating an important fact of the Q.L. method.

2. REVIEW OF PREVIOUS WORKS

We begin with closely related examples considered by Hutton and Nelson (1986), and by Heyde (1994). Hutton and Nelson construct the Q.S. estimating function when the observed process is a special semimartingale, i.e., the finite variation process in a decomposition of semimartingale is predictable. They introduce the Q.S. estimating function within a specified class of estimating functions and then prove that the proposed estimating function satisfies an optimal criterion. Heyde derives the Q.S. estimating function within a class of estimating functions for the model which can be written in the semimartingale form.

The work of Hutton and Nelson: Let $\{X_t, 0 \le t \le T\}$ be a continuous time d-dimensional stochastic process whose possible distributions are denoted by $(P_\theta), \theta \in \Theta \subseteq R^r$ such that $(\Omega, \mathcal{F}, P_\theta)$ is a complete probability space. Let $(\mathcal{F}_t^X)$ denote the natural filtration. Suppose that for each P_θ, the observed process $\{X_t, 0 \le t \le T\}$ is representable in the form

$$X_t = \int_0^t f_s(\theta)\, d\lambda_s + \nu_t(\theta), \text{ for all } t \in [0,T] \text{ and } \theta \in \Theta. \tag{2.1}$$

Here (λ_t) is a real, monotone increasing, right-continuous process with $\lambda_0 = 0$, while $(\nu_t(\theta), \mathcal{F}_t^X)$ is a *càdlàg* (right-continuous with left limit), locally square integrable martingale with the quadratic characteristic (or angle bracket) $< \nu(\theta) >_t$ given by

$$< \nu(\theta) >_t = \int_0^t \alpha_s(\theta)\, d\lambda_s \ , \tag{2.2}$$

where the process (λ_t), $\left(\alpha_t^{i,j}(\theta)\right)_{1 \le i,j \le d}$ and $(f_t^i(\theta))_{1 \le i \le d}$ are $(\mathcal{F}_t^X)$-predictable processes and $(f_t(\theta))$ is almost surely continuously differentiable with respect to

θ. This is the setting used by Hutton and Nelson (1986) and it covers many continuous time stochastic models.

They introduce the Q.S. estimating function $G_T^*(\theta)$ for (2.1) by defining

$$G_T^*(\theta) := \int_0^T \dot{f}_t(\theta)\alpha_t^+(\theta)\,dX_t - \int_0^T \dot{f}_t(\theta)\alpha_t^+(\theta)f_t(\theta)\,d\lambda_t \tag{2.3}$$

where $\alpha_t^+(\theta)$ denotes its Moore-Penrose inverse. From (2.1) and (2.3), it follows that

$$G_T^*(\theta) := \int_0^T \dot{f}_t(\theta)\alpha_t^+(\theta)\,d\nu_t(\theta) \text{ a.e. } P_\theta\ .$$

Then they prove that $G_T^*(\theta)$ is the Q.S. estimating function within the class of estimating functions $\mathcal{H}$ which will be precisely defined in Section 4 in connection with the filtering problems.

Heyde's Work: The model which Heyde (1994) considers can be written in the semimartingale form

$$dX_t = dA_t(\theta) + dM_t(\theta) \tag{2.4}$$

where the finite variation process $(A_t(\theta)), t \in$ [0,T], can be interpreted as the signal and a locally square integrable martingale $(M_t(\theta)), t \in$ [0,T], can be interpreted as the noise. Here he suggests that the Q.S. estimating function based on $\mathcal{H}$ is given by

$$\int_0^T d\bar{M}_t(\theta)(d < M(\theta) >_t)^+\,dM_t(\theta) \tag{2.5}$$

where $d\bar{M}_t(\theta) = E[d\dot{M}_t(\theta)|\mathcal{F}_{t-}]$, $\mathcal{F}_t$ is a filtration of past-history σ-field. But we can see that the integral (2.5) is not well defined. From (2.4), (2.5) can be written as

$$\int_0^T E(d\dot{A}_t(\theta)|\mathcal{F}_{t-})(d < M(\theta) >_t)^+\,(dA_t(\theta) - dX_t)\ . \tag{2.6}$$

Note that the integral (2.6) contains the signal process $(A_t(\theta))$. However, in practice, the signal process $(A_t(\theta))$ is not directly observable even if the true value of θ is known.

3. THE NONLINEAR FILTERING MODEL

Let $(\Omega, \mathcal{F}, P_\theta)$ be a complete probability space, $(\mathcal{F}_t), t \in [0,T]$, is an increasing family of sub-σ-fields of $\mathcal{F}$ and all P_θ-null sets belong to $\mathcal{F}_0$ where $\theta = (\theta_1, \dots, \theta_r) \in \Theta \subseteq R^r$.

Suppose that an observation process (X_t) can be written as

$$dX_t = dS_t(\theta) + dB_t(\theta) \tag{3.1}$$

where $(S_t(\theta))$ will be called the signal or system process and $(B_t(\theta))$ is the noise process. The signal $(S_t(\theta))$ is not directly observed in practice. Information concerning $(S_t(\theta))$ is obtained by observations on (X_t).

Let us make the following assumptions about the three processes (X_t), $(S_t(\theta))$ and $(B_t(\theta))$: for all $\theta \in \Theta$,

(C1) (X_t) is $(\mathcal{F}_t)$-adapted.

(C2) $S_0(\theta) = 0$ and $(S_t(\theta))$ has right-continuous sample functions and is of bounded variation (B.V.) in [0,T].

(C3) $E(Var_T S(\theta))^2 < \infty$, where $Var_T S(\theta)$ is the total variation of $S_u(\theta)$ in [0,T].

(C4) $B_0(\theta) = 0$, for every $t \in [0,T]$, $B_t(\theta)$ is square integrable and for $s < t$,

$$E[B_t(\theta) - B_s(\theta) | \mathcal{F}_s] = 0. \tag{3.2}$$

Note that it is not assumed that $(B_t(\theta))$ is $(\mathcal{F}_t)$-adapted. If that were so, $(S_t(\theta))$ would be $(\mathcal{F}_t)$-adaptable, i.e., observable.

(C5) $(B_t(\theta))$ has right-continuous paths.

The proof of the next result is to be found in Kallianpur (1980) (see p51).

Theorem 3.1 (Kallianpur) *Let $(\mathcal{F}_t), t \in [0,T]$, be an increasing family not assumed to be right-continuous. Let $\tilde{\mathcal{F}}_t = \mathcal{F}_{t+}$. Let $(U_t), t \in [0,T]$, be an integrable, increasing process not necessarily adapted to $(\mathcal{F}_t)$. Then there exists an integrable, increasing process $(\bar{U}_t)$ such that*

(i) $\bar{U}_t$ *is $(\mathcal{F}_t)$-measurable for each $t \in [0,T]$.*

(ii) $E[(U_t - U_s)|\mathcal{F}_s] = E[(\bar{U}_t - \bar{U}_s)|\mathcal{F}_s]$, *for $s, t \in [0,T]$, $s \le t$.*

Let us fix $\theta \in \Theta$. Write $S(\theta) = U(\theta) - V(\theta)$, where $U(\theta) = (U_t(\theta))$ and $V(\theta) = (V_t(\theta))$ are increasing processes, right-continuous and such that $U_0(\theta) = V_0(\theta) = 0$ (a.s.), $EU_t(\theta) < \infty$ and $EV_t(\theta) < \infty$ for each $t \in [0, T]$. From Theorem 3.1, we have

$$E[(U_T(\theta) - U_t(\theta))|\mathcal{F}_{t+}] = M_t(\theta) - \bar{U}_t(\theta) \tag{3.3}$$

where $(M_t(\theta))$ is $(\mathcal{F}_{t+})$-martingale with right-continuous sample paths and $(\bar{U}_t(\theta))$ is the uniquely determined, integrable increasing process which is predictable relative to $(\mathcal{F}_{t+})$. Similarly, we can obtain $V_t(\theta)$. Let $\bar{S}_t(\theta) = \bar{U}_t(\theta) - \bar{V}_t(\theta)$, where sometimes $(\bar{U}_t)$ is called the predictable compensator of (U_t), or the dual predictable projection of (U_t). Then define

$$\nu_t(\theta) = X_t - \bar{S}_t(\theta) \tag{3.4}$$

which is called the innovation process. To establish the basic propositionerties of the innovation process $(\nu_t(\theta))$, we need the following Lemmas whose proofs are given in Kallianpur (1980) (see p 195).

Lemma 3.1 (Kallianpur) *Let $(A_t), (0 \le t \le T)$ with $A_0 = 0$ be an increasing process (not necessarily adapted to $(\tilde{\mathcal{F}}_t)$), where $\tilde{\mathcal{F}}_t = \mathcal{F}_{t+}$. Suppose that $(\bar{A}_t)$ is the dual predictable projection of (A_t) relative to $(\tilde{\mathcal{F}}_t)$. If $E(A_T^2) < \infty$, then $E(\bar{A}_T^2) < \infty$.*

Lemma 3.2 (Kallianpur) *Let the integrable, increasing process $(A_t), (0 \le t \le T)$ with $A_0 = 0$ be a continuous process (not necessarily adapted to $(\tilde{\mathcal{F}}_t)$), where $\tilde{\mathcal{F}}_t = \mathcal{F}_{t+}$. Suppose that $(\bar{A}_t)$ is the dual predictable projection of (A_t) relative to $(\tilde{\mathcal{F}}_t)$. If $\bar{A}_T \le C < \infty$ (a.s.), then $E(A_T^2) < \infty$.*

Let the assumptions $(C1) - (C5)$ be satisfied in the model (3.1). Then the innovation process is a square integrable martingale.

Theorem 3.2. *$(\nu_t(\theta), \mathcal{F}_t), t \in [0, T]$, is a square integrable martingale with right continuous sample paths.*

Proof: For the dual predictable projection $\bar{S}_t(\theta)$ of $S_t(\theta)$, we have $E(Var_T\bar{S}_t(\theta))^2 < \infty$ by the assumption (C2) and Lemma 3.1. For $s < t$,

$$\begin{aligned} & E[\nu_t(\theta) - \nu_s(\theta)|\mathcal{F}_s] \\ = \; & E[S_t(\theta) - S_s(\theta)|\mathcal{F}_s] - E[\bar{S}_t(\theta) - \bar{S}_s(\theta)|\mathcal{F}_s] + E[B_t(\theta) - B_s(\theta)|\mathcal{F}_s] \\ = \; & 0. \end{aligned}$$

Moreover for every $t \in [0,T]$,

$$E[\nu_t(\theta)^2] \leq 3\left(E[S_t(\theta)^2] + E[\bar{S}_t(\theta)^2] + E[B_t(\theta)^2]\right).$$

Note that $E(Var_T S(\theta))^2 < \infty$ and $E(Var_T \bar{S}(\theta))^2 < \infty$ imply $E(S_t(\theta))^2 < \infty$ and $E(\bar{S}_t(\theta))^2 < \infty$, respectively, so that $E[\nu_t(\theta)^2] < \infty$ with the assumption (C4). Since $\bar{S}_t(\theta)$ is $(\mathcal{F}_{t+})$-predictable, it follows that $\nu_t(\theta)$ is $(\mathcal{F}_t)$-adapted. Hence the result follows. □

Remark 3.1: It will be assumed that (ν_t) is a locally square integrable martingale in the next section. Actually, if $S_t(\theta)$ is continuous and $E\left(Var_T S_t(\theta)\right) < \infty$ (not necessarily $E(Var_T S_t(\theta))^2 < \infty$), the innovation process is a locally square integrable martingale.

Proposition 3.1. *Let the above assumptions be satisfied except* $(C3)$. *If*

(*i*) $S_t(\theta)$ *is continuous,*

(*ii*) $E\left(Var_T S_t(\theta)\right) < \infty$ (*not necessarily* $E(Var_T S_t(\theta))^2 < \infty$),

then $(\nu_t(\theta), \mathcal{F}_t), t \in [0,T]$, *is a locally square integrable martingale with right continuous sample paths.*

Proof: Define the stopping times

$$\tau_N = \begin{cases} \inf\{t \leq T : Var_t \bar{S}(\theta) > N\} & \\ T & \text{if the above set is empty} \end{cases} \tag{3.5}$$

and let $B_t^N(\theta) = B_{t\wedge\tau_N}(\theta)$. Define $S_t^N(\theta) = S_{t\wedge\tau_N}(\theta)$ and $\bar{S}_t^N(\theta) = \bar{S}_{t\wedge\tau_N}(\theta)$. Then it is obvious that $(S_t^N(\theta))$ is of bounded variation in [0,T] and that $(\bar{S}_t^N(\theta))$ is the dual predictable projection relative to $\tilde{\mathcal{F}}_t$ of $(S_t^N(\theta))$. Hence we have

$$E[S_T^N(\theta) - S_t^N(\theta)|\tilde{\mathcal{F}}_t^N] = E[\bar{S}_T^N(\theta) - \bar{S}_t^N(\theta)|\tilde{\mathcal{F}}_t^N], \tag{3.6}$$

where $\tilde{\mathcal{F}}_t^N = \tilde{\mathcal{F}}_{t\wedge\tau_N}$. Let us define $X_t^N = X_{t\wedge\tau_N}$. Then, defining $(\nu_t^N(\theta))$ by

$$\nu_t^N(\theta) = X_t^N - \bar{S}_t^N(\theta), \tag{3.7}$$

we see that $\nu_t^N(\theta) = \nu_{t\wedge\tau_N}(\theta)$. Now note that $S_t^N(\theta) = U_t^N(\theta) - V_t^N(\theta)$ and $\bar{S}_t^N(\theta) = \bar{U}_t^N(\theta) - \bar{V}_t^N(\theta)$. From the definition of τ_N, it follows that $Var_T \bar{S}^N(\theta) \leq$

N, and so $\bar{U}_T^N(\theta) \leq N$ and $\bar{V}_T^N(\theta) \leq N$. It follows from the assumption (C4) and Lemma 3.2 that for every $t \in [0, T]$,

$$\begin{aligned} E[\nu_t^N(\theta)^2] &= E\left[(S_t^N(\theta) - \bar{S}_t^N(\theta) + B_t^N(\theta))^2\right] \\ &\leq K(E[U_T^N(\theta)^2] + E[V_T^N(\theta)^2] + E[\bar{U}_T^N(\theta)^2] + E[\bar{V}_T^N(\theta)^2] \\ &\quad + E[B_t^N(\theta)^2]) \\ &< \infty. \end{aligned} \tag{3.8}$$

Note that $(\tilde{\mathcal{F}}_t^N)$-predictability of $(\bar{S}_t^N(\theta))$ yields the fact that $(\bar{S}_t^N(\theta))$ is $(\mathcal{F}_t^N)$-adapted. Since (X_t) is $(\mathcal{F}_t)$-adapted and right-continuous, (X_t) is progressively measurable, and so (X_t^N) is $(\mathcal{F}_t^N)$-adapted. Hence $(\nu_t^N(\theta))$ is $(\mathcal{F}_t^N)$-adapted. Since $(\nu_t(\theta), \mathcal{F}_t)$ is a right-continuous martingale, it follows that $(\nu_t(\theta), \tilde{\mathcal{F}}_t)$ is a right-continuous martingale and hence $\nu_t(\theta) = E[\nu_T(\theta)|\tilde{\mathcal{F}}_t]$, which in turn implies that $(\nu_t(\theta))$ is uniformly integrable. By the optional sampling theorem, $\nu_s^N(\theta) = E[\nu_T^N(\theta)|\tilde{\mathcal{F}}_s^N]$. Hence $(\nu_t^N(\theta), \mathcal{F}_t^N)$ is a right-continuous martingale. Thus it follows from (3.9) that $(\nu_t^N(\theta), \mathcal{F}_t^N)$ is a square integrable martingale and so, $(\nu_t^N(\theta), \mathcal{F}_t)$ is a square integrable martingale. Hence $(\nu_t(\theta), \mathcal{F}_t)$ is a locally square integrable martingale.

□

Remark 3.2: To prove that $(\nu_t^i(\theta), \mathcal{F}_t)$ $(i = 1, \ldots, d)$ is a locally square integrable martingale, we use the same method as in the one-dimensional case. Only the definition of τ_N has to be slightly modified as follows:

$$\tau_N = \begin{cases} \inf\{t \leq T : |Var_t\bar{S}(\theta)| > N\} & \\ T & \text{if the above set is empty,} \end{cases}$$

where $|Var_t\bar{S}(\theta)|^2 = \sum_{i=1}^d (Var_t\bar{S}^i(\theta))^2$.

4. THE QUASI SCORE ESTIMATING FUNCTION

Our goal is to make inference about θ based on an d-dimensional observation process $(X_t), t \in [0, T]$, whose components satisfy (3.1). The dual predictable projection of $S_t(\theta)$ is $\bar{S}_t(\theta) = (\bar{S}_t^i(\theta))$ $(i = 1, \ldots, d)$, where $\bar{S}_t^i(\theta)$ is the dual predictable projection of the ith component $(S_t^i(\theta))$ of $(S_t(\theta))$. Now the d-dimensional observation process $(X_t), t \in [0, T]$, is representable in the form

$$dX_t = d\bar{S}_t(\theta) + d\nu_t(\theta). \tag{4.1}$$

where $(\nu_t^i(\theta), \mathcal{F}_t)$ $(i = 1, \dots, d)$ is a locally square integrable martingale. We consider parameter estimation problems for the model (4.1). Here the σ-field $\mathcal{F}_t$ is chosen to be the observation σ-field $\mathcal{F}_t^X$, satisfying the usual conditions. In order to derive the Q.S. estimating function on a suitably choosen sub-class of $\mathcal{G}$, we define the following class of estimating functions:

Definition 4.1. *An estimating function $G_T(\theta)$ is said to belong to class $\mathcal{H}$ if for all $t \in [0,T]$ and for all $\theta \in \Theta$, $G_T(\theta)$ can be expressed in the form*

$$G_T(\theta) = \left(\sum_{l=1}^{d} \int_0^T a_t^{1,l}(\theta)\, d\nu_t^l(\theta), \dots, \sum_{l=1}^{d} \int_0^T a_t^{r,l}(\theta)\, d\nu_t^l(\theta) \right) \tag{4.2}$$

where the following conditions are satisfied:

(i) $\left(a_t^{i,j}(\theta)\right)_{\substack{1\le i\le r\\ 1\le j\le d}}$ *is a $(\mathcal{F}_t^X)$-predictable process and is almost surely differentiable with respect to θ.*

(ii) $E[G_T(\theta)G_T(\theta)']$ *is nonsingular.*

(iii) For $i = 1, \dots, r$ and $j = 1, \dots, d$,

$$E\left[\int_0^T a_t^{i,j}(\theta)^2\, d < \nu^j(\theta) >_t\right] < \infty. \tag{4.3}$$

(iv) $E[\dot{G}_T(\theta)]$ *is nonsingular and*

$$\left(E\left[\frac{\partial G_{T,i}(\theta)}{\partial \theta_j}\right]\right)_{1\le i,j\le r} = -\left(E\left[\sum_{l=1}^{d} \int_0^T a_t^{i,l}(\theta)\, d_t \frac{\bar{S}_t^l(\theta)}{\partial \theta_j}\right]\right)_{1\le i,j\le r}. \tag{4.4}$$

Remark 4.3: From the condition (4.3) and propositionerties of the stochastic integral with respect to a locally square integrable martingale, $(G_t(\theta), \mathcal{F}_t^X)$ is a square integrable martingale. It follows from the above conditions that $\mathcal{H} = \{G_T(\theta)\} \subseteq \mathcal{G}$.

The following Lemma, whose proof is given in Kim (1998), will be used to obtain the Q.S. estimating function within $\mathcal{H}$.

Lemma 4.1. *Suppose that $(h_{t,i}(\theta)), t \in [0,T]$ and $\theta \in \Theta$, is a family of random variables for $i = 1, 2$ such that for all $\theta \in \Theta$,* (*i*) $(h_{t,i}(\theta))$ *is* $\mathcal{B}[0,T]\otimes\mathcal{F}$*-measurable.*

(*ii*) $(h_{t,i}(\theta))$ *is square integrable for each* $t\in[0,T]$ *such that*

$$E\left[\int_0^T |h_{t,i}(\theta)|^2\, d<\nu^i(\theta)>_t\right]<\infty,\ \text{for all } \theta\in\Theta \tag{4.5}$$

where $(\nu_t^i(\theta))$, $i=1,2$, *are right-continuous locally square integrable martingales with* $\nu_0^i(\theta)=0$.

Then

$$\begin{aligned} & E\left[\int_0^T E[h_{t,1}(\theta)|\mathcal{F}_{t-}]h_{t,2}(\theta)\, d<\nu^1(\theta),\nu^2(\theta)>_t\right] \\ = \ & E\left[\int_0^T E[h_{t,1}(\theta)|\mathcal{F}_{t-}]\, E[h_{t,2}(\theta)|\mathcal{F}_{t-}]d<\nu^1(\theta),\nu^2(\theta)>_t\right], \end{aligned} \tag{4.6}$$

for all $\theta\in\Theta$.

With the aid of Lemma 4.1, we now derive the Q.S. estimating function within $\mathcal{H}$.

Theorem 4.1. *Suppose that for all* $t\in[0,T]$, *the equations*

$$\frac{\partial\bar{S}_t^j(\theta)}{\partial\theta_k}=\sum_{l=1}^d\int_0^t h_s^{k,l}(\theta)\, d<\nu^j(\theta),\nu^l(\theta)>_s \tag{4.7}$$

for $k=1,\dots,r$ *and* $j=1,\dots,d$ *has a solution* $(h_t(\theta))=\left(h_t^{k,j}(\theta)\right)_{\substack{1\le k\le r\\ 1\le j\le d}}$ *with probability one. Further, assume that the solution* $(h_t(\theta))$ *satisfies: for all* $\theta\in\Theta$,

(*i*) $E[G_T(\theta)G_T(\theta)']$ *is nonsingular where*

$$G_T(\theta)=\left(\sum_{j=1}^d\int_0^T E(h_t^{1,j}(\theta)|\mathcal{F}_{t-}^X)\, d\nu_t^j(\theta),\dots,\sum_{j=1}^d\int_0^T E(h_t^{r,r}(\theta)|\mathcal{F}_{t-}^X)\, d\nu_t^j(\theta)\right)$$

(*ii*) *For* $k=1,\dots,r$ *and* $j=1,\dots,d$

$$E\left[\int_0^T h_t^{k,j}(\theta)^2\, d<\nu^j(\theta)>_t\right]<\infty.$$

(iii) For $k = 1, \ldots, r$ and $j = 1, \ldots, d$

$$E\left[\frac{\partial}{\partial\theta_i}\int_0^T E(h_t^{k,j}(\theta)|\mathcal{F}_{t-}^X)\,d\nu_t^j(\theta)\right] = -E\left[\int_0^T E(h_t^{k,j}(\theta)|\mathcal{F}_{t-}^X)\,d_t\frac{\partial\bar{S}_t^j(\theta)}{\partial\theta_i}\right].$$

Then the Q.S. estimating function within $\mathcal{H} = \{G_T(\theta)\}$ is given by

$$G_T^*(\theta) = \left(\sum_{j=1}^d\int_0^T E(h_t^{1,j}(\theta)|\mathcal{F}_{t-}^X)\,d\nu_t^j(\theta), \ldots, \sum_{j=1}^d\int_0^T E(h_t^{r,j}(\theta)|\mathcal{F}_{t-}^X)\,d\nu_t^j(\theta)\right). \tag{4.8}$$

Proof: It is obvious that $G_T^*(\theta) \in \mathcal{H} \subseteq \mathcal{G}$. By the definition of the Q.S. estimating function, we need to show that $I(G_T^*(\theta)) - I(G_T(\theta))$ is nonnegative-definite for all $G_T(\theta) \in \mathcal{H}$, where

$$I(G_T(\theta)) := (E[\dot{G}_T(\theta)])'\left(E[G_T(\theta)G_T(\theta)']\right)^{-1}(E[\dot{G}_T(\theta)]).$$

Note that $(G_t^*(\theta), \mathcal{F}_t^X)$ is a square integrable martingale and

$$\begin{aligned} & E\left[\int_0^T E(h_t^{k,j}(\theta)|\mathcal{F}_{t-}^X)\,d\nu_t^j(\theta)\int_0^T E(h_t^{k,j'}(\theta)|\mathcal{F}_{t-}^X)\,d\nu_t^{j'}(\theta)\right] \\ = \; & E\left[\int_0^T E(h_t^{k,j}(\theta)|\mathcal{F}_{t-}^X)E(h_t^{k,j'}(\theta)|\mathcal{F}_{t-}^X)\,d < \nu^j(\theta), \nu^{j'}(\theta) >_t\right], \end{aligned} \tag{4.9}$$

From the condition *(iii)* and Equation (4.7), it follows that

$$\begin{aligned} & E[\dot{G}_T^*(\theta)] \\ = \; & \left(E\left[\frac{\partial G_{T,i}^*(\theta)}{\partial\theta_j}\right]\right)_{1\le i,j\le r} \\ = \; & \left(E\left[\frac{\partial}{\partial\theta_j}\sum_{l=1}^d\int_0^T E(h_t^{i,l}(\theta)|\mathcal{F}_{t-}^X)\,d\nu_t^l(\theta)\right]\right)_{1\le i,j\le r} \\ = \; & -\left(E\left[\sum_{l=1}^d\int_0^T E(h_t^{i,l}(\theta)|\mathcal{F}_{t-}^X)\,d_t\frac{\partial\bar{S}_t^l(\theta)}{\partial\theta_j}\right]\right)_{1\le i,j\le r} \\ = \; & -\left(E\left[\sum_{l=1}^d\int_0^T E(h_t^{i,l}(\theta)|\mathcal{F}_{t-}^X)\sum_{l'=1}^d h_t^{j,l'}(\theta)\,d < \nu^l(\theta), \nu^{l'}(\theta) >_t\right]\right)_{1\le i,j\le r} \end{aligned} \tag{4.10}$$

Applying Lemma 4.1 to (4.10), we can show that (4.10) is equal to

$$-\left(E\left[\sum_{l=1}^d\sum_{l'=1}^d\int_0^T E(h_t^{i,l}(\theta)|\mathcal{F}_{t-}^X)E(h_t^{j,l'}(\theta)|\mathcal{F}_{t-})\,d < \nu^l(\theta), \nu^{l'}(\theta) >_t\right]\right)_{1\le i,j\le r}$$

It follows from Equation (4.9) that

$$\begin{aligned} & E[\dot{G}_T^*(\theta)] \\ = & -\left(E\left[\sum_{l=1}^d \int_0^T E(h_t^{i,l}(\theta)|\mathcal{F}_{t-}^X)\, d\nu_t^l(\theta) \sum_{l'=1}^d \int_0^T E(h_t^{j,l'}(\theta)|\mathcal{F}_{t-}^X)\, d\nu_t^{l'}(\theta)\right]\right)_{1\le i,j\le r} \end{aligned}$$

which is equal to $E[G_T^*(\theta)G_T^*(\theta)']$, i.e.,

$$E[\dot{G}_T^*(\theta)] = -E[G_T^*(\theta)G_T^*(\theta)']. \tag{4.11}$$

Hence we have $I(G_T^*(\theta)) = -E[G_T^*(\theta)G_T^*(\theta)']$. Let us choose $G_T(\theta) \in \mathcal{H}$ which is given by

$$G_T(\theta) = \left(\sum_{l=1}^d \int_0^T a_t^{1,1}(\theta)\, d\nu_t^l(\theta), \dots, \sum_{l=1}^d \int_0^T a_t^{r,l}(\theta)\, d\nu_t^j(\theta)\right).$$

As in the case of $E[\dot{G}_T^*(\theta)]$, we have

$$\begin{aligned} & E[\dot{G}_T(\theta)] \\ = & \left(E\left[\frac{\partial G_{T,i}(\theta)}{\partial \theta_j}\right]\right)_{1\le i,j\le r} \\ = & -\left(E\left[\sum_{l=1}^d \int_0^T a_t^{i,l}(\theta)\, d\nu_t^l(\theta) \sum_{l'=1}^d \int_0^T E(h_t^{j,l'}(\theta)|\mathcal{F}_{t-}^X)\, d\nu_t^{l'}(\theta)\right]\right)_{1\le i,j\le r} \end{aligned}$$

since $\left(a_t^{i,l}(\theta)\right)_{\substack{1\le i\le r \\ 1\le l\le d}}$ is $(\mathcal{F}_{t-})$-measurable for each $t > 0$. Hence

$$E[\dot{G}_T(\theta)] = -E[G_T(\theta)G_T^*(\theta)']. \tag{4.12}$$

The dispersion matrix of $\left(G_{T,1}(\theta), \dots, G_{T,r}(\theta), G_{T,1}^*(\theta), \dots, G_{T,r}^*(\theta)\right)'$ can be written as the partitioned matrix

$$D = \begin{pmatrix} E[G_T(\theta)G_T(\theta)'], & E[G_T(\theta)G_T^*(\theta)'] \\ (E[G_T(\theta)G_T^*(\theta)'])', & E[G_T^*(\theta)G_T^*(\theta)'] \end{pmatrix}$$

From (4.11) and (4.12), D is equal to

$$\begin{pmatrix} E[G_T(\theta)G_T(\theta)'], & -E[\dot{G}_T(\theta)] \\ -(E[\dot{G}_T(\theta)])', & E[G_T^*(\theta)G_T^*(\theta)'] \end{pmatrix}$$

Using the same techniques as in Rao (1965) (see 5a.3), we can show that

$$E[G_T(\theta)G_T(\theta)'] - E[\dot{G}_T(\theta)]\left(E[G_T^*(\theta)G_T^*(\theta)']\right)^{-1}\left(E[\dot{G}_T(\theta)]\right)'$$

is nonnegative-definite for all $G_T(\theta) \in \mathcal{H}$. Since all the matrices are nonsingular,

$$E[G_T^*(\theta)G_T^*(\theta)'] - E[\dot{G}_T(\theta)]\left(E[G_T(\theta)G_T(\theta)']\right)^{-1}\left(E[\dot{G}_T(\theta)]\right)'$$

also is nonnegative-definite, so that

$$I(G_T^*(\theta)) - I(G_T(\theta))$$

is nonnegative-definite for all $G_T(\theta) \in \mathcal{H}$. Hence the result follows. □

Examples will be presented to show the feasibility of Theorem 4.1.

Example 4.1: (The Filtering Problem in Standard Form) Assume that the signal process $(S_t(\theta)), 0 \le t \le T$, is of the form

$$S_t(\theta) = \int_0^t h_s(\theta)\, ds,$$

where $h_s(\theta)$ is progressively measurable,

$$\int_0^T E[h_s(\theta)^2]\, ds < \infty \quad \text{and} \quad \int_0^T E[\dot{h}_s(\theta)^2]\, ds < \infty.$$

Then Remark 8.1.1 in kallianpur (1980) yields the fact that

$$\bar{S}_t(\theta) = \int_0^t E(h_s(\theta)|\mathcal{F}_s^X)\, ds\ .$$

Hence the innovation process $\nu_t(\theta)$ is given by

$$\nu_t(\theta) = X_t - \int_0^t E(h_s(\theta)|\mathcal{F}_s^X)\, ds\ .$$

If we assume that $(B_t(\theta), \mathcal{G}_t)$ is a Wiener martingale, where $\mathcal{G}_t \supseteq \mathcal{F}_t^X$, then $(\nu_t(\theta), \mathcal{F}_t^X)$ is a Wiener martingale (see corollary 8.1.1 in Kallianpur (1980)). Now we will further assume that

(i) $\dot{\bar{S}}_t(\theta) = \int_0^t E[\dot{h}_s(\theta)|\mathcal{F}_{s-}^X]\, ds$ for almost all ω and all $t \in [0, T]$.

(ii) $\nu_t(\theta)$ is independent of θ. i.e., $< \nu(\theta) >_t = t$

First notice that

$$\dot{\bar{S}}_t(\theta) = \int_0^t E(\dot{h}_s(\theta)|\mathcal{F}_s^X)\, ds = \int_0^t E(\dot{h}_s(\theta)|\mathcal{F}_s^X)\, d < \nu(\theta) >_s .$$

From Theorem 4.1 we can show that the Q.S. estimating function within $\mathcal{H}$ has the form

$$\begin{aligned} G_T^*(\theta) &= \int_0^T E\left[E(\dot{h}_t(\theta)|\mathcal{F}_t^X)|\mathcal{F}_{t-}^X\right] d\nu_t(\theta) \\ &= \int_0^T E(\dot{h}_t(\theta)|\mathcal{F}_{t-}^X)(\, dX_t - E(\dot{h}_t(\theta)|\mathcal{F}_t^X)\, dt). \end{aligned}$$

Example 4.2: In Section 3, we see that the Q.S. estimating function for the model considered by Hütton and Nelson is given by

$$Q(\theta) := \int_0^T \dot{f}_t(\theta)\alpha_t^+(\theta)\, d\nu_t(\theta) .$$

Now we derive the Q.S. estimating function by using Theorem 4.1. For convenience, we consider the case of $d = 1$ and $r = 1$. The following conditions will be assumed:

(*i*) $|\dot{f}_t(\theta, \omega)| \leq g_t(\omega)$ for all $t \in (0, T)$ and almost all $\omega \in \Omega$, where $\int_0^T g_t\, d\lambda_s < \infty$ (a.s.).

(*ii*) $E\left[\int_0^T (\dot{f}_t(\theta)\alpha_t^+(\theta))^2\, d < \nu(\theta) >_t\right] < \infty$ for all $\theta \in \Theta$, where

$$\alpha_t^+(\theta) = \begin{cases} \alpha_t(\theta)^{-1} & \text{if } \alpha_t(\theta) > 0 \\ 0 & \text{if } \alpha_t(\theta) = 0. \end{cases}$$

Writing $\bar{S}_t(\theta) := \int_0^t f_s(\theta)\, d\lambda_s$, it then follows that

$$\dot{\bar{S}}_t(\theta) = \int_0^t \dot{f}_s(\theta)\, d\lambda_s = \int_0^t \dot{f}_s(\theta)\alpha_s^+(\theta)\, d < \nu(\theta) >_s$$

from (2.3) and the assumption (*i*). Hence the Q.S. estimating function within $\mathcal{H}$ is given by

$$\begin{aligned} G_T^*(\theta) &= \int_0^T E(\dot{f}_t(\theta)\alpha_s^+(\theta)|\mathcal{F}_{t-}^X)\, d\nu_t(\theta) \\ &= \int_0^T \dot{f}_t(\theta)\alpha_t^+(\theta)\, d\nu_t(\theta) \end{aligned}$$

because $(\dot{f}_t(\theta)\alpha_t^+(\theta))$ is $(\mathcal{F}_{t-}^X)$-measurable. Note that the stochastic integral $G_T^*(\theta)$ is well defined by the assumption (ii). The Q.L.E. $\hat{\theta}$ of θ is given by a solution to the estimating equation $G_T^*(\hat{\theta}) = 0$.

Example 4.3: The unknown real parameter θ is to be estimated from the observations $\{X_t, 0 \leq t \leq T\}$ with

$$dX_t = \theta X_{t-} d < \nu >_t + d\nu_t(\theta) \tag{4.13}$$

where $X_{t-} = \lim_{s \uparrow t} X_t$. Here $(\nu_t(\theta), \mathcal{F}_t^X)$ is a locally square integrable martingale and ν_t is independent of θ. This is the setting used by Liptser (1980)

The Least Squares Estimator (L.S.E.) $\hat{\theta}$ for θ is specified by the formula

$$\hat{\theta} = \frac{\int_0^T X_{t-}\, dX_t}{\int_0^T X_{t-}^2\, d < \nu >_t} . \tag{4.14}$$

We assume that

$$E\left[\int_0^T X_{t-}^2\, d < \nu >_t\right] < \infty. \tag{4.15}$$

Let $\bar{S}_t(\theta) = \int_0^t \theta X_{s-}\, d < \nu >_s$. Then since X_{t-} is $(\mathcal{F}_t)$-predictable and $\dot{\bar{S}}_t(\theta) = \int_0^t X_{s-}\, d < \nu >_s$, the Q.S. estimating function is given by

$$G_T^*(\theta) = \int_0^T X_{t-}\, d\nu_t(\theta) \tag{4.16}$$

which is well defined from (4.15). Hence the Q.L. equation for this model is given by

$$\int_0^T X_{t-}(dX_t - \theta X_{t-}\, d < \nu >_t) = 0. \tag{4.17}$$

Solving Equation (4.17), we obtain the Q.L.E. $\hat{\theta}$ of θ

$$\hat{\theta} = \frac{\int_0^T X_{t-}\, dX_t}{\int_0^T X_{t-}^2\, d < \nu >_t} ,$$

which is the same as the L.S.E.

REFERENCES

Godambe, V.P. and Heyde, C.C. (1987) . "Quasi-likelihood and optimal estimation," *International Statistical Review,* **55**, 231-244.

Heyde, C.C. (1994). "A quasi-likelihood approach to estimating parameters in diffusion-type processes," *Studies in Applied Probability* (*Journal of Applied Probability* Vol 31A), 283-290.

Hutton, J.E. and Nelson, P.I. (1986). "Quasi-likelihood estimation for semimartingales," *Stochastic Processes and their Applications,* **22**, 245-257.

Kallianpur, G. (1980). *Stochastic Filtering Theory*, Springer-Verlag.

Kim, Y.T. (1998). "Parameter Estimation in infinite-dimensional Stochastic Differential Equations," submitted for publication.

Liptser, R.S. (1980). "A strong law of large numbers for local martingales," *Stochastic*, 3:217-228.

Liptser, R.S. and Shiryaev, A.N. (1977). *Statistics of Random Peocesses*, Vol. 1 and 2, Springer-Verlag.

Rao, C.R. (1965). *Linear Statistical Inference and its Applications*, Wiley, New York.