Stochastics and Stochastics Reports, Vol. 56, pp. 63–74
Reprints available directly from the publisher
Photocopying permitted by license only

ON OPTIMALITY OF REGULAR PROJECTIVE ESTIMATORS FOR SEMIMARTINGALE MODELS III: ONE STEP IMPROVEMENTS

KACHA DZHAPARIDZE

Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam

PETER SPREIJ

Department of Econometrics, Free University, De Boelelaan 1105, 1081 HV Amsterdam

(*Recieved 11 November 1994; in final form 3 August 1995*)

In this paper we consider one step improvement techniques that yield optimal regular projective estimators.

KEY WORDS: Optimal estimator, asymptotic differentiability, one step improvement.

1 INTRODUCTION

In a previous paper [4] we discussed an optimality concept for certain parameter estimators, when the observed processes are semimartingales under each member of a class of probability measures. Estimators were there called optimal if they belong to the class of so called *admissible* estimators and if they have minimum *spread.*

In the present paper we pay attention to the question of constructing these optimal estimators in the sense that they are obtained by suitable transformations of some initial estimators. The latter are assumed to have certain consistency properties, but need not be admissible in the sense of [4].

This approach essentially dates back to Fisher [6] although a thorough investigation has first been given by LeCam for the iid case in [7] and for estimators based on more general likelihood functions in [8]. It is also the basis for the construction of iterative estimators, see e.g. [3].

Unlike the set up used by these authors, our approach is not likelihood based, but extends to a more general context, e.g. regression where usually apart from the regression function only moment conditions on the error process are specified. A key concept in the description of estimators is formed by certain (asymptotic) representations that estimators are supposed to fulfil.

[1]Centre for Mathematics and Computer Science, Kruislaan 413, 1098 SJ Amsterdam.
[2]Department of Econometrics, Free University, De Boelelaan 1105, 1081 HV Amsterdam.

The present paper is to be considered as a follow up of [4]. Hence, for undefined concepts and notation the reader is referred to [4] for further explanation.

The rest of the paper is organized as follows. In the next section we briefly review parametrizations of the model under consideration and provide some complements to the concept of *asymptotic weak differentiability*. Section 3 contains the main result which states that a particular estimator is optimal in the class of admissible estimators.

2 SMOOTH PARAMETRIZATIONS

2.1 *Parametrization*

Assume that one has a certain stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, where $\mathbb{P}$ is a set of probability measures and on this a multivariate adapted process X with values in $\mathbb{R}^d$, which we observe and which is assumed to be a semimartingale under each $P \in \mathbb{P}$. Denote by ν^P the compensator of the jump measure of X under P. Similarly A^P is the compensator of X under P and by $X^{c,P}$ or $\mathbb{M}^{c,P}$ we denote the continuous martingale part of X under P.

Like in [4] we assume to be given two classes of predictable processes $\mathscr{W}$ and $\mathscr{H}$, satisfying certain regularity properties, e.g. if $H \in \mathscr{H}$ then $W \in \mathscr{W}$ with $W(t,x) = H(t)x$. Notice that in the general set up of [4] the class of processes $\mathscr{W}$ is usually bigger than the one consisting of the processes W with $W(t,x) = H(t)x$ for $H \in \mathscr{H}$. This is trivially the case in a discrete time setting (see [4] for a detailed treatment). Moreover we assume the existence of a finite dimensional parametrization of the quotient space $[\mathbb{P}]$ of $\mathbb{P}$ under the equivalence relation defined by: P is equivalent to P' *iff* $H \cdot A^P = H \cdot A^{P'}$ for all $H \in \mathscr{H}$ and $W * \nu^P = W * \nu^{P'}$ for all $W \in \mathscr{W}$. That is, there exists a map $\vartheta : [\mathbb{P}] \to \mathbb{R}^k$, which is bijective onto an open subset Θ of $\mathbb{R}^k$. Under this assumption, which holds throughout the paper, we write $H \cdot A^\theta$ for $H \cdot A^P$ and $W * \nu^\theta$ for $W * \nu^P$ if $P \in \mathbb{P}^\theta := \vartheta^{-1}(\theta)$.

2.2 *Asymptotic weak differentiability, complements*

Next we turn to smoothness of the previously introduced parametrization. To that end we introduce the following notation. The tilde operator for given ν is defined for each $W \in \mathscr{W}$ by $\tilde{W} = W + (1-a)^+\hat{W}$ with $a = \hat{1}$, where in turn the hat operator on $\mathscr{W}$ is defined by $\hat{W}(t) = \int W(t,x)\nu(\{t\}, dx)$. When integrals of the type $W * \nu$ are parametrized by θ we will often write $\hat{W}^\theta$ and $\tilde{W}^\theta$.

For $H \in \mathscr{H}$ we write $H \cdot \mathbb{M}^{c,\theta}$ for $H \cdot \mathbb{M}^{c,P}$ if $P \in \mathbb{P}^\theta$. This is in agreement with the notation above, since $H \cdot \mathbb{M}^{c,P} = H \cdot X - H \cdot A^P - Hx * \mu + Hx * \nu^P = H \cdot X - H \cdot A^\theta - Hx * \mu + Hx * \nu^\theta$. For all $H \in \mathscr{H}$ and $W \in \mathscr{W}$ we write $M^\theta = M^\theta(H, W)$ for the martingale defined by $M^\theta = H \cdot \mathbb{M}^{c,\theta} + W * \mu - W * \nu^\theta$. $\bar{M}^\theta$ is the martingale, defined by $\bar{M}^\theta = b.\mathbb{M}^{c,\theta} + \widetilde{\lambda}^\theta * (\mu - \nu^\theta)$ (see definition 2.1 below).

We assume that $\langle M^\theta \rangle_t$ and $\langle \bar{M}^\theta \rangle_t$ are invertible for t large enough. Let then $\psi_t = \psi_t^\theta(H, W)$ be any matrix that satisfies the equality $\psi_t^T \psi_t = \langle M^\theta \rangle_t^{-1}$ and $\phi_t = \phi_t^\theta$ be any matrix that satisfies $\phi_t \phi_t^T = \langle \bar{M}^\theta \rangle_t^{-1}$. Similarly we write $\psi_t^W = \psi_t^\theta(0, W)$ and $\psi_t^H = \psi_t^\theta(H, 0)$.

We will assume that *asymptotic weak differentiability* in the sense of [4] holds (definition 2.1 below). Since the definition in [4] together with the discussion that follows it is ambiguous, we give a slightly different one, which is such that the ambiguity is removed. First we define $\mathring{A}^P = A^P - x * \nu^P$ and $H \cdot \mathring{A}^\theta = H \cdot A^\theta - Hx * \nu^\theta$.

Definition 2.1 The compensators A^θ and ν^θ are called asymptotically weakly differentiable with rate ϕ if there exist $b \in \mathscr{H}, \lambda \in \mathscr{W}$ (both possibly depending on θ) such that with ϕ satisfying $\phi\phi^T = \langle \tilde{M}^\theta \rangle^{-1}$ for all $u \in \mathbb{R}^m$, $H \in \mathscr{H}, W \in \mathscr{W}$ as $t \to \infty$ in all $\mathbb{P}^\theta$ probabilities:

$$\text{(i)} \qquad \psi_t^H \left[H \cdot \mathring{A}_t^{\theta+\phi_t u} - H \cdot \mathring{A}_t^\theta - \int_{[0,t]} Hd\langle \mathbb{M}^c \rangle b^T \phi_t u \right] \to 0 \tag{2.1}$$

$$\text{(ii)} \qquad \psi_t^W \left[W * \nu_t^{\theta+\phi_t u} - W * \nu_t^\theta - W\lambda^T * \nu_t^\theta \phi_t u \right] \to 0 \tag{2.2}$$

with ψ^H and ψ^W as above.

Although b, λ and $\mathbb{M}^c$ in general depend on θ, this dependence is often not explicitly written in order to avoid some cumbersome notation. Furthermore we often, especially in proofs, abbreviate the phrase "in all $\mathbb{P}^\theta$ probabilities" by "in probability". Notice that for two different parameter values θ and θ' the following relation holds.

$$M^{\theta'}(H, W) = M^\theta(H, W) + H \cdot (\mathring{A}^{\theta'} - \mathring{A}^\theta) + W * (\nu^{\theta'} - \nu^\theta) \tag{2.3}$$

Therefore we define the process $M_t^{\theta+\phi_t u}$ by substituting at time t in the right hand side of equation (2.3) $\theta + \phi_t u$ for θ'. If we write $M^\theta = M^\theta(H, W)$ and $\tilde{M}^\theta = M^\theta(b, \tilde{\lambda})$, like above, then definition 2.1 has an equivalent statement.

Proposition 2.2 *Expressions* (2.1) *and* (2.2) *in definition* 2.1 *hold for all H and W iff for all* $M^\theta = M^\theta(H, W)$

$$\psi_t[M_t^{\theta+\phi_t u} - M_t^\theta + \langle M^\theta, \tilde{M}^\theta \rangle_t \phi_t u] \to 0 \textit{ in all } \mathbb{P}^\theta \textit{ probabilities} \tag{2.4}$$

Proof Suppose that equation (2.4) holds for all martingales of the type $M^\theta = M^\theta(H, W)$, then we consider the two distinct cases $M^\theta(H, 0)$ and $M^\theta(0, W)$. This leads to equations (2.1) and (2.2) respectively. Conversely, assuming that (2.1) and (2.2) hold, we write the process in expression (2.4) by using (2.3) as

$$\psi_t(\psi_t^H)^{-1} \left[\psi_t^H \left[H \cdot \mathring{A}_t^{\theta+\phi_t u} - H \cdot \mathring{A}_t^\theta - \int_{[0,t]} Hd\langle \mathbb{M}^c \rangle b^T \phi_t u \right] \right]$$

$$+ \psi_t(\psi_t^W)^{-1} \left[\psi_t^W \left[W * \nu_t^{\theta+\phi_t u} - W * \nu_t^\theta - W\lambda^T * \nu_t^\theta \phi_t u \right] \right].$$

The result then follows, because $\psi_t(\psi_t^H)^{-1}$ and $\psi_t(\psi_t^W)^{-1}$ are bounded, because for instance $\psi_t(\psi_t^H)^{-1}(\psi_t^H)^{-T}\psi_t^T \leq I$. □

Recall that in definition 2.1 ϕ_t is any matrix that satisfies the equality $\phi_t\phi_t^T = \langle\tilde{M}\rangle_t^{-1}$. Obviously one wants that if asymptotic weak differentiability holds for a given (matrix valued) rate function ϕ it also holds for other asymptotically equivalent rate functions. For example, in the original definition in [4] we have taken the special ϕ_{0t} to be the symmetric positive square root of $\langle\tilde{M}\rangle_t^{-1}$. Then we have that this assumption holds for any other such ϕ *under the extra assumption* that we have definition 2.1 with u replaced with any adapted bounded random process $\{u_t\}$. Indeed, then with $u_t = \phi_{0t}^{-1}\phi_t u$ we have that $\phi_t u = \phi_{0t}u_t$ and $|u_t| = |u|$ since $\phi_{0t}^{-1}\phi_t$ is an orthogonal matrix.

Therefore we need the following assumption.

Assumption 2.2 *Expressions* (2.1) *and* (2.2) *in definition* 2.1 *hold for all H and W with u replaced by any random variable u_t such that the process $\{u_t\}$ is adapted and bounded. In particular all processes involved in* (2.1) *and* (2.2) *with u substituted by u_t are assumed to be adapted.*

An equivalent formulation of assumption 2.3 is of course that equation (2.4) holds with u replaced with u_t for an adapted bounded process $\{u_t\}$:

$$\psi_t[M_t^{\theta+\phi_t u_t} - M_t^\theta + \langle M^\theta, \tilde{M}^\theta\rangle_t\phi_t u_t] \to 0 \text{ in all } \mathbb{P}^\theta \text{ probabilities} \tag{2.5}$$

We mention a sufficient condition for equation (2.5) to hold. Let $B > 0$ and write

$$Z_t^\theta(B) = \sup_{|u|\leq B}\left|\psi_t\left[M_t^{\theta+\phi_t u} - M_t^\theta + \langle M^\theta, \tilde{M}^\theta\rangle_t\phi_t u\right]\right| \tag{2.6}$$

Then equation (2.5) holds if $Z_t^\theta(B)$ is measurable and $\lim_{t\to\infty}P(Z_t^\theta(B) > \varepsilon) = 0$ for all positive B and ε and for all $P \in \mathbb{P}^\theta$. See the appendix for a discussion of measurability issues connected with $Z_t^\theta(B)$ and with the substitution of u by a random u_t.

Let M_1 and M_2 be two locally square integrable martingales. We introduce (see [5]) the *correlation process* $\rho(M_1, M_2)$ as follows. Let ϕ_i be such that $\phi_i\phi_i^T = \langle M_i\rangle^+$ (Moore–Penrose inverse) for $i = 1, 2$. Then $\rho(M_1, M_2) = \phi_1^T\langle M_1, M_2\rangle\phi_2$. With this definition of the correlation process we can rephrase equation (2.4) as

$$\psi_t\left[M_t^{\theta+\phi_t u} - M_t^\theta\right] + \rho(M^\theta, \tilde{M}^\theta)u \to 0 \text{ in all } \mathbb{P}^\theta \text{ probabilities.}$$

The correlation process will show up again at various places in the sequel.

It should be noted that the weak derivatives b and λ are by no means unique. Suppose we have two other possible candidates for the weak derivatives, call them b^0 and λ^0. Correspondingly we have $\tilde{M}^{0\theta}$ instead of $\tilde{M}^\theta$. Let ϕ^0 be such that $\phi^0\phi^{0T} = \langle\tilde{M}^{0\theta}\rangle^{-1}$. So we assume that equations (2.1) and (2.1) hold with b, λ and ϕ replaced with b^0, λ^0 and ϕ^0, or equivalently equation (2.3) with $\tilde{M}^{0\theta}$ instead of $\tilde{M}^\theta$:

$$\psi_t[M_t^{\theta+\phi_t^0 u} - M_t^\theta + \langle M^\theta, \tilde{M}^{0\theta}\rangle_t\phi_t^0 u] \to 0 \text{ in all } \mathbb{P}^\theta \text{ probabilities.} \tag{2.7}$$

First we claim the following.

PROPOSITION 2.4 *Under assumption* 2.3 *there is no sequence* $\{t_n\}$ *in* $\mathbb{R}$, *tending to infinity, such that* $|(\phi^0_{t_n})^{-1}\phi_{t_n}| \wedge |\phi^{-1}_{t_n}\phi^0_{t_n}| \to 0$ *in all* $\mathbb{P}^\theta$ *probabilities.*

Proof Suppose that the contrary holds true. Then for some sequence $\{t_n\}$ we have for instance $(\phi^0_{t_n})^{-1}\phi_{t_n} \to 0$ in probability. Let $u_{t_n} = (\phi^0_{t_n})^{-1}\phi_{t_n}u$ for some fixed vector u. The sequence $\{u_{t_n}\}$ is clearly bounded. So we can insert this into equations (2.4) and (2.7) where we take for M the special choice $\tilde{M}^\theta$. Hence we get the following two convergence results in probability. Both

$$\phi^T_{t_n}\left[\tilde{M}^{\theta+\phi_{t_n}u}_{t_n} - \tilde{M}^\theta_{t_n} + \langle\tilde{M}^\theta\rangle_{t_n}\phi_{t_n}u\right] \to 0$$

and

$$\phi^T_{t_n}\left[\tilde{M}^{\theta+\phi_{t_n}u}_{t_n} - \tilde{M}^\theta_{t_n} + \langle\tilde{M}^\theta, \tilde{M}^{0\theta}\rangle_{t_n}\phi_{t_n}u\right] \to 0.$$

Substraction of the two yields $I - \rho(\tilde{M}^\theta, \tilde{M}^{0\theta})_{t_n}(\phi^0_{t_n})^{-1}\phi_{t_n} \to 0$. But this cannot happen since $\rho(\tilde{M}^\theta, \tilde{M}^{0\theta})_{t_n}\rho(\tilde{M}^{0\theta}, \tilde{M}^\theta)_{t_n} \leq I$. □

The conclusion is that ϕ and ϕ^0 are equivalent rate processes. Hence in equation (2.6) we can replace the ϕ^0 with ϕ. The next thing we will show is that the martingales $\tilde{M}^\theta$ and $\tilde{M}^{0\theta}$ are close in the following sense.

PROPOSITION 2.4 *In all* $\mathbb{P}^\theta$*-probabilities*

(i) $$\phi^T_t\langle\tilde{M}^\theta - \tilde{M}^{0\theta}\rangle_t\phi_t \to 0 \tag{2.8}$$

(ii) $$\phi^T_t\langle\tilde{M}^\theta, \tilde{M}^{0\theta}\rangle_t\phi_t \to I \tag{2.9}$$

(iii) $$\rho(\tilde{M}^\theta, \tilde{M}^{0\theta})\rho(\tilde{M}^{0\theta}, \tilde{M}^\theta) \to I \tag{2.10}$$

Proof Take in equations (2.4) and (2.7) M^θ to be equal to $\tilde{M}^\theta - \tilde{M}^{0\theta}$. Substraction of the two equations yields

$$\psi_t\langle\tilde{M}^\theta - \tilde{M}^{0\theta}\rangle\phi_t \to 0 \tag{2.11}$$

in probability, where $\psi^T_t\psi_t = \langle\tilde{M}^\theta - \tilde{M}^{0\theta}\rangle^+_t$. The first assertion follows by taking squares in equation (2.11).

Before we prove the other two assertions we introduce the short hand notation $\rho = \rho(\tilde{M}^\theta, \tilde{M}^{0\theta})$ and $R = (\phi^0)^{-1}\phi$. The process in the first assertion can then be written as

$$I - \rho R - R^T\rho^T + R^T R = (I - R^T\rho^T)(I - \rho R) + R^T(I - \rho^T\rho)R.$$

According to the first assertion, this process tends to zero in probability, and since it is the sum of two nonnegative processes, we obtain that the process $I - \rho R$ tends to zero in probability. But then $R^T_t R_t$ tends to I in probability.

Before proving the last assertion we rewrite the process in expression (2.8) as

$$(R_t^T - \rho_t)(R_t - \rho_t^T) + I - \rho_t \rho_t^T$$

from which it follows like above that $\rho_t \rho_t^T$ tends to I in probability. One can similarly prove that $\rho_t^T \rho_t$ tends to I in probability, because we can replace (2.8) with $\phi_t^{0T} \langle \tilde{M}^\theta - \tilde{M}^{0\theta} \rangle_t \phi_t^0 \to 0$. □

Remark The notation $R = (\phi^0)^{-1}\phi$ and $\rho = \rho(\tilde{M}^\theta, \tilde{M}^{0\theta})$ used in the above proof will be frequently used in the sequel.

Assume that some predictable processes b^0 and $\tilde{\lambda}^0$ are given and $\tilde{M}^{0\theta} = b^0.\mathbb{M}^{c\theta} + \tilde{\lambda}^0 * (\mu - \nu^\theta)$. Assume that the convergence in (2.8) takes place and that assumption 2.3 holds. Then also equation (2.7) is satisfied:

PROPOSITION 2.6 *Under assumption* 2.3 *and equation* (2.8) *also the convergence in* (2.7) *takes place.*

Proof As in the proof of the previous proposition, we know from equation (2.8) that $R = (\phi^0)^{-1}\phi$ is such that $R_t^T R_t$ tends to I in probability. Hence, under assumption 2.3 we may replace ϕ in equation (2.4) with ϕ^0. Hence the validity of (2.3) would follow from $\psi_t \langle M, \tilde{M}^{0\theta} - \tilde{M}^\theta \rangle_t \phi_t^0 u \to 0$ in probability. But this is guaranteed by the Kunita–Watanabe inequality since (2.8) holds by assumption. □

The interpretation is that under the assumptions made we can both use $\langle M^\theta, \tilde{M}^\theta \rangle$ and $\langle M^\theta, \tilde{M}^{0\theta} \rangle$ as a weak derivative of a martingale M^θ.

3 IMPROVED ESTIMATORS

Suppose that we are given an estimator θ^0 of θ, which is assumed to be an adapted process and moreover that is ϕ_t-consistent by which we mean that for all $\theta \in \Theta$ the process $\phi^{-1}(\theta^0 - \theta)$ is $\mathbb{P}^\theta$-tight:

$$\lim_{K \to \infty} \limsup_{t \to \infty} P(|\phi_t^{-1}(\theta_t^0 - \theta)| > K) = 0, \quad \text{for all} \quad P \in \mathbb{P}^\theta. \tag{3.1}$$

Notice however, that we don't require θ^0 to be admissible in the sense of [4].

Assume that we are also given an estimator Q_t of $\langle \tilde{M}^\theta \rangle^{-1}$ that is consistent in the following sense:

$$\phi_t^T Q_t^{-1} \phi_t \to I \tag{3.2}$$

in all $\mathbb{P}^\theta$-probabilities.

Similarly, we consider estimators b^0 of b and $\tilde{\lambda}^0$ of $\tilde{\lambda}^\theta$ that are assumed to be predictable processes belonging to $\mathscr{H}$ and $\mathscr{W}$ respectively, not depending on θ of course, that are consistent in the sense that if $\tilde{M}^{0\theta}$ is defined by $\tilde{M}^{0\theta} = M^\theta(b^0, \tilde{\lambda}^0) = b^0 \cdot \mathbb{M}^{c,\theta} + \tilde{\lambda}^0 * (\mu - \nu^\theta)$, then

$$\phi_t^T \langle \tilde{M}^\theta - \tilde{M}^{0\theta} \rangle \phi_t \to 0 \tag{3.3}$$

in all $\mathbb{P}^\theta$-probabilities. Notice that under the assumptions made, we can replace $\tilde{M}^\theta$ in equation (2.4) by $\tilde{M}^{0\theta}$ (cf. proposition 2.6).

Since λ as well as $\tilde{\lambda}^\theta$ in general depend on θ, a typical choice for $\widetilde{\lambda^0}_t$ is obtained by plugging in θ^0_{t-} in all places where θ appears in the expression for $\tilde{\lambda}^\theta$. Under fairly broad conditions (see the appendix) this yields the process $\widetilde{\lambda^0}$ predictable. A similar remark holds for b^0. Of course the consistency requirement has to be verified, but it often holds under a continuity condition. See the example at the end of this section. The next thing we do is the defining of an estimator $\hat{M}$ of the martingale $\tilde{M}^\theta$. $\hat{M}$ is, by the way, *not* a martingale itself. Recall equation (2.3) and define

$$\hat{M}_t = M_t^{\theta^0_t}(b^0, \widetilde{\lambda^0}) = M_t^\theta(b^0, \widetilde{\lambda^0}) + (b^0 \cdot (\mathring{A}^\theta - \mathring{A}^{\theta^0_t}))_t + (\widetilde{\lambda^0} * (\nu^\theta - \nu^{\theta^0_t}))_t. \qquad (3.4)$$

We impose the more stringent condition (2.6) mentioned in the previous section on the smoothness of the parametrization.

We will also need the following assumption on the asymptotic behaviour of all martingales of the form $M^\theta = M^\theta(H, W)$ for $H \in \mathscr{H}$ and $W \in \mathscr{W}$. If $\psi^T\psi = \langle M^\theta \rangle^{-1}$, then

$$\text{All the processes } \psi M^\theta \text{ are } \mathbb{P}^\theta\text{-tight.} \qquad (3.5)$$

We recall from Dzhaparidze & Spreij [4] the definition of an optimal regular estimator $\hat{\theta}$ of θ. It is such that its spread attains the lower bound $\langle \tilde{M}^\theta \rangle^{-1}$. Such an estimator is characterized (cf. [4], proposition 7.1.2) by satisfying

$$\langle \tilde{M}^\theta \rangle (\hat{\theta} - \theta) = \tilde{M}^\theta + \eta^\theta \qquad (3.6)$$

with $\phi_t^T \eta_t^\theta \to 0$ in all $\mathbb{P}^\theta$-probabilities.

Below we will need an auxiliary result on the representation of an optimal estimator under the condition that (3.3) or (2.8) holds (which roughly speaking implies that we can often replace $\tilde{M}^\theta$ with $\tilde{M}^{0\theta}$). It is the content of the following proposition.

PROPOSITION 3.1 *Let an estimator $\hat{\theta}$ be representable as*

$$\langle \tilde{M}^{0\theta} \rangle (\hat{\theta} - \theta) = \tilde{M}^{0\theta} + \eta^\theta \qquad (3.7)$$

with $\phi_t^0 T \eta_t^\theta \to 0$ in all $\mathbb{P}^\theta$-probabilities. Then $\hat{\theta}$ is a regular estimator and optimal in the sense that its spread asymptotically equals the lower bound $\langle \tilde{M}^\theta \rangle^{-1}$.

Proof Proving regularity is equivalent to proving the following statement (see section 6 in [4]): $\phi_t^{0T} \langle \tilde{M}^{0\theta}, \tilde{M}^{0\theta} - \tilde{M}^\theta \rangle_t \phi_t \to 0$. Using the notation of the previous section, this statement can be written as $R_t - \rho_t^T \to 0$. But we proved this already in the previous section (proposition 2.5).

According to its definition (see [4]) the spread of $\hat{\theta}$ is $\langle \tilde{M}^{0\theta} \rangle^{-1} = \phi^0 \phi^{0T} = \phi (R^T R)^{-1} \phi^T$ and the optimal spread is $\langle \tilde{M}^\theta \rangle^{-1} = \phi \phi^T$, which are asymptotically equal since $R^T R$ tends to I (see the proof of proposition 2.5). □

The main result of the paper is the following.

THEOREM 3.2 *Assume that* (2.6) *holds. Define the estimator* $\hat{\theta}_t$ *by*

$$\hat{\theta}_t = \theta_t^0 + Q_t \hat{M}_t. \tag{3.8}$$

Then $\hat{\theta}_t$ *is optimal in the sense that it satisfies equation* (3.7).

Proof We have to show that η^θ defined by

$$\eta^\theta = \langle \tilde{M}^{0\theta} \rangle (\hat{\theta} - \theta) - \tilde{M}^{0\theta} \tag{3.9}$$

is such that $\phi_t^{0T} \eta_t^\theta \to 0$ in all $\mathbb{P}^\theta$-probabilities. The proof is divided into a number of steps.

Step 1 Let ϕ^0 be such that $\phi^0(\phi^0)^{-T} = \langle \tilde{M}^{0\theta} \rangle$. Define ε^θ as $\varepsilon^\theta = \phi^{0T}[\hat{M} - \tilde{M}^{0\theta} + \langle \tilde{M}^{0\theta} \rangle (\theta_t^0 - \theta)]$. We claim that $\varepsilon_t^\theta \to 0$ in all $\mathbb{P}^\theta$-probabilities. So consider for $P \in \mathbb{P}^\theta P(|\varepsilon_t^\theta| > \delta) \leq P(|\varepsilon_t^\theta| > \delta, |u_t| \leq B) + P(|u_t| > B)$ with $u_t = (\phi_t^0)^{-1}(\theta_t^0 - \theta)$. Notice that $\{u_t\}$ is tight since $u_t = R_t(\phi_t)^{-1}(\theta_t^0 - \theta)$ and $R_t^T R_t \to I$. (Here and elsewhere in this proof convergence is always to be understood as convergence in all the $\mathbb{P}^\theta$-probabilities). The last probability can be made arbitrarily small by choosing B large enough on view of equation (3.1), whereas the former one tends to zero for any B in view of (2.6) with $H = b^0$ and $W = \tilde{\lambda}^0$.

Step 2 Consider the asymptotic behaviour of Q. Clearly not only (3.2) holds, but also $\phi_t^{-1} Q_t \phi_t^{-T} \to I$ and $(\phi_t^0)^{-1} Q_t (\phi^0)_t^{-T} \to I$. To see the latter, consider $(\phi_t^0)^{-1} Q_t (\phi^0)_t^{-T} - I = R_t(\phi_t^{-1} Q_t \phi_t^{-T} - (R_t^T R_t)^{-1}) R_t^T$. Call the term in outer parentheses α_t, then $\alpha_t \to 0$. Consequently we have $0 \leq R_t \alpha_t R_t^T R_t \alpha_t^T R_t^T \leq \operatorname{tr}(R_t^T R_t)^2 \cdot \alpha_t \alpha_t^T \to 0$. So we can write

$$(\phi_t^0)^{-1} Q_t = (I + \delta_t) \phi_t^{0T} \tag{3.10}$$

with $\delta_t \to 0$.

Step 3 Use the result of step 1 to write

$$\begin{aligned}
\phi^{0T} \eta^\theta &= \phi^{0T}[\langle \tilde{M}^{0\theta} \rangle (\hat{\theta} - \theta) - \tilde{M}^{0\theta}] \\
&= \phi^{0T}[\langle \tilde{M}^{0\theta} \rangle (\theta^0 - \theta + Q\hat{M}) - \tilde{M}^{0\theta}] \\
&= \phi^{0T}[\langle \tilde{M}^{0\theta} \rangle (\theta^0 - \theta + Q((\phi^0)^{-T} \varepsilon^\theta + \tilde{M}^{0\theta} - \langle \tilde{M}^{0\theta} \rangle (\theta^0 - \theta))) - \tilde{M}^{0\theta}] \\
&= (\phi^0)^{-1} Q (\phi^0)^{-T} \varepsilon^\theta + (\phi^0)^{-1}[I - Q\langle \tilde{M}^{0\theta} \rangle](\theta^0 - \theta) + [(\phi^0)^{-1} Q - \phi^{0T}] \tilde{M}^{0\theta}.
\end{aligned} \tag{3.11}$$

Use again $R = (\phi^0)^{-1}\phi$ and consider the first term in expression (3.11). It can then, making use of the result in step 2, be written as $(I + \delta)\varepsilon^\theta$. Hence $(I + \delta_t)\varepsilon_t^\theta \to 0$. Next we consider the second term in expression (3.11). We again use step 2 to write it as $-\delta(\phi^0)^{-1}(\theta^0 - \theta) = -\delta R \phi^{-1}(\theta^0 - \theta)$. Because $\delta_t \to 0$, $R_t^T R_t \to I$ and θ^0 is assumed to be ϕ-consistent, we conclude that the whole second term tends to zero.

Finally we look at the last term of (3.11). Rewrite it as $\delta\phi^{0T}\tilde{M}^{0\theta}$. Assumption (3.5) together with $\delta_t \to 0$ yields this tending to zero.

This finishes the proof of the theorem. □

Remark By using similar techniques as in the proof one can also show the following statement.

$$\phi_t^T[\hat{M}_t + \phi_t^{-1}(\theta_t^0 - \theta)] \to 0 \text{ in all } \mathbb{P}^\theta \text{ probabilities} \tag{3.12}$$

Remark Inspection of the proof reveals that one can imagine situations where it is not needed that θ^0 is ϕ_t-consistent. Suppose that one can say a little more about the convergence in (2.4), for instance that one can replace in this expression the process ϕ by another process r such that $\psi^T[M^{\theta+r_t u} - M^\theta + \langle M^\theta, \tilde{M}^\theta\rangle_t r_t u]$ is stochastically bounded (in all $\mathbb{P}^\theta$-probabilities), and such that $\psi^T[M^{\theta+u_t} - M^\theta + \langle M^\theta, \tilde{M}^\theta\rangle_t u_t] \to 0$ in probability if $r_t^{-1}u_t \to 0$ in probability. In many cases this implies $r_t^{-1}\phi_t \to 0$ in probability, so r_t converges slower to 0 then ϕ_t. Suppose then that θ^0 is such that $r_t^{-1}(\theta_t^0 - \theta) \to 0$ in probability. Then the ε^θ process in step 1 of the proof of the above theorem still tends to zero in probability under the present assumptions. There is however a price to pay for allowing slower rates of convergence for θ^0, which is imposing conditions on the behaviour of Q in order to have also the second term in (3.11) converging to zero. Clearly, tightness of $\{\delta_t(\phi_t^0)^{-1}r_t\}$ is what one needs. So, under this condition the content of theorem 3.12 remains true. We illustrate this remark by the following example.

Example Consider a counting process N with an intensity process under a measure P^θ of the form θf_t. Here θ is a positive parameter (to be estimated) and f a known positive Lebesgue-measurable function. We choose $\mathscr{W}$ to be the set of all processes of the form $W(t,x) = w_t x$ with w predictable and $\int_0^t w_s^2 f_s ds$ finite for all $t > 0$. Then all martingales M^θ are affine in θ, so we may take r_t to be identically 1. Let now θ^0 be any *strongly* consistent estimator of θ such that $\theta_t^0 > 0$ and $\int_0^t (\theta_s^0)^2 f_s ds$ finite for all $t > 0$ (no rates of convergence required so far). Choose $\lambda^0(s,x) = (\theta_{s-}^0)^{-1}x$. Take $Q_t^{-1} = \theta_t^0 \int_0^t (\theta_s^0)^{-2} f_s ds$, and assume that $\lim_{t\to\infty} F_t = \infty$ with $F_t = \infty$ with $F_t = \int_0^t f_s ds$. Then $Q_t\langle \tilde{M}^{0\theta}\rangle \to 1$ in probability and by strong consistency of θ^0 also $Q_t\langle \tilde{M}^{0\theta}\rangle \to 1$ in probability (even almost surely). An easy computation shows that the tightness condition on the convergence of the Q-process mentioned in the previous remark here comes down to tightness of the process $F^{1/2}(\theta^0 - \theta)$. Since in this example we can replace ϕ with $F^{-1/2}$, we have nothing gained compared to theorem 3.2 by not imposing a tightness condition on the behaviour of θ^0. However, if we assume a priori that $F^{1/4}(\theta^0 - \theta)$ is tight, then one can easily check that all the conditions mentioned in the last remark above are satisfied and with

$$\hat{M}_t = \int_0^t \frac{dN_s}{\theta_{s-}^0} - \theta_t^0 \int_0^t \frac{f_s}{\theta_s^0} ds$$

the estimator $\hat{\theta} = \theta^0 + Q\hat{M}$ is optimal according to the reamark above.

Another possibility is to take $Q_t^{-1} = \int_0^t (\theta_s^0)^{-1} f_s ds$. With the same $\hat{M}$ as above the $\hat{\theta}$ from equation (3.8) can now be represented as the solution of the following equation

$$d\hat{\theta}_t = \frac{Q_t}{\theta_{t-}^0}(dN_t - \hat{\theta}_t f_t dt) \tag{3.13}$$

while Q satisfies

$$\dot{Q}_t = -\frac{f_t}{\theta_t^0} Q_t^2 \tag{3.14}$$

Equations of this kind are encountered when considering recursive estimators. See for instance [10]. Hence the somewhat unusual expression for $\hat{M}$ is a consequence of the definition of $\hat{\theta}$ as solution of the system of stochastic differential equations (3.13) and (3.14). As a final remark we notice the following. Suppose we replace in (3.13) and in (3.14) θ_t^0 with $\hat{\theta}_t$ and we can prove consistency of $\hat{\theta}$ (see [10] for the appropriate techniques). It then follows form the preceding discussion that the thus obtained *recursive* estimator is optimal.

Example Consider a stationary situation and assume that the observed X is the sum of a continuous compensator and a continuous local martingale under each of the probability measures involved. Stationarity in this case means that under a suitable parametrization we have the following model

$$dX_t = a(\theta)dt + dW_t^\theta \tag{3.15}$$

where W^θ is a Wiener process under each P^θ and a is a known function, independent of time. It is easy to see that in this case $b = \dot{a}$ in assumption 2.1, assuming that a is an (ordinary) differentiable function of θ. If $\dot{a}$ is not vanishing, then it easily follows that $\phi_t = \dot{a}(\theta)^{-1} t^{-1/2}$. Let now θ^0 be a strongly consistent estimator of θ that is also $\sqrt{t}$-consistent. We define the predictable process b^0 by $b_t^0 = \dot{a}(\theta_{t-}^0)$. One can then show that if a is a continuously differentiable function, the P^θ limit for $t \to \infty$ of $t^{-1} \int_0^t (b_s^0 - \dot{a}(\theta))^2 ds$ is zero, hence the convergence in (3.3) takes place. By taking $Q_t^{-1} = \dot{a}(\theta_t^0)^2 t$ also (3.2) holds. So according to theorem 3.2 the estimator $\hat{\theta}$ defined by

$$\hat{\theta}_t = \theta_t^0 + \frac{\int_0^t b_s^0 dX_s - \dot{a}(\theta_t^0) \int_0^t b_s^0 ds}{\dot{a}(\theta_t^0)^2 t}$$

is optimal.

A MEASURABILITY ISSUES

In this appendix we set forth conditions under which suprema of random variables are measurable and discuss some other measurability issues. We follow the approach given by Pollard in [9], Appendix C, which in turn is based on chapter III of Dellacherie and Meyer [1].

Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let $\mathscr{B}^m$ and $\mathscr{B}$ be the Borel σ-fields on $\mathbb{R}^m$ and $\mathbb{R}$ respectively, and let $\mathscr{F} \otimes \mathscr{B}^m$ be the product σ-algebra on $\Omega \times \mathbb{R}^m$. Then we have

LEMMA A.1 *Assume that $Z : \Omega \times \mathbb{R}^m \to \mathbb{R}$ be measurable. Then the map $Z^* : \Omega \to \mathbb{R}$ given by $Z^*(\omega) = \sup\{Z(\omega, u) : u \in \mathbb{R}^m\}$ is measurable w.r.t $\mathscr{F}^P$, the completion of $\mathscr{F}$ for the probability measure P.*

Proof See Pollard [9], page 197. □

Suppose now that there is instead of a single measure P a whole family $\mathbb{P}$ of measures P is defined on $(\Omega, \mathscr{F})$, the usual situation in statistical problems. Assume that there is a α-finite measure μ that dominates this family, then we have in the notation introduced above

LEMMA A.2 *Assume that $Z : \Omega \times \mathbb{R}^m \to \mathbb{R}$ be measurable. The map Z^* is measurable w.r.t $\mathscr{F}$ if $\mathscr{F}$ is complete for some $P \in P$ or for μ.*

The proof is obvious.

Another commonly used possibility to attack measurability problems for suprema is to assume a separability condition for Z. However, similar problems as above appear here, since the definition of separability (see [2]) involves sets of measure zero. So in the presence of a family of measures it is not directly clear to which of those "measure zero" refers.

Probably the easiest way to get rid of these measurability problems is to assume that for each ω the map $Z(\omega, \cdot)$ is continuous (or piecewise continuous). Then suprema are determined by a fixed dense subset of $\mathbb{R}^m$ like the rationals.

Other measurability problems arise when we replace u in $Z(\omega, u)$ with a random vector, and we want the map $\omega \to Z(\omega, u(\omega))$ to be measurable. A sufficient condition for this to happen is that Z is jointly measurable in ω and u and that $\omega \to u(\omega)$ is measurable (cf. LeCam [8], restriction (M2)).

If we apply this result to the processes in section 2, we get for instance adaptiveness of the process in expression (2.5) under the following set of conditions.

ASSUMPTION A.3 *For each fixed $H \in \mathscr{H}$, $W \in \mathscr{W}$ and $t > 0$ the maps given by $(\omega, \theta) \mapsto \phi(\omega, \theta)$, $(\omega, \theta) \mapsto H \cdot \mathring{A}^{\theta_t}(\omega)$, $(\omega, \theta) \mapsto W * \nu_t^\theta(\omega)$, $\omega \mapsto u_t(\omega)$ are jointly $\mathscr{F}_t$-measurable with respect to all their arguments.*

References

[1] C. Dellacherie and P. A. Meyer, Probabilités et potential, Hermann (1975).
[2] J. L. Doob, *Stochastic Processes*, Wiley (1953).
[3] K. Dzhaparidze, On iterative procedures of asymptotic inference, *Statistica Neerlandica* **37** (4) (1990), 181–189.
[4] K. Dzhaparidze and P. J. C. Spreij, On optimality of regular projective estimators in semimartingale models, *Stochastics and Stochastics Reports* **43** (1993), 161–178.
[5] K. Dzhaparidze and P. J. C. Spreij, On correlation calculus for multivariate martingales, *Stochastic Processes and Their Applications* **46** (1993), 283–299.
[6] R. A. Fisher, Theory of statistical estimation, *Proc. Cambridge Philos. Soc.* **22** (1925), 700–725.

[7] L. LeCam, On the asymptotic theory of estimation and testing hypotheses, *Proceedings of the Third Berkeley Symposium on Mathematical Statist. and Probab.* (California Univ. Press) **I** (1956), 129–156.
[8] L. LeCam, Locally Asymptotically Normal Families of Distributions, Univ. California Publ. in Statist. (California Univ. Press) **3** (2) (1960), 37–98.
[9] D. Pollard, *Convergence of Stochastic Processes*, Springer (1984).
[10] P. J. C. Spreij, Recursive Approximate Maximum Likelihood Estimation for a Class of Counting Process Models, *J. of Multivariate Analysis* **39** (2) (1991), 236–245.