Stochastics and Stochastics Reports, Vol. 47, pp. 247–268
Reprints available directly from the publisher
Photocopying permitted by license only

ON OPTIMALITY OF REGULAR PROJECTIVE ESTIMATORS FOR SEMIMARTINGALE MODELS, PART II: ASYMPTOTICALLY LINEAR ESTIMATORS

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

(Received 25 May 1993; in final form 8 June 1993)

In this paper we consider estimators that (asymptotically) admit a so called linear representation. Using a parametrization of the model, that has been defined in a previous paper [1], and a certain notion of smoothness of the parametrization, it is possible to define a concept of optimality for these estimators and to characterize the optimal estimators. In contrast with the situation in [1], only the compensator is fully parametrized by the parameter we want to estimate. Embedding the problem under consideration in the previously developed framework then requires the introduction of several nuisance parameters, that are needed to describe certain stochastic integrals with respect to the compensator of the jump measure.

KEY WORDS: Semimartingale, compensator, quadratic variation, regular estimator, asymptotically linear estimator, spread.

1. INTRODUCTION

The purpose of the present paper is to present an optimality theory for estimators that admit a so called linear representation. This notion will be defined precisely in section 2. We have chosen to embed this theory within the framework that has been put forward in [1], which serves as the basic reference. This framework includes such notions as *regularity* of an estimator, *admissibility*, and a measure of the estimation error, called *spread*.

Following the set up in [1], we assume that we have at our disposal observations of a multivariate stochastic process X and a family of sets of probability measures. Each of these is (partially) described by a finite dimensional parameter θ. Write P^θ for each such a set. Then we assume that under each member of such a set X is a special semimartingale with decomposition $X = A(\theta) + \mathbb{M}(\theta)$. Here $A(\theta)$ is a predictable process, the compensator of X, which is assumed to be the same under each of the members of the set P^θ. This statistical problem is then to estimate the parameter and our concern is to describe the best possible estimator. Best here refers to minimum spread within a class of admissible estimators. We refer to section 6 and 7 of [1] for the background of this approach in a somewhat more general situation. The class of admissible estimators in the present paper is that of estimators, that can be represented after a suitable centering and scaling as a stochastic integral with respect to the basic martingales $\mathbb{M}(\theta)$. These estimators are called asymptotically linear and form a class which is in principle smaller than the one that is obtained by completely following the approach of [1]. The reason is that we have to introduce in the present paper some nuisance parameters in order to use the parametrizations of [1]. At the end of section 2 we also indicate a direct approach to the optimality results to be presented in section 4, which is still in the spirit of our previous paper [1], but avoids explicitely using that framework, and thus the introducing of certain nuisance parameters. We are deliberately not very precise at this point, since it involves some delicate considerations concerning parametrization, which will be dealt with in the next section. As in [1] the minimization problem only makes sense if the scaling factor is of a specific type, which is obtained by imposing regularity of the estimators under consideration. Under such conditions the minimization problem can easily be solved by application of the Kunita-Watanable inequality.

Furthermore we also show under what conditions the optimal estimators, also of the nuisance parameters, yield the optimal linear one.

2. SMOOTH PARAMETRIZATIONS

2.1 *Parametrization*

Assume that one has a certain stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t \geqslant 0}, \mathbb{P})$, where $\mathbb{P}$ is a set of probability measures and on this a multivariate adapted process X, which we observe and which is assumed to be a semimartingale under each $P \in \mathbb{P}$. Denote by $\nu = \nu(P)$ the compensator of the jump measure of X under P. Let $D^\nu(\omega) = \{t : \hat{1}_t^\nu = \int \nu(\{t\}, dx) > 0\}$. For each ν the set $D^\nu(\omega)$ is countable. Below we will assume that we are dealing with a smooth parametrization of $\hat{1}^\nu$. This entails that if $\nu = \nu(P)$ and $\nu' = \nu(P')$ for $P, P' \in \mathbb{P}$, are close in a sense to be defined below the same should hold for $\sum_{s \leqslant t}(\hat{1}_s^{\nu'} - \hat{1}_s^\nu)$. It is then natural to impose that the summation variable s in this expression runs through a countable set that is independent of ν' and ν. Therefore we will assume that for each ω we have that $D^\nu(\omega) = D^{\nu'}(\omega)$, for all ν and ν'. So we can write $D(\omega)$ for each of those sets and $C(\omega)$ for its complement. Introduce furthermore the sets $D = \{(\omega, t) : t \in D(\omega)\}$, and C its complement.

Consider the following situation. Suppose that we are given a (finite dimensional) parametrization of the compensator A only. The purpose then is to consider estimators of the parameter involved which admit a so called linear representation. By this we mean a representation of the form like equation (5.2.4) of [1], where the predictable integrands H and W satisfy the relation $W = Hx$. We impose that the parametrization of A is smooth in the sense of Definition 4.1.1 of [1], which means asymptotic weak differentiability in two classes of predictable processes $\mathscr{W}$ and $\mathscr{H}$. Therefore the first thing that we have to do is to specify the classes of predictable processes $\mathscr{H}$ and $\mathscr{W}$. $\mathscr{W}$ is the subspace of $L^2(d\nu)$, where we can write for each of its members $W(t,x) = W_{0t} + W_{1t}x$ and $\mathscr{H}$ is the subspace of $L^2(d\langle \mathbb{M}^c \rangle)$ such that the condition $Hx \in \mathscr{W}$ is satisfied for all $H \in \mathscr{H}$. Here $\mathbb{M}^c$ is the continuous part of $\mathbb{M}$. Unlike we did in section 1, we usually suppress the dependence on θ. Superscripts θ that are used in the sequel have another meaning (see the next subsection).

Since it is assumed that with W also $\hat{W}$ $(\hat{W} = \int W\nu(\{t\}, dx))$ belongs to $\mathscr{W}$, we take $W^C := W 1_C$ of the form $W^C(t,x) = W^C_{1t}x$, so that W^C doesn't contain a component that is independent of x, and that $W^C \in \mathscr{W}$. However $W^D := 1_D W = W - W^C$ will usually contain a nonvanishing term W_0.

As we mentioned above, we assume that a parametrization $A(\theta)$ of A is given. This involves among other things, that we work with a collection P^θ of sets of measures, such that under each member of a set P^θ, $A(\theta)$ is the compensator of X. See [1] for a precise definition of this parametrization. In particular the discontinuous part $\sum_{s\leqslant \cdot} \Delta A_s = \sum_{s\leqslant \cdot} \Delta A_s(\theta)$ is fully parametrized by θ. We want to stay within the framework that has been put forward in [1]. This means that we also should assume that $x*\nu$ is fully parametrized by θ. In many circumstances, when assuming the identifiability property that $A(\theta) = A(\theta')$ implies $\theta = \theta'$, it may happen that the parameter θ has a dimension which is too low in order to specify the process $x*\nu$ too. Therefore we will make the assumption that an extra parameter (vector) α is needed in order to give a full specification of $x*\nu$. So we write $x*\nu = x*\nu^{\theta,\alpha}$. But, since $\Delta A_t = \Delta A_t(\theta) = \int x\nu(\{t\}, dx\}$ is already fully described by θ, the only place where α appears in the parametrization of $x*\nu$ is in the integrals $1_C x*\nu = 1_C x*\nu^{\theta,\alpha}$. A similar consideration can be given for the parametrization of $\hat{1}^\nu$, which is also assumed to be fully parametrized if we follow the approach of [1]. Hence we assume that we need another extra parameter vector β in order to obtain a full description of $\hat{1}^\nu$ for which we can then write $\hat{1}^{\theta,\beta}$. We now also obtained a parametrization of $1_D*\nu = 1_D*\nu^{\theta,\beta}$. Notice that we do not need the parameter α to describe $1_D*\nu$, since α was only introduced to parametrize integrals w.r.t. ν of processes that are zero outside C. As for θ we will assume that also α and β live in a finite dimensional space, although it is possible to relax this assumption, which is beyond the scope of this paper.

Summarizing the preceding discussion, we conclude that with the aid of the triple parameter (θ, α, β) we have $A = A(\theta)$, $1_C x*\nu = 1_C x*\nu^{\theta,\alpha}$, $1_D*\nu = 1_D*\nu^{\theta,\beta}$. Nevertheless we will occasionally write $A = A(\bar{\theta})$ with $\bar{\theta} = (\theta, \alpha, \beta)$ and likewise we will use the notation $\nu(\bar{\theta})$.

2.2 *Asymptotic Weak Differentiability*

Next we turn to smoothness of the previously introduced parametrization. To that end we introduce the following notation. For all $H \in \mathscr{H}$ and $W \in \mathscr{W}$ we write $M = M(H, W)$ for the martingale defined by $M = H \cdot \mathbb{M}^c + W*(\mu - \nu)$. $\tilde{M}$ is the martingale, defined by $\tilde{M} = b \cdot \mathbb{M}^c + \tilde{\lambda}*(\mu - \nu)$. (see Assumption 2.1 below). This tilde operator is defined for each $W \in \mathscr{W}$ by $\tilde{W} = W + 1_{\{\alpha<1\}} \hat{W}/1 - \alpha$ with $a = \hat{1}$. See below for more details on $\tilde{M}$. We assume that $\langle M \rangle_t$ and $\langle \tilde{M} \rangle_t$ are invertible for t large enough. Let then ψ_t be any matrix that satisfies the equality $\psi_t^T \psi_t = \langle M \rangle_t^{-1}$ and ϕ_t be any matrix that satisfies $\phi_t \phi_t^T = \langle \tilde{M} \rangle_t^{-1}$. We will assume that the following weak differentiability in the sense of [1] holds.

Assumption 2.1 *There exist (cf. [1], Definition 4.1.1) $b \in \mathscr{H}$, $\lambda \in \mathscr{W}$ such that for all fixed u, $H \in \mathscr{H}$, $W \in \mathscr{W}$ in all P^θ probabilities*:

(i) $\psi_t[H \cdot A_t(\bar{\theta} + \phi_t u) - H \cdot A_t(\bar{\theta}) - (\int_{[0,t]} Hd\langle \mathbb{M}^c \rangle b^T + Hx\lambda^T * \nu_t)\phi_t u] \to 0$

(ii) $\psi_t[W*(\nu(\bar{\theta} + \phi_t u)_t - \nu(\bar{\theta})_t) - W\lambda^T * \nu(\bar{\theta})_t \phi_t u] \to 0$

with ψ and ϕ as above.

In the sequel we abbreviate the phrase "in all $P^{\bar{\theta}}$ probabilities" by "in probability". Of course b, λ and $\mathbb{M}^c$ depend on $\bar{\theta}$, but this dependence is not explicitely written in order to avoid some cumbersome notation. Different from the notation in [1], we use superscripts to distinguish between the different components of b and λ related to θ, α, β we write $b^T = [b^{\theta T}, b^{\alpha T}, b^{\beta T}]$ and a similar decomposition of λ. Furthermore we write

$$\phi_t = \begin{bmatrix} \phi_t^{\theta\theta} & \phi_t^{\theta,\alpha\beta} \\ \phi_t^{\alpha\beta,\theta} & \phi_t^{\alpha\beta,\alpha\beta} \end{bmatrix}, \quad u^T = [u^{\theta T}, u^{\alpha T}, u^{\beta T}] = [u^{\theta T}, u^{\alpha\beta T}].$$

Following the discussion in [1] on the specific choice of ϕ and ψ, we impose that (i)–(iii) of the above assumption are also valid if $\phi_t u$ is replaced by $\phi_t u_t$ provided that $\{|u_t|\}$ is bounded. So we assume a uniform version of differentiability.

Recall that in this assumption ϕ_t is any matrix that satisfies the equality $\phi_t \phi_t^T = \langle \tilde{M} \rangle_t^{-1}$.
In the original definition in [1] we have taken the special ϕ_t^0 to be the symmetric positive square root of $\langle \tilde{M} \rangle_t^{-1}$, which is assumed to exist. Then we have that this assumption holds for any other such ϕ because with $u_t = (\phi_t^0)^{-1}\phi_t u$ we have that $\phi_t u = \phi_t^0 u_t$ and $|u_t| = |u|$ since $(\phi_t^0)^{-1}\phi_t$ is an orthogonal matrix.
Take now in particular

$$\phi_t = \begin{bmatrix} \phi_t^{\theta\theta} & 0 \\ \phi_t^{\alpha\beta,\theta} & \phi_t^{\alpha\beta,\alpha\beta} \end{bmatrix}, \quad u = \begin{bmatrix} 0 \\ u^{\alpha\beta} \end{bmatrix}.$$

Here the block decompositions are such that the size of $\phi_t^{\theta\theta}$ corresponds to the dimension of θ, etc. For these choices we have $\bar{\theta} + \phi_t u = \bar{\theta} + \begin{bmatrix} 0 \\ \phi_t^{\alpha\beta,\alpha\beta} u^{\alpha\beta} \end{bmatrix}$. Since $A_t(\bar{\theta})$ only depends on $\bar{\theta}$ through θ, we have that $A_t(\bar{\theta} + \phi_t u) - A_t(\bar{\theta}) = 0$. By using this in the expression in (i) of Assumption 2.1 we see that we may assume that for all H

$$\int H d\langle \mathbb{M}^c \rangle [b^{\alpha T}, b^{\beta T}] + Hx[\lambda^{\alpha T}, \lambda^{\beta T}] * \nu = 0. \tag{2.1}$$

Similar considerations ($1_C x * \nu$ and $x * \nu$ do not depend on β, $1_D * \nu$ does not depend on α) lead to

$$x\lambda^{\beta T} * \nu = 0 \tag{2.2}$$

$$1_C x\lambda^{\beta T} * \nu = 0 \tag{2.3}$$

$$1_D \lambda^{\alpha T} * \nu = 0 \tag{2.4}$$

It then follows from Eqs. (2.1) and (2.3) that $\int d\langle \mathbb{M}^c \rangle b^{\beta T} = 0$. We also get from (2.4) that $1_D \lambda^\alpha = 0$ and from $1_C \lambda^\beta = 1_C \lambda_1^\beta x$ and from (2.4) that $1_C \lambda^\beta = 0$. From the same equations it also easily follows that

$$\int d\langle \mathbb{M}^c\rangle b^{\alpha T} + xx^T\lambda_1^{\alpha T} * v = 0 \tag{2.5}$$

$$1_D(x\lambda_0^{\beta T} + xx^T\lambda_1^{\beta T}) * v = 0 \tag{2.6}$$

$$1_D\lambda_{1t}^\beta = -\lambda_{0t}^\beta \hat{x}^T(\widehat{xx^T})^{-1}1_D \tag{2.7}$$

In particular, we have $\hat{\lambda}_\alpha = 0$, and hence $\tilde{\lambda}^\alpha = \lambda^\alpha$. In Eq. (2.7) it is assumed that the inverse exists. If this happens to be not true, one can replace it with the Moore-Penrose inverse. It is easy to show that if one defines λ_{1t}^β in this way, Eq. (2.6) is again satisfied. All the computations below are valid with the obvious changes if this substitution is carried out. Henceforth we will not bother about this and assume that this inverse indeed exists.

It is sometimes more convenient to write $\langle \mathbb{M}^c\rangle$ as $c\cdot v$ where c is a square matrix valued predictable process and v a real increasing predictable process.

In terms of Proposition 7.1.2 of [1] $\tilde{M}$ introduced above is the optimal admissible estimation martingale for $\bar{\theta}$. Next we give some explicit computations.

$$\tilde{M} = \begin{bmatrix} b^\theta\cdot\mathbb{M}^c \\ b^\alpha\cdot\mathbb{M}^c \\ 0 \end{bmatrix} + \begin{bmatrix} \tilde{\lambda}^\theta * (\mu - v) \\ \lambda^\alpha * (\mu - v) \\ \tilde{\lambda}^\beta * (\mu - v) \end{bmatrix} = \begin{bmatrix} \tilde{M}^\theta \\ \tilde{M}^\alpha \\ \tilde{M}^\beta \end{bmatrix} = \begin{bmatrix} \tilde{M}^\theta \\ \tilde{M}^{\alpha\beta} \end{bmatrix} \tag{2.8}$$

and its predictable covariation process is given by $\langle \tilde{M}\rangle =$

$$\begin{bmatrix} b^\theta c b^{\theta T}\cdot v + \lambda^\theta\lambda^{\theta T} * v + \sum 1_{\{a<1\}}\dfrac{\hat{\lambda}^\theta\hat{\lambda}^{\theta T}}{1-a} & b^\theta c b^{\alpha T}\cdot v + \lambda^\theta\lambda^{\alpha T} * v & \lambda^\theta\lambda^{\beta T} * v + \sum 1_{\{a<1\}}\dfrac{\hat{\lambda}^\theta\hat{\lambda}^{\beta T}}{1-a} \\ b^\theta c b^{\alpha T}\cdot v + \lambda^\theta\lambda^{\alpha T} * v & b^\alpha c b^{\alpha T}\cdot v + \lambda^\alpha\lambda^{\alpha T} * v & 0 \\ \lambda^\beta\lambda^{\theta T} * v + \sum 1_{\{a<1\}}\dfrac{\hat{\lambda}^\beta\hat{\lambda}^{\theta T}}{1-a} & 0 & \lambda^\beta\lambda^{\beta T} * v + \sum 1_{\{a<1\}}\dfrac{\hat{\lambda}^\beta\hat{\lambda}^{\beta T}}{1-a} \end{bmatrix}$$

where we used the fact that $\langle \tilde{M}^\alpha, \tilde{M}^\beta\rangle = 0$.

LEMMA 2.2 *On D it holds that*

(i) $$\Delta\langle \tilde{M}^\beta\rangle = \lambda_0^\beta\lambda_0^{\beta T}1_{\{a<1\}}\frac{(a - \hat{x}^T(\widehat{xx^T})^{-1}\hat{x})(1 - \hat{x}^T(\widehat{xx^T})^{-1}\hat{x})}{1-a} \tag{2.9}$$

(ii) $$\Delta\langle \tilde{M}^\theta, \tilde{M}^\beta\rangle = (\lambda_0^\theta + \lambda_1^\theta\hat{x})\lambda_0^{\beta T}1_{\{a<1\}}\frac{a - \hat{x}^T(\widehat{xx^T})^{-1}\hat{x}}{1-a} \tag{2.10}$$

On C it holds that

(iii) $$\langle \tilde{M}^\alpha\rangle = \int b^\alpha d\langle \mathbb{M}^c\rangle(b^\alpha - \lambda_1^\alpha)^T = \int(\lambda_1^\alpha - b^\alpha)\,d\langle \mathbb{M}^d\rangle\lambda_1^{\alpha T} \tag{2.11}$$

(iv) $$1_C\cdot\langle \tilde{M}^\theta, \tilde{M}^\alpha\rangle = (\lambda_1^\theta - b^\theta)xx^T\lambda_1^{\alpha T} * v = \int 1_C(\lambda_1^\theta - b^\theta)\,d\langle \mathbb{M}^d\rangle\lambda_1^{\alpha T} \tag{2.12}$$

Proof (i) and (ii) follow from Eq. (2.7) and (iii) and (iv) follow from Eqs. (2.5) and (2.4).

2.3 *Explicit Computation of the Weak Derivative of $A(\theta)$*

In this subsection we will derive an explicit expression of the weak derivative of $A(\theta)$ as a stochastic integral with respect to the quadratic variation of $\mathbb{M}$. More precisely, we will show, under Assumption 2.1, the existence of a predictable process L satisfying the relation $\dot{A} = L\cdot\langle \mathbb{M} \rangle$, where $\dot{A} = \dot{A}(\theta)$ denotes the asymptotic weak derivative of $A(\theta)$. Introduce the martingale $\tilde{\tilde{M}}^\theta$defined by

$$\tilde{\tilde{M}}^\theta = [I, -\kappa]\cdot\begin{bmatrix} \tilde{M}^\theta \\ \tilde{M}^{\alpha\beta} \end{bmatrix} \tag{2.13}$$

where κ is any solution of the equation

$$\langle \tilde{M}^\theta, \tilde{M}^{\alpha\beta} \rangle = \kappa\cdot\langle \tilde{M}^{\alpha\beta} \rangle. \tag{2.14}$$

We know from [2] that $\kappa\cdot\tilde{M}^{\alpha\beta}$ is uniquely defined (up to indistinguishability). Write $\kappa = [\kappa^C, \kappa^D]$, with $\kappa^C = 1_C\kappa$. From Lemma 2.2 it then follows that on C we have

$$(\lambda_1^\theta - b^\theta)\, d\langle \mathbb{M}^d \rangle \lambda_1^{\alpha T} = \kappa^C(\lambda_1^\alpha - b^\alpha)\, d\langle \mathbb{M}^d \rangle \lambda_1^{\alpha T} \tag{2.15}$$

and on D we have

$$(\lambda_0^\theta + \lambda_1^\theta \hat{x})\lambda_0^{\beta T} 1_{\{a<1\}} = \kappa^D \lambda_0^\beta \lambda_0^{\beta T} 1_{\{a<1\}}(1 - \hat{x}^T(\widehat{xx^T})^{-1}\hat{x}) \tag{2.16}$$

Define the following three processes.

$$\dot{A}^C = b^\theta\cdot\langle X^C \rangle + 1_C\lambda_1^\theta\cdot(xx^T * \nu) \tag{2.17}$$

$$\dot{A}^D = \sum_{s\leqslant\cdot} [\lambda_{1s}^\theta \widehat{xx_s^T} + \lambda_{0s}^\theta \hat{x}_s^T] \tag{2.18}$$

$$\dot{A} = \dot{A}^C + \dot{A}^D. \tag{2.19}$$

The following proposition gives an expression for $\dot{A}$, where the integrand (L) involved will play later on the role of the optimal scoring function, when we treat optimality of estimators. Clearly $\dot{A} = \dot{A}(\theta)$ is to be understood as the derivative of A with respect to θ in the sense of Eq. (2.25) below. Again explicit dependence on θ is suppressed in our notation.

PROPOSITION 2.3 (a) *There exists a predictable process L such that $\dot{A} = L\cdot\langle \mathbb{M} \rangle$ and a locally square integrable martingale N such that $\tilde{\tilde{M}}^\theta = \kappa\cdot\tilde{M}^{\alpha\beta} + L\cdot\mathbb{M} + N$, where the martingales on the right hand side of this identity are mutually orthogonal.*

(b) *If, moreover, for all (t, ω) outside an evanescent set λ_{1t}^α or b_t^α has full column rank, then $N 1_C = 0$ and if λ_{0t}^β has full column rank, then $N 1_D = 0$.*

Proof (a) In view of Lemma A.1 from the appendix, there exists a predictable process ξ such that $\langle \mathbb{M}^c \rangle = \xi \cdot \langle \mathbb{M} \rangle$. Writing $\mathbb{M}^d$ for the discontinuous martingale appearing in the Doob-Meyer decomposition of X, we also have that $\langle \mathbb{M}^d \rangle = (I - \xi) \cdot \langle \mathbb{M} \rangle$. We define the process L as follows.
Take $L = L^C + L^D$, with

$$L^C = 1_C(b^\theta \xi + \lambda_1^\theta (I - \xi)) \tag{2.20}$$

$$L^D = 1_D \left(\lambda_1^\theta + \frac{(\lambda_0^\theta + \lambda_1^\theta \hat{x}) \hat{x}^T (\widehat{xx^T})^{-1}}{1 - \hat{x}^T (\widehat{xx^T})^{-1} \hat{x}} \right) \tag{2.21}$$

It is easy to verify, that with this choice of L the identity $L \cdot \langle \mathbb{M} \rangle = \dot{A}$ holds. Again using the given expression for L, one easily verifies orthogonality of the martingales involved.

(b) Assume that λ_1^β is a full column rank process. Then we get from Eq. (2.15) that κ^C satisfies

$$(\lambda_1^\theta - b^\theta) d\langle \mathbb{M}^d \rangle = \kappa^C (\lambda_1^\alpha - b^\alpha) d\langle \mathbb{M}^d \rangle \tag{2.22}$$

So, restricted to the set C, we have $1_C \cdot \tilde{M}^\theta - 1_C \kappa \cdot \tilde{M}^{\alpha\beta} = 1_C \tilde{M}^\theta - \kappa^C \cdot \tilde{M}^\alpha = (b^\theta - \kappa^C b^\alpha) \cdot \mathbb{M}^c + 1_C(\lambda_1^\theta - \kappa^C \lambda_1^\alpha) \cdot \mathbb{M}^d = (b^\theta - \kappa^C b^\alpha) \cdot \mathbb{M} + 1_C(\lambda_1^\theta - \kappa^C \lambda_1^\alpha - b^\theta + \kappa^C b^\alpha) \cdot \mathbb{M}^d$. The purely discontinuous martingale on the right hand side of this Eq. is zero, because its brackets are zero in view of Eq. (2.22). So we conclude that on C

$$1_C \cdot \tilde{\tilde{M}}^\theta = 1_C \cdot \tilde{M}^\theta - \kappa^C \cdot \tilde{M}^\alpha = 1_C (b^\theta - \kappa^C b^\alpha) \cdot \mathbb{M}.$$

The next step is to show that $1_C \tilde{\tilde{M}}^\theta = 1_C L \cdot \mathbb{M}$. Take the difference and use the given expression for L to get $1_C[(b^\theta - \lambda_1^\theta)(I^\theta - \xi) - \kappa^C b^\alpha] \cdot \mathbb{M}$. Take now the brackets of this martingale with $\mathbb{M}$:

$$\begin{aligned}
\langle 1_C \tilde{\tilde{M}} - 1_C L \cdot \mathbb{M}, \mathbb{M} \rangle &= 1_C (b^\theta - \lambda_1^\theta)(I - \xi) \cdot \langle \mathbb{M} \rangle - \kappa^C b^\alpha \cdot \langle \mathbb{M} \rangle \\
&= 1_C (b^\theta - \lambda_1^\theta) \cdot \langle \mathbb{M}^d \rangle - \kappa^C b^\alpha \cdot \langle \mathbb{M}^c \rangle - \kappa^C b^\alpha \cdot \langle \mathbb{M}^d \rangle \\
&= \kappa^C (b^\alpha - \lambda_1^\alpha) \cdot \langle \mathbb{M}^d \rangle - \kappa^C b^\alpha \cdot \langle \mathbb{M}^c \rangle - \kappa^C b^\alpha \cdot \langle \mathbb{M}^d \rangle \\
&= \kappa^C \cdot (-\lambda_1^\alpha \cdot \langle \mathbb{M}^d \rangle - b^\alpha \cdot \langle \mathbb{M}^c \rangle) \\
&= 0
\end{aligned}$$

Here the last equality follows from Eq. (2.5) and the one before that from Eq. (2.22).

The analysis on D is as follows. By assumption λ_0^β is a full rank process. Hence one obtains from Eq. (2.16) the following identity

$$(\lambda_0^\theta + \lambda_1^\theta \hat{x}) 1_{\{a<1\}} = \kappa^D \lambda_0^\beta 1_{\{a<1\}} (1 - r), \tag{2.23}$$

with $r = \hat{x}^T(xx^T)^{-1}\hat{x}$. Use now Eq. (2.7) to write $\lambda_1^\beta \hat{x} + \lambda_0^\beta = \lambda_0^\beta(1-r)$. Then

$$\widetilde{\lambda^\theta - \kappa^D\lambda^\beta} = (\lambda_1^\theta - \kappa^D\lambda_1^\beta)x + 1_{\{a<1\}}\frac{\lambda_1^\theta\hat{x} + \lambda_0^\theta - \kappa^D(\lambda_1^\beta\hat{x} + \lambda_0^\beta)}{1-a}$$

$$= (\lambda_1^\theta - \kappa^D\lambda_1^\beta)x + 1_{\{a<1\}}\frac{\lambda_1^\theta\hat{x} + \lambda_0^\theta - \kappa^D\lambda_0^\beta(1-r)}{1-a}$$

$$= (\lambda_1^\theta - \kappa^D\lambda_1^\beta)x,$$

where we used in the last equality (2.23). Using again Eq. (2.7) and Eq. (2.23), we get that $\lambda_1^\theta - \kappa^D\lambda_1^\beta = L$. Hence $\tilde{M}^\theta - \kappa^D\cdot\tilde{M}^\beta = (\widetilde{\lambda^\theta - \kappa^D\lambda^\beta})*(\mu - \nu) = L\cdot\mathbb{M}$. ∎

For all locally square integrable multivariate martingales m_1 and m_2 we denote by $c(m_1, m_2)$ the process $\langle m_1\rangle - \langle m_1, m_2\rangle\langle m_2\rangle^+\langle m_2, m_1\rangle$. (Cf. [3] for some properties of this process). In particular we use the notation c for the process defined by $c_t = c(\tilde{M}^\theta, M^{\alpha\beta})_t$.

By taking the matrix ϕ_t in Assumption 2.1 of the form $\begin{bmatrix}\phi_t^\theta & 0\\ * & *\end{bmatrix}$, one obtains $\phi_t^\theta = c_t^{-1/2}$, with $c_t^{1/2}$ any square root of c_t. This easily follows from the appendix. Using our convention to denote by u^θ the first component of u, Eq. (2.1) and Proposition 2.3, we can now reformulate Assumption 2.1 (i) as

$$\psi_t[H\cdot A_t(\bar{\theta} + \phi_t u) - H\cdot A_t(\bar{\theta}) - \int_{[0,t]} Hd\langle\mathbb{M}\rangle L^T c_t^{-1/2}u^\theta] \to 0 \tag{2.24}$$

Observe also that in Eq. (2.24) $A_t(\bar{\theta} + \phi_t u)$ is nothing else but $A_t(\theta + c_t^{1/2}u^\theta)$ and that $A(\bar{\theta}) = A(\theta)$, so we can rewrite Eq. (2.24) as

$$\psi_t[H\cdot A_t(\theta + c_t^{-1/2}u) - H\cdot A_t(\theta) - \int_{[0,t]} Hd\langle\mathbb{M}\rangle L^T c_t^{-1/2}u] \to 0. \tag{2.25}$$

In Eq. (2.25) we dropped the superscript θ for notational convenience.

Remark Introduce $\bar{\kappa} = \langle\tilde{M}^\theta, \tilde{M}^{\alpha\beta}\rangle\langle\tilde{M}^{\alpha\beta}\rangle^+$. Then $c = \langle\tilde{M}^\theta\rangle - \bar{\kappa}\langle\tilde{M}^{\alpha\beta}\rangle\bar{\kappa}^T = \langle\kappa\cdot\tilde{M}^{\alpha\beta}\rangle + \langle L\cdot\mathbb{M}\rangle + \langle N\rangle - \bar{\kappa}\langle\tilde{M}^{\alpha\beta}\rangle\bar{\kappa}^T = c(\kappa\cdot\tilde{M}^{\alpha\beta}, \tilde{M}^{\alpha\beta}) + \langle L\cdot\mathbb{M}\rangle + \langle N\rangle$. Hence it follows that $c \geqslant \langle L\cdot\mathbb{M}\rangle$. Therefore Eq. (2.25) is implied by the stronger statement

$$\psi_t[H\cdot A_t(\theta + \phi_t^0 u) - H\cdot A_t(\theta) - \int_{[0,t]} Hd\langle\mathbb{M}\rangle L^T\phi_t^0 u] \to 0, \tag{2.26}$$

which is obtained by replacing in (2.25) the process $c^{-1/2}$ by ϕ^0 with $\phi^0(\phi^0)^T = \langle L\cdot\mathbb{M}\rangle^{-1}$. Notice that there is equivalence between the two formulations (2.25) and (2.26) if $\{\phi_t^0 c_t^{1/2}\}$ is tight.

Hence it is possible to avoid the explicit introduction of the nuisance parameters α and β, if one directly starts with a smoothness assumption on $A(\theta)$ in the sense that

(2.26) holds instead of Assumption 2.1. Eq. (2.26) enables us to derive in section 4 that the optimal score function used in the representation of a linear regular estimator of θ is equal to L.

3. REGULAR ESTIMATORS

3.1 A Special Case

This subsection serves as an appetizer for the more general approach that we undertake in the next one. To simplify matters we consider here an analysis on the set C only, or assume that $D = \emptyset$. So the parametrization involves the parameters θ and α only. Assume that we have at our disposal a pair of (jointly) regular estimators $\hat{\theta}$ and $\hat{\alpha}$. This means that we have a representation of these estimators of the form

$$B\begin{bmatrix}\hat{\theta}-\theta \\ \hat{\alpha}-\alpha\end{bmatrix} = M + \eta = \begin{bmatrix}M_1 \\ M_2\end{bmatrix} + \eta \tag{3.1}$$

Here η is a remainder term that is small compared to the martingale M in the sense of Proposition 6.3.1 of [1] and $B = \langle M, \tilde{M}\rangle$. Decompose B into blocks B_{ij} of appropriate sizes and solve this Eq. to obtain a representation for $\hat{\theta}$:

$$\bar{B}(\hat{\theta}-\theta) = M_1 - B_{12}B_{22}^{-1}M_2 + \bar{\eta} \tag{3.2}$$

The matrix $\bar{B}$ figuring in Eq. (3.1) is equal to $B_{11} - B_{12}B_{22}^{-1}B_{21}$. Notice that the right hand side of this Eq. is in general not expressible as a martingale and it is not necessarily the case that its martingale part is independent of the nuisance parameter α. (β doesn't play a role here). Hence in order to have an admissible representation of θ, which loosely speaking amounts to saying that the derivative of the martingale part of this right hand side w.r.t. α is zero, we have to impose extra conditions. Consider first a simple case.

Assume that $M_1 = H\cdot\mathbb{M}$ for some $H \in \mathscr{H}$. Then $B_{12} = \langle M_1, \tilde{M}^{\alpha}\rangle = 0$ in view of Eq. (2.4). This immediately leads to Eq. (3.4) below.

In the more general case where M_1 is not of this specific form, we proceed as follows. We make the following assumption for the rest of this section, unless the contrary is explicitely stated. We assume that $B_{12}B_{22}^{-1}$ can be chosen to be a constant matrix, so independent of time, Γ say. This is a situation that one often encounters in practice, for instance in the situation in which X has stationary increments. Then the predictable variation processes grow with t. But it also holds for the simple case described above, since in that case $\Gamma = 0$. The term $\bar{\eta}$ in Eq. (3.2) is again a remainder term which is small compared to the martingale part as can easily be proven from (3.1). Indeed, notice that $\bar{\eta} = \eta_1 - \Gamma\eta_2$. Then we get from the appendix (Lemma A.2)

$$\bar{\eta}^T\langle M_1 - \Gamma M_2\rangle^{-1}\bar{\eta} \leqslant \eta^T\langle M\rangle^{-1}\eta,$$

which tends to zero in probability.

Write now $M_i = H_i \cdot \mathbb{M}^c + W_i * (\mu - \nu)$, with $W_i \in \mathscr{W}$. Because we have assumed that $D = \varnothing$, we know that W_i is of the form $W_i = W_{i1} x$, and thus $M_i = H_i \cdot \mathbb{M}^c + W_{i1} \cdot \mathbb{M}^d$. With a little abuse of notation we write from now on $M_i = H_i \cdot \mathbb{M}^c + W_i \cdot \mathbb{M}^d$. So the right hand side of Eq. (3.2) becomes $M_1 - \Gamma M_2 + \bar{\eta}$ or $(H_1 - \Gamma H_2) \cdot \mathbb{M}^c + (W_1 - \Gamma W_2) \cdot \mathbb{M}^d + \bar{\eta}$. Since it is the purpose of this paper to study asymptotically linear estimators we now impose that $\hat{\theta}$ is asymptotically linear by which we mean that the martingale that appears in the representation (3.2) is a stochastic integral which respect to the basic martingale $\mathbb{M}$. This is guaranteed if

$$H_1 - \Gamma H_2 = W_1 - \Gamma W_2 \tag{3.3}$$

Introduce now the following process in $\mathscr{H}: H = H_1 - \Gamma H_2$. Then, using the last relation between the H_i and the W_i, we obtain that $M_1 = H \cdot \mathbb{M} + \Gamma M_2$. Use now also that from the assumption that we started with regular estimators, the processes B_{ij} can be (at least asymptotically) identified with the predictable cross brackets of the M_i with the optimal martingales of the previous section. Then we get the following two identities.

$$B = \begin{bmatrix} I & \Gamma \\ 0 & I \end{bmatrix} \begin{bmatrix} \langle H \cdot \mathbb{M}, \tilde{M}^\theta \rangle & 0 \\ \langle M_2, \tilde{M}^\theta \rangle & \langle M_2, \tilde{M}^\alpha \rangle \end{bmatrix} = \begin{bmatrix} I & \Gamma \\ 0 & I \end{bmatrix} B_0$$

and

$$M = \begin{bmatrix} I & \Gamma \\ 0 & I \end{bmatrix} \begin{bmatrix} H \cdot \mathbb{M} \\ M_2 \end{bmatrix}$$

Hence the representation (3.1) is equivalent to

$$B_0 \begin{bmatrix} \hat{\theta} - \theta \\ \hat{\alpha} - \alpha \end{bmatrix} = \begin{bmatrix} H \cdot \mathbb{M} \\ M_2 \end{bmatrix} + \eta \tag{3.4}$$

One obtains from (3.4), or directly from (3.2), that in particular the following holds under the condition that Eq. (3.3) is satisfied

$$\bar{B}(\hat{\theta} - \theta) = H \cdot \mathbb{M} + \bar{\eta} \tag{3.5}$$

So Eq. (3.3) gives a sufficient condition to derive from (3.1), that $\hat{\theta}$ satisfies a linear representation.

In the next section we will treat optimal estimators. In the present section this reduces to the following. An estimator $\hat{\theta}$ will be called optimal if its spread is minimal. This quantity is defined as the matrix

$$\bar{B}^{-1} \langle H \cdot \mathbb{M} \rangle \bar{B}^{-T} \tag{3.6}$$

Notice that this quantity is equal to the 11-block of $B^{-1} \langle M \rangle B^{-T}$, where B and M are as in Eq. (3.1). The minimization problem is then to find the H that gives the minimal value of this quantity. In order to find the lower bound for the spread of $\hat{\theta}$ and the martingale for which this lower bound is attained, we observe that we can write $\langle H\cdot\mathbb{M}, \tilde{M}^\theta\rangle$ as $\langle H\cdot\mathbb{M}, L\cdot\mathbb{M}\rangle$, where L is as in Proposition 2.3. Hence a simple application of Schwartz' inequality tells us that the optimal scoring function $H = L$. See section 4.2 for the precise formulation.

It is perhaps tempting to suspect that the solution of this (asymptotic) minimization problem, which is L, is such that the martingale, $\begin{bmatrix} L\cdot\mathbb{M} \\ \tilde{M}^\alpha \end{bmatrix}$ is the martingale that minimizes the spread to joint estimators of θ and α, which is $B^{-1}\langle M\rangle B^{-T}$, when the first component M_1 of M is restricted to be of the form $H\cdot\mathbb{M}$. This conjecture turns out to be false. The reason for this is that the martingales $\tilde{M}^\theta$ and $\tilde{M}^\alpha$ are not orthogonal. Notice however that $L\cdot\mathbb{M}$ is the projection of $\tilde{M}^\theta$ on the space of martingales of the form $H\cdot\mathbb{M}$. See section 5 for an example, that illustrates this claim.

3.2 The General Case

In this subsection and in the next section we will generalize the ideas of section 4.1 by extending the analysis to the case where D is not necessarily empty and by dropping the assumption that the matrix Γ can be taken as a constant. Our starting point in this section is again an Eq. like (3.1). Suppose that we have at our disposal a joint estimator of θ, α, β. Write $\xi^T = (\alpha^T, \beta^T)$ and assume that this estimator satisfies

$$B\begin{bmatrix} \hat{\theta}-\theta \\ \hat{\xi}-\xi \end{bmatrix} = M + \eta = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} + \eta \tag{3.7}$$

with $B = \langle M, \tilde{M}\rangle$. First let $\gamma\cdot M_2 + H\cdot\mathbb{M}$ be the orthogonal projection of M_1 on the linear space of martingales spanned by M_2 and $\mathbb{M}$, and let N be the projection error. So $N = M_1 - \gamma\cdot M_2 - H\cdot\mathbb{M}$. Assume that γ and N obey the following assumption.

ASSUMPTION 3.1 *With N and γ as above and Γ such that $\Gamma\langle M_2, \tilde{M}^\xi\rangle = \gamma\cdot\langle M_2, \tilde{M}^\xi\rangle$ it holds that*

$$tr\left(\langle H\cdot\mathbb{M}\rangle_t^{-1}\left(\int_{[0,t]}(\gamma_s-\Gamma_t)d\langle M_2\rangle_s(\gamma_s-\Gamma_t)^T\right)+\langle N\rangle_t\right)\to 0 \tag{3.8}$$

in probability as t tends to ∞.

Notice that in the situation where $\gamma = 0$, so if $M_1 = H\cdot\mathbb{M}$, this assumption is trivially satisfied. As a side remark we mention the following. In the previous subsection we dealt with the case where $D = \emptyset$. If we assume, using the same notation, that $H_2 - W_2$ has full column rank (which is usually the case if the dimension of the observations doesn't exceed the dimension of α), then it follows that γ and H are such that $H_1 - \gamma H_2 = W_1 - \gamma W_2 = H$, as one can easily verify, simply by computing the projection of M_1. In this case $N = 0$.

Rewrite M_1 as $H\cdot\mathbb{M} + \gamma\cdot M_2 + N$. Solving Eq. (3.7) for $\hat{\theta}-\theta$ yields again an equation of the form (3.2). It reads

$$\bar{B}(\hat{\theta}-\theta) = H\cdot\mathbb{M} + R + N + \bar{\eta}. \tag{3.9}$$

Here we denoted $\gamma \cdot M_2 - \Gamma M_2$ by R. We require that R is a remainder term as well, by which we mean that its generalized covariance, which naturally takes the following form $\int_{[0,t]}(\gamma_s - \Gamma_t) d\langle M_2 \rangle_s (\gamma_s - \Gamma_t)^T$), is small compared to $\langle H \cdot \mathbb{M} \rangle$. But this follows from Assumption 3.1 above. Similarly we have that also N and $\bar{\eta}$ can be viewed as remainder terms.

Since we want that our estimators are regular, we need that R, N and $\bar{\eta}$ are also small (compared to $H \cdot \mathbb{M}$) in shrinking neighbourhoods of (θ, ξ). This enables us to identify $\bar{B}$ with $\langle H \cdot \mathbb{M}, L \cdot \mathbb{M} \rangle$, at least asymptotically. The precise result is given in Eq. (3.11) below. Recall that in the previous subsection we encountered a situation where actually equality between these processes holds.

PROPOSITION 3.2 *Assume that the joint estimator of θ and ξ satisfies Eq. (3.7) with $B = \langle M, \tilde{M} \rangle$ (hence it is regular). Then under Assumption 3.1:*

$$\bar{B}(\hat{\theta} - \theta) = H \cdot \mathbb{M} + \bar{\eta} \tag{3.10}$$

where $\bar{\eta}$ is a remainder term, and $\bar{B}$ satisfies

$$\varepsilon_t = \psi_t(\bar{B}_t - \langle H \cdot \mathbb{M}, L \cdot \mathbb{M} \rangle_t) c_t^{-1/2} \to 0 \tag{3.11}$$

Proof The fact that the $\bar{\eta}$ in Eq. (3.10) is small compared to $H \cdot \mathbb{M}$ has already been shown in the paragraph that precedes this proposition. The only thing that still needs a proof is Eq. (3.11). Start from Eq. (3.9). Actually all the terms appearing there depend on $\bar{\theta}$. In particular we now write $R^{\bar{\theta}}$ instead of R. Consider then the difference

$$R^{\bar{\theta}} - R^{\bar{\theta} + \phi u} = [I \ \ -\Gamma] \begin{bmatrix} \gamma \cdot M_2^{\bar{\theta}} - \gamma \cdot M_2^{\bar{\theta} + \phi u} \\ M_2^{\bar{\theta}} - M_2^{\bar{\theta} + \phi u} \end{bmatrix} \tag{3.12}$$

Write the last factor in this Eq. as $m^{\bar{\theta}} - m^{\bar{\theta} + \phi u}$. Let Ψ be such that $\Psi^T \Psi = \langle m^{\bar{\theta}} \rangle^{-1}$. A calculation shows that

$$[I \ \ -\Gamma_t] \langle m \rangle_t \begin{bmatrix} I \\ -\Gamma_t^T \end{bmatrix} = \int_{[0,t]} (\gamma_s - \Gamma_t) d\langle M_2 \rangle_s (\gamma_s - \Gamma_t)^T \tag{3.13}$$

From assumption 2.1 we know that $\Psi(m^{\bar{\theta}} - m^{\bar{\theta} + \phi u} - \langle m^{\bar{\theta}}, \tilde{M} \rangle \phi u) = o_P(1)$. Consider now

$$\psi(R^{\bar{\theta}} - R^{\bar{\theta} + \phi}) = \psi [I \ \ -\Gamma] \Psi^{-1} o_P(1) + \psi [I \ \ -\Gamma] \langle m^{\bar{\theta}}, \tilde{M} \rangle \phi u \tag{3.14}$$

Equation (3.13) together with Assumption 3.1 shows that the first term in the last Eq. tends to zero in probability. To analyze the behaviour of the second term, we consider

$$\psi [I \ \ -\Gamma] \langle m^{\bar{\theta}}, \tilde{M} \rangle \phi \phi^T \langle \tilde{M}, m^{\bar{\theta}} \rangle \begin{bmatrix} I \\ -\Gamma^T \end{bmatrix} \psi^T \tag{3.15}$$

which equals, evaluated at time t,

$$\psi_t \int_0^t (\gamma_s - \Gamma_t) d\langle M_2, \tilde{M} \rangle_s \phi_t \phi_t^T \int_0^t d\langle \tilde{M}, M_2 \rangle_s (\gamma_s - \Gamma_t)^T \psi_t^T \tag{3.16}$$

where we used again Eq. (3.13). Using the fact that $\langle \tilde{M} \rangle_t^{-1} = \phi_t \phi_t^T$ and a version of the Kunita-Watanabe inequality, we see that this is majorized by

$$\psi_t \int_0^t (\gamma_s - \Gamma_t) d\langle M_2 \rangle_s (\gamma_s - \Gamma_t)^T \psi_t^T \to 0$$

which tends to zero in probability in view of Assumption 3.1. A similar analysis applies to the terms N and $\bar{\eta}$ in Eq. (3.9). Now we return to Eq. (2.9). We have seen above that under our assumption the asymptotic derivative of $R + N + \bar{\eta}$ is zero, both with respect to θ and ξ. Hence carrying out differentiation of (3.9) w.r.t θ and ξ we obtain with ϕ as above

$$\psi_t([\bar{B}_t \quad 0] - \langle H \cdot \mathbb{M}, \tilde{M} \rangle_t) \phi_t \to 0 \tag{3.17}$$

Observe now that $\langle H \cdot \mathbb{M}, \tilde{M}^\xi \rangle = 0$, in view of Eqs. (2.5) and (2.6). Then it follows that Eq. (3.17) implies Eq. (3.11).

Of special interest—see the next section, where we discuss spread and optimality of estimators—is the case where we may replace in Eq. (3.11) c_t with $\langle L \cdot \mathbb{M} \rangle_t$, in which case we have with $\phi^0 \phi^{0T} = \langle L \cdot \mathbb{M} \rangle^{-1}$

$$\varepsilon_t = \psi_t(\bar{B}_t - \langle H \cdot \mathbb{M}, L \cdot \mathbb{M} \rangle_t) \phi_t^0 \to 0 \tag{3.18}$$

in probability. This happens if $\{\phi_t^0 c_t^{-1/2}\}$ is tight.

Proposition 3.2 makes the following definition clear.

DEFINITION 3.3 *Assume that the family of compensators $\{A(\theta)\}$ is such that Eq. (2.26) holds. An estimator $\hat{\theta}$ of θ is called a regular asymptotically linear estimator, if it is representable as in Eq. (3.10), where $\bar{B}$ satisfies Eq. (3.18).*

Putting Proposition 3.2 and Definition 3.3 together we obtain that under Assumptions 2.1 and 3.1 and if $\{\phi_t^0 c_t^{-1/2}\}$ is tight, then a regular estimator of θ and ξ yields a regular asymptotically linear estimator of θ.

4. OPTIMALITY

4.1 *A Cramer–Rao Type Bound*

The purpose of the present section is to discuss optimality of linear estimators of θ and to define a suitable notion of what optimality means in the present context. We will parallel to a certain extend the approach in [1]. Therefore we need the definition of *spread*. The starting point of the analysis is the following. Assume that we are given a joint estimator $\hat{\bar{\theta}}$ of $\bar{\theta}$ and assume that it is regular in the sense of Definition 6.1.1 of [1] and that it satisfies (3.7). Assume moreover that Assumption 3.1 is satisfied, so that the estimator of θ has the linear representation of Eq. (3.10). In this case the spread Σ of $\hat{\theta}$ around θ, which is by definition the 11-block of $B^{-1}\langle M \rangle B^{-T}$ can be taken to be $\bar{B}^{-1}\langle H \cdot \mathbb{M} \rangle \bar{B}^{-T}$. Of course the fact that the spread takes the form of this last expression immediately follows from (3.10). But it is also a direct consequence of Assumption 3.1. This can be seen as follows.

Let as before the process ψ be such that $\psi^T\psi = \langle H\cdot \mathbb{M}\rangle^{-1}$. It is an easy computation to show that the 11-block $B^{-1}\langle M\rangle B^{-T}$ is equal to

$$\bar{B}^{-1}(\langle M_1\rangle - \Gamma\langle M_2, M_1\rangle - \langle M_1, M_2\rangle\Gamma^T + \Gamma\langle M_2\rangle\Gamma^T)\bar{B}^{-T}.$$

So it is sufficient to prove that $\langle H\cdot\mathbb{M}\rangle$ is asymptotically equivalent to

$$\langle M_1\rangle - \Gamma\langle M_2, M_1\rangle - \langle M_1, M_2\rangle\Gamma^T + \Gamma\langle M_2\rangle\Gamma^T,$$

which means that

$$\psi_t(\langle M_1\rangle_t - \Gamma_t\langle M_2, M_1\rangle_t - \langle M_1, M_2\rangle_t\Gamma_t^T + \Gamma_t\langle M_2\rangle_t\Gamma_t^T)\psi_t^T \to 0$$

in probability. Use now the decomposition (in the notation of the previous section) $M_1 = \gamma\cdot M_2 + H\cdot\mathbb{M} + N$ to write the last expression as

$$\psi_t\Bigg(\int_{[0,t]}(\gamma_s - \Gamma_t)d\langle M_2\rangle_s(\gamma_s - \Gamma_t)^T + \langle N\rangle_t + \langle H\cdot\mathbb{M}\rangle_t$$
$$+\int_{[0,t]}(\gamma_s - \Gamma_s)d\langle M_2, H\cdot\mathbb{M}\rangle_s + \int_{[0,t]} d\langle M_2, H\cdot\mathbb{M}\rangle_s^T(\gamma_s - \Gamma_t)^T\Bigg)\psi_t^T.$$

Clearly the first two terms tend to zero in probability in view of Assumption 3.1, but also the last two by a version of the Kunita-Watanable inequality.

Let now $\mathscr{B}$ be defined by $\mathscr{B} = \psi\bar{B}$. Then we have similar to Proposition 6.1.4 in [1] the following.

PROPOSITION 4.1 (i) *For all symmetric positive definite matrices δ the event*

$$\mathscr{B}_t(\Sigma_t - c_t^{-1})\mathscr{B}_t^T \geqslant -\delta \tag{4.1}$$

takes place with probability tending to 1 *for $t\to\infty$.*

(ii) *If moreover the process $c^{1/2}\langle L\cdot\mathbb{M}\rangle^{-1}c^{1/2}$ is bounded in probability, then we may replace c_t in Eq. (4.1) with $\langle L\cdot\mathbb{M}\rangle_t$.*

Proof From Eq. (3.11) we get $\bar{B} = \langle H\cdot\mathbb{M}, L\cdot\mathbb{M}\rangle + \psi^{-1}\varepsilon c^{1/2}$. Introduce ϕ^0 such that $\phi^0\phi^{0T} = \langle L\cdot\mathbb{M}\rangle^{-1}$. Since $c \geqslant \langle L\cdot\mathbb{M}\rangle$, it holds that $\phi^{0T}c\phi^0 \geqslant I$. Furthermore, write $\rho = \psi\langle H\cdot\mathbb{M}, L\cdot\mathbb{M}\rangle\phi$. Then $\rho\rho^T \leqslant I$. Observe now that (4.1) is equivalent to $\mathscr{B}_t c_t^{-1}\mathscr{B}_t^T \leqslant I + \delta$. Hence by using the introduced notation, we get that Eq. (4.1) is equivalent to $(\varepsilon_t + \rho_t(\phi_t^0)^{-1}c_t^{-1/2})(\varepsilon_t + \rho_t(\phi_t^0)^{-1}c_t^{-1/2})^T \leqslant I + \delta$. Assertion (i) now follows since $\varepsilon_t \to 0$ in probability.

In order to prove the second assertion, we will use that differentiability as in Assumption 2.1 and (hence) in Eq. (2.24) holds uniformly in u. So we can take as u^θ in Eq. (2.24) $c_t^{1/2}\phi_t^0 u$, for some fixed u, since this is now a (in probability) bounded process.

Consequently, we have that Eq. (3.18) holds and hence $\mathscr{B}_t(\Sigma_t - \langle L\cdot\mathbb{M}\rangle_t^{-1})\mathscr{B}_t^T \geqslant -\delta$ now reduces to $(\varepsilon_t + \rho_t)(\varepsilon_t + \rho_t)^T \leqslant I + \delta$. ∎

4.2 Optimal Estimators

Throughout this subsection we assume that the process $c^{1/2}\langle L\cdot\mathbb{M}\rangle^{-1}c^{1/2}$ is bounded in probability (or equivalently, that $tr(\langle L\cdot\mathbb{M}\rangle^{-1}c)$ is bounded in probability).

DEFINITION 4.2 (i) *An asymptotically linear regular estimator $\hat{\theta}$ (cf. definition 3.3) is said to be optimal if its spread is given by $\langle L\cdot\mathbb{M}\rangle^{-1}$.*

(ii) *Any joint estimator $\hat{\bar{\theta}}$, that is regular in the sense that it satisfies Eq. (3.7), is said to yield the optimal asymptotically linear estimator if its first component $\hat{\theta}$ is as in* (i).

THEOREM 4.3 *A regular estimator $\hat{\theta}$ is an optimal linear estimator of θ iff $\hat{\theta}$ can be represented as*

$$\langle L\cdot\mathbb{M}\rangle(\hat{\theta}-\theta) = L\cdot\mathbb{M} + \eta \tag{4.2}$$

with a remainder term η.

Proof Assume that Eq. (4.2) holds. Then clearly the spread of $\hat{\theta}$ can be taken as $\langle L\cdot\mathbb{M}\rangle$.

Conversely, assume that $\hat{\theta}$ has a linear representation and that its spread can be taken as $\langle L\cdot\mathbb{M}\rangle$. Then $\bar{B}^{-1}\langle H\cdot\mathbb{M}\rangle\,\bar{B}^{-T} = \langle L\cdot\mathbb{M}\rangle$. In view of Eq. (3.11), this equality can be rewritten as $(\varepsilon+\rho)^T(\varepsilon+\rho) = I$, where we used the same notation as in the proof of the second assertion of Proposition 4.1. Since $\varepsilon_t \to 0$ in probability, we conclude that $\rho_t \to I$ in probability. As in the proof of Proposition 7.1.2 of [1] we conclude that $\hat{\theta}$ can be represented as in Eq. (4.2).

COROLLARY 4.4 *The optimal estimator of $\bar{\theta}$ yields the best optimal asymptotically linear estimator of θ iff $c_t^{1/2}\langle L\cdot\mathbb{M}\rangle^{-1}c_t^{1/2} \to I$ in probability.*

Proof Observe first that Assumption 3.1 applied to the pertaining case $M = \tilde{M}$ is equivalent to $c_t^{1/2}\langle L\cdot\mathbb{M}\rangle^{-1}c_t^{1/2} \to I$ in probability. This follows from the orthogonal decomposition of M^θ in Proposition 2.3. Hence if this condition is satisfied, then $\hat{\theta}$ automatically has a linear representation as in Eq. (4.2). So clearly this condition is sufficient. In order to prove necessity, we proceed as follows. Notice that for $M = \tilde{M}$ Eq. (4.2) takes the special form

$$c(\hat{\theta}-\theta) = \tilde{M}^\theta - \Gamma\tilde{M}^{\alpha\beta} + \eta \tag{4.3}$$

Hence the spread of $\hat{\theta}$ is now equal to c^{-1}. But since by assumption the very same $\hat{\theta}$ also yields the best linear estimator of θ, its spread must necessarily asymptotically equal $\langle L\cdot\mathbb{M}\rangle$, which proves the necessity. ∎

As a side remark we notice the following. Denote by $\hat{\theta}_L$ the optimal linear estimator of θ. Assume (somewhat artificially) that additionally there is another regular estimator $\tilde{\theta}$ available that is representable as

$$\tilde{B}(\tilde{\theta}-\theta) = \kappa\cdot\tilde{M}^{\alpha\beta} + N - \Gamma\tilde{M}^{\alpha\beta}. \tag{4.4}$$

Here κ and N are as in Proposition 2.3, and $\Gamma = \langle \tilde{M}^\theta, \tilde{M}^{\alpha\beta} \rangle \langle \tilde{M}^{\alpha\beta} \rangle^{-1} = \bar{\kappa}$. Regularity of $\tilde{\theta}$ is implied by taking $\tilde{B} = c - \langle L \cdot \mathbb{M} \rangle$. This follows by taking the brackets of the martingales in the right hand side of Eq. (4.4) with $\tilde{M}^\theta$ and using the orthogonality properties mentioned in Proposition 2.3. Assume for simplicity that all the brackets involved are deterministic. In this case the spread of the involved estimators is nothing else but the covariance. A computation then shows, that the spread of $\tilde{\theta}$ can be taken as $(c - \langle L \cdot \mathbb{M} \rangle)^{-1}$ and that $\hat{\theta}_L$ and $\tilde{\theta}$ are asymptotically orthogonal in the sense that their covariance equals zero. Consider now a convex combination, with matrices P and $I - P$ as weights, of these estimators. One wants to find the optimal convex combination, where optimality refers to minimum spread, so minimum covariance. Then one sees that here the optimal weight is $P = c^{-1} \langle L \cdot \mathbb{M} \rangle$, and the optimal linear combination is then the estimator $\hat{\theta}$, that satisfies Eq. (4.3). Hence if the condition in the last corollary is satisfied, then we see that one cannot improve on the spread of the best linear estimator. And conversely if one cannot improve on the spread of the best linear estimator, then the overall best estimator of θ, which is the one that satisfies (4.3), enjoys a linear representation.

5. EXAMPLES

5.1 Example in a Quasi Left Continuous Situation

Let N be a counting process and Z a diffusion process that satisfy the following pair of stochastic differential equations:

$$dN = \alpha\theta f \, dt + dm \tag{5.1}$$

$$dZ = (1 - \alpha)\theta f \, dt + dW \tag{5.2}$$

Here $\alpha \in (0, 1)$, $\theta > 0$, f is a locally square integrable function on $[0, \infty]$ and W a standard Brownian motion. Assume that only the sum X of N and Z is observed. Notice that X obeys the equation

$$dX = \theta f \, dt + d\mathbb{M} \tag{5.3}$$

Introduce the functions F and G as follows: $F_t = \int_0^t f_s \, ds$, $G_t = \int_0^t f_s^2 \, ds$.
In the previously employed notation we have (as can easily be verified)

$$b^\theta = (1 - \alpha) f \qquad \lambda_1^\theta = \theta^{-1}$$

$$b^\alpha = -\theta f \qquad \lambda_1^\alpha = \alpha^{-1}$$

$$\tilde{M}^\theta = (1 - \alpha) f \cdot W + \theta^{-1} m \quad \tilde{M}^\alpha = -\theta f \cdot W + \alpha^{-1} m$$

Hence a simple computation yields that

$$\kappa = \frac{\alpha}{\theta} - \frac{\alpha f}{\alpha\theta f + 1} \tag{5.4}$$

$$\bar{\kappa} = \frac{\alpha}{\theta} - \frac{\alpha G}{\alpha\theta G + F} \tag{5.5}$$

Here κ and $\bar{\kappa}$ are as in section 2. Now it is easy to compute

$$\tilde{\tilde{M}}^{\theta} = \frac{f}{\alpha\theta f + 1} \cdot \mathbb{M} = L \cdot \mathbb{M} \tag{5.6}$$

$$\langle \tilde{\tilde{M}}^{\theta} \rangle_t = \int_0^t \frac{f^2}{\alpha\theta f + 1} ds \tag{5.7}$$

$$= \frac{F_t}{\alpha\theta} - \frac{t}{(\alpha\theta)^2} + \frac{1}{(\alpha\theta)^2} \int_0^t \frac{1}{1 + \alpha\theta f} ds \tag{5.8}$$

Moreover

$$c(\tilde{M}^{\theta}, \tilde{M}^{\alpha}\rangle = \langle \tilde{M}^{\theta} \rangle - \bar{\kappa}^2 \langle \tilde{M}^{\alpha} \rangle = \frac{GF}{\alpha\theta G + F}$$

Consider first the case in which $f = 1$. Then

$$\langle \tilde{\tilde{M}}^{\theta} \rangle_t = c(\tilde{M}^{\theta}, \tilde{M}^{\alpha})_t = \frac{t}{\alpha\theta + 1}$$

Hence the maximum likelihood estimator of θ and the optimal linear estimator of θ have the same representation, which is in this case

$$t(\hat{\theta}_t - \theta) = \mathbb{M}_t + \eta_t$$

Notice that the concrete estimator $\hat{\theta}_t = X_t/t$ obeys this representation with $\eta = 0$ and it is optimal linear estimator.

Specializing to the case where $f_t = \sin t + 1$, one easily computes that $\lim_{t\to\infty} F_t/t = 1$ and $\lim_{t\to\infty} G_t/t = \frac{3}{2}$.

By also making use of the identity $\int_0^{2\pi} 1/a + b\sin x\, dx = 2\pi/\sqrt{a^2 - b^2}$ for $a > |b|$, one obtains

$$\lim_{t\to\infty} \frac{\langle \tilde{\tilde{M}}^{\theta} \rangle_t}{t} = \frac{1}{\alpha\theta} - \frac{1}{(\alpha\theta)^2} + \frac{1}{(\alpha\theta)^2 \sqrt{2\alpha\theta + 1}}$$

and

$$\lim_{t\to\infty} \frac{c(\tilde{M}^{\theta}, \tilde{M}^{\alpha})_t}{t} = \frac{1}{\alpha\theta + 2/3}$$

Hence

$$\lim_{t\to\infty} \frac{\langle \tilde{\tilde{M}}^{\theta} \rangle_t}{c(\tilde{M}^{\theta}, \tilde{M}^{\alpha})_t} = h(\alpha\theta)$$

for $h(\alpha\theta) = \alpha\theta + \frac{2}{3}/(\alpha\theta)^2(\alpha\theta - 1 + 1/\sqrt{2\alpha\theta + 1})$. So the conditions of Corollary 4.4 are not satisfied. It can be shown that $h(\alpha\theta)$ is minimal for $\alpha\theta = \frac{3}{4} + \sqrt{17}/4$ and that the minimal value approximately equals 0.9622. So if one works with the estimator that is the optimal linear one, there is no big loss in efficiency if one compares its spread with that of the maximum likelihood estimator. For the last one we have in this example the representation

$$t(\hat{\theta} - \theta) = \frac{2}{3} f \cdot W + m + \eta$$

from which we see that it is not an asymptotically linear estimator.

Notice that in this example κ keeps on oscillating, but that $\bar{\kappa}$ tends to a constant. This is the reason why $c(\tilde{M}^{\theta}, \tilde{M}^{\alpha})$ and $\langle \tilde{\tilde{M}}^{\theta} \rangle$ are not asymptotically equivalent.

Somewhat beyond the scope of the present paper is the following variant on the first example given in this subsection. Take the function f equal to 1, but replace the constant parameter α by an unknown function with range [0, 1]. Consider least squares estimators of θ, which are those that minimize the quadratic form

$$-2\theta \int_{[0,t]} r_s \, dX_s + \theta^2 \int_{[0,t]} r_s \, ds$$

Here r is a weight function that belongs to $L^1_{\mathrm{loc}} \cap L^2_{\mathrm{loc}}$. The solution is

$$\hat{\theta}(r)_t = \frac{\int_{[0,t]} r_s \, dX_s}{\int_{[0,t]} r_s \, ds} = \theta + \frac{\int_{[0,t]} r_s \, d\mathbb{M}_s}{\int_{[0,t]} r_s \, ds}$$

Clearly all estimators of this kind are regular. Moreover all regular asymptotically linear estimators are of this kind. Indeed assume that $\hat{\theta}$ is a regular asymptotically linear estimator. So it satisfies the representation $B(\hat{\theta} - \theta) = H \cdot \mathbb{M} = \eta$, and also in a suitable sequence of shrinking neighbourhoods of θ. Then it follows that one can take B to be asymptotically equal to $\int_{[0,t]} H_s \, ds$. Hence this estimator asymptotically coincides with the least squares estimator with weight function H. It is easy to compute

$$var\, \hat{\theta}(r)_t = \frac{\int_{[0,t]} r_s^2 (1 + \theta\alpha_s) \, ds}{\left(\int_{[0,t]} r_s \, ds \right)^2}.$$

In order to carry out some sort of worst case analysis one wants to minimize (over r)

$$\sup_{\alpha} var\, \hat{\theta}(r)_t = \frac{\int_{[0,t]} r_s^2(1+\theta)\,ds}{\left(\int_{[0,t]} r_s\,ds\right)^2}.$$

Clearly the minimum is obtained by taking $r = 1$, which corresponds to the estimator $\hat{\theta}_t = X_t/t$, and is equal to $(1+\theta)/t$.

Notice that also, $\sup_\alpha \inf_r var\, \hat{\theta}(r)_t = (1+\theta)/t$. So X_t/t can be considered as a kind of minimax estimator.

5.2 Example in Discrete Time

Let $X_t = \sum_{s \leqslant t} \Delta X_s$ with $P(\Delta X_t = \begin{cases} +1 \\ 0 \\ -1 \end{cases} | \mathscr{F}_{t-1}) = \begin{cases} p_1 \\ 1-(p_1+p_2), \\ p_2 \end{cases}$ where $t = 0, 1, \ldots$.
Here $p_i > 0$ and $p_1 + p_2 < 1$ and $\mathscr{F}_t = \sigma(X_1, \ldots, X_t)$. Introduce the alternative parametrization with $\theta = p_1 - p_2$, $\beta = p_1 + p_2$, and parameter space $\{(\theta, \beta): |\theta| \leqslant \beta \leqslant 1\}$.

One easily verifies that maximum likelihood estimators of θ and β are given by X_t/t and $1/t\int_{[0,t]}\int_{\{-1,+1\}} \mu(ds, dx)$ respectively, where μ is the jump measure associated with X. For this example a computation yields that

$$\lambda = \frac{1}{\beta^2 - \theta^2}\begin{bmatrix} -\theta & \beta \\ \beta & \theta \end{bmatrix}\begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$\tilde{\lambda} = \frac{1}{\beta^2 - \theta^2}\begin{bmatrix} -\theta & \beta \\ \beta & \theta \end{bmatrix}\begin{bmatrix} \dfrac{1}{1-\beta} \\ x + \dfrac{\theta}{1-\beta} \end{bmatrix}$$

It is then a simple computation to show that

$$\langle \tilde{M} \rangle = \frac{1}{\beta^2 - \theta^2}\begin{bmatrix} \beta & -\theta \\ -\theta & \dfrac{\beta - \theta^2}{1-\beta} \end{bmatrix} t.$$

Hence $\kappa = \bar{\kappa} = -\theta(1-\beta)/(\beta-\theta)^2$. So the optimal linear estimation martingale becomes

$$\approx{M}^\theta = \frac{x*(\mu - \nu)}{\beta - \theta^2}$$

and its predictable quadratic variation process is given by $t/(\beta-\theta)^2$. So in this case the maximum likehood estimator of θ is also the optimal linear estimator of θ, as could be expected. Both then have the exact representation

$$t(\hat{\theta}-\theta)=x*(\mu-\nu)=\mathbb{M}.$$

5.3 A Counterexample

The purpose of this subsection is to give an example to illustrate the last paragraph of section 3.1. We return to the example of subsection 5.1, but we make the following change. For Z we have the equation

$$dZ_t=(1-\alpha)\theta\,dt+\sigma_t\,dW_t. \tag{5.9}$$

The function f is taken to be 1, and $\sigma_t^2=1+\sin^2 t$. Before we curry out some calculations for the specific example at hand, we present some convenient general formulae.

Let $M=\begin{bmatrix}M_1\\M_2\end{bmatrix}$, with $M_1=L\cdot\mathbb{M}$. Let $\tilde{M}$ be the optimal martingale from Eq. (2.8), and $N=\begin{bmatrix}N_1\\N_2\end{bmatrix}$, with $N_1=M_1=L\cdot\mathbb{M}$ and $N_2=\tilde{M}^\alpha$. Write $m=\tilde{M}^\theta-N_1$. Consider now regular estimators that satisfy

$$\langle N,\hat{M}\rangle\begin{bmatrix}\hat{\theta}_0-\theta\\\hat{\alpha}_0-\alpha\end{bmatrix}=N=\begin{bmatrix}L\cdot\mathbb{M}\\\tilde{M}^\alpha\end{bmatrix}$$

and

$$\langle M,\hat{M}\rangle\begin{bmatrix}\hat{\theta}-\theta\\\hat{\alpha}-\alpha\end{bmatrix}=M=\begin{bmatrix}L\cdot\mathbb{M}\\M_2\end{bmatrix}$$

Denote their spreads by S_0^{-1} and S^{-1} respectively. Then a computation shows that

$$\begin{aligned}S_0&:=\langle\tilde{M},N\rangle\langle N\rangle^{-1}\langle N,\tilde{M}\rangle\\&=\frac{1}{\langle\tilde{M}^\alpha\rangle}\begin{bmatrix}\langle\tilde{M}^\theta,\tilde{M}^\alpha\rangle\\\langle\tilde{M}^\alpha\rangle\end{bmatrix}[\langle\tilde{M}^\theta,\tilde{M}^\alpha\rangle\langle\tilde{M}^\alpha\rangle]+\begin{bmatrix}\langle N_1\rangle&0\\0&0\end{bmatrix}\end{aligned}$$

Furthermore

$$\begin{aligned}S&:=\langle\tilde{M},M\rangle\langle M\rangle^{-1}\langle M,\tilde{M}\rangle\\&=\frac{1}{c(M_2,N_1)}\begin{bmatrix}\langle M_2,m\rangle\\\langle M_2,\tilde{M}^\alpha\rangle\end{bmatrix}[\langle M_2,m\rangle\langle M_2,\tilde{M}^\alpha\rangle]+\begin{bmatrix}\langle N_1\rangle&0\\0&0\end{bmatrix}\end{aligned}$$

Hence

$$S_0 - S = \frac{1}{\langle \tilde{M}^\alpha \rangle} \begin{bmatrix} \langle \tilde{M}^\theta, \tilde{M}^\alpha \rangle \\ \langle \tilde{M}^\alpha \rangle \end{bmatrix} [\langle \tilde{M}^\theta, \tilde{M}^\alpha \rangle \langle \tilde{M}^\alpha \rangle]$$

$$- \frac{1}{c(M_2, N_1)} \begin{bmatrix} \langle M_2, m \rangle \\ \langle M_2, \tilde{M}^\alpha \rangle \end{bmatrix} [\langle M_2, m \rangle \langle M_2, \tilde{M}^\alpha \rangle]$$

Now, if N minimizes the spread over all regular admissible estimators, where the first component of the martingale involved in the representation of such estimators is of the form $H \cdot \mathbb{M}$, then we should have that $S_0 - S$ is nonegative definite (at least asymptotically). Clearly the difference $S_0 - S$ is of the form $uu^T - vv^T$, which is $\geqslant 0$ if and only if $u = \lambda v$ for some λ with $|\lambda| \geqslant 1$.

Now we return to the specific example, mentioned in the beginning of this subsection. We will show that there is an example of a martingale M_2, such that $S_0 - S \geqslant 0$ doesn't hold, not even asymptotically for $t \to \infty$. Let $M_2 = H\mathbb{M}^c + W\mathbb{M}^d$, where H and W are real constants and $H \neq W$. First we compute the following limits.

$$\lim_{t \to \infty} \frac{\langle \tilde{M}^\theta, \tilde{M}^\alpha \rangle_t}{t} = -\frac{\theta(1-\alpha)}{\sqrt{2}} + 1.$$

$$[1 \lim_{t \to \infty} \frac{\langle \tilde{M}^\alpha \rangle_t}{t} = \frac{\theta}{\alpha} + \frac{\theta^2}{\sqrt{2}}$$

$$\lim_{t \to \infty} \frac{\langle M_2, m \rangle_t}{t} = (W - H)\left(\alpha - \frac{\alpha\theta}{\sqrt{(\alpha\theta + 1)(\alpha\theta + 2)}}\right)$$

$$\lim_{t \to \infty} \frac{\langle \tilde{M}^\alpha, M_2 \rangle_t}{t} = -(W - H)\theta$$

Clearly the vectors u and v that appear as one writes $\lim_{t \to \infty}(S_{0t} - S_t)/t = uu^T - vv^T$ are linearly independent and hence $\lim_{t \to \infty}(S_{0t} - S_t)/t \geqslant 0$ is not true in this case.

References

[1] K. Dzhaparidze and P. J. C. Spreij, On optimality of regular projective estimators in semimartingale models, CWI report BS-9029, **43** (1993), pp. 161–178.

[2] K. Dzhaparidze and P. J. C. Spreij, The strong law of large numbers for multivariate martingales with deterministic quadratic variation *Stochastics and Stochastics Reports*, Vol. **42** (1993), pp. 53–65.

[3] K. Dzhaparidze and P. J. C. Spreij, On correlation calculus for multivariate martingales, to appear in *Stochastic Processes and Their Applications*, **46** (1993), pp. 283–299.

A. APPENDIX

LEMMA A.1 *Let P, Q be symmetric nonnegative matrices of the same order such that $P \leqslant Q$. Then there exists a matrix ξ such that $P = \xi Q$.*

Proof Define $\xi = PQ^+$. We claim that $P - \xi Q = P(I - Q^+Q) = 0$. This can be seen as follows. Let u be an arbitrary vector. Then there exist vectors x, y such that $u = Q^+x + y$ and $y \in Ker\, Q = Ker\, Q^+$. Then $(P - \xi Q)u = P(I - Q^+Q)Q^+x + P(I - Q^+Q)y = PQ^+x - PQ^+x + Py - Q^+Qy = Py$. But $Py = 0$, because $Ker\, Q \subset Ker\, P$. ∎

LEMMA A.2 *Let* $\begin{bmatrix} P & Q \\ Q^T & R \end{bmatrix} > 0$ *and symmetric. Then* $[u^T, v^T]\begin{bmatrix} P & Q \\ Q^T & R \end{bmatrix}^{-1}\begin{bmatrix} u \\ v \end{bmatrix} = u^TP^{-1}u + (v - Q^TP^{-1}u)^T(R - Q^TP^{-1}Q)^{-1}(v - Q^TP^{-1}u)$.

Proof Use the decomposition

$$\begin{bmatrix} P & Q \\ Q^T & R \end{bmatrix} = \begin{bmatrix} I & 0 \\ Q^TP^{-1} & I \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & R - Q^TP^{-1}Q \end{bmatrix}\begin{bmatrix} I & P^{-1}Q \\ 0 & I \end{bmatrix}$$

to write

$$\begin{bmatrix} P & Q \\ Q^T & R \end{bmatrix}^{-1} = \begin{bmatrix} I & -P^{-1}Q \\ 0 & I \end{bmatrix}\begin{bmatrix} P^{-1} & 0 \\ 0 & (R - Q^TP^{-1}Q)^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -Q^TP^{-1} & I \end{bmatrix}$$

The result now immediately follows. ∎