The Annals of Statistics
2001, Vol. 29, No. 5, 1508–1536

ON SEQUENTIAL ESTIMATION OF PARAMETERS IN SEMIMARTINGALE REGRESSION MODELS WITH CONTINUOUS TIME PARAMETER

BY L. GALTCHOUK AND V. KONEV[1]

Strasbourg University and Tomsk University

We consider the problem of parameter estimation for multidimensional continuous-time linear stochastic regression models with an arbitrary finite number of unknown parameters and with martingale noise. The main result of the paper claims that the unknown parameters can be estimated with prescribed mean-square precision in this general model providing a unified description of both discrete and continuous time process. Among the conditions on the regressors there is one bounding the growth of the maximal eigenvalue of the design matrix with respect to its minimal eigenvalue. This condition is slightly stronger as compared with the corresponding conditions usually imposed on the regressors in asymptotic investigations but still it enables one to consider models with different behavior of the eigenvalues. The construction makes use of a two-step procedure based on the modified least-squares estimators and special stopping rules.

1. Introduction. In recent years many papers have been devoted to estimation of parameters in linear stochastic regression models, specified either by stochastic difference or by stochastic differential equations. These models include the Gauss-Markov processes and autoregression processes and arise in different applications: time series analysis, adaptive stochastic control problems, on-line identification of dynamic systems, geophysics, financial mathematics and so on. The most popular estimation schemes for stochastic regression models are based on the least-squares and the maximum likelihood methods and the properties of estimates are usually studied under the assumption that the sample-size tends to infinity. It is well-known that asymptotic properties of least-squares estimates in autoregression and linear regression models heavily depend on the values of unknown parameters and on the behavior of the eigenvalues of a design matrix [we refer the reader to Anderson and Taylor (1976), Lai, Robbins and Wei (1979) and Lai and Wei (1982) for details and further references]. The results on asymptotic theory of estimates of parameters in linear stochastic differential equations driven by the Wiener processes are given in Arato, Kolmogorov and Sinai (1962), Le Breton (1977), Liptser and Shiryaev (1977) and Jankunas and Khasminskii (1997), among others. Estimation problems for stochastic differential equations driven by martingale noise are considered in Novikov (1984), Christopeit (1986) and Darwich and Le Breton (1991).

Received March 1999; revised June 2000.

[1]Supported by Russian Fund of Fundamental Investigations Grant 00-01-00880.

AMS 2000 *subject classifications*. 62L12, 62M09.

Key words and phrases. Weighted least-squares estimators, sequential procedure, estimators with prescribed precision, stochastic regression, semimartingales, stopping times.

Several recent papers have shown that many difficulties in the theory of estimation for stochastic regression models can be overcome by a sequential approach. It turns out to be helpful both in asymptotic and non-asymptotic studies. One of the ways to improve the asymptotic properties of the least-squares estimates in discrete-time processes is connected with a special choice of stopping rules instead of a fixed sample size. Lai and Siegmund (1983) have proved that the least-squares estimate of an autoregressive parameter of a first-order autoregression process becomes asymptotically uniformly normal on the interval $[-1, 1]$ if one uses a special stopping rule. Shiryaev and Spokoiny (1997) have established that this estimate has one universal limit distribution for all admissible values of the unknown parameter in the case of gaussian noises. For futher asymptotic results and references on estimation in AR(1) we refer the reader to Sriram (1988), Aras (1990), Greenwood and Shiryaev (1992), Konev and Pergamenshchikov (1997), among others.

A different but closely related application of sequential analysis consists in constructing point estimators with prescribed mean-square precision which provide non-asymptotic solutions to the estimation problems. Liptser and Shiryaev (1977) [see also Novikov (1971)] have put forward the idea of using the maximum likelihood estimator with a special stopping time for the problem of estimating the drift coefficient of a scalar diffusion process. This estimator is unbiased and has a prescribed mean-square precision. Borisov and Konev (1977) have proposed an unbiased guaranteed least-squares estimate for a parameter for an AR(1), which, besides the introduction of the special stopping time, requires a certain modification of the estimate itself. Such an approach can be extended to estimate linear parameters in multivariate discrete and continuous time processes provided that the number of unknown parameters does not exceed the dimension of the process [see Borisov and Konev (1977) and Konev and Vorobeichikov (1980)]. The unbiased guaranteed estimates in this case can be constructed on the basis of generalized least-squares estimates with a proper choice of a weight matrix.

Melnikov and Novikov (1988) and Melnikov (1996) considered the problem of guaranteed estimation of parameters in multivariate regression models with martingale noise. These models provide a unified description of both discrete and continuous time processes, specified either by stochastic difference or stochastic differential equations. However, the applicability of the proposed procedure is also restricted to the case when the number of unknown parameters does not exceed the dimension of the observed process. This restriction rules out a broad class of processes, for example, scalar autoregressive processes of order greater than one.

The present paper considers the problem of sequential point estimation of parameters in multivariate stochastic regression models with martingale noise and any finite number of unknown parameters. The proposed procedure enables us to estimate the parameters with any prescribed mean square accuracy under some conditions on the regressors. The procedure is based on the weighted least-squares method with a special choice of weight matrices and it is one-step if the number of unknown parameters doesn't exceed the dimension of the observation process and it is two-steps in a general case. In the first case, the construction of the procedure is similar to that proposed by Konev and Vorobeichikov (1980) and in the second case it is close to that proposed in Konev and Pergamenshchikov (1981), Konev and Lai (1995) and Galtchouk and Konev (1997). Among conditions on the regressors there is one bounding the growth of the maximum eigenvalue of a design matrix with respect to its minimal eigenvalue. This condition is slightly stronger than those usually imposed in asymptotic investigations but it makes possible to consider models with different behavior of the eigenvalues.

The paper is arranged as follows. In Section 2 the model is described and different particular examples are given. Section 3 provides justification of the weighted least-squares estimator (LSE) of the unknown parameter which is based on discrete-time approximations. The weighted LSE is the basis of constructions of guaranteed estimators. In Section 4, the unbiased prescribed precision estimators for multidimensional processes are constructed (see Theorem 1). It is assumed that the number of unknown parameters does not exceed the dimension of the process under observation and in this case the sequential procedure is one-step. The sequential two-step estimation procedure for the general case with an arbitrary number of unknown parameters is studied in Section 5 (Theorem 2). The results are proved under minimal assumptions on the regressors. Section 6 includes auxiliary propositions.

2. Model.

Examples. Let $(\Omega, \mathscr{F}, \mathbf{F} = (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ be a filtered probability space satisfying the usual conditions: the filtration $\mathbf{F} = (\mathscr{F}_t)_{t\geq 0}$ is right continuous, that is $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s$, and the σ-algebra $\mathscr{F}_0$ contains all $\mathbf{P}$-null sets. Consider the observation process $X = (X(t))_{t\geq 0}$ specified by the stochastic regression model

$$X(t) = X(0) + \int_0^t \Phi'(s)\theta da(s) + m(t), \qquad t \geq 0, \tag{2.1}$$

where $\int_0^t = \int_{]0,t]}$; the prime denotes transposition; $X, m \in \mathbb{R}^n$; Φ is a predictable $p \times n$ matrix (the matrix of stochastic regressors); $\theta \in \mathbb{R}^p$ is the vector of unknown parameters; $m(t) = (m_1(t), \dots, m_n(t))'$, $m(0) = 0$, is a noise which is a locally square integrable martingale with the trajectories continuous on the right and having left-side limits:

$$a(t) = \langle m, m\rangle(t) = \operatorname{tr}\ \left((\langle m_i, m_j\rangle)_{1\leq i,j\leq n}\right)(t) = \sum_{i=1}^{n} \langle m_i, m_i\rangle(t).$$

Recall that $(\langle m_i, m_j\rangle)_{1\leq i,j\leq n}$ is the matrix of predictable local quadratic variations of m [the predictable process of bounded variation $\langle m_i, m_j\rangle$ is such that the process $m_i m_j - \langle m_i, m_j\rangle$ is a local martingale; see Galtchouk (1975, 1976), Jacod (1979)]. Let

$$B(t)=\left(\frac{d\langle m_i,m_j\rangle}{da}(t)\right),\quad B^c(t)=\left(\frac{d\langle m_i^c,m_j^c\rangle}{d\langle m^c,m^c\rangle}(t)\right),$$

$$B^d(t)=\left(\frac{d\langle m_i^d,m_j^d\rangle}{d<m^d,m^d>}(t)\right),\qquad 1\le i,\,j\le n,$$

$$\begin{aligned}\langle m^c,m^c\rangle(t)&=\sum_{i=1}^n\langle m_i^c,m_i^c\rangle(t),\\ \langle m^d,m^d\rangle(t)&=\sum_{i=1}^n\langle m_i^d,m_i^d\rangle(t);\end{aligned}\tag{2.2}$$

here $m=m^c+m^d$ is the orthogonal decomposition of the vector-valued martingale m into a continuous martingale and a purely discontinuous martingale. Note that $\langle m_i^c,m_j^d\rangle=0$, for all $1\le i,\,j\le n$,

$$\begin{aligned}a(t)&=\langle m^c,m^c\rangle(t)+\langle m^d,m^d\rangle(t),\\ B(t)d\langle m,m\rangle(t)&=B^c(t)d\langle m^c,m^c\rangle(t)\\ &\quad+B^d(t)d\langle m^d,m^d\rangle(t).\end{aligned}\tag{2.3}$$

The observation process X is well defined, if for all $t\ge 0$,

$$\int_0^t\|\Phi(s)\|da(s)<\infty\qquad\text{a.s.}$$

The goal of this paper is to construct an estimator of the unknown parameter θ in model (2.1) with prescribed mean-square precision, that is for any given positive number h we have to define an estimator θ_h^* such that

$$\mathbf{E}\|\theta_h^*-\theta\|^2\le\text{const}\cdot h^{-1}.$$

The process X is a semimartingale which includes both the discrete-time and continuous-time (random and deterministic) regression models.

Consider some examples.

EXAMPLE 2.1. The discrete-time linear stochastic regression

$$X_n=\theta_1Z_{n,1}+\cdots+\theta_pZ_{n,p}+\varepsilon_n$$

where $(\varepsilon_n)_{n\ge1}$ is a square integrable martingale-difference with respect to its natural filtration $(\mathscr{F}_n)_{n\ge1}$, $\mathscr{F}_0=\{\Omega,\varnothing\}$:

$$\mathbf{E}\varepsilon_n^2<\infty,\ \mathbf{E}[\varepsilon_n\mid\mathscr{F}_{n-1}]=0,\ a_n=\mathbf{E}[\varepsilon_n^2\mid\mathscr{F}_{n-1}]>0,\qquad n\ge1;$$

$(Z_{n,1},\dots,Z_{n,p})$ are consecutively determined $\mathscr{F}_{n-1}$-measurable random vectors.

This model is a particular case of (2.1): it suffices to set

$$X(t) = \sum_{i=1}^{[t]} X_i, \quad m(t) = \sum_{i=1}^{[t]} \varepsilon_i, \quad a(t) = \sum_{i=1}^{[t]} a_i,$$

$$\Phi(t) = (\Phi_1(t), \dots, \Phi_p(t))', \;\; \Phi_i(t) = Z_{[t],i}/a_i,$$

where $[t]$ is the integer part of t.

This model includes an autoregression $AR(p)$ model

$$X_k = \theta_1 X_{k-1} + \cdots + \theta_p X_{k-p} + \varepsilon_k, \qquad k \geq 0, \tag{2.4}$$

where the initial values $X_{-1}, \dots, X_{-p}$ are given.

EXAMPLE 2.2. Let $Z(t)$ be a stationary Gaussian process in continuous time with the rational spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|P(\sqrt{-1}\lambda)\right|^{-2}, \qquad -\infty < \lambda < \infty,$$
$$P(z) = z^p + \theta_1 z^{p-1} + \cdots + \theta_p,$$

where the coefficients $\theta_1, \dots, \theta_p$ are unknown and all roots of the polynomial $P(z)$ have negative real parts. This process satisfies the equation [see Arato (1982)]:

$$dZ^{p-1}(t) + \left[\theta_1 Z^{p-1}(t) + \cdots + \theta_p Z(t)\right] dt = \sigma dw(t).$$

Setting

$$\Phi'(t) = (-Z^{(p-1)}(t), \dots, -Z(t)), \;\; X(t) = Z^{(p-1)}(t)$$

we obtain the model of type (2.1) :

$$dX(t) = \Phi'(t)\theta dt + \sigma dw(t),$$

where $w(t)$ is the standard Wiener process.

EXAMPLE 2.3. The Itô process

$$dX(t) = \Phi'(t)\theta dt + \sigma(t) dw(t),$$

where Φ, σ are predictable $p \times n$ and $n \times r$-matrices respectively, w is the standard r-dimensional Wiener process, $B = \sigma\sigma'$. This model is widely used in stochastic control, dynamic input-output systems, time-series analysis, geophysics and so on.

EXAMPLE 2.4. The point process (p_t) is the particular case of (2.1): $p_t = \sum_i I_{(T_i \leq t)}$,

$$dX(t) = \lambda(t)\theta da(t) + dm(t),$$

where $(T_i)_{i\geq 1}$ is an increasing sequence of stopping times, $\theta \in \mathbb{R}$, $\lambda(t)$ is a positive predictable process such that

$$m(t) = p(t) - \int_0^t \lambda(s)\theta da(s)$$

is a locally square integrable martingale.

EXAMPLE 2.5. If $da(t) = dt$ in the previous model, we obtain a doubly stochastic Poisson process.

REMARK 1. The integral in (2.1) is defined with respect to the process $a(t)$ which is assumed to be the trace of the matrix of predictable local quadratic variation of a locally square integrable martingale m. One can take, of course, any other process which is absolutely continuous with respect to the process $a(t)$ as the integrator.

3. Weighted LSE for semimartingale models in continuous time. In this section, we propose to use special discrete-time approximation schemes to construct and justify the weighted LSE in continuous-time regression models. The least-squares method is well developed for discrete-time regression models for which one has a reasonable functional to measure the quality of estimates– the sum of squares of residuals taken over all observations. For continuous-time regression models such a natural functional is not available because the direct extension by making use of the discrete-time approximations breaks down: the limit of the corresponding sum of squares of residuals does not exist in any reasonable sense. The idea not to use the sum of squares as a starting point in the least-squares method for continuous-time models does not seem to be attractive either. However, usually authors prefer to avoid the above-mentioned difficulties and write LSE by analogy with those for the discrete-time models [see Novikov (1984), Christopeit (1986), Le Breton and Musiela (1987), Melnikov and Novikov (1988), Melnikov (1996)]. We propose to construct LSE for continuous-time models as follows:

(i) perform the time discretization of the model (2.1);
(ii) construct the sum of squares of residuals separating the sum of addends connected with the jumps of the process $a(t)$ caused by the predictable time jumps of the observed process X and the sum of addends connected with the continuous part of the process a; in the second sum each addend is supplied with a special normalization factor;
(iii) consider LSE for the discrete-time approximation and find the limit of these estimators as the size of partitions tends to zero.

Let $(X_t)_{t\geq 0}$ be an observation process starting from X_0 and specified by the stochastic differential equation (2.1). Let $\widetilde{W} = \widetilde{W}(t)$, $t \geq 0$, be some predictable symmetric positive definite weight matrix of size $n \times n$. Let $a = a^c + a^d$ be the decomposition of the predictable increasing process a into continuous and purely discontinuous parts:

$$a_t^d = \sum_{s\le t} \Delta a_s, \qquad a_t^c = a_t - a_t^d, \quad \Delta a_s = a_s - a_{s-}.$$

Introduce a loss function

$$\begin{aligned} I_\delta(\theta) = {} & \sum_{t_k<T} \big(\Delta X_{t_k}^d - \Phi'(t_k)\theta\Delta a_{t_k}^d\big)'\widetilde{W}(t_k)\big(\Delta X_{t_k}^d - \Phi'(t_k)\theta\Delta a_{t_k}^d\big) \\ & + \sum_{t_k<T} (\Delta a_{t_k}^c)^{\oplus}\big(\Delta X_{t_k}^c - \Phi'(t_k)\theta\Delta a_{t_k}^c\big)'\widetilde{W}(t_k)\big(\Delta X_{t_k}^c - \Phi'(t_k)\theta\Delta a_{t_k}^c\big), \end{aligned}$$

where $\delta = \{t_0, t_1, \ldots\}$, $0 = t_0 < t_1 < \ldots < t_n = T$, is some partition of the interval $[0, T]$; $x^{\oplus} = x^{-1}$, if $x \ne 0$, $x^{\oplus} = 0$, otherwise

$$\Delta X_{t_k}^d = X_{t_{k+1}}^d - X_{t_k}^d, \quad \Delta a_{t_k}^c = a_{t_{k+1}}^c - a_{t_k}^c, \quad \Delta a_{t_k}^d = a_{t_{k+1}}^d - a_{t_k}^d,$$

$$X_t^d = \int_0^t I_{\{\Delta a_s \ne 0\}} dX_s, \quad X_t^c = X_t - X_t^d;$$

I_A is the indicator function of a set A.

It must be noted that the process X^d has jumps at the same times as the process a^d and the process X^c has no jumps at the predictable stopping times.

For the fixed partition δ and weight matrix $\widetilde{W}$ one can find an estimator $\widehat{\theta}_\delta$ which minimizes the loss function $I_\delta(\theta)$. We have

$$\begin{aligned} \nabla_\theta I_\delta(\theta) = {} & -2\sum_{t_k<T} \Phi\widetilde{W}(t_k)\Delta X_{t_k}^d \Delta a_{t_k}^d + 2\sum_{t_k<T} \Phi\widetilde{W}\Phi'(t_k)(\Delta a_{t_k}^d)^2\theta \\ & -2\sum_{t_k<T} \Phi\widetilde{W}(t_k)\Delta X_{t_k}^c \Delta a_{t_k}^c (\Delta a_{t_k}^c)^{\oplus} + 2\sum_{t_k<T} \Phi\widetilde{W}\Phi'(t_k)\theta\Delta a_{t_k}^c, \end{aligned}$$

where ∇_θ is the gradient with respect to θ. The equation $\nabla_\theta I_\delta(\theta) = 0$ yields the estimator

$$\begin{aligned} \widehat{\theta}_\delta = {} & \left[\sum_{t_k<T} \Phi\widetilde{W}\Phi'(t_k)(\Delta a_{t_k}^d)^2 + \sum_{t_k<T} \Phi\widetilde{W}\Phi'(t_k)\Delta a_{t_k}^c\right]^{-1} \\ & \times \left[\sum_{t_k<T} \Phi\widetilde{W}(t_k)\Delta a_{t_k}^d \Delta X_{t_k}^d + \sum_{t_k<T} \Phi\widetilde{W}(t_k)\Delta X_{t_k}^c\right], \end{aligned}$$

where we make use of the equality $\Delta X_{t_k}^c \Delta a_{t_k}^c (\Delta a_{t_k}^c)^{\oplus} = \Delta X_{t_k}^c$, which is true, because the process X^c does not change on constancy intervals of the process a^c [see Stricker (1981)].

Taking a sequence of partitions $\delta_n = \{t_0^n, t_1^n, \ldots\}$, $0 = t_0^n < t_1^n < \ldots < t_n^n = T$, such that $\max_k(t_{k+1}^n - t_k^n) \to 0$ as $n \to \infty$ we obtain the following result:

PROPOSITION 1. *Let, for all $t > 0$,*

$$\int_0^t r[\Phi \widetilde{W} \Phi'](s)(\Delta a_s + I_{\{\Delta a_s = 0\}}) da_s < \infty \qquad a.s.,$$

$$\int_0^t tr[\Phi \widetilde{W} B \widetilde{W} \Phi'](s)((\Delta a_s)^2 + I_{\{\Delta a_s = 0\}}) da_s < \infty \qquad a.s.$$

and the matrix

$$\int_0^T \Phi \widetilde{W} \Phi'(s)(\Delta a_s + I_{\{\Delta a_s = 0\}}) da_s$$

be invertible for sufficiently large T a.s.

Then $\widehat{\theta}_{\delta_n} \to \widehat{\theta}_T$ in probability as $n \to \infty$ and $\max_k (t_{k+1}^n - t_k^n) \to 0$, where

$$\begin{aligned} \widehat{\theta}_T = &\left[\int_0^T \Phi \widetilde{W} \Phi'(s) \Delta a_s da_s + \int_0^T \Phi \widetilde{W} \Phi'(s) da_s^c \right]^{-1} \\ &\times \left[\int_0^T \Phi \widetilde{W} \Delta a_s dX_s^d + \int_0^T \Phi \widetilde{W}(s) dX_s^c \right]. \end{aligned} \tag{3.1}$$

By making use of the equalities

$$\Delta a_s dX_s = \Delta a_s dX_s^d, \qquad dX_s^c = I_{\{\Delta a_s = 0\}} dX_s,$$

we can rewrite $\widehat{\theta}_T$ as follows:

$$\widehat{\theta}_T = \left[\int_0^T \Phi(s) W(s) \Phi'(s) da_s \right]^{-1} \int_0^T \Phi(s) W(s) dX_s, \tag{3.2}$$

where

$$W(s) = \widetilde{W}(s)(\Delta a_s + I_{\{\Delta a_s = 0\}}).$$

The estimator $\widehat{\theta}_T$ is called the *weighted least-squares estimator.*

REMARK 2. If the matrix B in model (2.1) is non-degenerate, we can put $\widetilde{W}(t) = B^{-1}(t)(\Delta a_t + I_{\{\Delta a_t = 0\}})^{-1}$. In this case we obtain the *least-squares estimator* used by a number of authors.

REMARK 3. For discrete-time deterministic regression models with $\Delta a_k = 1, k = 0, 1, \ldots$ and $\widetilde{W}(k) = B^{-1}(k)$, the estimator $\widehat{\theta}_T$ is known to have the minimal variance [the Gauss-Markov theorem; see, e.g., Rao (1968)].

If $\Delta a_k \neq 1$, then $\widehat{\theta}_T$ is a LSE with special weights.

REMARK 4. The second factors in the right-hand sides of (3.1)–(3.2) are stochastic integrals with respect to the vector-valued semimartingales. By (2.1)

$$\int_0^T \Phi(s) W(s) dX_s = \int_0^T \Phi(s) W(s) \Phi'(s) da_s + \int_0^T \Phi(s) W(s) dm_s.$$

These integrals are well defined under the given conditions [see Galtchouk (1975, 1976), Jacod (1979)].

REMARK 5. It is noteworthy that the normalizing factor $(\Delta a^c_{t_k})^{\oplus}$ is used only for those addends in the loss function $I_\delta(\theta)$ which are connected with the component X^c of the observed process. The addends connected with the component X^d do not need any normalization. Note that Christopeit (1986) used a special normalization to obtain quasi-least-squares estimators.

4. Unbiased prescribed precision estimation for multidimensional processes. Let us consider the model (2.1) under the condition that both the process X_t and the vector θ are multidimensional but the dimension of unknown parameter vector θ does not exceed the dimension of X_t:

$$\dim\,\theta = p \le n = \dim\,X_t \qquad \forall t \ge 0.$$

In this case one can construct unbiased sequential estimators for θ with prescribed mean-square error by using the special weight matrix W and stopping rules. Our estimation scheme is similar to that proposed by Konev and Vorobejchikov (1980) for multidimensional discrete-time processes and later used by Mel'nikov and Novikov (1988) for semimartingale models.

Assume that the matrices B and $\Phi B^{-1}\Phi'$ are positive definite $d\mathbf{P} \times da$-a.e.

We begin with the weighted LSE

$$\widehat{\theta}_T = \left[\int_0^T \Phi(s)W(s)\Phi'(s)da_s\right]^{-1}\int_0^T \Phi(s)W(s)dX_s. \tag{4.1}$$

Let the weight matrix $W(t)$ be such that

$$\Phi W\Phi'(t) = c(t)I, \tag{4.2}$$

$$r[\Phi WBW\Phi'](t) \le c(t), \tag{4.3}$$

where $c(t)$ is a positive predictable function. Equation (4.2) is satisfied for

$$W(t) = c(t)B^{-1}\Phi'(\Phi B^{-1}\Phi')^{-2}\Phi B^{-1}(t). \tag{4.4}$$

Substituting this function in inequality (4.3) yields

$$c^2(t)r[(\Phi B^{-1}\Phi')^{-1}](t) \le c(t).$$

Let

$$c(t) = \{r[(\Phi B^{-1}\Phi')^{-1}](t)\}^{-1}. \tag{4.5}$$

Conditions (4.2), (4.3) enable us to invert the matrix in (4.1) and reduce the problem of constructing a sequential estimator for the vector θ to the scalar case.

For each $h > 0$ we introduce a predictable stopping time τ_h as

$$\tau_h = \inf\left\{t \ge 0 : \int_0^t \frac{da_s}{r[(\Phi B^{-1}\Phi')^{-1}](s)} \ge h\right\} \tag{4.6}$$

and a random variable η_h, with values in $[0,1]$, uniquely determined from the equation

$$\int_{]0,\tau_h[} \frac{da_s}{r[(\Phi B^{-1}\Phi')^{-1}](s)} + \eta_h \frac{\Delta a_{\tau_h}}{r[(\Phi B^{-1}\Phi')^{-1}](\tau_h)} = h. \tag{4.7}$$

The random variable η_h is $\mathscr{F}_{\tau_h -}$-mesurable.

On the basis of the estimator (4.1) with the weight matrix given by (4.4), (4.5), we define the sequential estimator for the vector θ as

$$\begin{aligned}\theta^*(h) &= h^{-1}\left[\int_{]0,\tau_h[} \Phi W(s)dX_s + \eta_h \Phi W(\tau_h)\Delta X_{\tau_h}\right] \\ &= h^{-1}\int_{]0,\tau_h]} \Phi W(s)(I_{]0,\tau_h[}(s) + \eta_h I_{\{\tau_h\}}(s))dX_s,\end{aligned} \tag{4.8}$$

where

$$W(t) = [r[(\Phi B^{-1}\Phi')^{-1}](t)]^{-1} B^{-1}\Phi'(\Phi B^{-1}\Phi')^{-2}\Phi B^{-1}(t). \tag{4.9}$$

This estimator has the following properties.

THEOREM 1. *Let matrices B and $\Phi B^{-1}\Phi'$ be not degenerate $d\mathbf{P}\times da$- a.e., the integral*

$$\int_0^t \frac{da_s}{r[(\Phi B^{-1}\Phi')^{-1}](s)} \tag{4.10}$$

be finite for $0 < t < \infty$ a.s. and converging to $+\infty$ as $t \to +\infty$ a.s.

Then, for each $h > 0$,

$$\tau_h < \infty \qquad a.s.,$$

$$\mathbf{E}_\theta \theta^*(h) = \theta,$$

$$\mathbf{E}_\theta ||\theta^*(h) - \theta||^2 \le h^{-1},$$

where $\mathbf{E}_\theta$ denotes the average by the distribution $\mathbf{P}_\theta$ of the process X with the given parameter θ.

PROOF. Since the integral (4.10) tends to $+\infty$, as $t \to \infty$, the stopping time τ_h is finite a.s. for all $h > 0$.

From (2.1) and (4.8) we have

$$\theta^*(h) = h^{-1}\left[\int_{]0,\tau_h]} \Phi W \Phi'(s)(I_{]0,\tau_h[}(s) + \eta_h I_{\{\tau_h\}}(s))\theta da_s + \widetilde{m}_{\tau_h}\right],$$

where

$$\widetilde{m}_{\tau_h} = \int_{]0,\tau_h]} \Phi W(s)(I_{]0,\tau_h[}(s) + \eta_h I_{\{\tau_h\}}(s))dm_s.$$

By (4.4)–(4.7),

$$\theta^*(h) = \theta + h^{-1}\widetilde{m}_{\tau_h}. \tag{4.11}$$

Since m is a locally square integrable martingale, then in view of (4.9),

$$
\begin{aligned}
&\operatorname{tr}\left[(<\widetilde{m}_i,\widetilde{m}_j>)_{1\le i,j\le n}\right](\tau_h)\\
&\quad=\int_{]0,\tau_h]} r[\Phi W B W \Phi'](s)(I_{]0,\tau_h[}(s)+\eta_h I_{\{\tau_h\}}(s))^2 da_s\\
&\quad=\int_{]0,\tau_h]} \frac{(I_{]0,\tau_h[}(s)+\eta_h I_{\{\tau_h\}}(s))^2 da_s}{r[(\Phi B^{-1}\Phi')^{-1}](s)}\\
&\quad\le\int_{]0,\tau_h]} \frac{(I_{]0,\tau_h[}(s)+\eta_h I_{\{\tau_h\}}(s)) da_s}{r[(\Phi B^{-1}\Phi')^{-1}](s)}=h.
\end{aligned}
\tag{4.12}
$$

Hence $\mathbf{E}_\theta \widetilde{m}_{\tau_h}=0$ and from (4.11) we obtain $\mathbf{E}_\theta\theta^*(h)=\theta$.

Further from (4.9), (4.11), (4.12), it follows that

$$\mathbf{E}_\theta||\theta^*(h)-\theta||^2=h^{-2}\mathbf{E}_\theta \operatorname{tr}\left[(<\widetilde{m}_i,\widetilde{m}_j>)_{1\le i,j\le n}\right](\tau_h)\le h^{-1}. \qquad \square$$

REMARK 6. If the matrix B is degenerate, the inverse matrix B^{-1} in the sequential procedure can be replaced by some positive definite symmetric matrix Γ^{-1} such that $B\le\Gamma$ and $\Phi\Gamma^{-1}\Phi'$ is not degenerate $d\mathbf{P}\times da$-a.e. Then the sequential design $(\tau_h,\theta^*(h))$ is defined by formulae

$$
\begin{aligned}
\tau_h&=\inf\left\{t\ge 0:\int_0^t \frac{da_s}{r[(\Phi\Gamma^{-1}\Phi')^{-1}](s)}\ge h\right\},\\
\theta^*(h)&=h^{-1}\int_{]0,\tau_h]}\frac{(\Phi\Gamma^{-1}\Phi')^{-1}\Phi\Gamma^{-1}(s)}{r[(\Phi\Gamma^{-1}\Phi')^{-1}](s)}(I_{]0,\tau_h[}(s)+\eta_h I_{\{\tau_h\}}(s))dX_s,
\end{aligned}
$$

where η_h is a multiplier which is determined from the equation

$$\int_{]0,\tau_h]}\frac{(I_{]0,\tau_h[}(s)+\eta_h I_{\{\tau_h\}}(s))da_s}{r[(\Phi\Gamma^{-1}\Phi')^{-1}](s)}=h.$$

If the integral

$$\int_0^t \frac{da_s}{r[(\Phi\Gamma^{-1}\Phi')^{-1}](s)}$$

is finite for $0<t<\infty$ and converges to $+\infty$ as $t\to+\infty$ a.s, then the assertion of Theorem 1 holds true for this sequential design.

REMARK 7. The unbiased guaranteed estimator considered in this section has been constructed under the assumption that the matrix $\Phi B^{-1}\Phi'$ (or $\Phi\Gamma^{-1}\Phi'$) is invertible. This condition is very restrictive and is not satisfied if $\dim\theta>\dim X$. This is the case, for example, when (X_t) is a scalar autoregressive process $AR(p)$ of order $p>1$.

5. Construction of the two-step sequential procedure in the case of an arbitrary number of unknown parameters. In the case when the number of unknown parameters in model (2.1) is arbitrary, the guaranteed estimator for θ can also be constructed on the basis of the weighted LSE, defined by (3.2). It is convenient to rewrite this estimate as

$$\theta(t) = M^{-1}(t) \int_0^t \Psi(s) W^{1/2}(s) dX_s, \tag{5.1}$$

where

$$M(t) = \int_0^t \Psi(s)\Psi'(s) da_s, \ \ \Psi(t) = \Phi W^{1/2}(t),$$

the inverse matrix $M^{-1}(t)$ is assumed to exist. The matrix M is called *the information matrix* or *design matrix.*

In the sequel the following conditions are imposed on the regressors and on the weight matrix W:

(A_1) The regressors matrix-valued function Φ is predictable and such that for all $t \geq 0$

$$\int_0^t \|\Phi(s)\| da(s) < \infty \qquad \text{a.s.}$$

(A_2) The weight matrix W is such that

$$W^{1/2} B W^{1/2} \leq I, \qquad dP \times da\text{-a.e.}$$

where I is the identity matrix of size $n \times n$.

(A_3) Both integrals in (5.1) are well defined. This is true if for all $t \geq 0$

$$\int_0^t \operatorname{tr} [\Psi\Psi'](s) \max(1, \Delta a_s + I_{\{\Delta a_s = 0\}}) da_s < \infty \qquad \text{a.s.}$$

Let $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ denote the smallest and the largest eigenvalues of the matrix M.

(A_4) $\lim_{t\to\infty} \lambda_{\min}(M(t)) = +\infty$, a.s.

(A_5) There exists δ, $0 < \delta < 1$, such that

$$\liminf_{t\to\infty} \lambda_{\min}^{\delta}(M(t)) / \ln \lambda_{\max}(M(t)) > 0 \qquad \text{a.s.}$$

The procedure is constructed in two steps.

Step 1. Let $(C_j)_{j\geq 1}$, $(\beta_j)_{j\geq 1}$ be two sequences of positive numbers such that

$$C_j \uparrow \infty, \qquad \sum_{j\geq 1} \beta_j < \infty, \qquad \sum_{j\geq 1} \beta_j C_j^{\frac{1-\delta}{\delta}} = \infty.$$

Here δ, $0 < \delta < 1$, is the same as in condition (A_5).

By virtue of condition (A_4), for any given positive constant C_0 we can define the a.s. finite stopping time T as

$$T = \inf\{t \geq 0 : \ \lambda_{\min}(M(t)) \geq C_0\}, \qquad \inf\{\varnothing\} = +\infty. \tag{5.2}$$

Next we introduce the sequence of stopping times τ_j, $j \geq 1$, as

$$\tau_j = \inf\left\{ t \geq T:\ C_0^{-1}\int_0^T \|\Psi\Psi'\|(s)da(s) + r\int_T^t \Psi' M^{-1}\Psi(s)da(s) \geq C_j \right\}, \tag{5.3}$$

and the sequence of estimators

$$\theta_j = \theta(\tau_j) = M^{-1}(\tau_j)\int_0^{\tau_j} \Psi W^{1/2}(s)dX_s.$$

On the basis of these estimators we define the desired sequential estimators of the unknown vector θ by applying a special smoothing procedure.

Step 2. Let us define the estimator θ_h^* as a weighted average of estimators θ_j:

$$\theta_h^* = \left[\sum_{j=1}^{\sigma(h)} b_j\right]^{-1} \sum_{j=1}^{\sigma(h)} b_j \theta_j, \tag{5.4}$$

where h is a positive parameter; $\sigma(h)$ is the stopping time given by

$$\sigma(h) = \inf\left(n \geq 1:\ \sum_{j=1}^{n} b_j \geq h\right),$$

$$b_j = \beta_j / [C_j tr[M^{-1}(\tau_j)]].$$

Denote

$$N(h) = \tau_{\sigma(h)}.$$

The main result is the following.

THEOREM 2. *Let the regressor matrix-valued function* Φ *in model* (2.1) *and the weight matrix* W *be predictable and satisfy conditions* (A_1)–(A_5). *Then the sequential design* $(N(h), \theta_h^*)$ *has the following properties*: *for any* $h > 0$,

$$\text{(i)} \qquad N(h) < \infty \qquad a.s.,$$

$$\text{(ii)} \qquad \mathbf{E}_\theta \|\theta_h^* - \theta\|^2 \leq h^{-1}\sum_{j=1}^{\infty} \beta_j (1 + pC_j^{-1}).$$

PROOF. By condition (A_4) we have $T < \infty$ a.s. From the definition of τ_j and Lemma 5 we have $\tau_j < \infty$ a.s. and $\tau_j \uparrow +\infty$, as $j \to \infty$. Therefore, the inequality $N(h) < \infty$ is true provided that

$$\sum_{j \geq 1} b_j = +\infty \qquad \text{a.s.} \tag{5.5}$$

Let us verify this equality. From the definitions of b_j and τ_j it follows that

$$\begin{aligned} b_j &= \beta_j/[C_j tr[M^{-1}(\tau_j)]] \geq \beta_j/[C_j p\lambda_{\max}(M^{-1}(\tau_j))] \\ &= \beta_j(pC_j)^{-1}\lambda_{\min}(M(\tau_j)) = \beta_j p^{-1} C_j^{1/\delta-1}\big(\lambda^{\delta}_{\min}(M(\tau_j))/C_j\big)^{1/\delta} \\ &\geq \beta_j p^{-1} C_j^{1/\delta-1} \\ &\quad\times\left(\lambda^{\delta}_{\min}(M(\tau_j))/[g(T)+r\int_{]T,\tau_j]}\Psi'(u)M^{-1}(u)\Psi(u)da(u)]\right)^{1/\delta}, \end{aligned}$$

where

$$g(T) = C_0^{-1}\int_0^T \|\Psi\|^2(u)da(u).$$

By making use of Lemma 3 we obtain

$$b_j \geq \beta_j p^{-1} C_j^{1/\delta-1}\left(\lambda^{\delta}_{\min}\big(M(\tau_j)\big)/[g(T)+L\ln\lambda_{\max}\big(M(\tau_j)\big)]\right)^{1/\delta},$$

where L is some positive constant. From this, the properties of β_j, C_j and condition (A_5) we obtain (5.5).

Further we have

$$\theta_j - \theta = M^{-1}(\tau_j)Y(\tau_j),$$

where

$$Y(t) = \int_0^t \Psi W^{1/2}(s)dm(s).$$

From this it follows that

$$\begin{aligned} \|\theta_j-\theta\|^2 &= \|M^{-1/2}(\tau_j)M^{-1/2}(\tau_j)Y(\tau_j)\|^2 \\ &\leq \|M^{-1/2}(\tau_j)\|^2\|M^{-1/2}(\tau_j)Y(\tau_j)\|^2 = Q(\tau_j)\mathrm{tr}\,[M^{-1}(\tau_j)], \end{aligned}$$

where

$$Q(\tau_j) = Y'(\tau_j)M^{-1}(\tau_j)Y(\tau_j). \tag{5.6}$$

Taking into account the definition of b_j and applying the Cauchy-Schwarz inequality we obtain

$$\begin{aligned} \|\theta^*(h)-\theta\|^2 &\leq \sum_{i=1}^{\sigma(h)} b_i\left(\sum_{j=1}^{\sigma(h)} b_j\|\theta_j-\theta\|^2\right)\left(\sum_{j=1}^{\sigma(h)} b_j\right)^{-2} \\ &\leq h^{-1}\sum_{j\geq 1} b_j\|\theta_j-\theta\|^2 \leq h^{-1}\sum_{j\geq 1} b_j Q(\tau_j)\mathrm{tr}\,[M^{-1}(\tau_j)] \\ &= h^{-1}\sum_{j\geq 1}\beta_j Q(\tau_j)/C_j. \end{aligned}$$

Hence,

$$\mathbf{E}\|\theta^*(h)-\theta\|^2 \le h^{-1}\sum_{j\ge 1}\beta_j \mathbf{E}Q(\tau_j)/C_j.$$

From this and Lemma 6 we obtain the desired result. □

EXAMPLE 5.1. Consider a non-explosive autoregression process (2.4), that is all roots of its characteristic polynomial

$$\mathscr{P}(z) = z^p - \theta_1 z^{p-1} - \cdots - \theta_p$$

lie on or inside the unite circle on the complex plane. Assume that the martingale-difference (ε_n) satisfy the conditions

$$\mathbf{E}(\varepsilon_n^2 \mid \mathscr{F}_{n-1}) = 1 \qquad \text{a.s.},$$
$$\sup_{n\ge 1}\mathbf{E}(\mid \varepsilon_n \mid^{\alpha} \mid \mathscr{F}_{n-1}) < \infty \qquad \text{a.s.}$$

for some $\alpha > 2$. Since the process X is scalar, $\mathrm{Var}(\varepsilon_n) = 1$, for all $n \ge 1$, and the process a is step-wise with the unit jumps, then we can put the weight matrix $W(t) = 1$ for all t in the basic estimator (5.1). One can easily verify that in this case conditions (A_1)–(A_3) are satisfied. To verify conditions (A_4), (A_5) we can apply Theorem 3 by Lai and Wei (1983) which yields

$$\liminf_{n\to\infty} n^{-1}\lambda_{\min}(M(n)) > 0 \tag{5.7}$$

and

$$\begin{aligned}\lambda_{\max}(M(n)) &= O(n) \qquad \text{a.s. if } \varrho = 0\\ &= O(n^{2\varrho}\log\log n) \qquad \text{a.s. if } \varrho \ge 1,\end{aligned} \tag{5.8}$$

where

$$M(n) = \sum_{k=1}^{n}\Phi_{k-1}\Phi'_{k-1},\ \ \Phi_k = (X_k, X_{k-1}, \dots, X_{k-p+1})',$$

$\varrho = 0$ if all roots of $\mathscr{P}(z)$ lie inside the unit circle and otherwise ϱ is the largest multiplicity of all the distinct roots on the unit circle. The property (5.8) implies condition (A_4). Further, by (5.7), (5.8),

$$\liminf_{n\to\infty}\lambda^{\delta}_{\min}(M(n))/\ln\lambda_{\max}(M(n)) \ge C\liminf_{n\to\infty} n^{\delta}/\ln n = \infty$$

for $\varrho = 0$ and all $0 < \delta < 1$. The same limiting relationship is true for the case $\varrho \ge 1$. Hence condition (A_5) is also satisfied and Theorem 2 is true for this autoregression model.

6. Auxiliary propositions.

6.1. *The Itô formula.* Let $Z = (Z(t))_{t\geq 0}$ be a semimartingale, $Z(t) \in \mathbb{R}^N$, $Z(t) = m(t) + A(t)$, where $m = (m_1, \ldots, m_N)'$ is a local martingale, $A = (A_1, \ldots, A_N)'$ is a process of finite variation on all compacts on the line. All processes are continuous on the right having left-side limits. Let $F : \mathbb{R}^N \to \mathbb{R}^1$ be a function, $F \in C^2(\mathbb{R}^N)$. Then

$$\begin{aligned} F(Z(t)) = F(Z(0)) &+ \int_0^t \big(\nabla_z F(Z(s-)), dm(s)\big) + \int_0^t \big(\nabla_z F(Z(s-)), dA(s)\big) \\ &+ \tfrac{1}{2} \textstyle\int_0^t r[\nabla_z \nabla_z F(Z(s-)) B^c(s)] d < m^c, m^c > (s) \\ &+ \sum_{0<s\leq t} [F(Z(s)) - F(Z(s-)) - \big(\nabla_z F(Z(s-)), \Delta Z(s)\big)] \qquad \text{a.s.,} \end{aligned}$$

where $Z(t-)$ denotes the limit of Z on the left at the time t, $\nabla_z = (\frac{\partial}{\partial z_1}, \ldots, \frac{\partial}{\partial z_N})'$; (x, y) is a scalar product of vectors x, y; m^c is the continuous component of the martingale m; and

$$B^c = \left(\frac{d\langle m_i^c, m_j^c \rangle}{d\langle m^c, m^c \rangle} \right)_{1\leq i,j\leq N}, \qquad \langle m^c, m^c \rangle = \sum_{i=1}^N \langle m_i^c, m_i^c \rangle.$$

The proof of this formula can be found, for example, in Meyer (1975).

LEMMA 1. *Let B be a $p \times p$ matrix and W be a $p \times 1$ non-zero vector of real numbers. If the matrix $A = B + WW'$ is non-singular, then*

$$W'A^{-1}W = (| A | - | B |)/ | A |,$$

where $| A |$ is the determinant of A.

This result is given (without proof) in Lemma 2 in the work of Lai and Wei (1982). For the sake of convenience, we reproduce the proof.

PROOF. We need the following equality:

$$|I - xy'| = 1 - y'x,$$

where I is the $p \times p$ identity matrix, x and y are $p \times 1$ vectors. To verify this equality in the case when vectors x, y are non-zero, it suffices to notice that $\lambda = 1$ is the eigenvalue of the matrix $I - xy'$ with multiplicity $p - 1$ and $\lambda = 1 - y'x$ is its eigenvalue with multiplicity one corresponding to the eigenvector x. The desired result easily follows from the equality

$$| B |=| A - WW' |=| A || I - A^{-1}WW' |=| A | (1 - W'A^{-1}W). \qquad \square$$

LEMMA 2. *Let B be a $p \times p$ matrix and Φ be a $p \times p$ symmetric non-negative definite non-zero matrix of real numbers. If the matrix $A = B + \Phi$ is non singular and $rank\Phi = r$, then*

$$tr[A^{-1}\Phi] = \sum_{i=1}^{r}[\mid A \mid - \mid A - \lambda_i e_i e_i' \mid]/\mid A \mid,$$

where (λ_i) and (e_i) are the eigenvalues and the eigenvectors of matrix Φ respectively. If, besides, B is symmetric non-negative definite, then

$$tr[A^{-1}\Phi] \leq r[\mid A \mid - \mid B \mid]/\mid A \mid \leq r.$$

PROOF. The matrix Φ can be written as

$$\Phi = \sum_{i=1}^{r} \lambda_i e_i e_i'.$$

Therefore

$$tr[A^{-1}\Phi] = \sum_{i=1}^{r} \lambda_i tr A^{-1} e_i e_i' = \sum_{i=1}^{r} \lambda_i e_i' \left(B + \sum_{j=1}^{r} \lambda_j e_j e_j' \right)^{-1} e_i.$$

By applying Lemma 1 we obtain

$$r[A^{-1}\Phi] = \sum_{i=1}^{r}[\mid A \mid - \mid A - \lambda_i e_i e_i' \mid]/\mid A \mid.$$

It remains to notice that $\mid A - \lambda_i e_i e_i' \mid \geq \mid B \mid, i = 1, 2, \ldots r,$ if B is symmetric non-negative definite. Hence Lemma 2. □

LEMMA 3. *Let the regressor matrix Φ in* (2.1) *and the weight matrix W satisfy conditions* (A_1), (A_3) *and T be defined as in* (5.2).

Then for any $t \geq T$

$$tr\left(\int_T^t \Psi'(u)\widetilde{M}^{-1}(u)\Psi(u)da^c(u)\right) = \int_T^t \frac{d \mid \widetilde{M}(u) \mid}{\mid \widetilde{M}(u) \mid} \qquad a.s., \tag{6.1}$$

where

$$\begin{aligned} \widetilde{M}(t) &= M(T) + \textstyle\int_T^t \Psi\Psi'(u)da^c(u), \\ \Psi(t) &= \Phi W^{1/2}(t). \end{aligned} \tag{6.2}$$

PROOF. First we verify that the integral in the left-hand side of equality (6.1) is well-defined, that is, for all $t \geq T$,

$$\int_T^t \|\Psi'(u)\widetilde{M}^{-1}(u)\Psi(u)\| da^c(u) < \infty \qquad \text{a.s.} \tag{6.3}$$

By the inequality

$$\|\Psi'(u)\widetilde{M}^{-1}(u)\Psi(u)\| \leq \|\Psi'\Psi(u)\| tr[\widetilde{M}^{-1}(u)]$$

we have

$$\begin{aligned}\int_T^t \|\Psi'(u)\widetilde{M}^{-1}(u)\Psi(u)\|da^c(u) &\le \int_T^t tr[\widetilde{M}^{-1}(u)]\|\Psi(u)\|^2 da^c(u)\\ &\le p\int_T^t \lambda_{\max}(\widetilde{M}^{-1}(u))tr[\Psi\Psi'(u)]da^c(u)\\ &\le p\lambda_{\min}^{-1}(M(T))\int_T^t tr[\Psi'\Psi(u)]da(u).\end{aligned}$$

In view of condition (A_3), we obtain (6.3).

Equality (6.1) is equivalent to the one for the differentials:

$$\frac{d\mid \widetilde{M}(u)\mid}{\mid \widetilde{M}(u)\mid} = tr[\Psi'(u)\widetilde{M}^{-1}(u)\Psi(u)]da^c(u).$$

Let us find $d\mid \widetilde{M}(u)\mid$. By the definition of a determinant we have

$$\mid \widetilde{M}(t)\mid = \sum_{(i_1,\dots,i_p)} (-1)^{[i_1,\dots,i_p]}\langle\widetilde{M}\rangle_{i_1,1}(t)\cdots\langle\widetilde{M}\rangle_{i_p,p}(t),$$

where $\langle\widetilde{M}\rangle_{ik}(t)$ is the (i,k)th element of the matrix $\widetilde{M}(t)$ and the summation is taken over all permutations $(i_1,\dots,i_p)$ of numbers $1,\dots,p$, and $[i_1,\dots,i_p]$ denotes the number of inversions in a permutation $(i_1,\dots,i_p)$.

Since the matrix-valued process $\widetilde{M}(t)$ is continuous with bounded variation then by the Itô formula we obtain

$$d\prod_{l=1}^p\langle\widetilde{M}\rangle_{i_l,l}(t) = \sum_{k=1}^p\left(\prod_{\substack{l=1\\ l\neq k}}^p\langle\widetilde{M}\rangle_{i_l,l}(t)\right)d\langle\widetilde{M}\rangle_{i_k,k}(t)$$

and, hence,

$$d\mid \widetilde{M}(t)\mid = \sum_{k=1}^p\sum_{(i_1,\dots,i_p)}(-1)^{[i_1,\dots,i_p]}\left(\prod_{\substack{l=1\\ l\neq k}}^p\langle\widetilde{M}\rangle_{i_l,l}(t)\right)d\langle\widetilde{M}\rangle_{i_k,k}(t).$$

By (6.2)

$$d\langle\widetilde{M}\rangle_{ik}(t) = <\Psi\Psi'>_{ik}(t)da^c(t)$$

and therefore

$$d\mid \widetilde{M}(t)\mid = \sum_{k=1}^p\mid \widetilde{M}^{(k)}(t)\mid da^c(t),$$

where $|\widetilde{M}^{(k)}(t)|$ is the determinant which is obtained from $\widetilde{M}$ by replacing the kth column by the column-vector $(\langle\Psi\Psi'\rangle_{1k},\dots,\langle\Psi\Psi'\rangle_{pk})'$. Decomposing the determinant $|\widetilde{M}^{(k)}(t)|$ by the elements of kth column yields

$$|\widetilde{M}^{(k)}(t)|=\sum_{i=1}^{p}\widetilde{M}_{ik}(t)\langle\Psi\Psi'\rangle_{ik}(t),$$

where $\widetilde{M}_{ik}(t)$ is the algebraic adjoint for the element $\langle\widetilde{M}\rangle_{ik}(t)$ of the matrix $\widetilde{M}(t)$. Thus

$$d\,|\widetilde{M}(t)|=\sum_{k=1}^{p}\sum_{i=1}^{p}\widetilde{M}_{ik}(t)\langle\Psi\Psi'\rangle_{ik}(t)\,da^c(t).$$

From here it follows that

$$\frac{d\,|\widetilde{M}(t)|}{|\widetilde{M}(t)|}=tr[\Psi'(t)\widetilde{M}^{-1}(t)\Psi(t)]\,da^c(t). \qquad \square$$

REMARK 8. A result similar to Lemma 3 is given in the paper of Christopeit [(1986), Lemma 3] under weaker assumption that the matrix-valued process $\widetilde{M}(t)$ is continuous on the left. However, by applying the Itô formula to the determinant $|\widetilde{M}(t)|$ one can easily make sure that the formula (6.1) is not true if the process $\widetilde{M}(t)$ admits jumps. Note that for this reason (and also because of a mistake in the proof) the corollary to the above-mentioned Lemma 3 does not hold without stronger assumptions. It worth noting that the above-given Itô formula is obtained for processes continuous on the right. For processes continuous on the left the Itô formula is also true [see Galtchouk (1980)].

LEMMA 4. *Under assumptions* (A_1) *and* (A_3) *the following inequalities are satisfied*:

$$\begin{aligned}\ln\frac{|M(t)|}{|M(T)|}&\le\int_T^t\frac{d\,|M(u)|}{|M(u-)|}\\&=\ln\frac{|M(t)|}{|M(T)|}+\sum_{T<s\le t}\left[\frac{\Delta|M(s)|}{|M(s-)|}-\ln\left(1+\frac{\Delta|M(s)|}{|M(s-)|}\right)\right]\qquad a.s.\end{aligned}$$

If $M(t)$ is a continuous matrix-valued process, then

$$\ln\frac{|M(t)|}{|M(T)|}=\int_T^t\frac{d\,|M(u)|}{|M(u)|}.$$

PROOF. By applying the Itô formula to the process $\ln|M(t)|$, $t\ge T$, we obtain

$$\begin{aligned}\ln|M(t)|=\ln|M(T)|&+\int_T^t\frac{d\,|M(u)|}{|M(u-)|}\\&+\sum_{T<s\le t}\left[\ln\frac{|M(s)|}{|M(s-)|}-\frac{\Delta|M(s)|}{|M(s-)|}\right]\ \text{a.s.}\end{aligned}$$

From here using the inequality $\ln(1+x) \le x$, $x \ge 0$, we come to the desired result. $\square$

LEMMA 5. *Under assumptions* (A_1), (A_3) *and* (A_4),

$$tr \int_T^t \Psi'(u) M^{-1}(u) \Psi(u)\, da(u) = O(\ln \lambda_{\max}(M(t))), \qquad t \to \infty \text{ a.s.}, \tag{6.4}$$

$$\lim_{t\to\infty} r \int_T^t \Psi'(u) M^{-1}(u) \Psi(u)\, da(u) = +\infty \qquad \text{a.s.}, \tag{6.5}$$

where $f(t) = O(g(t))$, $t \to \infty$, *means that there exist* $t_0 > T$ *and* $0 < C < \infty$ *such that* $| f(t) | \le C \mid g(t) \mid$ *for all* $t \ge t_0$.

PROOF. We have

$$\begin{aligned} tr \int_T^t \Psi'(u) M^{-1}(u) \Psi(u)\, da(u) &= tr \int_T^t \Psi'(u) M^{-1}(u-) \Psi(u)\, da^c(u) \\ &\quad + tr \sum_{T<u\le t} \Psi'(u) M^{-1}(u) \Psi(u) \Delta a(u). \end{aligned} \tag{6.6}$$

Let us introduce the process

$$\widetilde{M}(t) = M(T) + \int_T^t \Psi(u)\Psi'(u)\, da^c(u).$$

This process is continuous and satisfies the inequality $\widetilde{M}(t) \le M(t)$ which implies $M^{-1}(t) \le \widetilde{M}^{-1}(t)$. From this and Lemmas 3 and 4 it follows that for all $t \ge T$,

$$\begin{aligned} tr \int_T^t \Psi'(u) M^{-1}(u-) \Psi(u)\, da^c(u) &\le tr \int_T^t \Psi'(u) \widetilde{M}^{-1}(u) \Psi(u)\, da^c(u) \\ &= \int_T^t \frac{d \mid \widetilde{M}(u) \mid}{\mid \widetilde{M}(u) \mid} \\ &= \ln \frac{\mid \widetilde{M}(t) \mid}{\mid \widetilde{M}(T) \mid} \le \ln \frac{\mid M(t) \mid}{\mid M(T) \mid}. \end{aligned} \tag{6.7}$$

Now we find the upper bound for the second addend in the right-hand side of (6.6). Denoting

$$\hat{M}(t) = M(T) + \sum_{T<u\le t} \Psi(u)\Psi'(u)\Delta a(u)$$

and applying Lemma 2, we obtain

$$\begin{aligned} & tr \sum_{T<u\le t} \Psi'(u) M^{-1}(u) \Psi(u) \Delta a(u)]] \\ &\quad \le tr \sum_{T<s\le t} \Psi'(u) \hat{M}^{-1}(u) \Psi(u) \Delta a(u) \\ &\quad \le p \sum_{T<s\le t} [\mid \hat{M}(s) \mid - \mid \hat{M}(s-) \mid] / \mid \hat{M}(s) \mid \\ &\quad \le p \sum_{T<s\le t} \int_{|\hat{M}(s-)|}^{|\hat{M}(s)|} \frac{dx}{x} \\ &\quad \le p \int_{|\hat{M}(T)|}^{|\hat{M}(t)|} \frac{dx}{x} = p \ln \frac{\mid \hat{M}(t) \mid}{\mid \hat{M}(T) \mid} \le p \ln \frac{\mid M(t) \mid}{\mid M(T) \mid}. \end{aligned}$$

Substituting this estimate and (6.7) in (6.6) yields

$$tr \int_T^t \Psi'(u) M^{-1}(u) \Psi(u) da(u) \le (p+1) \ln \frac{|M(t)|}{|M(T)|} \le p(p+1) \ln \frac{\lambda_{\max}(M(t))}{|M(T)|}.$$

From this in view of condition (A_4) we obtain (6.4).

Now we verify (6.5). The integrand in (6.5) can be estimated from below by

$$\begin{aligned} tr[\Psi'(u) M^{-1}(u) \Psi(u)] &\ge \lambda_{\max}(\Psi'(u) M^{-1}(u) \Psi(u)) \\ &\ge \lambda_{\min}(M^{-1}(u)) \sup_{z \ne 0} \frac{\|\Psi(u) z\|^2}{\|z\|^2} \\ &\ge C \lambda_{\max}^{-1}(M(u)) \|\Psi(u)\|^2 \\ &\ge C \|\Psi(u)\|^2 / tr M(u) = C \|\Psi(u)\|^2 / V(u), \end{aligned}$$

where

$$V(u) = \int_0^u \|\Psi(s)\|^2 da(s),$$

and C is some positive constant. Hence,

$$\lim_{t\to\infty} tr \int_T^t \Psi'(u) M^{-1}(u) \Psi(u) da(u) \ge C \lim_{t\to\infty} \int_T^t \frac{\|\Psi(u)\|^2 da(u)}{V(u)} = C \lim_{t\to\infty} \int_T^t \frac{dV(u)}{V(u)}.$$

Assume that (6.5) is not true. Then with positive probability

$$\int_T^\infty \frac{dV(u)}{V(u)} < \infty.$$

From here it follows that

$$\lim_{t\to\infty} \frac{V(t-)}{V(t)} = 1,$$

and there exists such $T_1 > T$ that for all $t \ge T_1$

$$V(t-)/V(t) \ge 1/2.$$

By making use of this inequality and Lemma 4 we obtain

$$\begin{aligned} +\infty > \lim_{t\to\infty} \int_T^t \frac{dV(u)}{V(u)} &\ge \lim_{t\to\infty} \int_{T_1}^t \frac{dV(u)}{V(u-)} \frac{V(u-)}{V(u)} \\ &\ge 2^{-1} \lim_{t\to\infty} \int_{T_1}^t \frac{dV(u)}{V(u-)} \ge 2^{-1} \lim_{t\to\infty} \ln \frac{V(t)}{V(T_1)}. \end{aligned}$$

Thus, with positive probability,

$$\lim_{t\to\infty} \ln \lambda_{\min}(M(t)) \le \lim_{t\to\infty} \ln r(M(t)) = \lim_{t\to\infty} \ln V(t) < +\infty.$$

This contradicts to the condition (A_4). $\square$

LEMMA 6. *Under the assumptions* (A_1)–(A_5) *the function* $Q(\tau_j)$ *in* (5.6) *satisfies the inequality*

$$\mathbf{E}Q(\tau_j) \le C_j + p, \qquad j \ge 1,$$

where the sequence $(C_j)_{j\ge 1}$ *is the same as in* (5.3).

PROOF. Let us introduce the processes

$$\begin{aligned} Y(t) &= \big(y_1(t), \dots, y_p(t)\big)' = \int_0^t \Psi W^{1/2}(s)\,dm(s), \\ Z(t) &= \big(Y'(t), U_1'(t), \dots, U_p'(t)\big)', \\ F(Z(t)) &= Y'(t)M^{-1}(t)Y(t), \end{aligned} \tag{6.8}$$

where $U_i(t)$ is the ith column of the matrix $M^{-1}(t)$. Note that $Z(t)$ is a $(p+1)p \times 1$-dimensional vector semimartingale. In this notation we have

$$Q(\tau_j) = F(Z(\tau_j)). \tag{6.9}$$

Let us calculate the stochastic differential of the process $F(Z)$ by applying the Itô formula. The process $F(Z(t))$ can be written as

$$\begin{aligned} F(Z(t)) &= Y'(t)[U_1 \dots U_p]Y(t) \\ &= \big(Y'(t)U_1(t), \dots, Y'(t)U_p(t)\big)Y(t) \\ &= \sum_{i=1}^{p} Y'(t)U_i(t)y_i(t). \end{aligned} \tag{6.10}$$

For the function $F : \mathbb{R}^d \to \mathbb{R}^1$, $d = p(p+1)$ and the semimartingale Z defined by (6.8), the Itô formula has the form

$$\begin{aligned} F(Z(t)) = F(Z(T)) &+ \int_T^t \big(\nabla_y F(Z(s-)), dY(s)\big) \\ &+ \int_T^t \sum_{i=1}^{p} \big(\nabla_{u_i} F(Z(s-)), dU_i(s)\big) \\ &+ 2^{-1} \int_T^t tr[\nabla_y \nabla_y F(Z(s-))\Psi W^{1/2} B^c W^{1/2} \Psi'(s)] \\ &\qquad \times d < m^c, m^c > (s) \\ &+ \sum_{T<s\le t} [F(Z(s)) - F(Z(s-)) - \big(\nabla_z F(Z(s-)), \Delta Z(s)\big)], \end{aligned} \tag{6.11}$$

where $\nabla_y = (\frac{\partial}{\partial y_1}, \dots, \frac{\partial}{\partial y_p})'$, $(x, y) = y'x$ is a scalar product of vectors x, y. By (6.10) we obtain

$$\begin{aligned} \frac{\partial F}{\partial y_k} &= \sum_{i=1}^{p} \frac{\partial}{\partial y_k}[Y'(t)U_i(t)]y_i(t) + \sum_{i=1}^{p} Y'(t)U_i(t)\delta_{ik} \\ &= \sum_{i=1}^{p} u_{ik}(t)y_i(t) + Y'(t)U_k(t), \end{aligned}$$

where $U_i(t) = (u_{i1}(t), \dots, u_{ip}(t))'$. Therefore

$$
\begin{aligned}
&\nabla_y F(Z(t)) \\
&\quad = \sum_{i=1}^{p} (u_{i1}(t), \dots, u_{ip}(t))' y_i(t) + \big(Y'(t)U_1(t), \dots, Y'(t)U_p(t)\big)' \\
&\quad = \sum_{i=1}^{p} U_i(t) y_i(t) + M^{-1}Y(t) = 2M^{-1}Y(t).
\end{aligned} \tag{6.12}
$$

Further we have

$$
\begin{aligned}
\nabla_{u_k} F(Z(t)) &= \sum_{i=1}^{p} [\nabla_{u_k} Y'(t) U_i(t)] y_i(t) = \sum_{i=1}^{p} Y(t)\delta_{ik} y_i(t) = Y(t) y_k(t); \\
\nabla_y \nabla_y F &= \left(\frac{\partial^2 F}{\partial y_k \partial y_j} \right)_{1 \le k, j \le p}; \\
\frac{\partial^2 F(Z(t))}{\partial y_k \partial y_j} &= \frac{\partial}{\partial y_j} \frac{\partial F(Z(t))}{\partial y_k} = \frac{\partial}{\partial y_j} \left[\sum_{i=1}^{p} u_{ik}(t) y_i(t) + Y'(t) U_k(t) \right] \\
&= \sum_{i=1}^{p} u_{ik}(t) \delta_{ij} + u_{kj}(t) = 2u_{kj}(t).
\end{aligned} \tag{6.13}
$$

Thus

$$
\nabla_y \nabla_y F = 2M^{-1}(t).
$$

From this and (6.11)–(6.13), it follows that

$$
F(Z(t)) = F(Z(T)) + 2\mu_t + I_1(t) + I_2(t) + I_3(t), \tag{6.14}
$$

where

$$
\begin{aligned}
\mu_t &= \int_T^t \big(M^{-1}Y(s-), dY(s)\big) \\
&= \int_T^t \big(M^{-1}Y(s-), \Psi(s) W^{1/2}(s) dm(s)\big), \\
I_1(t) &= \sum_{i=1}^{p} \int_T^t (y_i(s-)\big(Y(s-), dU_i(s)\big), \\
I_2(t) &= \int_T^t tr[M^{-1}(s-)\Psi W^{1/2} B^c W^{1/2} \Psi'(s)] d < m^c, m^c > (s), \\
I_3(t) &= \sum_{T < s \le t} [F(Z(s)) - F(Z(s-)) - 2Y'(s-)M^{-1}(s-)\Delta Y(s) \\
&\qquad\qquad - Y'(s-)\Delta M^{-1}(s) Y(s-)].
\end{aligned} \tag{6.15}
$$

In order to study I_1 in (6.15) we need to find the differential for $M^{-1}(t)$. We have

$$
M^{-1}(t) = \left[\int_0^t \Psi\Psi'(s) da(s) \right]^{-1}, dM(t) = \Psi\Psi'(t) da(t)
$$

and

$$MM^{-1}(t) = I,$$

where I is the identity matrix. From here by the Itô formula

$$d[MM^{-1}(t)] = [dM(t)]M^{-1}(t) + M(t-)dM^{-1}(t) = 0$$

and, hence,

$$dM^{-1}(t) = -M^{-1}(t-)[dM(t)]M^{-1}(t) = -M^{-1}(t-)\Psi\Psi'(t)M^{-1}(t)da(t). \tag{6.16}$$

By (6.15) and (6.16),

$$\begin{aligned} I_1 &= \sum_{i=1}^{p}\sum_{k=1}^{p}\int_T^t y_i(s-)y_k(s-)d < M^{-1} >_{ik}(s) \\ &= \int_T^t Y'(s-)d[M^{-1}(s)]Y(s-) \\ &= -\int_T^t Y'(s-)M^{-1}(s-)\Psi(s)\Psi'(s)M^{-1}(s)Y(s-)da^c(s) \\ &\quad + \sum_{T<s\le t} Y'(s-)\Delta M^{-1}(s)Y(s). \end{aligned} \tag{6.17}$$

Due to the continuity of the process a^c the matrix $M^{-1}(s)$ can be changed to $M^{-1}(s-)$ in the last integral. The matrix $M^{-1}(s)$ is non-increasing because the matrix $M(s)$ is non-decreasing. Therefore the matrix $\Delta M^{-1}(s) \le 0$, and the right-hand side of (6.17) is non-positive. Thus

$$I_1(t) \le 0 \qquad \text{a.s. for all } t \ge T.$$

Consider the term $I_3(t)$ in (6.14). We have

$$\begin{aligned} I_3(t) &= \sum_{T<s\le t} [Y'(s)M^{-1}(s)Y(s) - Y'(s-)M^{-1}(s-)Y(s-) \\ &\qquad -2Y'(s-)M^{-1}(s-)\Delta Y(s) - Y'(s-)\Delta M^{-1}(s)Y(s-)] \\ &= \sum_{T<s\le t} [2Y'(s-)\Delta M^{-1}(s)\Delta Y(s) \\ &\qquad +\Delta Y'(s)M^{-1}(s-)\Delta Y(s) + \Delta Y'(s)\Delta M^{-1}(s)\Delta Y(s)] \\ &= \sum_{T<s\le t} [\Delta Y'(s)M^{-1}(s)\Delta Y(s) + 2\nu_t], \end{aligned}$$

where

$$\nu_t = \sum_{T<s\le t} [Y'(s-)\Delta M^{-1}(s)\Delta Y(s)]. \tag{6.18}$$

From (6.14), (6.15), (6.17) and (6.18), it follows that

$$\begin{aligned} F(Z(t)) &\le F(Z(T)) \\ &\quad + \int_T^t tr[M^{-1}(s-)\Psi W^{1/2}B^c W^{1/2}\Psi'(s)]d < m^c, m^c > (s) \\ &\quad +A(t) + 2\mu(t) + 2\nu(t) + \delta(t), \end{aligned} \tag{6.19}$$

where

$$\delta(t) = \sum_{T<s\le t} \Delta Y'(s) M^{-1}(s) \Delta Y(s) - A(t),$$

$\delta(t)$ is a local martingale, and $A(t)$ is the increasing predictable process in the Doob-Meyer decomposition of the submartingale $\sum_{T<s\le t} \Delta Y'(s)M^{-1}(s)\Delta Y(s)$:

$$A(t) = \int_T^t tr[W^{1/2}\Psi' M^{-1}\Psi W^{1/2} B^d(s)]d < m^d, m^d > (s),$$
$$= \int_T^t tr[M^{-1}\Psi W^{1/2} B^d W^{1/2}\Psi'(s)]d < m^d, m^d > (s).$$

The process $A(t)$ is well-defined, because by conditions (A_2) and (A_3) we have

$$\begin{aligned} A(t) &\le \int_T^t tr[M^{-1}\Psi\Psi'(s)]d < m^d, m^d > (s) \\ &\le \int_T^t tr[M^{-1}(s)\Psi(s)\Psi'(s)]da(s) \le [trM(T)]^{-1} \int_T^t \|\Psi(s)\Psi'(s)\| da(s) < \infty \end{aligned}$$

for all $t \ge T$.

Let us verify that the processes μ, ν in (6.19) are locally square integrable martingales. Their predictable quadratic variations are given by the formulae

$$\begin{aligned} \langle \mu, \mu \rangle (t) &= \int_T^t Y'(s-)M^{-1}(s-)\Psi(s) \\ &\qquad \times W^{1/2}(s)B(s)W^{1/2}(s)\Psi'(s)M^{-1}(s-)Y(s-)da(s), \\ \langle \nu, \nu \rangle (t) &= \int_T^t [Y'(s-)\Delta M^{-1}(s)\Psi(s)W^{1/2}(s)B^d(s) \\ &\qquad \times W^{1/2}(s)\Psi'(s)\Delta M^{-1}(s)Y(s-)d\langle m^d, m^d\rangle(s)]. \end{aligned}$$

By condition (A_2) and the Cauchy-Schwarz inequality,

$$\begin{aligned} \langle \nu, \nu \rangle (t) &\le \int_T^t Y'(s-)\Delta M^{-1}(s)\Psi(s)\Psi'(s)\Delta M^{-1}(s)Y(s-)da(s) \\ &= \int_T^t \left(Y'(s-)\Delta M^{-1}(s)\Psi(s)\right)^2 da(s) \\ &\le \int_T^t \|\Delta M^{-1/2}(s)Y(s-)\|^2 \|\Delta M^{-1/2}(s)\Psi(s)\|^2 da(s) \\ &\le (trM^{-1}(T))^2 \int_T^t \|Y(s-)\|^2 \|\Psi(s)\Psi'(s)\| da(s) < \infty \qquad \text{a.s.} \end{aligned}$$

for all $t \ge T$ due to condition (A_3) and the continuity of process $(Y(t-))_{t\ge T}$ on the left. In a similar way one can verify that $\langle \mu, \mu \rangle (t) < \infty$ a.s. for all $t \ge T$.

For each $k > 0$ we introduce the stopping time σ_k as

$$\sigma_k = \inf\{t > T : A(t) + \langle \mu, \mu \rangle (t) + \langle \nu, \nu \rangle (t) \ge k\}.$$

The stopping times σ_k are predictable and $\sigma_k \uparrow \infty$ a.s. as $k \to \infty$. There exist stopping times σ'_k such that

$$A(\sigma'_k) + \langle \mu, \mu \rangle(\sigma'_k) + \langle \nu, \nu \rangle(\sigma'_k) \le k, \qquad \sigma'_k \uparrow \infty$$

a.s. as $k \to \infty$ [see Dellacherie (1972)]. Then from (6.19) and condition (A_2) we obtain, for $k > 0$,

$$\begin{aligned}
&\mathbf{E}F(Z(t \wedge \sigma'_k)) \\
&\quad \le \mathbf{E}\biggl[F(Z(T)) + \int_T^{t\wedge\sigma'_k} tr[M^{-1}(s-)\Psi W^{1/2} B^c W^{1/2} \Psi'(s)] d\langle m^c, m^c\rangle(s) \\
&\qquad + \int_T^{t\wedge\sigma'_k} tr[M^{-1}(s)\Psi(s) W^{1/2}(s) B^d(s) W^{1/2}(s) \Psi'(s)] d\langle m^d, m^d\rangle(s)\biggr] \\
&\quad \le \mathbf{E}\biggl[F(Z(T)) + \int_T^{t\wedge\sigma'_k} tr[M^{-1}(s-)\Psi(s)\Psi'(s)] d\langle m^c, m^c\rangle(s) \\
&\qquad\qquad + \int_T^{t\wedge\sigma'_k} tr[M^{-1}(s)\Psi(s)\Psi'(s)] d\langle m^d, m^d\rangle(s)\biggr].
\end{aligned}$$

Letting $t = \tau_j$, taking the limit as k to $+\infty$ and applying the monotone convergence theorem, we obtain

$$\begin{aligned}
\mathbf{E}Q(\tau_j) &= \mathbf{E}(Y'(\tau_j) M^{-1}(\tau_j) Y(\tau_j)) \\
&\le \mathbf{E}\biggl[Y'(T) M^{-1}(T) Y(T) \\
&\qquad + \int_T^{\tau_j} tr[M^{-1}(s-)\Psi(s)\Psi'(s)] d\langle m^c, m^c\rangle(s) \\
&\qquad + \int_T^{\tau_j} tr[M^{-1}(s)\Psi(s)\Psi'(s)] d\langle m^d, m^d\rangle(s)\biggr].
\end{aligned}$$

Hence

$$\mathbf{E}Q(\tau_j) \le \mathbf{E}\Bigl(Y' M^{-1} Y(T) + \int_T^{\tau_j} tr[M^{-1}(s)\Psi(s)\Psi'(s)] da(s)\Bigr).$$

Now we can estimate $\mathbf{E}Q(\tau_j)$. We have

$$\begin{aligned}
\mathbf{E}Q(\tau_j) \le \ &\mathbf{E}\biggl[Y' M^{-1} Y(T) + \int_{]T,\tau_j[} tr[\Psi'(s) M^{-1}(s) \Psi(s)] da(s) \\
&\qquad + tr[\Psi'(\tau_j) M^{-1}(\tau_j) \Psi(\tau_j)] \Delta a(\tau_j)\biggr].
\end{aligned} \tag{6.20}$$

By the definition of stopping time T in (5.2) and condition (A_2),

$$\begin{aligned}
\mathbf{E}Y'(T) M^{-1}(T) Y(T) &\le \mathbf{E}\lambda_{\max}(M^{-1}(T)) Y'(T) Y(T) \\
&= \mathbf{E}\lambda_{\min}^{-1}(M(T)) \|Y(T)\|^2 \\
&\le C_0^{-1} \mathbf{E} \int_0^T \|\Psi(s)\|^2 da(s).
\end{aligned}$$

Combining this inequality and (6.20) yields

$$
\begin{aligned}
\mathbf{E}Q(\tau_j) \le \mathbf{E}\Bigg(& C_0^{-1}\int_0^T \|\Psi(s)\|^2 da(s) \\
& + \int_{]T,\tau_j[} tr[\Psi'(s)M^{-1}(s)\Psi(s)]da(s) \\
& + tr[\Psi'(\tau_j)M^{-1}(\tau_j)\Psi(\tau_j)]\Delta a(\tau_j)\Bigg) \\
\le C_j + \mathbf{E}\big(& tr[M^{-1}(\tau_j)\Psi(\tau_j)\Psi'(\tau_j)]\Delta a(\tau_j)\big).
\end{aligned} \tag{6.21}
$$

By Lemma 2,

$$
\begin{aligned}
& tr[M^{-1}(\tau_j)\Psi(\tau_j)\Psi'(\tau_j)\Delta a(\tau_j)] \\
& \quad \le r[\mid M(\tau_j) \mid - \mid M(\tau_j) - \Psi(\tau_j)\Psi'(\tau_j)\Delta a(\tau_j) \mid]/[\mid M(\tau_j) \mid] \\
& \quad \le r \le p,
\end{aligned}
$$

where r is the rank of the matrix $\Psi(\tau_j)\Psi'(\tau_j)$. By substituting this estimate in (6.21), we come to the assertion of Lemma 6. $\square$

Acknowledgments. We are grateful to referees, an Associate Editor and the Editor James Berger for helpful comments, as well as to Barry Arnold for careful reading of the manuscript and comments which improved the presentation.

REFERENCES

ANDERSON, T. W. and TAYLOR, J. B. (1976). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.* **4** 788–790.

ARAS, G. (1990). Sequential estimation of the mean of a first-order autoregressive process. *Comm. Statist. Theory Methods* **19** 1639–1652.

ARATO, M. (1982). *Linear stochastic systems with constant coefficients. A statistical approach. Lecture Notes in Control Inform. Sci.* **45** Springer, New York.

ARATO, M., KOLMOGOROV, A. N. and SINAI, Y. G. (1962). On parameter estimation of a complex stationary gaussian process. *Doklady Acad. USSR* **146** 747-750.

BORISOV, V. Z. and KONEV V. V. (1977). On sequential parameter estimation in discrete time processes. *Automat. Remote Control* **38** 58–64.

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

DARWICH, A. R. and LE BRETON, A. (1991). About asymptotic behavior of multidimensional Gaussian martingales in normal linear regression. *Statist. Probab. Lett.* **12** 317–321.

DELLACHERIE, C. (1972). *Capacites et processus stochastiques.* Springer, New York.

GALTCHOUK, L. I. (1975). The structure of a class of martingales. *Proceedings of the Seminar on Random Processes* 7–32. Academy of Sciences of Lithuania, Vilnus.

GALTCHOUK, L. I. (1976). Representation of some martingales. *Theory Probab. Appl.* **21** 613–620.

GALTCHOUK, L. I. (1980). Optional martingales. *Math. Sbornik* **112** 483–521.

GALTCHOUK, L. I. and MALJUTOV, M.B. (1995). One bound for the mean duration of sequential testing homogeneity. *Contributions to Statistics. Moda 4- Advances in Model-Oriented Data Analysis* 49–56. Springer, New York.

GALTCHOUK, L. I. and KONEV, V. I. (1997). On sequential estimation of parameters in continuous time stochastic regression. *Statistics and Control of Stochastic Processes. The R. Lipster Festschrift* 123–138. World Scientific, Singapore.

GREENWOOD, P. and SHIRYAEV, A. N. (1992). Asymptotic minimaxity of a sequential estimator for a first order autoregressive model. *Stochastics* **38** 49–65.

JACOD, J. (1979). *Calcul stochastique et problemes de martingales. Lecture Notes in Math.* **714**. Springer, New York.

JAUKUNAS, A. and KHASMINSKII, R. Z. (1997). Estimation of parameters of linear homogeneous stochastic differential equations. *Stochastic Processes Appl.* **72** 205–219.

KONEV, V. V. (1985). *Sequential Parameter Estimation of Stochastic Dynamical Systems.* Tomsk Univ. Publishing House, Tomsk. (In Russian.)

KONEV, V. V. and LAI, T. L. (1995). Estimators with prescribed precision in stochastic regression models. *Sequential Anal.* **14** 179–192.

KONEV, V. V. and PERGAMENSHCHIKOV S. (1981). Sequential identification plans for dynamic systems. *Automation Remote Control* **42** 917–924.

KONEV, V. V. and PERGAMENSHCHIKOV, S. (1992). Sequential estimation of parameters in unstable stochastic system with guaranteed mean-square accuracy. *Problems Inform. Trans.* **28** 35–48.

KONEV, V. V. and PERGAMENSHCHIKOV,S. (1997). On guaranted estimation of the mean of an autoregressive process. *Ann. Statist.* **25** 2127–2163.

KONEV, V. V. and VOROBEICHIKOV, S. E. (1980). On sequential identification of stochastic systems. *Izvestia Acad. USSR Technical Cybern.* **N4** 176–182.

LAI, T. L., ROBBINS, H. and WEI, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* **9** 343–361.

LAI, T. L. and WEI, C. Z. (1982). Least square estimates in stochastic regression regression models with application to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166.

LAI, T. L. and WEI, C. Z. (1983). Asymptotic properties of general autoregressive models and strong consistency of least square estimates of their parameters. *J. Multivariate Anal.* **13** 1–23.

LAI, T. L. and ZIEGMUND, D. (1983). Fixed-accuracy estimation of an autoregressive parameter. *Ann. Statist.* **11** 478–485.

LE BRETON, A. (1977). Parameter estimation in a linear stochastic differential equation. *Proceedings of the 7th Prague Conference* 15–32. Chechoslovak Academy of Science, Prague.

LE BRETON, A. and MUSIELA, M. (1987). Strong consistency of LS-estimates in linear regression models driven by semimartingales. *J. Multivariate Anal.* **23** 77–92.

LIPSTER, R. SH. and SHIRYAEV, A. N. (1977). *Statistics of Random Processes. I, II.* Springer, New York.

LIPSTER, R. SH. and SHIRYAEV, A. N. (1989). *Theory of Martingales.* Kluwer, Dordrecht.

MELNIKOV, A. V. (1996). Stochastic differential equations: nonsmoothness of coeffiicients, regression models and stochastic approximation. *Russian Math. Surveys* **51** 43–136.

MELNIKOV, A. V. and NOVIKOV, A. A. (1988). Sequential inference with fixed accuracy for semimartigales. *Theory Probab. Appl.* **33** 480–494.

MEYER, P. A. (1975). *Un cours sur les integrales stochastiques. Seminaire de Probabilites X. Lecture Notes in Math.* **511.** Springer, New York.

NOVIKOV, A. A. (1984). Consistency of least squares estimates in regression models with martingale errors. *Statistics and Control of Stochastic Processes* (N. V. Krylov, R. S. Liptser and A. A. Novikov, eds.) 389–409. Springer, New York.

RAO, C. R. (1968). *Linear Statistical Inferences and Its Applications.* Wiley, New York.

SHIRYAEV, A. N. and SPOKOINY, V. G. (1997). On sequential estimation of an autoregressive parameter. *Stochastics Stochastic Reports* **60** 219–240.

SRIRAM, T. N. (1988). Sequential estimation of the autoregressive parameter in a first-order autoregressive process. *Sequential Anal.* **7** 53–74.

STRICKER, CH. (1981). Les intervalles de constance de $\langle X, X \rangle$. *Seminaire de Probabilites XVI. Lecture Notes in Math.* **920** 219–220. Springer, New York.

DEPARTMENT OF MATHEMATICS
STRASBOURG UNIVERSITY
7, RUE RENE DESCARTES
67084, STRASBOURG
FRANCE
E-MAIL: galtchou@math.u-strasbg.fr

DEPARTMENT OF APPLIED MATHEMATICS
TOMSK UNIVERSITY
36, LENIN STR.
634050, TOMSK
RUSSIA
E-MAIL: vvkonev@vmm.tsu.tomsk.ru