Georgian Mathematical Journal
Volume 10 (2003), Number 2, 271–288

GENERAL M-ESTIMATORS IN THE PRESENCE OF NUISANCE PARAMETERS. SKEW PROJECTION TECHNIQUE

N. LAZRIEVA AND T. TORONJADZE

Abstract. A multidimensional parametric filtered statistical model is considered, the notions of the regularity and ergodicity of the model, as well as the notion of a general M-estimator are studied. A skew projection technique is proposed, by which we construct an M-estimator of the parameter of interest in the presence of nuisance parameters, which is asymptotically normal with the zero mean and the same covariance matrix as the corresponding block of the asymptotic covariance matrix of the full parameter's M-estimator.

2000 Mathematics Subject Classification: 65F12.
Key words and phrases: General M-estimator, nuisance parameter, projection.

1. Introduction

The well-known problem of the statistical estimation theory concerns the construction of estimators of parameter of interest in the presence of nuisance parameters. The projection technique is a frequently used for constructing efficient in the Fisher sense estimators in various observations (e.g., i.i.d., regression, etc.) schemes [1], [10], [3].

On the other hand, the key object in robust statistics is M-estimator, introduced by Huber [4] in the case of i.i.d. observations (see also [2], [8] for filtered statistical models). Hence to construct robust estimators in the presence of nuisance parameters it seems necessary to construct M-estimators in such a situation. For these purposes one needs to adopt the projection technique for M-estimators. As it will be shown below one need to use a skew projection of partial scores on the remainder terms of a full score (see (3.4)) rather than an orthogonal projection as in the case of MLE construction.

In the present paper a multidimensional parametric filtered statistical model is considered, notions of the regularity and ergodicity of models, a notion of general M-estimator are introduced (see Section 2).

In Section 3 a skew projection technique is proposed, which allows us to construct an estimator of a parameter of interest solving a d-dimensional equation (d is dimension of a parameter of interest) instead of a $(d+m)$-dimensional estimational equation for a M-estimator of the full $(d+m)$-dimensional parameter (see, e.g., (3.10) and (3.3), respectively, for $d=1$).

ISSN 1072-947X / \$8.00 /

In Theorem 3.1 we prove that under some regularity and ergodicity conditions the constructed estimator is asymptotically normal with the zero mean and the same covariance matrix as the corresponding $(d \times d)$-dimensional block of the asymptotic covariance matrix of the M-estimator of the full parameter (see (3.32) for $d = 1$).

In Theorem 3.2 we establish asymptotic properties of the estimator of a parameter of interest constructed by the one-step approximation procedure, where the score martingale constructed by skew projection is used (see (3.33) for $d = 1$).

Appendix is devoted to formulation of the key Lemma which is the base of the proofs of Theorems.

2. Specification of the Model, Regularity, Ergodicity

We consider a filtered statistical model

$$\mathcal{E} = \big(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0},\ \{P_\theta,\ \theta \in \Theta \subset R^{d+m}\}, P\big), \quad d \geq 1, \quad m \geq 1, \tag{2.1}$$

where $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ is a stochastic basis satisfying the usual conditions, $\{P_\theta,\ \theta \in \Theta \subset R^{d+m}\}$ is a family of probability measures on $(\Omega, \mathcal{F})$ such that $P_\theta \overset{loc}{\sim} P$ for each $\theta \in \Theta$, Θ is an open subset of R^{d+m}. Denote by

$$\rho(\theta) = \left(\rho_t(\theta) = \frac{d\,P_{\theta,t}}{d\,P_t}\right)_{t\geq 0}$$

the density (likelihood) ratio process, where

$$P_{\theta,t} = P_\theta \mid \mathcal{F}_t, \quad P_t = P \mid \mathcal{F}_t$$

are, respectively, the restriction of measures P_θ and P on the σ-algebra $\mathcal{F}_t$.

In the sequel we assume for simplicity that $P_{\theta,0} = P_0$ for all $\theta \in \Theta$.

As is well-known [9], there exists $M(\theta) = (M_t(\theta))_{t\geq 0} \in \mathcal{M}_{loc}(P)$ such that

$$\rho(\theta) = \mathcal{E}(M(\theta)), \tag{2.2}$$

where $\mathcal{E}(\cdot)$ is the Dolean exponential, i.e.,

$$\mathcal{E}_t(M(\theta)) = \exp\left\{M_t(\theta) - \frac{1}{2}\langle M^c(\theta)\rangle_t\right\} \prod_{s\leq t} \big(1 + \Delta M_s(\theta)\big)\, e^{-\Delta M_s(\theta)}$$

(for all unexplained notation see, e.g., [9], [5]).

Then

$$\ln \rho(\theta) = M(\theta) - \frac{1}{2}\,\langle M^c(\theta)\rangle + \sum_{s\leq \cdot} \big(\ln(1 + \Delta M_s(\theta)) - \Delta M_s(\theta)\big). \tag{2.3}$$

Definition 2.1. We say that model (2.1) is regular if the following conditions are satisfied:

(1) for each (t, ω) the mapping $\theta \rightsquigarrow M_t(\omega, \theta)$ is differentiable and $\dot{M}(\theta) \in \mathcal{M}_{loc}(P)$, where $\dot{M}(\theta) = \Big(\frac{\partial}{\partial\theta_1} M(\theta), \dots, \frac{\partial}{\partial\theta_{d+m}} M(\theta)\Big)'$,[1]

[1] In the sequel all vectors are assumed to be column-vectors, "$'$" is transposition sign.

(2) for each (t,ω) the mapping $\theta \rightsquigarrow \rho_t(\omega,\theta)$ is differentiable and

$$\frac{\partial}{\partial\theta}\ln\rho(\theta) = L\left(\dot M(\theta), M(\theta)\right),$$

i.e., $\frac{\partial}{\partial\theta_i}\ln\rho(\theta) = L\left(\frac{\partial}{\partial\theta_i}M(\theta), M(\theta)\right)$, where for $m, M \in \mathcal{M}_{loc}(P)$ $L(m,M)$ is the Girsanov transformation defined as

$$\begin{aligned} L(m,M) :&= m - (1+\Delta M)^{-1}\circ[m,M] \\ &= m - \langle m^c, M^c\rangle - \sum_{s\le\cdot}\frac{\Delta m_s\,\Delta M_s}{1+\Delta M_s}\,, \end{aligned} \tag{2.4}$$

(3) $L(\dot M(\theta), M(\theta)) \in \mathcal{M}^2_{loc}(P_\theta)$.

Recall that if $m\in\mathcal{M}_{loc}(P)$, Q is some measure such that $\frac{dQ}{dP} = \mathcal{E}(M)$, then $L(m,M)\in\mathcal{M}_{loc}(Q)$ (see, e.g., [5]).

Define the $(d+m)\times(d+m)$-matrix-valued Fisher information process and the Fisher information matrix as follows:

$$\widehat{I}_t(\theta) = \langle L(\dot M(\theta), M(\theta))\rangle_t, \quad t\ge 0,$$
$$I_t(\theta) = E_\theta\,\widehat{I}_t(\theta), \quad t\ge 0,$$

where E_θ stands for expectation w.r.t. P_θ.

Remark 2.1. Consider the case where model (2.1) is associated with a k-dimensional, $k\ge 1$, $\mathbb{F}$-adapted CADLAG process $X=(X_t)_{t\ge 0}$ in the following way: for each $\theta\in\Theta$ P_θ is a unique measure on $(\Omega,\mathcal{F})$ such that X is a P_θ-semimartingale with the triplet of predictable characteristics $(B(\theta), C(\theta), \nu_\theta)$ w.r.t. the standard truncation function $h(x) = xI_{\{|x|\le 1\}}$, where $|\cdot|$ is the Euclidean norm in R^k.

In what follows we assume that $P_{\theta'} \overset{loc}{\sim} P_\theta$ for each pair $(\theta,\theta')\in R^{d+m}\times R^{d+m}$ and all measures P_θ coincide on $\mathcal{F}_0$.

Fix some $\theta^*\in\Theta$ and let $P := P_{\theta^*}$, $(B,C,\nu) := (B(\theta^*), C(\theta^*), \nu_{\theta^*})$. Without loss of generality we can assume that there are a predictable process $c=(c_{ij})_{i,j\le k}$ with values in the set of all symmetric nonnegative definite $k\times k$-matrices and an increasing continuous $\mathbb{F}$-adapted process $A=(A_t)_{t\ge 0}$ such that

$$C = c\circ A \quad P\text{-a.s.}$$

The following relationship between triplets $(B(\theta), C(\theta), \nu_\theta)$ and (B,C,ν) is well known (see, e.g., [9]): there exists a $\widetilde{\mathcal{P}}$-measurable positive function $Y(\theta) = \left\{Y(\omega,t,x;\theta) : (\omega,t,x)\in\Omega\times R_+\times R^k\right\}$ and a predictable k-dimensional process $\beta(\theta) = (\beta_t(\theta))_{t\ge 0}$ with

$$\left|h(x)(Y(\theta)-1)\right| * \nu \in A^+_{loc}(P_\theta),$$
$$\left(\beta'(\theta)c\beta(\theta)\right)\circ A \in A^+_{loc}(P_\theta)$$

such that

(1) $B(\theta) = B + c\beta(\theta)\circ A + h(x)(Y(\theta)-1)*\nu$,
(2) $C(\theta) = C$,

(3) $\nu_\theta = Y(\theta) \cdot \nu$,
and, in addition, $a_t := \nu\big(\{t\}, R^k\big) = 1 \Rightarrow a_t(\theta) := \nu_\theta\big(\{t\}, R^k\big) = \widehat{Y}_t(\theta) = 1$. Here for a given $\widetilde{\mathcal{P}}$-measurable function W, $\widehat{W}_t = \int W(t,x)\nu(\{t\}, dx)$.

It is also well known (see, e.g., [6], [9]) that the P_θ-local martingale $M(\theta)$ in (2.2) can be writing explicitly as

$$M(\theta) = \beta'(\theta) \cdot X^c + \left(Y(\theta) - 1 + \frac{\widehat{Y}(\theta) - a}{1-a}\, I_{(0<a<1)}\right) * (\mu - \nu), \tag{2.5}$$

where X^c is the continuous martingale part of X under P.

Suppose that the following conditions are satisfied:

1) for all (ω, t, x) the functions $\beta : \theta \rightsquigarrow \beta_t(\omega, \theta)$ and $Y : \theta \rightsquigarrow Y(\omega, t, x; \theta)$ are differentiable (notation: $\dot\beta(\theta) = \frac{\partial}{\partial\theta}\,\beta(\theta) = (\dot\beta_{ij}(\theta))_{i\le k,\, j\le d+m}$ with $\dot\beta_{ij}(\theta) = \frac{\partial}{\partial\theta_j}\,\beta_i(\theta)$, $\dot Y(\theta) = \frac{\partial}{\partial\theta}\,Y(\theta) = \Big(\overset{(1)}{Y}(\theta), \dots, \overset{(d+m)}{Y}(\theta)\Big)'$ with $\overset{(j)}{Y}(\theta) = \frac{\partial}{\partial\theta_j}\,Y(\theta)$, $1 \le j \le d+m$,
2) $\dot M(\theta) := \frac{\partial}{\partial\theta}\,M(\theta) = \dot\beta'(\theta) \cdot X^c + \Big(\dot Y(\theta) + \frac{\dot a(\theta)}{1-a}\Big) * (\mu - \nu) \in \mathcal{M}_{loc}(P)$,
3) $\frac{\partial}{\partial\theta}\,\ln\rho(\theta) = L\left(\dot M(\theta), M(\theta)\right) \in \mathcal{M}^2_{loc}(P_\theta)$.

Then model (2.1) associated with the semimartingale X is regular in the sense of Definition 2.1.

It is not difficult to verify that

$$L(\theta) := L\left(\dot M, M(\theta)\right) = \dot\beta'(\theta) \cdot (X^c - c\beta(\theta) \circ A) + \Phi(\theta) * (\mu - \nu_\theta),$$

where

$$\Phi(\theta) = \frac{\dot Y(\theta)}{Y(\theta)} + \frac{\dot a(\theta)}{1 - a(\theta)}$$

and $I_{\{a(\theta)=1\}}\dot a(\theta) = 0$.

Another definition of the regularity of a statistical model associated with the semimartingale X is given in [7]. Namely, suppose that the following conditions are satisfied: for each $\theta \in \Theta$ there exist a predictable $k \times (d+m)$-matrix-valued process $\dot\beta(\theta)$ and a predictable $(d+m)$-dimensional vector function $W(\theta)$ with

$$\left|\dot\beta'(\theta)\, c\, \dot\beta(\theta)\right| \circ A_t < \infty, \quad |W(\theta)|^2 * \nu_t < \infty, \quad t \ge 0, \quad P\text{-a.s.}$$

and such that for all $t \geq 0$ we have

$$
\begin{aligned}
&(1)\quad \Big[\Big(\beta(\theta') - \beta(\theta) - \dot{\beta}(\theta)(\theta'-\theta)\Big)' \times \\
&\qquad\qquad c\Big(\beta(\theta') - \beta(\theta) - \dot{\beta}(\theta)(\theta'-\theta)\Big)\Big] \circ A_t \Big/ |\theta'-\theta|^2 \xrightarrow{P_\theta} 0, \\
&(2)\quad \left(\sqrt{\frac{Y(\theta')}{Y(\theta)}} - 1 - \frac{1}{2} W'(\theta)(\theta'-\theta)\right)^2 * \nu_{\theta,t} \Big/ |\theta'-\theta|^2 \xrightarrow{P_\theta} 0, \\
&(3)\quad \Bigg[\sum_{\substack{s \leq t \\ a_s(\theta)<1}} \Bigg(\sqrt{1-a_s(\theta')} - \sqrt{1-a_s(\theta)} \\
&\qquad\qquad + \frac{1}{2} \frac{\widehat{W}^\theta_T}{\sqrt{1-a_s(\theta)}} (\theta'-\theta)\Bigg)^2\Bigg] \Big/ |\theta'-\theta| \xrightarrow{P_\theta} = 0
\end{aligned}
\tag{2.6}
$$

as $\theta' \to \theta$, where $\widehat{W}^\theta_T(\theta) = \int W(t,x,\theta)\nu_\theta(\{t\},dx)$.

In this case as it is proved in [7] model (2.1) is regular in the following sense: at each point $\theta \in \Theta$ the process $\sqrt{\frac{\rho(\theta')}{\rho(\theta)}}$ is locally (in time) differentiable w.r.t. θ' with derivative $(d+m)$-dimensional process $\widetilde{L}(\theta) = (\widetilde{L}_t(\theta))_{t\geq 0} \in \mathcal{M}^2_{loc}(P_\theta)$ defined as follows:

$$
\widetilde{L}(\theta) = \dot{\beta}'(\theta) \cdot (X^c - c\beta(\theta) \circ A) + \left(W(\theta) + \frac{\widehat{W}^\theta(\theta)}{1-a(\theta)}\right) * (\mu - \nu_\theta).
$$

The Fisher information $(d+m) \times (d+m)$-matrix-valued process is defined as

$$
\widehat{I}(\theta) := \langle \widetilde{L}(\theta), \widetilde{L}(\theta)\rangle = \Big(\dot{\beta}'(\theta) c \beta(\theta)\Big) \circ A + W(\theta)W(*\theta)' * \nu + \sum_{s\leq \cdot} \frac{\widehat{W}^\theta_s(\theta)\widehat{W}^{\theta'}_s(\theta)}{1-a_s(\theta)} .
$$

Denote

$$
\widetilde{\Phi}(\theta) = W(\theta) + \frac{\widehat{W}^\theta(\theta)}{1-a(\theta)} .
$$

Then

$$
\widetilde{L}(\theta) = \dot{\beta}'(\theta) \cdot (X^c - c\beta(\theta) \circ A) + \widetilde{\Phi}(\theta) * (\mu - \nu_\theta).
$$

It should be noticed that if the model is regular in both above-given senses, then

$$
W(\theta) = \frac{\dot{Y}(\theta)}{Y(\theta)}, \quad \widehat{W}^\theta(\theta) = \dot{a}(\theta), \quad \Phi(\theta) = \widetilde{\Phi}(\theta).
$$

To preserve these notation for the second variant of regularity (regularity in the Jacod sense) let us formally define $\dot{Y}(\theta)$ and $\dot{a}(\theta)$ as

$$
\dot{Y}(\theta) = 2\sqrt{Y(\theta)}\, \frac{\partial}{\partial\theta} \sqrt{Y(\theta)}, \quad \dot{a}(\theta) = -2\sqrt{1-a(\theta)}\, \frac{\partial}{\partial\theta} \sqrt{1-a(\theta)}\,, \tag{2.7}
$$

where $\frac{\partial}{\partial\theta}\sqrt{Y(\theta)} = W(\theta)\sqrt{Y(\theta)}$, $\frac{\partial}{\partial\theta}\sqrt{1-a(\theta)} = -\frac{1}{2}\frac{\widehat{W}^\theta(\theta)}{\sqrt{1-a(\theta)}}$.

Then we have $W(\theta)=\frac{\dot Y(\theta)}{Y(\theta)}$, $\widehat{W}(\theta)=\dot a(\theta)$. $\dot M(\theta)$ can also be defined formally as

$$\dot M(\theta)=\dot\beta'(\theta)\cdot X^c+\left(\dot Y(\theta)+\frac{\dot a(\theta)}{1-a(\theta)}\right)*(\mu-\nu),$$

where $\dot\beta(\theta)$ is defined from (1) of (2.6) and $\dot Y(\theta)$ and $\dot a(\theta)$ from (2.7). In this case it is not difficult to verify that

$$\widetilde{L}(\theta)=L\left(\dot M(\theta),M(\theta)\right).$$

Many problems of asymptotic statistics have a natural setting in the scheme of series. We will now give the definition of the ergodicity of statistical models in terms of the scheme of series.

For this consider the scheme of series of regular statistical models

$$\mathcal{E}=(\mathcal{E}_n)_{n\ge1}=\big(\Omega^n,\mathcal{F}^n,\,\mathbb{F}^n=(\mathcal{F}^n_t)_{0\le t\le T},\,\{P^n_\theta,\,\theta\in\Theta\subset R^{d\times m}\},\,P^n\big)_{n\ge1},\quad T>0,$$

where it is assumed that $P^n_\theta\sim P^n$ for each $n\ge1$, $d\ge1$, $m\ge1$, and $\theta\in\Theta$.

Definition 2.2. We say that the series of filtered statistical models $\mathcal{E}=(\mathcal{E}_n)_{n\ge1}$ is ergodic if there exists a numerical sequence $(C_n)_{n\ge1}$ such that $C_n>0$, $\lim\limits_{n\to\infty}C_n=0$ and for each $\theta\in\Theta\subset R^{d\times m}$

$$C_n^2\widehat{I}^n_T(\theta)\xrightarrow{P^n_\theta}I_T(\theta),\quad C_n^2I^n_T(\theta)\to I_T(\theta)\quad\text{as}\quad n\to\infty,$$

where $I_T(\theta)$ is a finite positive definite matrix. In the sequel the subscript "T" will be omitted.

Denote $\ell^n_j(\theta)=L\big(\overset{(j)}{M}{}^n(\theta),M^n(\theta)\big)$, where $\overset{(j)}{M}{}^n(\theta)=\frac{\partial}{\partial\theta_j}M^n(\theta)$, $j\le d+m$. Obviously,

$$\widehat{I}^n_{ij}(\theta)=\langle\ell^n_i(\theta),\ell^n_j(\theta)\rangle,\quad \widehat{I}^n_{ij}(\theta)=E_\theta\langle\ell^n_i(\theta),\ell^n_j(\theta)\rangle$$

and as $n\to\infty$

$$C_n^2\langle\ell^n_i(\theta),\ell^n_j(\theta)\rangle\xrightarrow{P^n_\theta}I_{ij}(\theta),\quad C_n^2E_\theta\langle\ell^n_i(\theta),\ell^n_j(\theta)\rangle\to I_{ij}(\theta),\quad i,j=\overline{1,d+m}.$$

M-estimators. If model (2.1) is regular, the maximum likelihood (ML) equation (w.r.t. (θ)) takes the form

$$L_t\left(\dot M(\theta),M(\theta)\right)=0,\quad t\ge0.$$

Let $\{m(\theta),\ \theta\in\Theta\subset R^{d+m}\}$ be a family of $(d+m)$-dimensional P-local martingales such that $L(m(\theta),M(\theta))\in\mathcal{M}^2_{loc}(P_\theta)$. Consider the following (estimational) equation w.r.t. θ:

$$L\left(m(\theta),M(\theta)\right)=0. \tag{2.8}$$

Definition 2.3. Any solution $(\theta_t)_{t\ge0}$ of equation (2.8) is called an M-estimator.

To preserve the classical terminology we will say that the family $\{m(\theta),\ \theta \in \Theta\}$ of P-martingales define the M-estimator, and P_θ-martingale $L(m(\theta), M(\theta))$ is an influence (score) martingale.

We say that the family $\{m(\theta),\ \theta \in \Theta\}$ is regularly related to the regular model (2.1) if the following conditions are satisfied:

(1) for all (ω, t) the mapping $\theta \rightsquigarrow m(\omega, t, \theta)$ is continuously differentiable and $\overset{(i)}{m}(\theta) \in \mathcal{M}_{loc}(P)$, where $\overset{(i)}{m}(\theta) = \frac{\partial}{\partial \theta_i} m(\theta)$. Denote $\dot{m}(\theta) = \big(\overset{(1)}{m}(\theta), \dots, \overset{(d+m)}{m}(\theta)\big)'$.

(2) for all (ω, t) the mapping $\theta \rightsquigarrow L_t(m(\theta), M(\theta))$ is continuously differentiable and

$$\frac{\partial}{\partial \theta} L(m(\theta), M(\theta)) = L\left(\dot{m}(\theta), M(\theta)\right) - \left[L(m(\theta), M(\theta)),\ L\left(\dot{M}(\theta), M(\theta)\right)\right],$$

where $\frac{\partial}{\partial \theta} L(m(\theta), M(\theta)) := \big(\overset{(1)}{L}(m(\theta), M(\theta)), \dots, \overset{(d+m)}{L}(m(\theta), M(\theta))\big)'$ is a $(d+m) \times (d+m)$ matrix-valued process,

$$\overset{(i)}{L}(m(\theta), M(\theta)) = \frac{\partial}{\partial \theta_i} L(m(\theta), M(\theta))$$

and for $M, N \in M_{loc}(P)$, $[M, N] = \langle M^c, N^c \rangle + \sum_{s \le \cdot} \Delta M_s \Delta N_s$, M^c is the continuous martingale part of M. Note that $\frac{\partial}{\partial \theta_i} L(m_j(\theta), M(\theta)) = L\big(\overset{(i)}{m}_j(\theta), M(\theta)\big) - \big[L(m_j(\theta), M(\theta)), L\big(\overset{(i)}{M}(\theta), M(\theta)\big)\big]$.

3. Skew Projection Technique

In this section, for exposition simplicity, we consider the case $d = 1$, $\Theta = R^{m+1}$.

Consider the series of regular statistical models

$$\mathcal{E} = (\mathcal{E}^n)_{n \ge 1} = \big(\Omega^n, \mathcal{F}^n, \mathbb{F}^n = (\mathcal{F}^n_t)_{t \in [0,T]}, \{P^n_\theta,\ \theta \in R^{m+1}\}, P^n\big)_{n \ge 1}, \quad m \ge 1, \quad T > 0.$$

Represent $\theta \in R^{m+1}$ in the form $\theta = (\theta_1, \underline{\theta}'_2)'$ where $\theta_1, \theta_1 \in R^1$, is a parameter of interest, $\underline{\theta}_2 = (\theta_2, \dots, \theta_{m+1})'$ is a nuisance parameter.

Let for each $n \ge 1$ $\{m^n(\theta),\ \theta \in R^{m+1}\}$ be a family of $(m+1)$-dimensional P^n-martingales regularly related to the model $\mathcal{E}^n$.

Denote

$$h^n_i(\theta) = L\left(m^n_i(\theta), M^n(\theta)\right), \quad n \ge 1, \quad i = \overline{1, m+1}.$$

In the sequel we assume that for each $\theta \in R^{m+1}$, $h_i^n(\theta) \in \mathcal{M}_{loc}^2(P_\theta^n)$, $i = \overline{1, m+1}$, $n \geq 1$, and the following ergodicity conditions are satisfied:

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 \langle h_i^n(\theta), h_j^n(\theta)\rangle = \Gamma_{ij}(\theta), \tag{3.1}$$

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 \langle h_i^n(\theta), \ell_j^n(\theta)\rangle = \gamma_{ij}(\theta), \tag{3.2}$$

$i, j = \overline{1, m+1}$, where $\Gamma(\theta) = (\Gamma_{ij}(\theta))$ and $\frac{\gamma(\theta)+\gamma'(\theta)}{2}$ with $\gamma(\theta) = (\gamma_{ij}(\theta))$ are finite positive matrices continuous in θ.

We assume also that the Lindeberg condition **L** is satisfied:

L. for each $\varepsilon > 0$

$$P_\theta^n\text{-}\lim_{n\to\infty} \iint_0^T |x|^2\, I_{\{|x|>\varepsilon\}} \nu_\theta^n(ds, dx) = 0,$$

where ν_θ^n is a P_θ^n-compensator of the jump measure of the $2(m+1)$-dimensional P_θ^n-martingale $C_n\big(L(\dot{M}^n(\theta), M^n(\theta)),\ L(m^n(\theta), M^n(\theta))\big)$.

Further, consider the following estimational equation

$$L\big(m^n(\theta), M^n(\theta)\big) = 0, \quad n \geq 1. \tag{3.3}$$

Under suitable regularity conditions it can be proved (see Lemma A from the Appendix) that there exists a sequence $(\theta_n)_{n\geq 1}$ of asymptotic solutions of equation (3.3) (M-estimators) such that

$$\mathcal{L}\big\{C_n^{-1}(\theta_n - \theta) \mid P_\theta^n\big\} \Rightarrow N\big(0, \gamma^{-1}(\theta)\Gamma(\theta)(\gamma^{-1}(\theta))'\big) \quad \text{as} \quad n \to \infty.$$

The skew projection technique proposed in this section allows us to construct an estimator $(\overline{\theta}_{1,n})_{n\geq 1}$ of the parameter of interest θ_1, solving a one-dimensional equation (see equation (3.10) below), which is asymptotically normal with zero mean and the same variance as the first component of the $(m+1)$-dimensional estimator $(\theta_n)_{n\geq 1}$ of θ, i.e.,

$$\mathcal{L}\big\{C_n^{-1}(\overline{\theta}_{1,n} - \theta_1) \mid P_\theta^n\big\} \Rightarrow N\left(0, \left(\gamma^{-1}(\theta)\Gamma(\theta)\left(\gamma^{-1}(\theta)\right)'\right)_{11}\right).$$

Under the skew projection we mean the projection of $h_1^n(\theta)$ onto $\underline{h}_2^n(\theta) = (h_2^n(\theta), \dots, h_{m+1}^n(\theta))'$ in the direction defined by the relations

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 \langle h_1^n(\theta) - b'\underline{h}_2^n(\theta), \ell_j^n(\theta)\rangle = 0, \quad j = \overline{2, m+1}, \tag{3.4}$$

where $b \in R^m$.

(3.4) implies that

$$b' = \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta),$$

where $\underline{\gamma}_{12}(\theta)$ and $\underline{\gamma}_{22}(\theta)$ correspond to the partition of $\gamma(\theta)$:

$$\gamma(\theta) = \begin{pmatrix} \gamma_{11}(\theta) & \underline{\gamma}_{12}(\theta) \\ \underline{\gamma}_{21}(\theta) & \underline{\gamma}_{22}(\theta) \end{pmatrix}.$$

Remark 3.1. Roughly speaking, the direction defined by (3.4) is orthogonal to the linear space spanned by $\underline{\ell}_2^n(\theta) = (\ell_2^n(\theta), \dots, \ell_{m+1}^n(\theta))'$ rather than by $\underline{h}_2^n(\theta)$.

In the case of MLE (i.e., when $h_i^n(\theta) = \ell_i^n(\theta)$, $i = \overline{1, m+1}$) we come to the well-known orthogonal projection of $\ell_1^n(\theta)$ onto $\underline{\ell}_2^n(\theta)$ (see, e.g., [10]) and arrive to the efficient score function for θ_1:

$$\ell_1^n(\theta) - \underline{I}_{12}(\theta)\underline{I}_{22}^{-1}(\theta)\underline{\ell}_2^n(\theta),$$

where $\underline{I}_{12}(\theta)$ and $\underline{I}_{22}(\theta)$ correspond to the partition of

$$I(\theta) = \begin{pmatrix} I_{11}(\theta) & \underline{I}_{12}(\theta) \\ \underline{I}_{21}(\theta) & \underline{I}_{22}(\theta) \end{pmatrix}.$$

Denote

$$\begin{aligned} H_n(\theta) &= h_1^n(\theta) - \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{h}_2^n(\theta), \\ n^n(\theta) &= m_1^n(\theta) - \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{m}_2^n(\theta), \end{aligned} \tag{3.5}$$

where $\underline{m}_2^n(\theta) = (m_2^n(\theta), \dots m_{m+1}^2(\theta))'$. Then

$$H_n(\theta) = L\big(n^n(\theta), M^n(\theta)\big). \tag{3.6}$$

It should be noted that

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 \langle H_n(\theta), \underline{\ell}_2^n(\theta)\rangle = 0. \tag{3.7}$$

Remark 3.2. One can consider the case when $\theta \in \Theta \subset R^{d+m}$, and assume that $\theta = (\underline{\theta}_1, \underline{\theta}_2)$, where $\underline{\theta}_1 = (\theta_1, \dots, \theta_d)'$ is a parameter of interest, $\underline{\theta}_2 = (\theta_{d+1}, \dots, \theta_{d+m})'$ is a nuisance parameter.

In this case

$$H_n(\theta) = \underline{h}_1^n(\theta) - \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{h}_2^n(\theta), \tag{3.8}$$

where $\underline{h}_1^n(\theta) = (h_1^n(\theta), \dots, h_d^n(\theta))'$, $\underline{h}_2^n(\theta) = (h_{d+1}^n(\theta), \dots, h_{d+m}^n(\theta))'$ and $\underline{\gamma}_{12}(\theta)$, $\underline{\gamma}_{22}(\theta)$ and $\underline{\gamma}_{21}(\theta)$ corresponds to the partition of $\gamma(\theta) = (\gamma_{ij}(\theta))_{i,j=\overline{1,d+m}}$,

$$\gamma(\theta) = \begin{pmatrix} \overset{d\times d}{\underline{\gamma}_{11}(\theta)} & \overset{d\times m}{\underline{\gamma}_{12}(\theta)} \\ \underset{m\times d}{\underline{\gamma}_{21}(\theta)} & \underset{m\times m}{\underline{\gamma}_{22}(\theta)} \end{pmatrix}.$$

Now let $(\underline{\theta}_{2,n})_{n\ge 1}$ be a C_n-consistent estimator of $\underline{\theta}_2$, i.e.,

$$C_n^{-1}(\underline{\theta}_{2,n} - \underline{\theta}_2) = O_{P_\theta^n}(1). \tag{3.9}$$

Consider the following estimational equation (w.r.t. θ_1)

$$H_n(\theta_1, \underline{\theta}_{2,n}) = 0. \tag{3.10}$$

Our goal is to study the problem of solvability of equation (3.10) as well as the asymptotic properties of solutions.

Introduce the conditions:

(1) For each $n \ge 1$ the family $\{m^n(\theta),\ \theta \in R^{m+1}\}$ is regularly related to $\mathcal{E}^n$.
(2) The mapping $\theta \rightsquigarrow \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)$ is continuously differentiable.

(3) For any $\theta \in R^{m+1}$ and $x \in R^1$ there exists a function $\Delta(x,\theta)$ such that

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 H_n(x,\underline{\theta}_{\,2,n}) = \Delta(x,\theta)$$

and the equation (w.r.t. x)

$$\Delta(x,\theta)=0$$

has a unique solution $\theta_1^* = b(\theta) \in R^1$.

(4) For each $\theta \in R^{m+1}$

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 L_n\big(\overset{(i)}{n}{}^n(\theta), M^n(\theta)\big)=0, \quad i=\overline{1,m+1},$$

where $\overset{(i)}{n}{}^n(\theta) = \frac{\partial}{\partial\theta_i}\, n^n(\theta)$.

(5) For any $\theta \in R^{m+1}$, $N>0$ and $\rho>0$

$$\overline{\lim_{n\to\infty}}\, P_\theta^n\bigg\{\sup_{y:|y-\underline{\theta}_{\,2}|\le C_nN} C_n^2\big|\overset{(i)}{H}_n(\theta_1,y)-\overset{(i)}{H}_n(\theta_1,\underline{\theta}_{\,2})\big|>\rho\bigg\}=0, \quad i=\overline{1,m+1}.$$

(6) For any $\theta \in R^{m+1}$, $N>0$ and $\rho>0$

$$\overline{\lim_{n\to\infty}}\, P_\theta^n\bigg\{\sup_{\substack{x\in R^1,\, y\in R^m\\ |x-\theta_1|\le r\\ |y-\underline{\theta}_{\,2}|\le C_nN}} C_n^2\big|\overset{(1)}{H}_n(x,y)-\overset{(1)}{H}_n(\theta_1,y)\big|>\rho\bigg\}=0.$$

Remark 3.3. If $L\big(\overset{(i)}{n}{}^n(\theta), M^n(\theta)\big) \in \mathcal{M}_{loc}^2(P_\theta^n)$, then condition (4) will be satisfied if for any $\theta \in R^{m+1}$, $N>0$ and $\rho>0$

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2\big\langle L\big(\overset{(i)}{n}{}^n(\theta), M^n(\theta)\big)\big\rangle = 0$$

(see, e.g., [6]).

Theorem 3.1. *Let conditions* (1)–(6) *be satisfied. Then for each* $\theta \in R^{m+1}$, $\theta=(\theta_1,\underline{\theta}_{\,2}')'$, *there exists a sequence* $(\overline{\theta}_{1,n})_{n\ge1}$ *of random variables such that*

I. $\lim\limits_{n\to\infty} P_\theta^n\big\{H_n(\overline{\theta}_{1,n},\underline{\theta}_{\,2,n})=0\big\}=1$,

II. $P_\theta^n\text{-}\lim\limits_{n\to\infty}\overline{\theta}_{1,n}=\theta_1$,

III. *if there exists another sequence* $\widetilde{\theta}_{1,n}$ *with properties* I, II, *then*

$$\lim_{n\to\infty} P_\theta^n\big\{\widetilde{\theta}_{1,n}=\overline{\theta}_{1,n}\big\}=1,$$

IV. *if the sequence of distributions* $\mathcal{L}\{C_nH_n(\theta)\mid P_\theta^n\}$ *converges weakly to a certain distribution* Φ, *then*

$$\mathcal{L}\big\{\gamma_p(\theta)C_n^{-1}(\overline{\theta}_{1,n}-\theta_1)\mid P_\theta^n\big\}\Rightarrow\Phi,$$

where

$$\gamma_p(\theta)=\gamma_{11}(\theta)-\underline{\gamma}_{\,12}(\theta)\underline{\gamma}_{\,22}^{-1}(\theta)\underline{\gamma}_{\,21}(\theta)>0. \tag{3.11}$$

Proof. Keeping Remark A of Appendix in mind refer to Lemma A and put $\ell = m+1$, $k=1$, $Q^n_\theta = P^n_\theta$, $L_n(\theta) = L_n(\theta_1) = H_n(\theta_1, \underline{\theta}_{2,n})$, $\theta = (\theta_1, \underline{\theta}'_2)'$.

Now in order to prove the theorem we have to verify that all conditions of Lemma A are satisfied (for $L_n(\theta_1)$).

Since $L_n(\theta_1) = H_n(\theta_1, \underline{\theta}_{2,n})$ the continuous differentiability of $L_n(\theta_1)$ in θ_1 is a trivial consequence of (3.8), conditions (1) and (2). Hence condition b) of Lemma A is satisfied.

Further, let us show that for any $\theta \in R^{m+1}$, $\theta = (\theta_1, \underline{\theta}'_2)'$,

$$P^n_\theta\text{-}\lim_{n\to\infty} C_n\big(L_n(\theta_1) - H_n(\theta)\big) = 0. \tag{3.12}$$

Using the Taylor expansion, we obtain

$$\begin{aligned} C^2_n\big(L_n(\theta_1) - H_n(\theta)\big) &= C_n\big(H_n(\theta_1, \underline{\theta}_{2,n}) - H_n(\theta_1, \underline{\theta}_2)\big) \\ = C^2_n \overset{*}{H}{}'_n(\theta) C^{-1}_n(\underline{\theta}_{2,n} - \underline{\theta}_2) &+ C^2_n\big(\overset{*}{H}{}'_n(\theta_1, v_n) - \overset{*}{H}{}'_n(\theta_1, \underline{\theta}_2)\big) C^{-1}_n(\underline{\theta}_{2,n} - \underline{\theta}_2), \end{aligned}$$

where $\overset{*}{H}_n(\theta) = \big(\overset{(2)}{H}_n(\theta), \dots, \overset{(m+1)}{H_n}(\theta)\big)$, $v_n = \underline{\theta}_2 + \alpha(\underline{\theta}_{2,n} - \underline{\theta}_2)$, random variable $\alpha \in [0,1]$.

Consequently, since $C^{-1}_n(\underline{\theta}_{2,n} - \underline{\theta}_2) = O_{P^n_\theta}(1)$ for (3.12) it is sufficient to show that

$$P^n_\theta\text{-}\lim_{n\to\infty} C^2_n \overset{*}{H}_n(\theta) = 0 \tag{3.13}$$

and

$$P^n_\theta\text{-}\lim_{n\to\infty} C^2_n\big(\overset{*}{H}_n(\theta_1, v_n) - \overset{*}{H}(\theta_1, \underline{\theta}_2)\big) = 0. \tag{3.14}$$

From (3.6) and conditions (1), (2) we have

$$C^2_n \overset{*}{H}_n(\theta) = C^2_n L\big(\overset{*}{n}{}^n(\theta), M^n(\theta)\big) - C^2_n\big[H_n(\theta), \underline{\ell}^{\,n}_2(\theta)\big].$$

Condition **L** and (3.7) ensure that (see, e.g., [6])

$$P^n_\theta\text{-}\lim_{n\to\infty} C^2_n\big[H_n(\theta), \underline{\ell}^{\,n}_2(\theta)\big] = P^n_\theta\text{-}\lim_{n\to\infty} C^2_n\langle H_n(\theta), \underline{\ell}^{\,n}_2(\theta)\rangle = 0.$$

Now (3.13) directly follows from condition (4).

As for (3.14) observe that for any $N > 0$ and $\rho > 0$

$$\begin{aligned} &\overline{\lim_{n\to\infty}}\, P^n_\theta\big\{C^2_n\big|\overset{*}{H}_n(\theta_1, v_n) - \overset{*}{H}_n(\theta_1, \underline{\theta}_2)\big| > \rho\big\} \\ &= \overline{\lim_{n\to\infty}}\, P^n_\theta\big\{C^2_n\big|\overset{*}{H}_n(\theta_1, v_n) - \overset{*}{H}_n(\theta_1, \underline{\theta}_2)\big| > \rho,\ C^{-1}_n|\underline{\theta}_{2,n} - \underline{\theta}_2| \le N\big\} \\ &\quad + \overline{\lim_{n\to\infty}}\, P^n_\theta\big\{C^2_n\big|\overset{*}{H}_n(\theta_1, v_n) - \overset{*}{H}_n(\theta_1, \underline{\theta}_2)\big| > \rho,\ C^{-1}_n|\underline{\theta}_{2,n} - \underline{\theta}_2| > N\big\} \\ &\le \overline{\lim_{n\to\infty}}\, P^n_\theta\Big\{\sup_{y:|y-\underline{\theta}_2|\le C_n N} C^2_n\big|\overset{*}{H}_n(\theta_1, v_n) - \overset{*}{H}_n(\theta_1, \underline{\theta}_2)\big| > \rho\Big\} \\ &\quad + \overline{\lim_{n\to\infty}}\, P^n_\theta\big\{C^{-1}_n|\underline{\theta}_{2,n} - \underline{\theta}_2| > N\big\}. \end{aligned}$$

From this inequality using condition (5) we obtain

$$\begin{aligned}\overline{\lim_{n\to\infty}}\, P_\theta^n\Big\{ \sup_{y:|y-\underline{\theta}_2|\le C_nN} C_n^2\big|\overset{*}{H}_n(\theta_1,v_n)-\overset{*}{H}(\theta_1,\underline{\theta}_2)\big|>\rho\Big\}\\ \le \overline{\lim_{n\to\infty}}\, P_\theta^n\big\{C_n^{-1}|\underline{\theta}_{2,n}-\underline{\theta}_2|>N\big\}.\end{aligned}\tag{3.15}$$

Letting $N\to\infty$ in (3.15) and taking into account (3.9) we get (3.14). Thus (3.12) is proved.

From (3.12) it follows that

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2L_n(\theta_1)=P_\theta^n\text{-}\lim_{n\to\infty} C_n^2H_n(\theta).\tag{3.16}$$

On the other hand, by using the ergodicity conditions (3.1) and (3.2) one can easily check that

$$P_\theta^n\text{-}\lim_{n\to\infty}\langle C_nH_n(\theta)\rangle=P_\theta^n\text{-}\lim_{n\to\infty} C_n^2\langle H_n(\theta)\rangle=\Gamma_p(\theta),\tag{3.17}$$

where

$$\begin{aligned}\Gamma_p(\theta)=\Gamma_{11}(\theta)-2\underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{\Gamma}'_{12}(\theta)\\ +\underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}\underline{\Gamma}_{22}(\theta)\big(\underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\big)',\end{aligned}\tag{3.18}$$

$\underline{\Gamma}_{12}(\theta)$ and $\underline{\Gamma}_{22}(\theta)$ correspond to the partition of $\Gamma(\theta)$:

$$\Gamma(\theta)=\begin{pmatrix}\Gamma_{11}(\theta) & \underline{\Gamma}_{12}(\theta)\\ \underline{\Gamma}_{21}(\theta) & \underline{\Gamma}_{22}(\theta)\end{pmatrix}.$$

Recall that $\underline{\Gamma}'_{12}(\theta)=\underline{\Gamma}_{21}(\theta)$.

From (3.17) it follows that

$$P_\theta^n\text{-}\lim_{n\to\infty}\langle C_n^2H_n(\theta)\rangle=P_\theta^n\text{-}\lim_{n\to\infty} C_n^4\langle H_n(\theta)\rangle=0,$$

from which we obtain that (see, e.g., [9])

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2H_n(\theta)=0.\tag{3.19}$$

Now combining (3.12) and (3.19) we get

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2L_n(\theta_1)=\Delta(\theta_1,\theta)=0$$

and in view of condition (3) we aventually conclude that condition c) of Lemma A is satisfied with $\theta_1^*(\theta)=\theta_1$.

The next step is to verify condition d) of Lemma A.

By arguments similar to those we have used in the proof of (3.12) one can show that

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2\big(\dot L_n(\theta_1)-\overset{(1)}{H}_n(\theta)\big)=0.\tag{3.20}$$

Indeed, for any $\rho > 0$ we have

$$\begin{aligned} &P_\theta^n\bigl\{C_n^2\bigl|\dot L_n(\theta_1) - \overset{(1)}{H}_n(\theta)\bigr| > \rho\bigr\} \\ &\quad= P_\theta^n\bigl\{C_n^2\bigl|\overset{(1)}{H}_n(\theta_1, \underline{\theta}_{\,2}, n) - \overset{(1)}{H}_n(\theta_1, \underline{\theta}_{\,2})\bigr| > \rho\bigr\} \\ &\quad\le P_\theta^n\Bigl\{\sup_{|y-\underline{\theta}_{\,2}|\le C_n N} C_n^2\bigl|\overset{(1)}{H}_n(\theta_1, y) - \overset{(1)}{H}_n(\theta_1, \underline{\theta}_{\,2})\bigr| > \rho\Bigr\} \\ &\qquad+ P_\theta^n\bigl\{C_n^{-1}|\underline{\theta}_{\,2,n} - \underline{\theta}_{\,2}| > N\bigr\}. \end{aligned}$$

Therefore (3.14) follows from conditions (5) and (3.9).

At the same time, by

$$\overset{(1)}{H}_n(\theta) = L\bigl(\overset{(1)}{n}{}^n(\theta), M^n(\theta)\bigr) - \bigl[H_n(\theta), \ell_1^n(\theta)\bigr]$$

condition (4) together with condition **L** imply

$$\begin{aligned} P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 \overset{(1)}{H}_n(\theta) &= -P_\theta^n\text{-}\lim_{n\to\infty} C_n^2\bigl[H_n(\theta), \ell_1^n)\bigr] \\ &= -P_\theta^n\text{-}\lim_{n\to\infty} C_n^2\langle h_1^n(\theta) - \underline{\gamma}_{\,12}\underline{\gamma}_{\,22}^{-1}\underline{h}_{\,2}^{\,n}, \ell_1^n\rangle \\ &= \gamma_p(\theta), \end{aligned}$$

where $\gamma_p(\theta) > 0$ is defined by (3.11). In view of this equality we conclude that

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 \dot L_n(\theta_1) = -\gamma_p(\theta) < 0. \tag{3.21}$$

Thus condition d) of Lemma A is satisfied.

It remains to check condition e) of Lemma A.

For any $\rho > 0$, $r > 0$ and $N > 0$ we have

$$\begin{aligned} &\varlimsup_{n\to\infty} P_\theta^n\Bigl\{\sup_{|x_1-\theta_1|\le r} C_n^2\bigl|\dot L_n(x_1) - \dot L_n(\theta_1)\bigr| > \rho\Bigr\} \\ &= \varlimsup_{n\to\infty} P_\theta^n\Bigl\{\sup_{|x_1-\theta_1|\le r} C_n^2\bigl|\overset{*}{H}_n(x_1, \underline{\theta}_{\,2,n}) - \overset{*}{H}_n(\theta_1, \underline{\theta}_{\,2,n})\bigr| > \rho\Bigr\} \\ &\le \varlimsup_{n\to\infty} P_\theta^n\Bigl\{\sup_{\substack{|x_1-\theta_1|\le r \\ |y-\underline{\theta}_{\,2}|\le C_n N}} C_n^2\bigl|\overset{*}{H}_n(x_1, y) - \overset{*}{H}_n(\theta_1, y)\bigr| > \rho\Bigr\} \\ &\quad+ \varlimsup_{n\to\infty} P_\theta^n\bigl\{C_n^{-1}|\underline{\theta}_{\,2,n} - \underline{\theta}_{\,2}| > N\bigr\}, \end{aligned}$$

which in view of (3.9) and condition (6) implies that

$$\lim_{r\to 0}\varlimsup_{n\to\infty} P_\theta^n\Bigl\{\sup_{|x_1-\theta_1|\le r}\bigl|\dot L_n(x_1) - \dot L_n(\theta_1)\bigr| > \rho\Bigr\} = 0.$$

Thus condition e) of Lemma A is also satisfied. □

Corollary 3.1. *Under the conditions of Theorem* 3.1

$$\mathcal{L}\bigl\{C_n^{-1}(\overline{\theta}_{1,n} - \theta_1) \mid P_\theta^n\bigr\} \Rightarrow N\bigl(0, \Gamma_p(\theta)/\gamma_p^2(\theta)\bigr), \tag{3.22}$$

where $\gamma_p(\theta)$ *and* $\Gamma_p(\theta)$ *are defined by* (3.11) *and* (3.18), *respectively.*

Proof. Using CLT for locally square integrable martingales [6], from (3.17), condition $\mathbf{L}$ we have

$$\mathcal{L}\big\{C_n H_n(\theta) \mid P_\theta^n\big\} \Rightarrow N\big(0, \Gamma_p(\theta)\big), \tag{3.23}$$

from which, taking into account (3.12), we get

$$\mathcal{L}\big\{C_n L_n(\theta_1) \mid P_\theta^n\big\} \Rightarrow N\big(0, \Gamma_p(\theta)\big). \tag{3.24}$$

Finally, combining (3.24) with assertion IV of Theorem 3.1 we obtain the desirable convergence (3.22). □

Remark 3.4. As it has been mentioned above, if we consider system (3.3) of estimational equations

$$h_i^n(\theta) = 0, \quad i = \overline{1, m+1}, \quad n \geq 1,$$

and suppose that all conditions of Lemma A with $\ell = m+1$, $Q_\theta^n = P_\theta^n$, $L_n(\theta) = (h_1^n(\theta), \dots, h_{m+1}^n(\theta))'$ are satisfied, then all assertions of Lemma A hold true, i.e., for any $\theta \in R^{m+1}$ there exists a sequence $(\widehat{\theta}_n)_{n\geq 1}$ with properties I and II (of Lemma A) and such that

$$\mathcal{L}\big\{\gamma(\theta) C_n^{-1}(\widehat{\theta}_n - \theta) \mid P_\theta^n\big\} \Rightarrow \Phi \quad \text{as} \quad n \to \infty, \tag{3.25}$$

where Φ is defined as

$$\mathcal{L}\big\{C_n (h_1^n(\theta), \dots, h_{m+1}^n(\theta))' \mid P_\theta^n\big\} \Rightarrow \Phi \quad \text{as} \quad n \to \infty. \tag{3.26}$$

Now the ergodicity conditions, condition $\mathbf{L}$ and CLT for locally square integrable martingales imply that

$$\Phi = N(0, \Gamma_p(\theta)), \tag{3.27}$$

and, hence with (3.25) and (3.27) in mind it follows that

$$\mathcal{L}\big\{C_n^{-1}(\widehat{\theta}_n - \theta) \mid P_\theta^n\big\} \Rightarrow N\big(0, \gamma^{-1}(\theta)\Gamma_p(\theta)(\gamma^{-1}(\theta))'\big)$$

as $n \to \infty$. In particular,

$$\mathcal{L}\big\{C_n^{-1}(\widehat{\theta}_{1,n} - \theta_1) \mid P_\theta^n\big\} \Rightarrow N\left(0, \left(\gamma^{-1}(\theta)\Gamma_p(\theta)(\gamma^{-1}(\theta))'\right)_{11}\right). \tag{3.28}$$

On the other hand, from (3.25), (3.26) and (3.27) it directly follows that

$$\mathcal{L}\big\{\xi_1^n(\theta) - \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{\xi}_2^n(\theta) \mid P_\theta^n\big\} \Rightarrow \widetilde{\Phi}, \tag{3.29}$$

where

$$\xi_1^n(\theta) = \gamma_{11}(\theta) C_n^{-1}(\widehat{\theta}_{1,n} - \theta_1) + \underline{\gamma}_{12}(\theta) C_n^{-1}(\underline{\widehat{\theta}}_{2,n} - \underline{\theta}_2),$$
$$\underline{\xi}_2^n(\theta) = \underline{\gamma}_{21}(\theta) C_n^{-1}(\widehat{\theta}_{1,n} - \theta_1) + \underline{\gamma}_{22}(\theta) C_n^{-1}(\underline{\widehat{\theta}}_{2,n} - \underline{\theta}_2),$$

and $\widetilde{\Phi}$ is a weak limit of the sequence of distributions

$$\mathcal{L}\big\{C_n(h_1^n(\theta) - \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{h}_2^n(\theta) \mid P_\theta^n\big\} = \mathcal{L}\big\{C_n H_n(\theta) \mid P_\theta^n\big\} \quad \text{as} \quad n \to \infty,$$

i.e. (see (3.23)),

$$\widetilde{\Phi} = N(0, \Gamma_p(\theta)). \tag{3.30}$$

But

$$\xi_1^n(\theta) - \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{\xi}_2^n(\theta) = \big(\gamma_{11}(\theta) - \underline{\gamma}_{12}(\theta)\underline{\gamma}_{22}^{-1}(\theta)\underline{\gamma}_{21}(\theta)\big)C_n^{-1}(\widehat{\theta}_{1,n} - \theta_1)$$
$$= \gamma_p(\theta)C_n^{-1}(\widehat{\theta}_{1,n} - \theta_1).$$

Adding to this equality (3.29) and (3.30) we obtain

$$\mathcal{L}\big\{C_n^{-1}(\widehat{\theta}_{1,n} - \theta_1) \mid P_\theta^n\big\} \Rightarrow N\left(0, \frac{\Gamma_p(\theta)}{(\gamma_p(\theta))^2}\right). \tag{3.31}$$

Now after comparing (3.28) and (3.31), one can conclude that

$$\left(\gamma^{-1}(\theta)\Gamma(\theta)\left(\gamma^{-1}(\theta)\right)'\right)_{11} = \frac{\Gamma_p(\theta)}{(\gamma_p(\theta))^2} \tag{3.32}$$

and thus the estimator $(\overline{\theta}_{1,n})_{n\geq 1}$ constructed by the skew projection technique has the same asymptotic variance as the first component $(\widehat{\theta}_{1,n})_{n\geq 1}$ of the estimator $(\widehat{\theta}_n)_{n\geq 1}$ constructed by solving the full system (3.3).

If equation (3.10) has a unique solution $\overline{\theta}_{1,n}$, then Theorem 3.1 establishes the asymptotic distribution of the normed sequence $(C_n^{-1}(\overline{\theta}_{1,n} - \theta_1))_{n\geq 1}$ (see (3.22)).

To avoid the problem of the solvability of equation (3.10) one can consider the following one-step approximation procedure: let $(\theta_n)_{n\geq 1}$ be a C_n-consistent estimator of a full parameter $\theta \in R^{m+1}$. i.e., $C_n^{-1}(\theta_n - \theta) = O_{P_\theta^n}(1)$.

Define an estimator $\overline{\theta}_{1,n}$ of θ_1 by the one-step procedure

$$\overline{\theta}_{1,n} = \theta_{1,n} + \gamma_p^{-1}(\theta_n)C_n^2 H_n(\theta_n). \tag{3.33}$$

Theorem 3.2. *Let conditions* (1), (2), (4) *and* (5) *of Theorem* 3.1 *be satisfied as well as the condition*

(6′) *for any* $\theta \in R^{m+1}$, $N > 0$ *and* $\rho > 0$

$$\varlimsup_{n\to\infty} P_\theta^n\Bigg\{\sup_{\substack{x:|x-\theta_1|\leq C_n N\\ y:|y-\underline{\theta}_2|\leq C_n N}} C_n^2\big|\overset{(1)}{H}_n(x,y) - \overset{(1)}{H}_n(\theta_1,y)\big| > \rho\Bigg\} = 0.$$

Then,

$$\mathcal{L}\big\{C_n^{-1}(\overline{\theta}_{1,n} - \theta_1) \mid P_\theta^n\big\} \Rightarrow N\left(0, \frac{\Gamma_p(\theta)}{\gamma_p^2(\theta)}\right)$$

as $n \to \infty$.

Proof. (3.33) yields

$$\begin{aligned} C_n^{-1}(\overline{\theta}_{1,n} - \theta_1) &= C_n^{-1}(\theta_{1,n} - \theta_1) + \gamma_p^{-1}(\theta)C_n H_n(\theta_n) \\ &\quad + \left(\gamma_p^{-1}(\theta_n) - \gamma_p^{-1}(\theta)\right)C_n H_n(\theta_n) \\ &= C_n^{-1}(\theta_{1,n} - \theta_1) + \gamma_p^{-1}(\theta)C_n L_n(\theta_{1,n}) \\ &\quad + \left(\gamma_p^{-1}(\theta_n) - \gamma_p^{-1}(\theta)\right)C_n L_n(\theta_{1,n}), \end{aligned} \tag{3.34}$$

where $L_n(\theta_1) = H_n(\theta_1, \underline{\theta}_{2,n})$.

Using the Taylor expansion, we have

$$L_n(\theta_{1,n}) = L_n(\theta) + \dot{L}_n(x_n)(\theta_{1,n} - \theta_1), \tag{3.35}$$

where $\dot{L}(\theta_1) = \frac{\partial}{\partial\theta} L_n(\theta_1) = \overset{(1)}{H}_n(\theta_1, \underline{\theta}_{\,2,n})$, $x_n = \theta_1 - \alpha(\theta_{1,n} - \theta_1)$, $\alpha \in [0,1]$.

Substituting (3.35) in (3.34), we obtain

$$C_n^{-1}(\theta_{1,n} - \theta_1) = \gamma_p^{-1}(\theta) C_n L_n(\theta_1) + \varepsilon_n(x_n, \theta) + \delta(\theta_n, \theta), \tag{3.36}$$

where

$$\begin{aligned}
\varepsilon_n(x_n, \theta) &= \gamma_p^{-1}(\theta)\big(\gamma_p(\theta) + C_n^2 \dot{L}(x_n)\big) C_n^{-1}(\theta_{1,n} - \theta_1), \\
\delta_n(\theta_n, \theta) &= \big(\gamma_p^{-1}(\theta_n) - \gamma_p^{-}(\theta)\big)\, C_n L_n(\theta_{1,n}).
\end{aligned}$$

Now if we prove that

$$P_\theta^n\text{-}\lim_{n\to\infty} \varepsilon_n(x_n, \theta) = 0, \tag{3.37}$$

$$P_\theta^n\text{-}\lim_{n\to\infty} \delta_n(\theta_n, \theta) = 0, \tag{3.38}$$

then the desirable convergence (the assertion of Theorem 3.2) follows from (3.36) and (3.24).

First we prove (3.37).

Using a standard technique we have for any $\rho > 0$ and $N > 0$

$$\begin{aligned}
&\overline{\lim_{n\to\infty}}\, P_\theta^n\big\{C_n^2 \big|\dot{L}_n(x_n) - \dot{L}_n(\theta_1)\big| > \rho\big\} \\
&= \overline{\lim_{n\to\infty}}\, P_\theta^n\big\{C_n^2 \big|\overset{(1)}{H}_n(x_n, \underline{\theta}_{\,2,n}) - \overset{(1)}{H}_n(\theta_1, \underline{\theta}_{\,2,n})\big| > \rho\big\} \\
&\le \overline{\lim_{n\to\infty}}\, P_\theta^n\Big\{\sup_{\substack{x: |x-\theta_1| \le C_n N \\ y: |y - \underline{\theta}_{\,2}| \le C_n N}} C_n^2 \big|\overset{(1)}{H}_n(x_n, y) - \overset{(1)}{H}_n(\theta_1, y)\big| > \rho\Big\} \\
&\quad + \overline{\lim_{n\to\infty}}\, P_\theta^n\big\{C_n^{-1} \mid |\underline{\theta}_{\,2,n} - \underline{\theta}_{\,2}| > N\big\} \\
&= \overline{\lim_{n\to\infty}}\, P_\theta^n\big\{C_n^{-1} \mid |\underline{\theta}_{\,2,n} - \underline{\theta}_{\,2}| > N\big\},
\end{aligned} \tag{3.39}$$

where the last equality follows from condition (6′). Letting $N \to \infty$ in (3.39) we obtain

$$P_\theta^n\text{-}\lim_{n\to\infty} C_n^2 \big|\dot{L}_n(x_n) - \dot{L}_n(\theta_1)\big| = 0. \tag{3.40}$$

In view of (3.21), (3.40) implies

$$P_\theta^n\text{-}\lim_{n\to\infty} \big(\gamma_p(\theta) + C_n^2 \dot{L}_n(x_n)\big) = 0 \tag{3.41}$$

from which taking into account that $C_n^{-1}(\theta_{1,n} - \theta_1) = O_{P_\theta^n}(1)$ we get (3.37).

As for (3.38) in view of (3.41) and (3.24) it evidently follows from (3.35) that the sequence $(C_n L_n(\theta_{1,n}))_{n\ge 1}$ is bounded in probability. It remains to notice that $\gamma_p^{-1}(\theta)$ is continuous in θ. Theorem is proved. □

Appendix

Let for every $\theta \in \Theta \subset R^\ell$, $\ell \geq 1$, a sequence of probability measures $(Q^n_\theta)_{n\geq 1}$, $(Q^n_\theta \sim P^n)$ and a ℓ-dimensional random vectors $L_n(\theta)$, $n \geq 1$, as well as a sequence of positive numbers $(C_n)_{n\geq 1}$ be given on a measurable space $(\Omega^n, \mathcal{F}^n_T)$.

We are interested in the of solvability of the system of equations

$$L_n(\theta) = 0 \tag{A.1}$$

and asymptotic behavior of solutions as $n \to \infty$.

The following lemma is proved in [2].

Lemma A. *Let the following conditions be satisfied:*

a) $\lim\limits_{n\to\infty} C_n = 0$;

b) *for each $n \geq 1$ the mapping $\theta \rightsquigarrow L_n(\theta)$ is continuously differentiable P^n-a.s.;*

c) *for each pair (θ, y), $\theta \in \Theta$, $y \in \Theta$ there exists a function $\Delta_Q(\theta, y)$ such that*

$$Q^n_\theta\text{-}\lim_{n\to\infty} C^2_n L_n(y) = \Delta_Q(\theta, y)$$

and the equation (w.r.t. y)

$$\Delta_Q(\theta, y) = 0$$

has a unique solution $\theta^* = b(\theta) \in \Theta$;

d) *for each $\theta \in \Theta$*

$$Q^n_\theta\text{-}\lim_{n\to\infty} C^2_n \dot{L}_n(\theta^*) = -\gamma_Q(\theta),$$

where $\gamma_Q(\theta)$ is a positive definite matrix, $\dot{L}_n(\theta) = \Big(\frac{\partial}{\partial\theta_i} L_{n,j}(\theta)\Big)_{i,j=\overline{1,d}}$;

e) *for each $\theta \in \Theta$ and $\rho > 0$*

$$\lim_{r\to 0} \varlimsup_{n\to\infty} Q^n_\theta \Big\{ \sup_{y: |y-\theta^*|\leq r} |\dot{L}_n(y) - \dot{L}_n(\theta^*)| > \rho \Big\} = 0.$$

Then for any $\theta \in \Theta$ there exists a sequence of random vectors $\widehat{\theta} = (\widehat{\theta}_n)_{n\geq 1}$ taking values in Θ such that

I. $\lim\limits_{n\to\infty} Q^n_\theta\{L_n(\widehat{\theta}_n) = 0\} = 1$,

II. $Q^n_\theta\text{-}\lim\limits_{n\to\infty} \widehat{\theta}_n = \theta^*$,

III. *if $\widetilde{\theta} = (\widetilde{\theta}_n)_{n\geq 1}$ is another sequence with properties* I *and* II, *then*

$$\lim_{n\to\infty} Q^n_\theta\{\widetilde{\theta}_n = \widehat{\theta}_n\} = 1,$$

IV. *if the sequence of distributions $\mathcal{L}\{C_n L_n(\theta^*) \mid Q^n_\theta\}$ converges weakly to a certain distribution Φ, then*

$$\mathcal{L}\big\{\gamma_Q(\theta) C^{-1}_n (\widehat{\theta}_n - \theta^*) \mid Q^n_\theta\big\} \Rightarrow \Phi \quad \text{as} \quad n \to \infty.$$

Remark. By the same arguments as used in the proof of Lemma A one can easily verify that all assertions of Lemma A hold true of instead of ℓ-dimensional function $L_n(\theta)$ we consider a k-dimensional, $k<\ell$, function $L_n(\theta)$ such that $L_n(\theta)=L_n(\theta_1,\dots,\theta_k)$, $\theta=(\theta_1,\dots,\theta_k,\dots,\theta_\ell)'$, with $L_n(\theta_1,\theta_2,\dots,\theta_k)$ being continuously differentiable in each point θ_i, $i\le k$, and assume that all conditions of Lemma A are satisfied with k-dimensional function $\Delta_Q(\theta,y)=\Delta_Q(\theta,y_1,y_2,\dots,y_k)$, $y=(y_1,\dots,y_k,\dots,y_\ell)\in\Theta$, and

$$\dot{L}(\theta)=\left(\frac{\partial}{\partial\theta_i}\,L_{n,j}(\theta_1,\dots,\theta_k)\right)_{i,j\le k}.$$

It should be noticed that a function $L_n(\theta)$ is defined on projection of set $\Theta\subset R^k\times R^{\ell-k}$ onto R^k.

Acknowledgement

The research was supported by INTAS Grant No. 99–00559 and by Grant of Georgian Academy of Sciences No. 10.1.

References

1. P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, Efficient and adaptive estimation for semiparametric models. *Springer, New York,* 1998.
2. R. J. Chitachvili, N. L. Lazrieva, and T. A. Toronjadze, Asymptotic theory of M-estimators in general statistical models, Parts I, II. *Centrum voor Wiskunde en Informatica, Reporis* BS–R9019, BS–R9020, *June, Amsterdam,* 1990.
3. J. H. Friedman and W. Stuetzle, Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**(1981), No. 376, 817–823.
4. P. Huber, Robust statistics. *Willey, New York,* 1981.
5. J. Jacod, Calcul stochastique et problémes de martingales. *Lecture Notes in Mathematics,* 714. *Springer, Berlin,* 1979.
6. J. Jacod and A. N. Shiryaev, Limit theorems for stochastic processes. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*]*,* 288. *Springer-Verlag, Berlin,* 1987.
7. J. Jacod, Regularity, partial regularity, partial information process for a filtered statistical model. *Probab. Theory Related Fields* **86**(1990), No. 3, 305–335.
8. N. Lazrieva and T. Toronjadze, Robust estimators in discrete time statistical models. Contiguous alternatives. *Proc. A. Razmadze Math. Inst.* **115**(1997), 59–97.
9. R. S. Liptser and A. N. Shiryaev, Theory of martingales. (Russian) *Nauka, Moscow,* 1986.
10. J. A. Wellner, Semiparametric models: progress and problems. *Centrum voor Wiskunde en Informatica, Report* MS–R8614, *November, Amsterdam,* 1986.

(Received 10.01.2003)

Authors' address:

A. Razmadze Mathematical Institute
Georgian Academy of Sciences
1, M. Aleksidze St., Tbilisi 0193, Georgia