Pruscha, Wellisch:

Asymptotic behaviour of estimation equations with functional nuisance or working parameter

Sonderforschungsbereich 386, Paper 79 (1997)

Asymptotic behaviour of estimation equations with functional nuisance or working parameter

H.Pruscha
Mathematical Institute
University of Munich*

U.Wellisch
Mathematical Institute
University of Munich

SUMMARY

We are concerned with the asymptotic theory of semiparametric estimation equations. We are dealing with estimation equations which have a parametric component of interest and a functional (nonparametric) nuisance component. We give sufficient conditions for the existence and the asymptotic normality of a consistent estimation equation estimator for the parameter of interest. These conditions concern the asymptotic distribution of the estimation function and of its derivative as well as the effect of the functional nuisance part in the estimation equation. In order to treat the nonparametric component we introduce a general differential calculus and a general mean value theorem. For the nonparametric part in the estimation equation we distinguish two cases: the situation of a (classical) nuisance parameter and the case of a so called working parameter. As a special case we get regularity conditions for estimation equations with finite dimensional nuisance or working parameter. As an example we present the semiparametric linear regression model.

Some key words: Asymptotic normality; Consistent estimation equation estimator; Hadamard differentiation; Nuisance parameter; Semiparametric estimation equation; Semiparametric linear regression; Working parameter.

1 INTRODUCTION

The starting point of our investigations is an estimation equation of the form $U_n(\theta, \alpha) = 0$. It contains a finite dimensional parameter θ being of primary interest and a functional parameter α. The latter may play the role of a nuisance parameter (in the classical sense) or that of a working parameter (coming into statistical use with Liang and Zeger, 1986). A nonparametric estimator $\hat{\alpha}_n$

*Theresienstr.39, D-80333 Munich, Germany

is assumed to be given showing a certain kind of limit behaviour, the special type of the estimator being of no regard. For estimators $\hat{\theta}_n$ of θ which solve (asymptotically) the estimation equation we will prove consistency and asymptotic normality.

A special feature of the present paper is a consequent functionally orientated approach. The Taylor method—well established for finite dimensional spaces—is carried out in functional spaces and is employed in proving the asymptotic results. This program seems to be more direct and flexible than that of Severini and Wong (1992) and others, but the price are more involved regularity conditions. It bears some connections with van der Vaart and Wellner (1996, sec. 3.3). To perform this program an appropriate differential calculus is presented in sec. 2. Hadamard derivatives—a notion between Fréchet and Gâteaux derivatives—turn out to be most suitable to prove a mean value theorem, which will be our main tool of analysis. Some probabilistic notations in normed spaces can be found in the appendix.

A further characteristic of the present approach is the strict separation into the field of inference on one side and of statistical modelling on the other: The semiparametric inference in sec. 3 is independent of model assumptions and is based on conditions on the asymptotic behaviour of U_n, U_n' and $\hat{\alpha}_n$. It is general enough to allow (i) matrix norming (ii) unequal limit matrices in connection with U_n and U_n' (iii) inclusion of external variables (iv) dependencies in the sequence of observations. The proofs are sketched only, their complete versions will be given in a future paper by Wellisch.

In sec. 4 we demonstrate how the techniques work in the special case of a semiparametric linear regression model for possibly dependent response variables. For more substantial results on the model side we have to refer to a forthcoming paper. In the case of a finite dimensional working parameter α our technique is similar to that of Liang and Zeger (1986), Murphy and Li (1995).

2 DIFFERENTIAL CALCULUS IN TOPOLOGICAL LINEAR SPACES

2.1 $\mathcal{M}$-derivatives

Let $(E, O(E))$, $(F, O(F))$ be topological $\mathbb{R}$-linear Hausdorff spaces (TLS), $a \in A \in O(E)$ and $f : A \to F$.

Definition: f is said to be differentiable at a in the direction of $x \in E$ if the limit

$$f_a'(x) := \lim_{\epsilon \to 0} \epsilon^{-1}[f(a + \epsilon x) - f(a)]$$

exists. If this is the case, we write $f \in D(a, F; \to x)$ and $f_a'(x) \in F$ will be called the directional derivative of f at a in the direction of x. We introduce the following definition

$$D(a, F; \to E) := \bigcap_{x \in E} D(a, F; \to x).$$

Let $\mathcal{M}$ be a class of subsets of E such that every singleton belongs to $\mathcal{M}$. Let $L(E, F)$ denote the continuous and linear mappings from E to F.

Definition: f is $\mathcal{M}$-differentiable at a if there exists $u \in L(E, F)$ such that

$$\lim_{\epsilon \to 0} \epsilon^{-1} r(f, a, \epsilon x) = 0$$

uniformly with respect to $x \in M$, for each $M \subset \mathcal{M}$. The remainder $r(f, a, x)$ is defined by

$$r(f, a, x) := f(a + x) - f(a) - u(x).$$

We write $f \in D_{\mathcal{M}}(a, F)$ and the mapping $u \equiv f'(a) \in L(E, F)$ is called the $\mathcal{M}$-derivative of f at a. A mapping f is called $\mathcal{M}$-differentiable in A if it is $\mathcal{M}$-differentiable for all $a \in A$ and we write $f \in D_{\mathcal{M}}(A, F)$.

Definition:

(i) When $\mathcal{M}$ is the class of all bounded subsets of E, f is said to be Fréchet differentiable at a. We write $f \in D_{\mathcal{F}}(a, F)$.

(ii) When $\mathcal{M}$ is the class of all sequentially compact subsets of E, f is said to be Hadamard differentiable at a. We write $f \in D_H(a, F)$.

(iii) When $\mathcal{M}$ is the class of all single point subsets of E, f is said to be Gâteaux differentiable at a. We write $f \in D_G(a, F)$.

Lemma 2.1.1 $D_{\mathcal{F}}(a, F) \subset D_H(a, F) \subset D_G(a, F) \subset D(a, F; \to E)$.

2.2 Fundamental properties

Lemma 2.2.1 Assume that $E, F \in TLS$, $a, x \in A \in O(E)$, $f \in D_{\mathcal{M}}(a, F)$ with the $\mathcal{M}$-derivative $f'(a)$ of f at a. Then we have

$$f'_a(x) = f'(a)(x).$$

By the lemma we see that the $\mathcal{M}$-derivative is uniquely determined.

Lemma 2.2.2 (*chain rule*) If $f \in D_H(a, F)$ and $g \in D_H(f(a), G)$ where $g : A_1 \to G \in TLS$ and $f(a) \in A_1 \in O(F)$, then

$$g \circ f \in D_H(a, G) \text{ and } (g \circ f)'(a) = g'(f(a)) \circ f'(a).$$

Fréchet differentiation has the chain rule (composition) property, too, but not Gâteaux differentiation.

Proof: See Yamamuro (1974, p. 11). $\square$

Definition: (*partial derivatives*) Let $E_1, E_2, F \in TLS$, $E \equiv E_1 \times E_2$, $a = (a_1, a_2) \in A$, an open subset of E, and $f : A \to F$. We consider classes $\mathcal{M}_i$ of subsets of $E_i\,(i = 1, 2)$ as in section 2.1 and put $\mathcal{M} := \{M_1 \times M_2 : M_i \in \mathcal{M}_i\ (i = 1, 2)\}$. f is said to be partially $\mathcal{M}$-differentiable at a in the first variable if the mapping $x_1 \mapsto f(x_1, a_2)$ of E_1 into F is $\mathcal{M}_1$-differentiable at a_1 and, if this is the case, the derivative is denoted by $\partial_1 f(a_1, a_2)$. The partial derivative $\partial_2 f(a_1, a_2)$ of f at a in the second variable is defined similary.
Note that by definition $\partial_i f(a_1, a_2) \in L(E_i, F)\,(i = 1, 2)$.

Lemma 2.2.3 If $f \in D_{\mathcal{M}}(a, F)$ then $\partial_i f(a_1, a_2), i = 1, 2$, exist and

$$f'(a_1, a_2)(x_1, x_2) = \partial_1 f(a_1, a_2)(x_1) + \partial_2 f(a_1, a_2)(x_2).$$

Proof: See Yamamuro (1974, p. 37). $\square$

2.3 Mean value theorem

For $a, b \in E$ we introduce the notation

$$[a, b] := \{x \in E : \exists t \in [0, 1] : x = a + t(b - a)\}.$$

Theorem 2.3.1 Assume that $E \in TLS$, $F \in LCS$ (locally convex Hausdorff space), $[a, a + x] \subset A \in O(E)$ and $f \in D(A, F; \to E)$. Let F^* be the dual space of F. Then for any $x^* \in F^*$ there exists $\xi \in (0, 1)$ such that

$$x^*[f(a + x) - f(a)] = x^*[f'(a + \xi x, x)].$$

Proof: We define the mapping

$$g(\xi) := x^*[f(a + \xi x)], \quad [0, 1] \to \mathbb{R}.$$

A calculation of the ordinary difference quotient shows that the mapping g is differentiable in [0,1] and that the equation

$$g'(\xi) = x^*[f'(a + \xi x, x)]$$

is valid for some $\xi \in [0, 1]$. Applying the mean value theorem for mappings from $\mathbb{R}$ to $\mathbb{R}$ to $g(1) - g(0)$ we conclude the proof. $\square$

In our situation of estimation equations we use the mean value theorem in the following way.

Corollary Let $E \in TLS$, $F = \mathbb{R}^d, d \in \mathbb{N}$, $[a, a + x] \subset A \in O(E)$ and $f \in D_{\mathcal{M}}(A, F)$. Then there exists $\xi_i \in (0, 1)$, $i = 1, \ldots, d$ such that

$$[f(a + x) - f(a)]_i = [f'(a + \xi_i x)(x)]_i$$

for $i = 1, \ldots, d$, where the index i denotes the i-th component of the vector.

Proof: Using lemma 2.2.1 and the fact that the coordinate projections are elements of $L(\mathbb{R}^d, \mathbb{R})$ we obtain the corollary directly from the mean value theorem. $\square$

Notation: Let $E \in TLS$, $F = \mathbb{R}^d$, $\xi_i \in (0,1)$ for $i = 1, \ldots, d$. We introduce the abbreviation
$f'(a+\xi^* x)(x) := p_1 \circ f'(a+\xi_1 x)(x) + p_2 \circ f'(a+\xi_2 x)(x) + \ldots + p_d \circ f'(a+\xi_d x)(x)$, where $p_1, \ldots, p_d$ denote the d coordinate projections, which for $i = 1, ..., d$ are defined by the $d \times d$-matrices

$$p_i := e_i \cdot e_i^T,$$

with e_i the i-th unit vector. Obviously, the mapping

$$f'(a+\xi^* x) := \sum_{i=1}^{d} p_i \circ f'(a+\xi_i x) : E \to F \tag{1}$$

is continuous and linear. Further we identify a mapping $f \in L(\mathbb{R}^d, \mathbb{R}^d)$ with its representing $d \times d-$matrix F and write

$$f(x) = F \cdot x \equiv f \cdot x. \tag{2}$$

In particular for $f \in L(\mathbb{R}, \mathbb{R})$ we write $f(x) = f \cdot x$ and we identify f with a suitable real number.

3 SEMIPARAMETRIC ESTIMATION EQUATIONS

3.1 Estimation function

Let Θ be an open subset of $\mathbb{R}^d$ and A an open subset of $T \in TLS$. We consider an estimation function

$$U_n : \Theta \times A \to \mathbb{R}^d. \tag{3}$$

For each $(\theta, \alpha) \in \Theta \times A$ the mapping $U_n(\theta, \alpha)$ is a measurable function of n random elements $X_1, \ldots, X_n$. With an estimator

$$\widehat{\alpha}_n(\theta; X_1, \ldots, X_n) \equiv \widehat{\alpha}_n : \Theta \to A$$

we can transform (3) to an estimation function which depends only on the parameter of interest $\theta \in \Theta$ and which is, for every $\theta \in \Theta$, a measurable function of the n random elements $X_1, \ldots, X_n$. With the sequence

$$f_n : \Theta \to \Theta \times A, \theta \mapsto (\theta, \widehat{\alpha}_n(\theta))$$

of mappings we obtain the estimation function

$$(U_n \circ f_n) : \Theta \to \mathbb{R}^d. \tag{4}$$

For the functional part $\alpha \in A$ in the estimation equation we distinguish the case of a nuisance parameter and the case of a working parameter (cf. Liang and Zeger (1986)). Both kind of parameters (nuisance or working) are unknown variables in the estimation equation. A nuisance parameter is connected with the underlying $X_1, \dots, X_n$, in the sense that it parametrizes their distribution: we introduce the notation $\mathbb{P}_{\theta,\alpha}$ for the underlying distribution in this case. A working parameter does not primarily parametrize the underlying distribution, and we write $\mathbb{P}_\theta$. It is connected with a given sequence $\widehat{\alpha}_n$ of estimators (mostly the $\mathbb{P}_\theta$-limit thereof, cf. Crowder (1995)).

3.2 Expansion by the mean value theorem

We assume that the estimation function U_n and the estimator $\widehat{\alpha}_n$ satisfy the conditions on differentiability

$$U_n \in D_H(\Theta \times A, \mathbb{R}^d), n \geq 1 \tag{D1}$$

$$\widehat{\alpha}_n \in D_H(\Theta, A), n \geq 1. \tag{D2}$$

Theorem 3.2.1 Let $\theta, \theta_0 \in \Theta$ with the property $[\theta, \theta_0] \subset \Theta$. Let $\alpha, \widehat{\alpha}_n(\theta_0) \in A$ with $[\alpha, \widehat{\alpha}_n(\theta_0)] \subset A$ for $n \geq 1$. Then the vector equation

$$\begin{aligned}(U_n \circ f_n)(\theta) = {} & U_n(\theta_0, \alpha) \\ & + \partial_2 U_n(\theta_0, \alpha_n^*)(\widehat{\alpha}_n(\theta_0) - \alpha) \\ & + [\partial_1 U_n(\theta_n^*, \widehat{\alpha}_n(\theta_n^*)) + \partial_2 U_n(\theta_n^*, \widehat{\alpha}_n(\theta_n^*)) \circ \widehat{\alpha}'_n(\theta_n^*)] \cdot (\theta - \theta_0),\end{aligned} \tag{5}$$

where $\theta_n^* := \theta + \xi_n^* \cdot (\theta_0 - \theta)$, $\xi_n^* \in (0,1)$ and $\alpha_n^* := \widehat{\alpha}_n(\theta_0) + \bar{\xi}_n^* \cdot (\alpha - \widehat{\alpha}_n(\theta_0))$, $\bar{\xi}_n^* \in (0,1)$, is valid. Note that we have used the notation (1) and the interpretation (2) in the third term of the right hand side of (5).

Proof: Applying the mean value theorem to the estimation function (4) we get the vector equation

$$(U_n \circ f_n)(\theta) = (U_n \circ f_n)(\theta_0) + (U_n \circ f_n)'(\theta_n^*) \cdot (\theta - \theta_0). \tag{6}$$

Applying the mean value theorem again to the first term of the right hand side of (6) and lemmata 2.2.2 and 2.2.3 to the second term we obtain the equation (5). □

3.3 Sufficient conditions for the existence of a consistent EE-estimator

Definition: A sequence $\widehat{\theta}_n \equiv \widehat{\theta}_n(X_1, \ldots, X_n), n \geq 1$, of d-dimensional random vectors is said to be a consistent estimation equation estimator (EEE) for θ of the estimation equation $U_n(\theta) = 0$, if for every $\theta \in \Theta \subset \mathbb{R}^d, \Theta$ open, and every $\epsilon > 0$ the convergence

$$\mathbb{P}\left(\left|\widehat{\theta}_n - \theta\right| \leq \epsilon, U_n(\widehat{\theta}_n) = 0\right) \to 1$$

holds for $n \to \infty$.

Let $\Gamma_n \equiv \Gamma_n(\theta, \alpha), n \geq 1$, be a sequence of regular $d \times d$-matrices with $\Gamma_n \to 0$ for $n \to \infty$ (each element). With $N_{n,s}(\theta) := \left\{\gamma \in \mathbb{R}^d : \left|\Gamma_n^{-T} \cdot (\gamma - \theta)\right| \leq s\right\}, s > 0$, we denote a neighbourhood of $\theta \in \Theta$. In the following we are using the $d \times d-$matrix

$$W_n(\theta_n^*) := \partial_1 U_n(\theta_n^*, \widehat{\alpha}_n(\theta_n^*)) + \partial_2 U_n(\theta_n^*, \widehat{\alpha}_n(\theta_n^*)) \circ \widehat{\alpha}_n^{'}(\theta_n^*), \tag{7}$$

cf. (1) and (2).

First we want to consider the case of an estimation equation with a nuisance parameter. We assume that the conditions (D1) and (D2) on differentiability are satisfied. Further we present three regularity conditions for the estimation function U_n and the estimator $\widehat{\alpha}_n$.

For all $(\theta, \alpha) \in \Theta \times A$ and all $n \geq 1$,

$$[\widehat{\alpha}_n(\theta), \alpha] \subset A$$

holds as well as the following:

$\boxed{\mathcal{U}}$ The sequence $\Gamma_n \, U_n(\theta, \alpha)$, $n \geq 1$, is $\mathbb{P}_{\theta,\alpha}$-stochastically bounded.

$\boxed{\mathcal{W}}$ There exists an $a > 0$ and, for all $\epsilon > 0$ and $s > 0$, an $n_0 \geq 1$ such that for all $n \geq n_0$

$$\mathbb{P}_{\theta,\alpha}\left(y^T \, \Gamma_n \, W_n\,(\theta_n^*)\, \Gamma_n^T \, y \leq -a, \; \forall \, \theta_n^* \in N_{n,s}(\theta), y \in \mathbb{R}^d, |y| = 1\right) \geq 1 - \epsilon.$$

$\boxed{\mathcal{N}}$ The sequence $\Gamma_n \, \partial_2 U_n(\theta, \alpha_n^*)(\widehat{\alpha}_n(\theta) - \alpha)$, $n \geq 1$, is $\mathbb{P}_{\theta,\alpha}$-stochastically bounded, for all $\alpha_n^* \in [\widehat{\alpha}_n(\theta), \alpha]$.

Note that in the conditions $\mathcal{W}$ and $\mathcal{N}$ the notation (1) is used.

Theorem 3.3.1 If the conditions $\mathcal{U}$, $\mathcal{W}$ and $\mathcal{N}$ are fulfilled, then there exists a consistent EEE $\widehat{\theta}_n, n \geq 1$, for $\theta \in \Theta$, of the estimation equation $U_n \circ f_n(\theta) = 0$.

Proof: We are following the proof of Pruscha (1996, p. 222-224), where the proof of existence of an EEE is done for an estimation equation without a nuisance parameter. The new idea in the situation of a functional nuisance parameter is the expansion by the mean value theorem for Hadamard differentiable mappings. Due to the nuisance parameter we get the extra term in (7). The additional regularity condition $\mathcal{N}$ is used to keep the term $\Gamma_n \partial_2 U_n(\theta, \alpha_n^*)(\widehat{\alpha}_n(\theta) - \alpha)$ small. □

In the situation of a working parameter we get a similar result. Again we assume that the conditions on differentiability (D1) and (D2) are satisfied. In addition to the existence of the estimator $\widehat{\alpha}_n(\theta)$ we assume that there exists some $\alpha = \alpha(\theta) \in A$ with the following property:
For all $\theta \in \Theta$ and all $n \geq 1$,

$$[\alpha, \widehat{\alpha}_n(\theta)] \subset A$$

holds as well as the regularity conditions $\mathcal{U}_w$, $\mathcal{W}_w$ and $\mathcal{N}_w$. Hereby, $\mathcal{U}_w$, $\mathcal{W}_w$, $\mathcal{N}_w$ are the same as $\mathcal{U}$, $\mathcal{W}$, $\mathcal{N}$, except that $\mathbb{P}_{\theta,\alpha}$ is replaced by $\mathbb{P}_\theta$.
Arguing like in the case of a nuisance parameter, we get the existence result.

Theorem 3.3.2 If the conditions $\mathcal{U}_w$, $\mathcal{W}_w$ and $\mathcal{N}_w$ are fulfilled, then there exists a consistent EEE $\widehat{\theta}_n, n \geq 1$, for $\theta \in \Theta$, of the estimation equation $U_n \circ f_n(\theta) = 0$.

Remarks:

(i) The proof of existence follows Billingsley (1961), Feigin (1975) and others. But we are concerned with the more general case of an estimation equation with functional nuisance or working parameter.

(ii) We can even prove the stronger Γ_n^{-T}-consistency property of the estimator $\widehat{\theta}_n$: The sequence $\Gamma_n^{-T}(\widehat{\theta}_n - \theta), n \geq 1$, is $\mathbb{P}_{\theta,\alpha}$-stochastically resp. $\mathbb{P}_\theta$-stochastically bounded.

3.4 Sufficient conditions for the asymptotic normality of a consistent EE-estimator

For the asymptotic normality of a consistent resp. Γ_n^{-T}-consistent EEE $\widehat{\theta}_n$ we need stronger conditions than the regularity conditions in section 3.3.
First we want to consider the case of an estimation equation with nuisance parameter. We assume that there exists an estimator $\widehat{\alpha}_n(\theta)$ for $\alpha \in A$ which fulfills the condition

$$[\widehat{\alpha}_n(\theta), \alpha] \subset A, \text{ for all } (\theta, \alpha) \in \Theta \times A \text{ and all } n \geq 1.$$

Again we suppose that the conditions (D1) and (D2) on differentiability are satisfied and that the following conditions hold for all $(\theta, \alpha) \in \Theta \times A$ and for $n \to \infty$.

$\boxed{\mathcal{U}^*}$
$$\Gamma_n U_n(\theta,\alpha) \xrightarrow{D_{\theta,\alpha}} \mathcal{N}_d(0,\Sigma(\theta,\alpha)),$$

where $\Sigma(\theta,\alpha)$ denotes a positive definit $d\times d$-matrix, which can functionally depend on the parameter θ and α.

$\boxed{\mathcal{W}^*}$
$$\Gamma_n W_n(\theta_n^*)\Gamma_n^T \xrightarrow{\mathbb{P}_{\theta,\alpha}} -B(\theta,\alpha),$$

for all sequences of d-dimensional random vectors θ_n^* which are Γ_n^{-T}-consistent, where $B(\theta,\alpha)$ denotes a positive definit $d\times d$-matrix, which can functionally depend on the parameter θ and α.

$\boxed{\mathcal{N}^*}$
$$\Gamma_n\,\partial_2 U_n(\theta,\alpha_n^*)(\widehat{\alpha}_n(\theta)-\alpha) \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0,$$

for all sequences of random elements α_n^* which have the property

$$\alpha_n^* \in [\widehat{\alpha}_n(\theta),\alpha] \text{ for all } n\geq 1.$$

Note that in the conditions $\mathcal{W}^*$ and $\mathcal{N}^*$ for every component of the vectors there can be different arguments θ_n^* resp. α_n^* (cf. notation (1)).

Lemma 3.4.1 The implications

$$\mathcal{U}^* \Rightarrow \mathcal{U}, \quad \mathcal{W}^* \Rightarrow \mathcal{W}, \quad \mathcal{N}^* \Rightarrow \mathcal{N}$$

are valid.

Lemma 3.4.1 and theorem 3.3.1 show that under the conditions $\mathcal{U}^*, \mathcal{W}^*$ and $\mathcal{N}^*$ there exists a Γ_n^{-T}-consistent EEE.

In the following lemma we present sufficient conditions for the regularity condition $\mathcal{N}^*$ concerning the estimator $\widehat{\alpha}_n(\theta)$. For that purpose we restrict the space T to be a normed $\mathbb{R}$-linear space.

Lemma 3.4.2 Let $(T,\|.\|_T)$ be a normed $\mathbb{R}$-linear space. Each of the following two conditions (a) and (b) is sufficient for the condition $\mathcal{N}^*$. It exists a sequence $J_n, n\geq 1$, of a.s. continuous, linear and bijective random mappings from T to T with

(a)
$$\widehat{\alpha}_n(\theta), n\geq 1, \text{ is } J_n^{-1}\text{-consistent}$$

and
$$\Gamma_n\partial_2 U_n(\theta,\alpha_n^*)\circ J_n \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0.$$

(b)
$$J_n^{-1}(\widehat{\alpha}_n(\theta)-\alpha) \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0$$

and
$$\Gamma_n\partial_2 U_n(\theta,\alpha_n^*)\circ J_n, n\geq 1, \text{ is } \mathbb{P}\text{-stochastically bounded.}$$

Theorem 3.4.1 If the conditions $\mathcal{U}^*, \mathcal{W}^*$ and $\mathcal{N}^*$ are fulfilled, then for a Γ_n^{-T}-consistent EEE $\widehat{\theta}_n, n \geq 1$, of $\theta \in \Theta$ the convergence in distribution

$$\Gamma_n^{-T}(\widehat{\theta}_n - \theta) \xrightarrow{D_{\theta,\alpha}} \mathcal{N}_d(0, B^{-1}(\theta,\alpha)\Sigma(\theta,\alpha)B^{-1}(\theta,\alpha))$$

holds for $n \to \infty$.

Proof: Let $\delta > 0$ be so small that the neighbourhood $N_\delta(\theta)$ lies completely in Θ. We introduce the sequence of sets

$$M_n := \{\widehat{\theta}_n \in N_\delta(\theta)\}, n \geq 1.$$

Note that on the sets M_n we have $[\widehat{\theta}_n, \theta] \subset \Theta$, further the convergence

$$1_{M_n} \xrightarrow{\mathbb{P}_{\theta,\alpha}} 1$$

is valid. The expansion by the mean value theorem, cf. (5), yields the vector equation

$$\begin{aligned} &1_{M_n}(U_n \circ f_n)(\widehat{\theta}_n) = \\ &\quad 1_{M_n}U_n(\theta,\alpha) + 1_{M_n}\partial_2 U_n(\theta,\alpha_n^*)(\widehat{\alpha}_n(\theta) - \alpha) + 1_{M_n}W_n(\theta_n^*)\cdot(\widehat{\theta}_n - \theta). \end{aligned} \tag{8}$$

Since the EEE $\widehat{\theta}_n$ is Γ_n^{-T}-consistent the random sequence θ_n^* is Γ_n^{-T}-consistent,too. The equation (8) together with the conditions $\mathcal{U}^*, \mathcal{W}^*$ and $\mathcal{N}^*$ and an argument of the Crámer type (cf. Pruscha (1996), Prop. B 3.9., p. 397) completes the proof. □

For estimation equations with working parameter we can modify the conditions like we have done in the section 3.3. Again, in addition we have to assume that there exists a suitable $\alpha = \alpha(\theta) \in A$, which now replaces the nuisance parameter. In this way we get analogous regularity conditions $\mathcal{U}_w^*, \mathcal{W}_w^*$ and $\mathcal{N}_w^*$ and an analogous asymptotic normality result.

Remarks:

(i) Alternatively, theorem 3.4.1 can be formulated with the expression 'Γ_n^{-T}-consistent' replaced by the term 'consistent'. Then in the condition $\mathcal{W}^*$ 'Γ_n^{-T}-consistent' has to be replaced by 'consistent'.

(ii) Finite dimensional nuisance resp. working parameters

$$\alpha \in \mathbb{R}^c \quad , c \in \mathbb{N},$$

can be treated as special cases of our functional approach. The conditions (D1) and (D2) on differentiability and the derivatives $\partial_2 U_n(\theta,\alpha)$ and $\widehat{\alpha}_n'$ can be read with the usual differential calculus. Note that these derivatives can be identified with suitable matrices and the composition of linear mappings can be identified with multiplication of matrices.

(iii) Finite dimensional working parameter are treated in Liang and Zeger (1986), Crowder (1995), Murphy and Li (1995). Our regularity conditions, together with condition (a) in lemma 3.4.2, correspond with the conditions given in Murphy and Li (1995).

(iv) Let us consider the special estimator

$$\widehat{\alpha}_n \equiv \widehat{\alpha}_n(X_1, ..., X_n), n \geq 1, \text{ for } \alpha \in A,$$

which does not functionally depend on $\theta \in \Theta$. Here (7) is of the reduced form

$$W_n(\theta_n^*) = \partial_1 U_n(\theta_n^*, \widehat{\alpha}_n).$$

(v) In the case of a working parameter and of a deterministic sequence $\widehat{\alpha}_n \equiv \alpha_0$ of constants, $\mathcal{N}_w^*$ is trivially fulfilled if we choose $\alpha = \alpha_0$.

(vi) The asymptotic covariance structure of the random vector $\Gamma_n^{-T}(\widehat{\theta}_n - \theta)$ is influenced by the nuisance parameter resp. working parameter. Therefore we get a direct effect of the nuisance resp. working parameter on the efficiency of the EEE $\widehat{\theta}_n$.

4 LEAST SQUARE ESTIMATION IN SEMIPARAMETRIC LINEAR REGRESSION MODELS

In this section we want to present an example which demonstrates the handling of the regularity conditions $\mathcal{U}^*, \mathcal{W}^*$ and $\mathcal{N}^*$. We consider the estimation equation which is derived from the least square approach in a semiparametric linear regression model. For this example we compute the expressions of section 3 and present sufficient conditions for the regularity conditions.

We are concerned with the model

$$Y_i = \alpha(X_i) + \theta \cdot Z_i + e_i \quad , i = 1, ..., n,$$

where $\alpha \in A$ denotes an unknown nuisance parameter in some function space and $\theta \in \Theta \subset \mathbb{R}, \Theta$ open, denotes the real valued parameter of interest. The regressors X_i and Z_i are real valued and possibly stochastic. Let e_i denote a sequence of noise variables, typically a sequence of i.i.d. random variables with mean zero. The least square method yields the estimation function

$$l_n(\theta, \alpha) \equiv -\frac{1}{2}\sum_{i=1}^{n}(Y_i - \alpha(X_i) - \theta \cdot Z_i)^2,$$

and differentiation gives the estimation equation $U_n(\theta, \alpha) = 0$, with

$$U_n(\theta, \alpha) \equiv \frac{d\, l_n(\theta, \alpha)}{d\, \theta} = \sum_{i=1}^{n} (Y_i - \alpha(X_i) - \theta \cdot Z_i) \cdot Z_i.$$

For the computation of the regularity condition $\mathcal{N}^*$ we need the partial derivative

$$\partial_2 U_n(\theta_0, \alpha_0) : A \to \mathbb{R}$$

for $(\theta_0, \alpha_0) \in \Theta \times A$. A simple calculation shows that the mapping

$$\partial_2 U_n(\theta_0, \alpha_0) : A \to \mathbb{R}, \quad \alpha \mapsto -\sum_{i=1}^{n} Z_i \cdot \alpha(X_i)$$

fulfills the equation

$$r(U_n(\theta_0, \cdot), \alpha_0, \epsilon x) = 0,$$

so the remainder term property (cf. definition in sec. 2.1) is fulfilled. Clearly the mapping $\partial_2 U_n(\theta_0, \alpha_0)$ is linear. To prove the continuity we have to specify the function space and its topology.

1. Let $A \equiv C[a, b]$ endowed with the topology induced by the sup-norm $\|\alpha\| \equiv \max_{a \le t \le b} |\alpha(t)|$. In this case the regressors $X_i, i = 1, ..., n$, are supposed to take a.s. values in the interval $[a, b]$. Hence the unequality

$$|\partial_2 U_n(\theta_0, \alpha_0)(\alpha)| \le \sum_{i=1}^{n} |Z_i| \cdot \|\alpha\|,$$

holds; that gives the boundedness of the mapping $\partial_2 U_n(\theta_0, \alpha_0)$, and so the continuity of the mapping $\partial_2 U_n(\theta_0, \alpha_0)$ is proved.

2. Let A be the space of all measurable mappings from $\mathbb{R}$ to $\mathbb{R}$ endowed with the trace topology induced by the initial topology of the evaluation mappings

$$\varphi_t : f \mapsto f(t), f \in \mathbb{R}^{\mathbb{R}}, t \in \mathbb{R}.$$

Note that this space is a $\mathbb{R}$-linear Hausdorff space, but the space is not normable. The subset A is open and due to the construction of the topology the functional $\partial_2 U_n(\theta_0, \alpha_0)$ is continuous.

For both choices 1. and 2. the condition (D1) on differentiability is fulfilled. Infact the mapping

$$U_n'(\theta_0, \alpha_0) : (\theta, \alpha) \mapsto -\sum_{i=1}^{n} (Z_i \cdot \alpha(X_i) + \theta \cdot Z_i^2),$$

with $(\theta_0, \alpha_0) \in \Theta \times A$ and $(\theta, \alpha) \in \Theta \times A$, is the $\mathcal{M}$-derivative of the mapping

$$U_n(\theta, \alpha) : (\theta, \alpha) \mapsto \sum_{i=1}^{n} (Y_i - \alpha(X_i) - \theta \cdot Z_i) \cdot Z_i.$$

Further we assume that there exists a suitable estimator $\widehat{\alpha}_n$, which satisfies the condition (D2) on differentiability.
Now we present the regularity conditions and various sufficient conditions. Let $\gamma(n), n \geq 1$, be a norming sequence of real numbers, tending to zero.

ad $\mathcal{N}^*$: Condition $\mathcal{N}^*$ has the form

$$-\gamma(n) \sum_{i=1}^{n} Z_i \cdot (\widehat{\alpha}_n(\theta, X_i) - \alpha(X_i)) \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0, \text{ for all } (\theta, \alpha) \in \Theta \times A.$$

Each of the following conditions (a) and (b) is sufficient for the condition $\mathcal{N}^*$.

(a) $\gamma(n) \sum_{i=1}^{n} |Z_i|$ is $\mathbb{P}_{\theta,\alpha}$-stochastically bounded.

$$\sup_{1 \leq i \leq n} |\widehat{\alpha}_n(\theta, X_i) - \alpha(X_i)| \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0.$$

(b) $\gamma(n)^2 \sum_{i=1}^{n} |Z_i|$ is $\mathbb{P}_{\theta,\alpha}$-stochastically bounded.

$$\sup_{1 \leq i \leq n} \gamma(n)^{-1} |\widehat{\alpha}_n(\theta, X_i) - \alpha(X_i)| \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0.$$

ad $\mathcal{W}^*$: With interpretation (2) and the notation

$$\alpha_1 := \widehat{\alpha}_n'(\theta_n^*)(1) \in A,$$

(7) can be written as

$$W_n(\theta_n^*) := -\sum_{i=1}^{n} Z_i^2 - \sum_{i=1}^{n} Z_i \cdot \alpha_1(X_i).$$

Now we can formulate the regularity condition $\mathcal{W}^*$ as follows:

$$-\gamma(n)^2 \sum_{i=1}^{n} Z_i^2 - \gamma(n)^2 \sum_{i=1}^{n} Z_i \cdot \alpha_1(X_i) \xrightarrow{\mathbb{P}_{\theta,\alpha}} -B(\theta, \alpha),$$

for all $(\theta, \alpha) \in \Theta \times A$ and for all sequences θ_n^*, $n \geq 1$, which are $\gamma(n)^{-1}$-consistent. Here $B(\theta, \alpha)$ denotes a positive real number functionally depending on the parameters θ and α.

The following pair of conditions

$$\gamma(n)^2 \sum_{i=1}^{n} Z_i^2 \xrightarrow{\mathbb{P}_{\theta,\alpha}} B(\theta, \alpha), \qquad (W_1^*)$$

$$\gamma(n)^2 \sum_{i=1}^{n} Z_i \cdot \alpha_1(X_i) \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0 \qquad (W_2^*)$$

is sufficient for $\mathcal{W}^*$. Note that in the case of an estimator $\widehat{\alpha}_n$ which is functionally independent of $\theta \in \Theta$ (c.f. remark (iv) in sec. 3.4), the condition (W_2^*) is always fulfilled.

ad $\mathcal{U}^*$: The regularity condition $\mathcal{U}^*$ takes the following form:

$$\gamma(n) \sum_{i=1}^{n} Z_i \cdot e_i \xrightarrow{D_{\theta,\alpha}} \mathcal{N}(0, \Sigma(\theta,\alpha)), \text{ for all } (\theta,\alpha) \in \Theta \times A.$$

Here $\Sigma(\theta,\alpha)$ denotes a positive real number which is functionally dependent on the parameter θ and α.
Using central limit theorems we present sufficient conditions for the regularity condition $\mathcal{U}^*$.

(U_1^*) $(e_i, Z_i), i \geq 1$, is a sequence of i.i.d. random vectors with $\mathbb{E}(e_i \cdot Z_i) = 0$ and $\operatorname{Var}(e_i \cdot Z_i) > 0$, for $i \geq 1$.

Under this assumption the usual central limit theorem yields the condition $\mathcal{U}^*$. Note that condition (U_1^*) does not cover models with deterministic regressors Z_i and models with autoregressive structure (e.g. $Z_i \equiv Y_{i-1}$). In order to enclose this kind of models we have to use central limit theorems for martingal difference sequences.

(U_2^*) $(e_1, e_2, ...; \mathcal{F}_0 \equiv \{\emptyset, \Omega\}, \mathcal{F}_1, \mathcal{F}_2, ...)$, with $\mathcal{F}_i := \{Z_1, ..., Z_{i+1}, e_1, ..., e_i\}$ for $i \geq 1$, is a martingal difference sequence fulfilling:

- $\exists \delta > 2 : \sup_{i\in\mathbb{N}} \mathbb{E}(|e_i|^\delta | \mathcal{F}_{i-1}) \leq M$ a.s.,
- $\exists c > 0 : \sigma_i^2 \equiv \operatorname{Var}(e_i|\mathcal{F}_i) \geq c$ a.s. $\forall i \in \mathbb{N}$,
- $\gamma(n)^2 \sum_{i=1}^n \sigma_i^2 Z_i^2 \xrightarrow{\mathbb{P}_{\theta,\alpha}} \Sigma(\theta,\alpha)$,
- $\max_{1\leq i\leq n} \gamma(n)|Z_i| \xrightarrow{\mathbb{P}_{\theta,\alpha}} 0$.

According to a central limit theorem for martingal difference sequences (cf. Dvoretzky (1970)) the assumption (U_2^*) is sufficient for $\mathcal{U}^*$.

Remarks

(i) The following situation is an important special case of the sufficient condition (U_2^*).
$e_i, i \in \mathbb{N}$, is a sequence of independent, centered random variables with

$$\operatorname{Var}(e_i) \equiv \sigma_i^2 \geq c > 0 \text{ for all } i \in \mathbb{N}.$$

Further e_i is independent of the random variables $Z_1, ..., Z_i$ for $i \in \mathbb{N}$.
Note that within these assumptions models with autoregressive structure can be treated.

(ii) Condition (U_2^*) covers the case of deterministic regressors Z_i.

(iii) For various nonparametric estimators $\widehat{\alpha}_n$ in (generalized) linear models the asymptotic normality of $\widehat{\theta}_n$ is proved under the i.i.d. assumption by Heckman (1986) (spline estimator), Severini and Staniswalis (1994) (weighted quasi-likelihood), Bickel, Klaassen, Ritov and Wellner (1993) (sec. 4.3), Mammen and van de Geer (1997) (penalized quasi-likelihood).

5 APPENDIX

Definition: Let $(T, \|.\|_T)$ be a normed $\mathbb{R}$-linear space and $x_n \in T, n \geq 1$, a sequence of random elements with values in T. The sequence $x_n \in T, n \geq 1$, is called $\mathbb{P}$-stochastically bounded if and only if

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}(\|x_n\|_T \geq M) = 0.$$

The sequence $x_n \in T, n \geq 1$, is said to be stochastically convergent to 0 if and only if

$$\forall \epsilon \geq 0 : \lim_{n\to\infty} \mathbb{P}(\|x_n\| > \epsilon) = 0.$$

Definition: Let T be a normed $\mathbb{R}$-linear space and $\theta \in T$ an unknown value. Let $\widehat{\theta}_n, n \geq 1$, be an estimator for θ. Further let

$$F_n : T \to T, n \geq 1,$$

be a sequence of measurable random mappings. If the sequence

$$F_n(\widehat{\theta}_n - \theta), n \geq 1,$$

is $\mathbb{P}$-stochastically bounded, the estimator $\widehat{\theta}_n, n \geq 1$, is said to be F_n-consistent.

Definition: Let $(T, \|.\|_T)$ be a normed $\mathbb{R}$-linear space and $f_n : T \to \mathbb{R}^d, n \geq 1$, a sequence of random mappings which are a.s. linear and continuous. The sequence f_n is said to be stochastically convergent to the zero mapping,

$$f_n \xrightarrow{\mathbb{P}} 0,$$

if and only if

$$\forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}(\|f_n\| \geq \epsilon) = 0,$$

where $\|.\|$ denotes the usual operator norm

$$\|f_n\| := \sup\{|f_n(x)| : x \in T, \|x\|_T = 1\}.$$

Acknowledgement: This work is supported by the Deutsche Forschungsgemeinschaft (SFB 386).

REFERENCES

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* The John Hopkins University Press.

Billingsley, P. (1961). *Statistical Inference for Markov Processes.* The University of Chicago Press.

Crowder, M. (1995). *On the use of a working correlation matrix in using generalised linear models for repeated measures.* Biometrika 82, 407-410.

Dvoretzky, A. (1970). *Asymptotic normality for sums of dependent random variables.* Proceedings of the Sixth Berkley Symposium on Mathematical Statistics and Probability, 513-535.

Feigin, P.D. (1975). *Maximum likelihood estimation for stochastic processes — a martingal approach.* Ph. D. Thesis, Australian National University.

Flett, T.M. (1980). *Differential Analysis.* Cambridge University Press, Cambridge.

Gill, R.D. (1989). *Non- and semiparametric maximum likelihood estimators and the von-Mises method (part I).* Scandinavian Journal of Statistics 16, 1069-1112.

Heckman, N.E. (1986). *Spline smoothing in a partly linear model.* Journal of the Royal Statistical Society B 48, 244-248.

Liang, K.-Y. and Zeger, S.L. (1986). *Longitudinal data analysis using generalized linear models.* Biometrika 73, 13-22.

Luenberger, D.G. (1969). *Optimization by Vector Space Methods.* Wiley N.Y.

Mammen, E. and van de Geer, S. (1997). *Penalized quasi-likelihood estimation in partial linear models.* The Annals of Statistics 25, 1014-1035.

Murphy, S. and Li, B. (1995). *Projected partial likelihood and its application to longitudinal data.* Biometrika 82, 399-406.

Pfanzagl, J. (1994). *Parametric Statistical Theory.* de Gruyter Berlin.

Pruscha, H. (1996). *Angewandte Methoden der Mathematischen Statistik.* Teubner Stuttgart.

Severini, T.A. and Wong, W.H. (1992). *Generalized profile likelihood and conditionally parametric models.* The Annals of Statistics 20, 1768-1802.

Severini, T.A. and Staniswalis J.G. (1994). *Quasi-likelihood estimation in semiparametric models.* Journal of the American Statistical Association 89, 501-511.

Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer N.Y.

Yamamuro, S. (1974). *Differential Calculus in Topological Linear Spaces.* Lecture Notes in Mathmatics, Vol.374. Springer N.Y.