Perturbation Invariant Estimates and Incidental Nuisance Parameters

Helmut Strasser

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 48
April 1996

http://statistik.wu-wien.ac.at/

Perturbation Invariant Estimates and Incidental Nuisance Parameters

Helmut Strasser

April 1996

Abstract

It is shown (Proposition (3.9)) that the asymptotic information bound which is valid for the estimation of a parameter in the structure (mixture) model remains valid in the functional model (incidental nuisance parameters) if only perturbation symmetric estimators (Definition (3.6)) are admitted. Perturbation symmetry is a property which is closely related to permutation symmetry (Theorem (3.4)). In particular, equicontinuous functions of empirical processes are perturbation symmetric (Theorem (3.3)). Thus, the results of this paper continue a discussion initiated by Bickel and Klaassen, [2], Pfanzagl, [14], and Strasser, [21], on permutation symmetry of estimators and the exclusion of superefficiency in the functional model.

Contents

[1] *AMS 1991 subject classifications.* Primary: 62F12; secondary: 62B15, 62C15

[2] *Key words and phrases.* Asymptotic efficiency of estimates, incidental nuisance parameters, structure model, functional model, permutation symmetric estimates, perturbation symmetric estimates.

1 Preliminaries

Overview

The first section of this paper is devoted to fixing the notation. In section 2 we try to explain the nuisance parameter problem and the difference between the views of the structure model and the functional model. Section 3 contains the statement and the discussion of the main results. Since we are not able to prove the main results without referring to abstract results of asymptotic decision theory, in section 4 we collect some basic facts which are not new but difficult to quote. Section 5 contains a detailed discussion of the proof of Proposition (2.5) and of the main result stated in Proposition (3.9). Section 6 contains proofs of minor theorems and auxiliary lemmas.

Basic Notation

Suppose that $\Theta \subseteq \mathbb{R}$ and $\Lambda \subseteq \mathbb{R}^p$ are open parameter sets and let $(P_{\theta\eta} : \theta \in \Theta, \eta \in \Lambda)$ be a family of p-measures on a sample space $(\Omega, \mathcal{A})$. Assume that the family $(P_{\theta\eta})$ is dominated by a σ-finite measure $\mu|\mathcal{A}$ and denote the μ-densities by $p_\theta(\omega, \eta)$. In this paper we will apply some regularity conditions for the family $(P_{\theta\eta})$ which are explicitly stated in section 5, (5.1), (5.2) and (5.3). In particular, these regularity conditions imply the existence of partial loglikelihood derivatives of the densities, denoted by

$$\ell_{1\theta} = \frac{\partial}{\partial\theta} \log \frac{dP_{\theta\eta}}{d\mu}, \quad \ell_{2\theta} = \frac{\partial}{\partial\eta} \log \frac{dP_{\theta\eta}}{d\mu}.$$

In the following we will call a family regular if these regularity conditions are satisfied.

Since we will deal with mixture models for the parameter η we consider p-measures $\Gamma|\mathcal{B}(\Lambda)$. In this case the symbol $R_{\theta\Gamma}|\mathcal{A} \otimes \mathcal{B}(\Lambda)$ denotes the mixture defined by

$$R_{\theta\Gamma}(A \times B) := \int_B P_{\theta\eta}(A)\,\Gamma(d\eta),\ A \in \mathcal{A},\ B \in \mathcal{B}(\Lambda).$$

The marginal distribution of $R_{\theta\Gamma}$ on $\mathcal{A}$ is denoted by $R'_{\theta\Gamma} := R_{\theta\Gamma}|\mathcal{A}$. If f is an $R_{\theta\Gamma}$-integrable function on $\Omega \times \Lambda$ then $E_{\theta\Gamma}(f|\mathcal{A})$ denotes the conditional expectation of f given $\mathcal{A}$ with respect to $R_{\theta\Gamma}$.

We are going to apply the terminology of tangent score functions (influence functions) of probability measures. If $P|\mathcal{A}$ is any p-measure then a function $g \in L_2(P)$ satisfying $E_P(g) = 0$ is called a tangent score function of P. The set of all tangent score functions of P is denoted by $L_2^*(P)$. A sequence (P_n) is tangent to P and has the tangent score function $g \in L_2^*(P)$ with respect to P if

$$\sqrt{\frac{dP_n}{dP}} = 1 + \frac{1}{2\sqrt{n}} g + r_n \quad \text{and} \quad \int r_n^2\, dP = o\Big(\frac{1}{n}\Big).$$

The symbol $\mathcal{N}(a, \sigma^2)$ denotes the normal distribution with mean a and variance σ^2.

The goal of the present paper is the proof of local asymptotic information bounds for estimators. For an easy statement of the main result let us agree upon some terminology.

Let us call a sequence of numbers $(\theta_n) \subseteq \Theta$ tangent to θ if $\theta_n = \theta + t/\sqrt{n}$. We need a similar concept for triangular arrays of parameters $\eta \in \Lambda$. Let (η_{ni}) be a triangular array of parameters in Λ. Another array (τ_{ni}) is called to be tangent to (η_{ni}) if $\tau_{ni} = \eta_{ni} + \psi(\eta_{ni})/\sqrt{n}$ for some function $\psi \in \mathcal{C}_b(\Lambda)$.

Our last terminological convention is concerned with loss functions for which asymptotic information bounds are valid. Let us call a loss function $W : \mathbb{R} \to [0, \infty)$ admissible if W is bounded and continuous, and if $W(x) = \ell(|x|)$ where $\ell : [0, \infty) \to \mathbb{R}$ is increasing and $\ell \not\equiv \sup \ell$ on $(0, \infty)$.

2 Introduction

The problem of estimating θ in the presence of a nuisance parameter η can be considered in two different ways.

In the so-called structure model (or mixture model) it is assumed that the values of the nuisance parameter η are realizations of independent and identically distributed random variables. If Γ denotes the distribution of such a random nuisance parameter and if $P_{\theta\eta}$ is the distribution of the observation ω then the mixture $R_{\theta\Gamma}$ is the distribution of the pair (ω, η). The risk of an $\mathcal{A}^n$-measurable estimator κ_n for estimating θ is then the expectation

$$\int W\Big(\sqrt{n}(\kappa_n - \theta)\Big)\, dR_{\theta\Gamma}'^n$$

where W is a loss function. From that point of view the nuisance parameter problem is an estimation problem with independent and identically distributed observations whose distribution $R_{\theta\Gamma}'^n$ depends on θ and on the unknown distribution Γ.

A different view is taken by the so-called functional model which is based on a particular sequence $\eta_{n1}, \eta_{n2}, \dots, \eta_{nn}$ of nuisance parameters (incidental nuisance parameters). Considering the functional model the risk of an $\mathcal{A}^n$-measurable estimator κ_n of θ is the expectation

$$\int W\Big(\sqrt{n}(\kappa_n - \theta)\Big)\, d\bigotimes_{i=1}^{n} P_{\theta\eta_{ni}}$$

where W again denotes a loss function. Viewed by the functional model the nuisance parameter problem is an estimation problem with independent but not necessarily identically distributed observations whose distributions depend on a unknown sequence $\eta_{n1}, \eta_{n2}, \dots, \eta_{nn}$.

Information Bounds

It is the goal of the present paper to establish local asymptotic information bounds for the functional model. For the structure model the problem of local asymptotic information bounds is more or less solved. A textbook presentation of this topic can be found in Pfanzagl and Wefelmeyer, [15], section 14.3, or Bickel, Klaassen, Ritov and Wellner, [1], section 4.5. For the functional model the information bound problem is a more complicated question. Recent papers on this topic are Bickel and Klaassen, [2], Pfanzagl, [14], and Strasser, [21].

There are several ways of stating a theorem on an asymptotic information bound. One way is to consider classes of estimators such that the superefficiency phenomenon is excluded. Median unbiased estimators or regular estimators are common examples of such classes. A second way is to state the information bound as being valid for arbitrary estimators. This requires a statement of the result in the form of a Hájek-LeCam alternative which is equivalent to an asymptotic admissibility and uniqueness assertion. This kind of stating an asymptotic information bound is due to LeCam, [6], and Hájek, [5]. This second type of statement implies both the better known asymptotic minimax results and the results on regular estimators as special cases. To our knowledge, however, the results concerning median unbiased estimators cannot be derived from the Hájek-LeCam alternative.

In this paper we prefer to state the information bounds in terms of a Hájek-LeCam alternative. The interesting case of median unbiased estimators for which our results are valid, too, will be discussed elsewhere.

The Information Bound for the Structure Model

Let us begin with a brief summary of the known results for the structure model.

The main result on local asymptotic information bounds for the structure model is that any asymptotically efficient estimator sequence has a uniquely determined stochastic expansion of order 1. Let us state this result a bit more precisely.

We start with a fixed distribution $\Gamma|\mathcal{B}(\Lambda)$ and assume that all Γ-continuous distributions are admitted as distributions of the nuisance parameter. That means, we are considering a so-called full mixture model. Then there exists a uniquely determined tangent score function $h^*_{\theta\Gamma} \in L^*_2(R'_{\theta\Gamma})$ (the so-called efficient influence function) whose variance

$$\sigma^2_\theta(\Gamma) := \int h^{*2}_{\theta\Gamma}\, dR'_{\theta\Gamma}$$

provides the optimal information bound.

(2.1) REMARK Let us give some details of the construction. A typical tangent sequence of $(R'^n_{\theta\Gamma})$ is gven by

$$\left(R'_{\theta+\frac{1}{\sqrt{n}},(1+\frac{1}{\sqrt{n}}k)\Gamma}\right)^n \tag{2.2}$$

where $k \in L^*_2(\Gamma)$ is a bounded tangent score function of Γ. The tangent score function of the sequence (2.2) is the conditional expectation $E_{\theta\Gamma}(\ell_{1\theta} + k\,|\,\mathcal{A})$. We choose $k^* \in L^*_2(\Gamma)$ in such a way that the variance of this score function is minimized if $k = k^*$. Our regularity conditions (see section 5) imply that the minimal variance is positive. Then the optimal score function is

$$h^*_{\theta\Gamma} = \frac{E_{\theta\Gamma}(\ell_{1\theta} + k^*\,|\,\mathcal{A})}{\int E_{\theta\Gamma}(\ell_{1\theta} + k^*\,|\,\mathcal{A})^2 dR'_{\theta\Gamma}} \tag{2.3}$$

and the minimal variance is

$$\sigma^2_\theta(\Gamma) = \frac{1}{\int E_{\theta\Gamma}(\ell_{1\theta} + k^*\,|\,\mathcal{A})^2 dR'_{\theta\Gamma}}. \tag{2.4}$$

Let us state the result on the information bound as a Hájek-LeCam alternative.

(2.5) PROPOSITION *Suppose that the family $(P_{\theta\eta})$ is regular and that W is an admissible loss function. Let $\Gamma|\mathcal{B}(\Lambda)$ be a probability distribution and let $h^*_{\theta\Gamma}$ and $\sigma^2_\theta(\Gamma)$ be defined by (2.3) and (2.4).*

Then any sequence (κ_n) of estimates satisfies one of the following alternatives (1) and (2):
Either
(1) for every sequence (θ_n), tangent to θ, and every sequence (Γ_n), tangent to Γ,

$$\lim_{n\to\infty} \int W\left(\sqrt{n}(\kappa_n - \theta_n)\right) dR'^n_{\theta_n\Gamma_n} \;=\; \int W\, d\mathcal{N}(0, \sigma^2_\theta(\Gamma)),$$

or
(2) there is a sequence (θ_n), tangent to θ, and a sequence (Γ_n), tangent to Γ, such that

$$\limsup_{n\to\infty} \int W\left(\sqrt{n}(\kappa_n - \theta_n)\right) dR'^n_{\theta_n\Gamma_n} \;>\; \int W\, d\mathcal{N}(0, \sigma^2_\theta(\Gamma)).$$

(3) Assertion (1) is true iff

$$\sqrt{n}(\kappa_n(\omega) - \theta) \;-\; \frac{1}{\sqrt{n}} \sum_{i=1}^n h^*_{\theta\Gamma}(\omega_i) \;\xrightarrow{R'^n_{\theta\Gamma}}\; 0.$$

The assertion of Proposition (2.5) is well-known (see Strasser, [19]). However, frequently it is stated in terms of a local asymptotic minimax property. Our formulation is stronger in that it contains an admissibility and a uniqueness assertion. For the actual construction of asymptotically efficient estimator sequences

(κ_n) the distribution Γ is to be estimated. This can be a difficult problem. For detailed information on the aspects of construction one should consult Pfanzagl, [13], in particular chapter 8. A particularly simple case is the following.

(2.6) DEFINITION *Let us call a family* $(P_{\theta\eta})$ *simple-structured if*

$$p_\theta(\omega, \eta) = q_\theta(\omega) v(T(\omega), \theta, \eta),$$

and if for every $\theta \in \Theta$ *the statistic* T *is complete for the family* $(P_{\theta\eta} : \eta \in \Lambda)$.

The famous example of Neyman and Scott, [12], is simple-structured. If the family $(P_{\theta\eta})$ is simple-structured then $h^*_{\theta\Gamma}$ does not depend on Γ and it is possible to construct asymptotically efficient estimator sequences by means of the conditional maximum likelihood method. For details confer Pfanzagl, [13], Remark 7.10.

Information Bounds for the Functional Model

Let us turn to the functional model.

For an experiment with n observations the nuisance parameters are given by a sequence $\eta_{n1}, \eta_{n2}, \ldots, \eta_{nn}$. Thus, in an asymptotic framework, we have to deal with triangular arrays (TAs) (η_{ni}) of nuisance parameters. As a special case we may consider TAs consisting of the first elements $\eta_{ni} = \eta_i,\ i = 1, \ldots, n$ of a given sequence (η_i).

An important technical question is for conditions which are to be imposed on the TAs (η_{ni}) in order to produce theorems on the functional model. The weakest condition under which the results of the present paper are valid requires that the TAs essentially live in compact subsets.

(2.7) DEFINITION *A TA* (η_{ni}) *is called tight if for every* $\epsilon > 0$ *there is a compact set* $K_\epsilon \subseteq \Lambda$ *such that*

$$\frac{1}{n}\sum_{i=1}^{n} 1_{K_\epsilon}(\eta_{ni}) > 1 - \epsilon \quad \text{for all } n \in \mathbb{N}.$$

Tightness is obviously a very weak condition on TAs. A seemingly stronger condition is weak convergence.

(2.8) DEFINITION *Let* $\Gamma|\mathcal{B}(\Lambda)$ *be a p-measure. A TA* (η_{ni}) *converges weakly to* Γ *if*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(\eta_{ni}) \;=\; \int f\, d\Gamma \quad \text{for every } f \in \mathcal{C}_b(\Lambda).$$

Weak convergence of a TA is only slightly stronger than tightness. In fact, every tight TA contains a weakly convergent subarray. For this reason we may develop the asymptotic theory for the functional model based on weakly convergent TAs. We will see later how these results can then easily be extended to the case of tight TAs.

Let (η_{ni}) be a TA which is weakly convergent to Γ. Then it is not very difficult to establish asymptotic information bounds for the estimation of θ. However, those straightforward information bounds are suffering the drawback that they only can be achieved by actual estimator sequences (κ_n) if the TA (η_{ni}) can be estimated.

(2.9) REMARK Let us give some details. A typical tangent sequence of $(\bigotimes_{i=1}^n P_{\theta\eta_{ni}})$ is given by

$$\Big(\bigotimes_{i=1}^{n} P_{\theta+\frac{1}{\sqrt{n}},\eta_{ni}+\frac{1}{\sqrt{ni}}\psi(\eta_{ni})}\Big) \tag{2.10}$$

where $\psi \in \mathcal{C}_b(\Lambda)$. The notion of a tangent score function can be carried over from the i.i.d. case to the case with non i.i.d. observations. This has been done by Strasser, [18]. Applications of this framework are given in Strasser, [19] and Strasser, [21]. The basic facts of this approach are repeated in the present paper in section 5. The sequence (2.10) has the tangent score function $\ell_{1\theta} + \psi\ell_{2\theta}$. We choose $\bar{\psi} \in L_2^*(\Lambda)$ in such a way that the variance of this score function is minimized for $\psi = \bar{\psi}$. Then the efficient score function for the estimation of θ is

$$\bar{h}_{\theta\Gamma} = \frac{\ell_{1\theta} + \bar{\psi}\ell_{2\theta}}{\int(\ell_{1\theta} + \bar{\psi}\ell_{2\theta})^2 dR_{\theta\Gamma}} \tag{2.11}$$

and the efficient variance is

$$\bar{\sigma}^2(\Gamma) = \int \bar{h}_{\theta\Gamma}^2 dR_{\theta\Gamma} = \frac{1}{\int(\ell_{1\theta} + \bar{\psi}\ell_{2\theta})^2 dR_{\theta\Gamma}}. \tag{2.12}$$

Typically we have (see 2.4)

$$\bar{\sigma}^2(\Gamma) < \sigma^2(\Gamma).$$

Although the bound $\bar{\sigma}^2(\Gamma)$ does only depend on Γ and not on the array (η_{ni}), an efficient estimator sequence requires some notion of (η_{ni}) since it is characterized by the expansion

$$\sqrt{n}(\kappa_n(\omega) - \theta) \; - \; \frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{h}_{\theta\Gamma}(\omega_i, \eta_{ni}) \;\overset{\bigotimes_{i=1}^n P_{\theta\eta_{ni}}}{\longrightarrow}\; 0.$$

However, in general it is not possible to construct estimator sequences having this expansion for a reasonable class of TAs. The reason is that the stochastic expansion does not only depend on the limit distribution Γ but also on the TA (η_{ni}) itself. It is not enough to estimate Γ.

Asymptotically Linear Estimators

The common idea to overcome the difficulties with the functional model is to impose restrictions on the class of estimator sequences that are admitted to the game. One hopes to find a class K of estimator sequences for which the information bound $\sigma_\theta^2(\Gamma)$ of the structure model remains valid in the functional model, too.

There are some technicalities to be discussed in this connection. Let $h^*_{\theta\Gamma}$ be the tangent vector which is optimal for the structure model. For this tangent vector to be used in the functional model it is necessary that the TA of functions $(\frac{1}{\sqrt{n}}h^*_{\theta\Gamma}(\omega_i))$ satisfies the Lindeberg condition for $(\bigotimes_{i=1}^n P_{\theta\eta_{ni}})$. A sufficient condition for this to hold for all tight TAs (η_{ni}) is that $\eta \mapsto E_{\theta\eta}(h^{*2}_{\theta\Gamma})$ is bounded and continuous, (cf. section 5, Definitions (5.4) and (5.5) and the subsequent remarks).

A preliminary result is that the information bound for the mixture model is valid for the functional model if we restrict the class of estimators to asymptotically linear estimator sequences.

(2.13) Definition *Let $\Gamma|\mathcal{B}(\Lambda)$ be a p-measure and assume that the TA (η_{ni}) is tight. An estimator sequence (κ_n) is called asmptotically linear (with respect to $(\bigotimes_{i=1}^n P_{\theta\eta_{ni}})$), if there exists an $\mathcal{A}$-measurable function h such that $\eta \mapsto P_{\theta\eta}(h^2)$ is bounded and continuous, and such that*

$$\sqrt{n}(\kappa_n(\omega) - \theta) \; - \; \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(h(\omega_i) - P_{\theta\eta_{ni}}(h)\Big) \;\overset{\bigotimes_{i=1}^n P_{\theta\eta_{ni}}}{\longrightarrow}\; 0. \tag{2.14}$$

The following assertion (2.18) concerning asymptotically linear estimators is not a very sophisticated one. In fact, it is a special case of a more general result by Bickel and Klaassen, [2], Proposition 1.1. The reason why we would like to state and prove it is the following. Our main results which are stated and proved in the following sections require the abstract mathematical framework of asymptotic decision theory. But the basic reason why the results are true is the validity of two simple lemmas, stated as Lemma (2.15) and Lemma (2.16), below. The proofs of our main results in the most general form might obscure the simplicity of the ideas behind. For the following Theorem (2.18), however, we are able to give a very simple proof which is based on the Cauchy-Schwarz inequality and where the role of the basic lemmas can be seen very clearly. Thus, the following theorem and its proof should be viewed as an introduction to the main results of this paper.

(2.15) LEMMA *Suppose that $(P_{\theta\eta})$ is regular. Then*

$$\Big\{E_{\theta\Gamma}(k|\mathcal{A}) : k \in L_2^*(\Gamma)\Big\} = \Big\{E_{\theta\Gamma}(\psi l_{2\theta}|\mathcal{A}) : \psi \in L_2(\Gamma)\Big\}$$

(2.16) LEMMA *Suppose that $(P_{\theta\eta})$ is regular. Let $h \in L_2(R'_{\theta\Gamma})$ and let $g_1,\ g_2 \in L_2(R_{\theta\Gamma})$ be such that $E_{\theta\eta}(g_i(.,\eta)) = 0$ for all $\eta \in \Lambda$ and $i = 1, 2$. If $E_{\theta\Gamma}(g_1|\mathcal{A}) = E_{\theta\Gamma}(g_2|\mathcal{A})$, then*

$$\int \Big(h - E_{\theta,.}(h)\Big)\, g_1\, dR_{\theta\Gamma} = \int \Big(h - E_{\theta,.}(h)\Big) g_2\, dR_{\theta\Gamma}$$

Lemmas (2.15) and (2.16) are proved in section 6. Let us prepare the assertion of Theorem (2.18) by some remarks.

(2.17) REMARK We are going to adopt an elementary point of view to asymptotic efficiency. First of all, any asymptotically linear estimator sequence (κ_n) is asymptotically normal. Let $\theta_n = \theta + \frac{t}{\sqrt{n}}$ and $\tau_{ni} = \eta_{ni} + \frac{1}{\sqrt{n}}\psi(\eta_{ni})$. In order to determine the limiting normal distribution let us denote

$$\begin{aligned} b &:= \int \Big(h - E_{\theta,.}(h)\Big)\, \ell_{1\theta}\, dR_{\theta\Gamma}, \\ c(\psi) &:= \int \Big(h - E_{\theta,.}(h)\Big)\, \psi\ell_{2\theta}\, dR_{\theta\Gamma}, \\ \sigma^2 &:= \int \Big(h - E_{\theta,.}(h)\Big)^2 dR_{\theta\Gamma}. \end{aligned}$$

Then it follows from the central limit theorem and LeCam's Third Lemma (Hájek, [4], VI,1.4) that

$$\mathcal{L}\Big(\sqrt{n}(\kappa_n - \theta)\,\Big|\, \bigotimes_{i=1}^n P_{\theta_n\tau_{ni}}\Big) \to \mathcal{N}(tb + c(\psi), \sigma^2), \quad \text{weakly.}$$

For an admissible loss function W this implies

$$\lim_{n\to\infty} \int W\Big(\sqrt{n}(\kappa_n - \theta_n)\Big) d\bigotimes_{i=1}^n P_{\theta_n\tau_{ni}} = \int W\Big(t(b-1) + c(\psi)\Big)\, d\mathcal{N}(0, \sigma^2).$$

Thus, if $b \neq 1$ or $c(\psi) \neq 0$ for some ψ then we may choose t or ψ in such a way that the asymptotic risk of our estimator sequence is arbitrarily close to $\sup W$. In this case the estimator sequence cannot be efficient. Such cases can only be excluded if $b = 1$ and $c(\psi) = 0$ for all ψ. These conditions are equivalent with (asymptotic) median unbiasedness or with regularity of the estimator sequence.

(2.18) THEOREM *Suppose that the family $(P_{\theta\eta})$ is regular, that the TA (η_{ni}) converges weakly to $\Gamma|\mathcal{B}(\Lambda)$ and that $\eta \mapsto E_{\theta\eta}(h_{\theta\Gamma}^{*2})$ is bounded and continuous. Let W be an admissible loss function.*

Then every sequence (κ_n) of median unbiased asymptotically linear estimators satisfies

$$\lim_{n\to\infty} \int W\Big(\sqrt{n}(\kappa_n - \theta_n)\Big) d\bigotimes_{i=1}^n P_{\theta_n\tau_{ni}} \geq \int W\, d\mathcal{N}(0, \sigma_\theta^2(\Gamma))$$

for every sequence (θ_n), tangent to θ, and every TA (τ_{ni}), tangent to (η_{ni}). Equality holds iff the expansion (2.14) of (κ_n) is based on $h = h^*_{\theta\Gamma}$.

Proof: Let $\theta_n = \theta + \frac{t}{\sqrt{n}}$ and $\tau_{ni} = \eta_{ni} + \frac{1}{\sqrt{n}}\psi(\eta_{ni})$. Let h be the function which defines the expansion (2.14) of (κ_n). By Remark (2.17) median unbiasedness implies

$$\int \Big(h - E_{\theta,\cdot}(h)\Big)(\ell_{1\theta} + \psi\ell_{2\theta})\, dR_{\theta\Gamma} = 1 \quad \text{for all } \psi. \tag{2.19}$$

Now, choose k^* according to equation (2.3). At this point the basic lemmas (2.15) and (2.16) get in on the act. By Lemma (2.15) there exists a function ψ^* such that

$$\begin{aligned} &\int \Big(h(\omega) - E_{\theta,\cdot}(h)\Big) E_{\theta\Gamma}(\ell_{1\theta} + k^*|\mathcal{A})\, dR_{\theta\Gamma} \\ &= \int \Big(h(\omega) - E_{\theta,\cdot}(h)\Big) E_{\theta\Gamma}(\ell_{1\theta} + \psi^*\ell_{2\theta}|\mathcal{A})\, dR_{\theta\Gamma}. \end{aligned}$$

By Lemma (2.16) we have

$$\begin{aligned} &\int \Big(h(\omega) - E_{\theta,\cdot}(h)\Big) E_{\theta\Gamma}(\ell_{1\theta} + \psi^*\ell_{2\theta}|\mathcal{A})\, dR_{\theta\Gamma} \\ &= \int \Big(h(\omega) - E_{\theta,\cdot}(h)\Big)(\ell_{1\theta} + \psi^*\ell_{2\theta})\, dR_{\theta\Gamma} = 1 \end{aligned}$$

which implies that

$$\int \Big(h(\omega) - E_{\theta,\cdot}(h)\Big) E_{\theta\Gamma}(\ell_{1\theta} + k^*|\mathcal{A})\, dR_{\theta\Gamma} = 1.$$

Now, the Cauchy-Schwarz inequality gives

$$\sigma^2 := \int \Big(h - E_{\theta,\cdot}(h)\Big)^2 dR_{\theta\Gamma} \geq \frac{1}{\int E_{\theta\Gamma}(\ell_{1\theta} + k^* \mid \mathcal{A})^2 dR'_{\theta\Gamma}} = \sigma^2_\theta(\Gamma).$$

This proves the assertion. □

Although this result might be sufficient for practical purposes a mathematicians mind tries to go beyond. The general problem is whether the asymptotic information bound for the mixture model remains valid for the functional model for broader classes than asymptotically linear estimators. Bickel and Klaassen, [2], prove that this is the case for estimator sequences which are permutation symmetric, asymptotically normal and regular. In [14], Pfanzagl studies the question whether the assertion can be true for the class of all permutation symmetric estimator sequences. However, by means of counterexamples he is able to show that this is not the case. There are examples of even simple-structured (Definition (2.6)) families with weakly convergent TAs (η_{ni}) and superefficient permutation symmetric estimator sequences.

One way to exclude superefficiency of permutation symmetric estimators could be to impose conditions on the TAs (η_{ni}). This is done in Strasser, [21]. It can be shown that for TAs of nuisance parameters which behave like very regular realizations of one-dimensional i.i.d. random variables, the desired assertion is valid for all permutation symmetric estimators. However, in view of Pfanzagl's counterexamples there is no hope to prove it for all weakly convergent TAs.

It is therefore natural to look into a different direction and try to modify the class of permutation symmetric estimators. This is done in the present paper.

3 The Main results

Desired Type of Results

We are now in a position to state the type of assertion which is to be proved in this paper. The central point is a class K of estimators for which the information bound $\sigma_\theta^2(\Gamma)$ of the mixture model remains valid in the functional model. The definition of the class K of estimator sequences is left open and will be filled later. So to speak, the following assertion is a frame of a theorem which becomes a theorem when K is specified. We choose a fixed array (η_{ni}) and consider the asymptotic risks for all tangent arrays (τ_{ni}). The assertion states the Hájek–LeCam alternative concerning asymptotic admissibility and uniqueness. Asymptotic efficiency with respect to all tangent arrays (τ_{ni}) is characterized by an asymptotic expansion under (η_{ni}).

(3.1) ASSERTION *Suppose that the family $(P_{\theta\eta})$ is regular and that W is an admissible loss function. Assume that the TA (η_{ni}) converges weakly to $\Gamma|\mathcal{B}(\Lambda)$ and that $\eta \mapsto E_{\theta\eta}(h_{\theta\Gamma}^{*2})$ is bounded and continuous.*

Then every sequence (κ_n) of estimates in the class K satisfies one of the following alternatives (1) and (2):
Either
(1) for every sequence (θ_n), tangent to θ, and every TA (τ_{ni}), tangent to (η_{ni}),

$$\lim_{n\to\infty} \int W\Big(\sqrt{n}(\kappa_n - \theta_n)\Big)\, d\bigotimes_{i=1}^n P_{\theta_n\tau_{ni}} \;=\; \int W\, d\mathcal{N}(0, \sigma_\theta^2(\Gamma)),$$

or
(2) there is a sequence (θ_n), tangent to θ, and a TA (τ_{ni}), tangent to (η_{ni}), such that

$$\limsup_{n\to\infty} \int W\Big(\sqrt{n}(\kappa_n - \theta_n)\Big)\, d\bigotimes_{i=1}^n P_{\theta_n\tau_{ni}} \;>\; \int W\, d\mathcal{N}(0, \sigma_\theta^2(\Gamma)).$$

(3) Assertion (1) is true iff

$$\sqrt{n}(\kappa_n(\omega) - \theta) \;-\; \frac{1}{\sqrt{n}}\sum_{i=1}^n h_{\theta\Gamma}^*(\omega_i) \quad \overset{\bigotimes_{i=1}^n P_{\theta\eta_{ni}}}{\longrightarrow} \quad 0.$$

An interesting question is whether the convergence of the TA (η_{ni}) can be dispensed with. The goal is to prove an assertion valid for all tight TAs. In specifying the information bound the role of Γ should be taken by the empirical distributions $E_n(\eta_n)$ of the rows $\eta_n = (\eta_{n1}, \eta_{n2}, \ldots, \eta_{nn})$ of the TA. However, since the tangent score function $h_{\theta\Gamma}^*$ depends on Γ such a result can only be obtained under additional assumptions on $h_{\theta\Gamma}^*$. To be explicit, we need that $\Gamma \mapsto \sigma_\theta^2(\Gamma)$ is continuous for the weak topology and $\eta \mapsto E_{\theta\eta}(h_{\theta\Gamma}^{*2})$ is bounded and continuous for every probability measure $\Gamma|\mathcal{B}(\Lambda)$. A simple sufficient condition for this to hold is that the family $(P_{\theta\eta})$ is simple-structured in the sense of Definition (2.6). Then $h_{\theta\Gamma}^*$ does not depend on Γ and we may write $h_\theta^* := h_{\theta\Gamma}^*$.

(3.2) ASSERTION *Suppose that the family $(P_{\theta\eta})$ is regular and simple-structured and that W is an admissible loss function. Assume that the TA (η_{ni}) is tight and that $\eta \mapsto E_{\theta\eta}(h_\theta^{*2})$ is bounded and continuous.*

Then every sequence (κ_n) of estimates in the class K satisfies one of the following alternatives (1) and (2):
Either
(1) for every sequence (θ_n), tangent to θ, and every TA (τ_{ni}), tangent to (η_{ni}),

$$\lim_{n\to\infty}\left(\int W\Big(\sqrt{n}(\kappa_n - \theta_n)\Big)\, d\bigotimes_{i=1}^n P_{\theta_n\tau_{ni}} \;-\; \int W\, d\mathcal{N}\Big(0, \sigma_\theta^2(E_n(\tau_n))\Big)\right) = 0,$$

or
(2) there is a sequence (θ_n), tangent to θ, and a TA (τ_{ni}), tangent to (η_{ni}), such that

$$\limsup_{n\to\infty}\left(\int W\Big(\sqrt{n}(\kappa_n-\theta_n)\Big)\,d\bigotimes_{i=1}^{n}P_{\theta_n\tau_{ni}}\;-\;\int W\,d\mathcal{N}\Big(0,\sigma_\theta^2(E_n(\tau_n))\Big)\right)>0.$$

(3) Assertion (1) is true iff

$$\sqrt{n}(\kappa_n(\omega)-\theta)\;-\;\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_\theta^*(\omega_i)\;\stackrel{\bigotimes_{i=1}^n P_{\theta\eta_{ni}}}{\longrightarrow}\;0.$$

It is easy to show that Assertion (3.1) implies Assertion (3.2) provided that Assertion (3.1) holds for arbitrary subsequences $\mathbb{N}_0\subseteq\mathbb{N}$, too. In fact, this is the case for the situations considered in this paper.

Thus, we are in a position to state the problem of information bounds for the functional model as follows: For which classes K of estimator sequences are the Assertions (3.1) and (3.2) valid ?

Functions of the Empirical Process

We know from Theorem (2.18) that the class of all asymptotically linear estimator sequences is a class of the desired type. Now, we are going to consider a considerably broader class with the same property. This will be the class of all estimator sequences which are equicontinuous functions of the empirical process. The reason why this class of estimators works is due to its behaviour under special contiguous sequences of probability measures which we will call perturbations. As a first step let us define what we call a perturbation of $(P_{\theta\eta})$. Simply speaking, perturbations are sequences of p–measures having particularly simple tangent vectors.

Let $\theta\in\Theta$ be fixed and consider the family $(P_{\theta\eta})_{\eta\in\Lambda}$. A (bounded and continuous) perturbation of that family is a sequence of families

$$Q_{n\eta}:=\Big(1+\frac{1}{\sqrt{n}}g(.,\eta)\Big)\,P_{\theta\eta},\eta\in\Lambda,\;n\in\mathbb{N},$$

where g is bounded and μ–continuous with respect to η. Let us call the function g the tangent vector of the perturbation.

Let $\Omega=\mathbb{R}$ and let $F_{\theta\eta}$ be the distribution function of $P_{\theta\eta}$. The (centered) empirical process for the sample size n is then given by

$$Z_{n\omega}(x):=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(1_{(-\infty,x]}(\omega_i)\;-\;F_{\theta\eta_{ni}}(x)\Big),\;x\in\mathbb{R}.$$

The empirical process is a maximal permutation symmetric statistic. It is therefore interesting to begin with the study of the behaviour of empirical processes under perturbations.

(3.3) THEOREM *Assume that $\eta\mapsto F_{\theta\eta}$ is continuous. Let $(Q_{n\eta})$ be a perturbation with the tangent vector g. If the TA (η_{ni}) converges weakly to Γ, then the distributions of the empirical processes $\mathcal{L}\Big(Z_n\mid\bigotimes_{i=1}^n Q_{n\eta_{ni}}\Big)$ converge weakly to a centered Gaussian process with covariance function*

$$C(x,y)=\int\Big(F_{\theta\eta}(x\cap y)-F_{\theta\eta}(x)F_{\theta\eta}(y)\Big)\,\Gamma(d\eta),\;(x,y)\in\mathbb{R}^2.$$

and mean value function

$$\mu(x)=\int_{-\infty}^{x}E_{\theta\Gamma}(g\,|\,\mathcal{A})\,dR'_{\theta\Gamma},\;x\in\mathbb{R}.$$

This theorem is a matter of routine and is proved in section 6.

Theorem (3.3) tells us which kind of perturbations do not influence the asymptotic distribution of the empirical processes. Obviously, these are exactly the perturbations satisfying $E_{\theta\Gamma}(g\,|\,\mathcal{A}) = 0$ $R'_{\theta\Gamma}$–a.e.

The following theorem states that sufficiently regular functions of the empirical process can be used to define classes of estimator sequences with the desired properties.

(3.4) THEOREM *Suppose that* $(P_{\theta\eta})$ *is regular and that* (η_{ni}) *is tight. Let* K *be the class of estimator sequences* (κ_n) *satisfying*

$$\sqrt{n}(\kappa_n - \theta) - \phi_n \circ Z_n \overset{\bigotimes_{i=1}^n P_{\theta\eta_{ni}}}{\longrightarrow} 0,$$

where (ϕ_n) *is an equicontinuous sequence of functions defined on the path space of the empirical process. Then assertions (3.1) and (3.2) are valid for* K.

We do not prove this theorem formally since it will turn out to be a special case of our main result on perturbation symmetric estimators (see Corollary (3.8) and Proposition (3.9)). Instead, let us give an intuitive argument which explains in an informal way why the assertion is valid.

Let (κ_n) be a sequence of estimators which is equivalent to an equicontinuous sequence of functions of the empirical process. Then all perturbations with tangents g satisfying $E_{\theta\Gamma}(g\,|\,\mathcal{A}) = 0$ $R'_{\theta\Gamma}$–a.e. do not influence the limit distribution. Only those components of perturbations which are orthogonal to such tangents g are of interest. The orthogonal space of the set of those tangents g is well–known to us: From Lemma (2.16) it follows that it is identical to the space of score functions of asymptotically linear estimators ! Thus, the reduction of estimators considered in Theorem (3.4) is equivalent to the reduction to asymptotically linear estimators. This is the intrinsic reason why Theorem (3.4) holds.

Perturbation Symmetric Estimators

To be an equicontinuous sequence of functions of the empirical process is a regularity condition but does not define a property which can be understood from a statistical point of view. Therefore we will isolate the statistical property which is responsible for the validity of Theorem (3.4). This property will be called perturbation invariance.

The basic idea of perturbation symmetry is to isolate the asymptotic invariance property of the empirical process and use it as a defining property for a class of estimator sequences. However, we found it intuitively more appealing to express perturbation symmetry in terms of probability measures instead of tangent score functions and without any reference to a limit distribution Γ. This is possible in view of the following lemma.

(3.5) LEMMA *Suppose that the TA* (η_{ni}) *converges weakly to* Γ *and that* $\eta \mapsto P_{\theta\eta}$ *is continuous. Assume that* $(Q^1_{n\eta})$ *and* $(Q^2_{n\eta})$ *are perturbations of* $(P_{\theta\eta})$ *with tangent score functions* g_1 *and* g_2*, respectively. Then*

$$\int \Big| E_{\theta\Gamma}(g_1 - g_2|\mathcal{A}) \Big| \, dR'_{\theta\Gamma} = \lim_{n\to\infty} \sqrt{n} \Big\| \frac{1}{n}\sum_{i=1}^n Q^1_{n\eta_{ni}} - \frac{1}{n}\sum_{i=1}^n Q^2_{n\eta_{ni}} \Big\|$$

Lemma (3.5) is proved in section 6. Combining this lemma with Theorem (3.3) we see that the distributions of the empirical processes cannot distinghuish between perturbations having similar average probabilities. This is the starting point for our definition of perturbation symmetry. For measuring the distance between p–measures we use the Dudley norm $\|.\|_D$.

(3.6) DEFINITION *Let* $\theta \in \Theta$ *and assume that* (η_{ni}) *is tight. A sequence* (T_n) *of statistics is called (asymptotically) perturbation symmetric (with respect to* $\theta \in \Theta$ *and* (η_{ni})*), if for every* $\epsilon > 0$ *there is a* *number* $\delta(\epsilon) > 0$ *such that*

$$\limsup_{n\to\infty} \sqrt{n} \Big\| \frac{1}{n}\sum_{i=1}^{n} Q^1_{n\eta_{ni}} \; - \; \frac{1}{n}\sum_{i=1}^{n} Q^2_{n\eta_{ni}} \Big\| < \delta(\epsilon)$$

implies

$$\limsup_{n\to\infty} \Big\| \mathcal{L}\Big(T_n \Big| \bigotimes_{i=1}^{n} Q^1_{n\eta_{ni}}\Big) - \mathcal{L}\Big(T_n \Big| \bigotimes_{i=1}^{n} Q^2_{n\eta_{ni}}\Big) \Big\|_D < \epsilon,$$

whenever $(Q^1_{n\eta})$ *and* $(Q^2_{n\eta})$ *are a (bounded and continuous) perturbations of* $(P_{\theta\eta})$.

Roughly speaking, the distributions of perturbation symmetric estimator sequences depend continuously on the average probabilities. We would like to have a closer look at the class K_{pts} of all perturbation symmetric estimator sequences. One could ask whether asymptotically linear estimators are perturbation invariant. This is indeed the case since the distributions of asymptotically linear estimator sequences are even linear functions of the average probabilities.

(3.7) COROLLARY *Suppose that* $(P_{\theta\eta})$ *is regular and that* (η_{ni}) *is tight. Then every asymptotically linear estimator sequence is perturbation symmetric.*

What is the relation between permutation symmetry and perturbation symmetry ? Obviously, every permutation symmetric statistic is a function of the empirical process. But this does not imply that perturbation symmetry can be carried from empirical processes to arbitrary permutation symmetric statistics. What we need is a kind of invariance principle. The following assertion is an immediate consequence of Theorem (3.3).

(3.8) COROLLARY *Suppose that* $(P_{\theta\eta})$ *is regular and that* (η_{ni}) *is tight. Suppose that an estimator sequence* (κ_n) *satisfies*

$$\sqrt{n}(\kappa_n - \theta) \; - \; \phi_n \circ Z_n \;\xrightarrow{\bigotimes_{i=1}^{n} P_{\theta\eta_{ni}}}\; 0,$$

where (ϕ_n) *is an equicontinuous sequence of functions defined on the path space of the empirical process. Then* (κ_n) *is perturbation symmetric.*

Now, let us state our main result.

(3.9) THEOREM *Assertions (3.1) and (3.2) are true for the class* K_{pts} *of all perturbation symmetric estimator sequences (at* $\theta \in \Theta$*).*

The proof of Proposition (3.9) is given in section 5.

4 The General Framework

The present paper is going to apply basic results of the abstract asymptotic decision theory. These results are due to LeCam, [7], and LeCam, [8]. Available textbook presentations are Millar, [11], Strasser, [17], LeCam, [9], van der Vaart, [22], and LeCam and Yang, [10].

For the complete proof of Propositions (2.5) and (3.9) there are three steps to be performed. In a first step the decision theoretic structure of the problem is settled for Gaussian shift experiments. The second step is concerned with the relation between finite sample size situations and limiting Gaussian shift experiments. For this step we will apply the general tools of asymptotic decision theory. Finally in a third step we have to apply the abstract results obtained by the first two steps to the particular nuisance parameter problem considered in this paper.

In this section we are going to give an overview over steps one and two. The third step is discussed in section 5.

The Gaussian Shift Case

Let $(H, <.,.>)$ be a Hilbert space and let $E = (P_h)_{h\in H}$ be a Gaussian shift experiment with parameter space H. This means that there is a linear stochastic process $(L(h))_{h\in H}$ such that

$$\frac{dP_h}{dP_0} = \exp\left(L(h) - \frac{1}{2}\|h\|^2\right),\ h \in H,$$

and $\mathcal{L}(L(h)|P_0) = \mathcal{N}(0, \|h\|^2),\ h \in H.$

We are considering the estimation problem of a linear function $f : T \to \mathbb{R}$, where $T \subseteq H$ is a closed, linear subspace. Our decision functions are substochastic transition kernels $\rho : \Omega \times \mathcal{B}(\mathbb{R}) \to [0, 1]$, whose risk for a loss function W is then defined by

$$\int\int W(x - f(h))\,\rho(\omega, dx)\,P_h(d\omega).$$

The classical result on this estimation problem is a theorem by Blyth, [3]. This result has been proved by Blyth for the case $H = \mathbb{R}$ and $P_h = \mathcal{N}(h, 1)$. Its extension to the general case, however, is easy (cf. Strasser, [17], Theorem (72.4)).

In the general case Blyth's theorem states the following: There is an estimator ρ^*, which is minimax and admissible, and which is uniquely determined by its distribution. This estimator can be found in the following way: The continuous linear function f can be represented as $f(h) =< h, h^* >$, $h \in H$, with some uniquely determined $h^* \in H$. Then the estimator ρ^* is given by $\rho^* = \epsilon_{L(h^*)}$. Let us give a precise statement of Blyth's theorem.

(4.1) THEOREM *Let W be an admissible loss function and let $h^* \in T$ be such that $f(h) =< h, h^* >$, $h \in T$. Then for every decision function ρ the following alternative holds:*
Either
(1) for all $h \in T$

$$\int\int W(x - f(h))\,\rho(\omega, dx)\,P_h(d\omega) = \int W\,d\mathcal{N}(0, \|h^*\|^2),$$

or
(2) there exists at least one $h \in T$ such that

$$\int\int W(x - f(h))\,\rho(\omega, dx)\,P_h(d\omega) > \int W\,d\mathcal{N}(0, \|h^*\|^2).$$

(3) Assertion (1) is valid iff $\rho = \epsilon_{L(h^*)}$.

Assertions (1) and (2) imply minimaxity and admissibility. Assertion (3) is concerned with uniqueness.

The most common situation in a semiparametric framework is as follows. The subspace T is of the form $T = \{th_0 + k : t \in \mathbb{R}, k \in G\}$, where $G \subseteq H$ is a closed linear subspace and $h_0 \notin G$. The linear function to be estimated is given by $f(th_0 + k) = t$, $t \in \mathbb{R}$, $k \in G$. Let pr_G be the orthogonal projection from H onto G. Then it is easy to see that

$$h^* = \frac{h_0 - \mathrm{pr}_G h_0}{\|h_0 - \mathrm{pr}_G h_0\|^2}.$$

This is exactly the method of computing so-called canonical gradients of functionals.

In section 5 we will apply Theorem (4.1) to obtain the optimal asymptotic information bound for the structure model. For the functional model we are, however, faced with another Hilbert space H than in the structure model and, thus, we are getting a different solution for the information bound problem. As mentioned earlier (see (2.9)) this solution turns out to be useless for the construction of estimator sequences. Our method for overcoming this dilemma is to narrow the class of decision functions by a symmetry condition. For the limiting Gaussian shift experiment this symmetry condition is an invariance condition.

We are, thus, faced with the problem of modifying the assertion of Blyth's theorem in order to cover situations with invariant decision functions. This is done in the following.

Assume that $f : T \to \mathbb{R}$ is a continuous linear function. Let $V \subseteq H$ be a closed linear subspace satisfying $T \cap V \subseteq \ker f$. Let us call a decision function V–invariant if $\mathcal{L}(\rho|P_{h_1}) = \mathcal{L}(\rho|P_{h_2})$ whenever $h_1 - h_2 \in V$. We will consider the problem of estimating f by a V–invariant decision function.

It is a natural idea to extend f to $T_1 := \overline{\text{span}}\{T \cup V\}$ by defining $f(h) := 0$ if $h \in V$, and then to apply Theorem (4.1) to T_1 and f. Let $h^{**} \in T_1$ be such that $f(h) =< h, h^{**} >$ for $h \in T_1$. Note that h^{**} is uniquely determined by the properties $h^{**} \in \overline{\text{span}}\{T \cup V\} \cap V^\perp$ and $f(h) =< h, h^{**} >$, $h \in T$. Then Theorem (4.1) can be applied with h^* replaced by h^{**} and T by T_1. This gives a minimax and admissibility assertion among all decision functions for the larger parameter space T_1. But for V–invariant decision functions assertions concerning T_1 are equivalent to assertions restricted to T. Thus, we obtain the following result.

(4.2) COROLLARY *Let W be an admissible loss function and assume that $T \cap V \subseteq \ker f$. Let $h^{**} \in V^\perp \cap \overline{\text{span}}\{T \cup V\}$ be such that $f(h) =< h, h^{**} >$, $h \in T$. Then for every V–invariant decision function ρ the following alternative holds:*
Either
(1) for all $h \in T$

$$\int\int W(x - f(h))\, \rho(\omega, dx)\, P_h(d\omega) = \int W\, d\mathcal{N}(0, \|h^{**}\|^2),$$

or
(2) there exists at least one $h \in T$ such that

$$\int\int W(x - f(h))\, \rho(\omega, dx)\, P_h(d\omega) > \int W\, d\mathcal{N}(0, \|h^{**}\|^2).$$

(3) Assertion (1) is valid iff $\rho = \epsilon_{L(h^{**})}$.

Corollary (4.2) is the basis of our main asymptotic result. It implies that every estimator sequence which has only V–invariant accumulation points obeys the information bound defined by h^{**}. Moreover, for the functional model we will define a subspace V of the full tangent space H such that the tangent vector h^{**} will coincide with the optimal tangent vector h^* in the structure model.

The Asymptotic Situation

Let us turn to the second step of the proving procedure. We have to translate the assertions of Theorem (4.1) and Corollary (4.2) into asymptotic assertions.

Let $(\Omega_n, \mathcal{A}_n)$, $n \in \mathbb{N}$, be a sequence of sample spaces and let $P_{n0}|\mathcal{A}_n$, $n \in \mathbb{N}$, be a sequence of p–measures. The local asymptotic decision theory considers the properties of an estimator sequence not only for the fixed sequence of p–measures (P_{n0}) but also for other sequences (Q_n) whose likelihood ratios $(\frac{dQ_n}{dP_{n0}})$ satisfy some specific conditions.

Those situations which are typical for the asymptotic study of structure and functional models can be described in the following way.

There is a basic Hilbert space H and a dense linear subspace $C \subseteq H$. For every $h \in C$ there is a sequence $(L_n(h))$ of random variables, depending linearly from $h \in C$, and satisfying $\mathcal{L}(L_n(h)|P_{n0}) \to \mathcal{N}(0, \|h\|^2)$, weakly. A sequence of p–measures (Q_n) is said to have a LAN–expansion for (P_{n0}) with $(L_n(h))$ if

$$\frac{dQ_n}{dP_{n0}} = \exp\Big(L_n(h) - \frac{1}{2}\|h\|^2 + r_n\Big), \quad \text{where } r_n \stackrel{P_{n0}}{\to} 0.$$

Any sequence of p–measures having a LAN–expansion with $(L_n(h))$ is denoted generically by (P_{nh}).

In situations with i.i.d. observations, e.g. for the structure model, it is possible to define $(L_n(h))$ even for all $h \in H$. In more general cases like the functional model, however, distinguishing between H and C is indispensable.

The main results of asymptotic decision theory make it possible to turn Theorem (4.1) into an asymptotic assertion (cf. Strasser, [17], Theorem (83.5)).

(4.3) THEOREM *(Hájek, [5], LeCam, [8]) Let W be an admissible loss function. Assume that $C \cap T$ is dense in T and let $h^* \in C \cap T$ be such that $f(h) =< h, h^* >, h \in T$. Then for every sequence of estimates (S_n) the following alternative holds:*
Either
(1) for all $h \in C \cap T$

$$\lim_{n\to\infty} \int W(S_n - f(h))\, dP_{nh} = \int W\, d\mathcal{N}(0, \|h^*\|^2),$$

or
(2) there exists at least one $h \in C \cap T$ such that

$$\limsup_{n\to\infty} \int W(S_n - f(h))\, dP_{nh} > \int W\, d\mathcal{N}(0, \|h^*\|^2).$$

(3) Assertion (1) is valid iff $S_n - L_n(h^*) \stackrel{P_{n0}}{\to} 0.$

The application of Theorem (4.3) to the situation of the structure model gives Proposition (2.5). This will be shown in section 5.

For mastering the functional model we need an asymptotic version of Corollary (4.2). First we have to modify the concept of V–invariance in order to cover estimator sequences.

(4.4) DEFINITION *An estimator sequence (S_n) is called asymptotically V–invariant if for every $\epsilon > 0$ there is a number $\delta(\epsilon) > 0$ such that*

$$\limsup_{n\to\infty} \|\mathcal{L}(S_n|P_{nh_1}) - \mathcal{L}(S_n|P_{nh_2})\|_D < \epsilon$$

if $h_1,\ h_2 \in C$ and dist$(h_1 - h_2, V) < \delta(\epsilon)$.

If an estimator sequence (S_n) is asymptotically V–invariant then all its weak accumulation points ρ are V–invariant. This implies the following asymptotic version of Corollary (4.2).

(4.5) COROLLARY *Let W be an admissible loss function and assume that $T \cap V \subseteq \ker f$. Assume that $C \cap T$ is dense in T and let $h^{**} \in C \cap \overline{\text{span}}\{T \cup V\} \cap V^\perp$ be such that $f(h) =< h, h^{**} >, h \in T$. Then for every sequence of asymptotically V–invariant estimates (S_n) the following alternative holds:*
Either
(1) for all $h \in C \cap T$

$$\lim_{n\to\infty} \int W(S_n - f(h))\, dP_{nh} = \int W\, d\mathcal{N}(0, \|h^{**}\|^2),$$

or
(2) there exists at least one $h \in C \cap T$ such that

$$\limsup_{n\to\infty} \int W(S_n - f(h))\, dP_{nh} > \int W\, d\mathcal{N}(0, \|h^{**}\|^2).$$

(3) Assertion (1) is valid iff $S_n - L_n(h^{**}) \stackrel{P_{n0}}{\to} 0.$

Proof: Apply LeCam's theorems, (LeCam, [8], cf. Strasser, [17], (62.3) and (63.6)), and Corollary (4.2). □

This is the abstract version of our main result Proposition (3.9). In section 5 we will apply Theorem (4.3) and Corollary (4.5) to the structure model and the functional model in order to obtain Propositions (2.5) and (3.9).

5 Discussion of Proofs

We will apply the following regularity conditions.

(5.1) CONDITION The family $(P_{\theta\eta} : \theta \in \Theta, \eta \in \Lambda)$ is continuously $L_2(\mu)$-differentiable with partial loglikelihood derivatives $l_{1\theta}$ and $l_{2\theta}$.

(5.2) CONDITION For every $\theta \in \Theta$ the functions $\eta \mapsto P_{\theta\eta}(l_{1\theta}^2)$ and $\eta \mapsto P_{\theta\eta}(l_{2\theta}^2)$ are continuous and bounded on Λ.

If $\Gamma|\mathcal{B}(\Lambda)$ is a p-measure then we call (θ,Γ) a regular pair if

$$E_{\theta\Gamma}(l_{1\theta}|\mathcal{A}) \notin \{E_{\theta\Gamma}(k|\mathcal{A}) : k \in L_2^*(\Gamma)\}.$$

Let us call a parameter point $\theta \in \Theta$ identifiable if (θ,Γ) is a regular pair for every $\Gamma|\mathcal{B}(\Lambda)$. If $(P_{\theta\eta})$ is simple-structured then θ is identifiable iff

$$\int \Big(\frac{\dot q_\theta}{q_\theta}\Big)^2 dP_{\theta\eta} > 0 \quad \text{for all } \eta \in \Lambda.$$

(Apply Pfanzagl and Wefelmeyer, [15], p. 234, (14.3.20).)

(5.3) CONDITION Every $\theta \in \Theta$ is identifiable.

The family $(P_{\theta\eta})$ is to be called regular if conditions (5.1), (5.2) and (5.3) are satisfied.

The Structure Model

We will briefly discuss how to obtain Proposition (2.5) from Theorem (4.3). We have to specify the choice of H, T, C, (P_{nh}), (L_n) and f.

Let $H := L_2^*(R'_{\theta\Gamma})$ and define

$$L_n(h)(\omega) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(\omega_i), \quad h \in L_2^*(R'_{\theta\Gamma}).$$

We have to consider the asymptotic properties of an estimator sequence $S_n := \sqrt{n}(\kappa_n - \theta)$. Our basic sequence of p-measures will be $P_{n0} := R'^n_{\theta\Gamma}$. We are interested in contiguous sequences which come up by variation of the parameters θ and Γ.

Let $\theta_n = \theta + \frac{1}{\sqrt{n}} t$ and let (Γ_n) be tangent to Γ with tangent vector $k \in L_2^*(\Gamma)$. Then it follows from conditions (5.1) and (5.2) that $(R'_{\theta_n\Gamma_n})$ is tangent to $R'_{\theta\Gamma}$ with tangent vector

$$h := E_{\theta\Gamma}(tl_{1\theta} + k|\mathcal{A}) \in H.$$

This implies that a LAN-expansion with $(L_n(h))$ is valid and with the notation of section 4 we may denote $P_{nh} := R'^n_{\theta_n\Gamma_n}$. Let us denote the closed hull of all such tangent vectors by

$$T_1(\theta,\Gamma) := \overline{\{E_{\theta\Gamma}(tl_{1\theta} + k|\mathcal{A}) : t \in \mathbb{R},\, k \in L_2^*(\Gamma)\}} \subseteq L_2^*(R'_{\theta\Gamma}).$$

This is the tangent space of the model. For the application of Theorem (4.3) we will define $T := T_1(\theta,\Gamma)$.

For efficiency the estimator sequence $S_n = \sqrt{n}(\kappa_n - \theta)$ applied to the parameters $\theta_n = \theta + \frac{1}{\sqrt{n}} t$ has to be concentrated as much as possible around t. Therefore the linear function f to be estimated is

$$f(h) := t \quad \text{if} \;\; h := E_{\theta\Gamma}(tl_{1\theta} + k|\mathcal{A}) \in H.$$

If (θ,Γ) is a regular pair then there is a uniquely determined tangent vector $h^*_{\theta\Gamma} \in T_1(\theta,\Gamma)$ such that

$$\int h^*_{\theta\Gamma} E_{\theta\Gamma}(tl_{1\theta} + k|\mathcal{A})\, dR'_{\theta\Gamma} = t \quad \text{for all } t \in \mathbb{R},\; k \in L_2^*(\Gamma).$$

(For the construction see Remark (2.1).) With this notation we are in the position to apply Theorem (4.3).

Proof: (of Proposition (2.5)) Apply Theorem (4.3) with the notation introduced above. □

The Functional Model

For the proof of Proposition (3.9) we would like to apply Corollary (4.5). For doing so we have to extend concepts like tangent score functions and tangent spaces to situations with independent but not identically distributed observations. This has been done by Strasser, [18]. A detailed discussion of the completely general situation with independent but not identically distributed observations is contained in Strasser, [20]. For the present purposes we will repeat the main steps.

Instead of a single p-measure for every parameter θ we are faced with a family $\mathbb{P}_\theta := (P_{\theta\eta})_{\eta\in\Lambda}$. We have to define tangent spaces of such families.

Define

$$H := L_2^*(\mathbb{P}_\theta, \Gamma) := \Big\{ g \in L_2(R_{\theta\Gamma}) : \int g(.,\eta)\, dP_{\theta\eta} = 0 \ \Gamma\text{-a.e.} \Big\}.$$

Note, that $L_2^*(\mathbb{P}_\theta, \Gamma)$ is a subspace of $L_2^*(R_{\theta\Gamma})$ but does not coincide with the latter space.

Let (η_{ni}) be a TA which converges weakly to Γ. In the functional model we again are interested in the asymptotic properties of estimator sequences $S_n := \sqrt{n}(\kappa_n - \theta)$. The basic sequence of p-measures in this case is $P_{n0} := \bigotimes_{i=1}^n P_{\theta\eta_{ni}}$ which is a product of different factors. In order to construct contiguous sequences we may not use all elements of $H = L_2^*(\mathbb{P}_\theta, \Gamma)$.

(5.4) DEFINITION *A function $g \in L_2^*(\mathbb{P}_\theta, \Gamma)$ satisfies condition (L) (Lindeberg condition) for (η_{ni}) if*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n P_{\theta\eta_{ni}}\Big(g(.,\eta_{ni})^2 \Big) < \infty, \quad \textit{and}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \int_{|g(.,\eta_{ni})| > \sqrt{n}\epsilon} g(.,\eta_{ni})^2\, dP_{\theta\eta_{ni}} = 0 \quad \textit{for all } \epsilon > 0.$$

For $h \in L_2^*(\mathbb{P}_\theta, \Gamma)$ define

$$L_n(h)(\omega) := \frac{1}{\sqrt{n}} \sum_{i=1}^n h(\omega_i, \eta_{ni}).$$

If h satisfies condition (L) for (η_{ni}) then we have

$$\mathcal{L}\Big(L_n(h) \Big| \bigotimes_{i=1}^n P_{\theta\eta_{ni}} \Big) \ \to\ \mathcal{N}(0, \|h\|_{\theta\Gamma}^2), \text{ weakly}.$$

We are now going to define a set of functions which is dense in $L_2^*(\mathbb{P}_\theta, \Gamma)$ for all $\Gamma|\mathcal{B}(\Lambda)$ and whose elements satisfy condition (L) for all tight TAs (η_{ni}).

(5.5) DEFINITION *Let $\mathcal{C}_b^*(\mathbb{P}_\theta)$ be the set of all measurable functions $g : \Omega \times \Lambda \to \mathbb{R}$ such that $\eta \mapsto g(.,\eta)\sqrt{p_\theta(.,\eta)}$ is a continuous and norm-bounded function from Λ to $L_2(\mu)$ and such that $E_{\theta\eta}(g(.,\eta)) = 0$ for all $\eta \in \Lambda$.*

The set $\mathcal{C}_b^*(\mathbb{P}_\theta)$ will take the role of C when we are going to apply Corollary (4.5).

If $(P_{\theta\eta})$ satisfies conditions (5.1) and (5.2) then the partial loglikelihood derivatives $l_{1\theta}$ and $l_{2\theta}$ are functions in $\mathcal{C}_b^*(\mathbb{P}_\theta)$. Moreover, an $\mathcal{A}$-measurable function h is in $\mathcal{C}_b^*(\mathbb{P}_\theta)$ iff $E_{\theta\eta}(h) = 0$, $\eta \in \Lambda$, and $\eta \mapsto E_{\theta\eta}(h^2)$ is continuous and bounded on Λ, (Lemma (6.1)). The main advantages of $\mathcal{C}_b^*(\mathbb{P}_\theta)$ are the properties which we have already announced. The set $\mathcal{C}_b^*(\mathbb{P}_\theta)$ is dense in $L_2^*(\mathbb{P}_\theta, \Gamma)$ for all $\Gamma|\mathcal{B}(\Lambda)$, (Lemma (6.2)), and all functions in $\mathcal{C}_b^*(\mathbb{P}_\theta)$ satisfy condition (L) for all tight TAs (η_{ni}), (Lemma (6.3)).

Let us consider p-measures which come up from $\bigotimes_{i=1}^n P_{\theta\eta_{ni}}$ by variation of the parameter θ and by variation of the TA (η_{ni}).

(5.6) LEMMA *Assume that $(P_{\theta\eta})$ satisfies conditions (5.1) and (5.2), and let (η_{ni}) be tight. Let $\theta_n = \theta + \frac{1}{\sqrt{n}} t$ and $\tau_{ni} = \eta_{ni} + \frac{1}{\sqrt{n}} \psi(\eta_{ni})$ for some $\psi \in \mathcal{C}_b(\Lambda)$. Then*

$$\sqrt{\frac{dP_{\theta_n\tau_{ni}}}{dP_{\theta\eta_{ni}}}} = 1 + \frac{1}{2\sqrt{n}} \Big(t l_{1\theta}(.,\eta_{ni}) + \psi(\eta_{ni}) l_{2\theta}(.,\eta_{ni}) \Big) + r_{ni},$$

where

$$\lim_{n\to\infty}\sum_{i=1}^{n}\int r_{ni}^2\,dP_{\theta\eta_{ni}}=0.$$

The proof of Lemma (5.6) is given in section 6. In view of Lemma (5.6) it is natural to call

$$h(\omega,\eta):=tl_{1\theta}(\omega,\eta)+\psi(\eta)l_{2\theta}(\omega,\eta)$$

the tangent vector of $(\bigotimes_{i=1}^n P_{\theta_n\tau_{ni}})$ with respect to $(\bigotimes_{i=1}^n P_{\theta\eta_{ni}})$. Let us denote the closed hull in $L_2^*(\mathbb{P}_\theta,\Gamma)$ of all such tangent vectors by

$$T_2(\theta,\Gamma):=\overline{\{tl_{1\theta}+\psi l_{2\theta}:t\in\mathbb{R},\psi\in L_2(\Gamma)\}}.$$

This is the tangent space of the model. If h satisfies condition (L) for (η_{ni}) then an LAN-expansion is valid.

(5.7) LEMMA *Assume that $(P_{\theta\eta})$ satisfies conditions (5.1) and (5.2), and that (η_{ni}) converges weakly to Γ. Let $\theta_n=\theta+\frac{1}{\sqrt{n}}t$ and $\tau_{ni}=\eta_{ni}+\frac{1}{\sqrt{n}}\psi(\eta_{ni})$ for some $\psi\in\mathcal{C}_b(\Lambda)$. Then*

$$\frac{d\bigotimes_{i=1}^n P_{\theta_n\tau_{ni}}}{d\bigotimes_{i=1}^n P_{\theta\eta_{ni}}}=\exp\Big(L_n(h)-\frac{1}{2}\int h^2\,dR_{\theta\Gamma}+r_n\Big),$$

where $h(\omega,\eta):=tl_{1\theta}\omega,\eta)+\psi(\eta)l_{2\theta}(\omega,\eta)$ and $r_n\overset{\bigotimes_{i=1}^n P_{\theta\eta_{ni}}}{\longrightarrow}0.$

The proof of Lemma (5.7) follows well-known patterns, (see Corollary 74.4 in Strasser, [17]).

Corollary (4.5) is based on an asymptotic invariance condition. We are now going to relate perturbation symmetry with such an invariance condition.

(5.8) LEMMA *Assume that (η_{ni}) converges weakly to Γ and consider the sequence of p-measures $(\bigotimes_{i=1}^n P_{\theta\eta_{ni}})$. If (κ_n) is a perturbation symmetric estimator sequence then $S_n=\sqrt{n}(\kappa_n-\theta)$ is asymptotically $V_{\theta\Gamma}$-invariant where*

$$V_{\theta\Gamma}:=\Big\{g\in L_2^*(\mathbb{P}_\theta,\Gamma):E_{\theta\Gamma}(g|\mathcal{A})=0\ R'_{\theta\Gamma}\text{-a.e.}\Big\}.$$

Proof: Apply Lemma (3.5) to Definition (3.6). □

With these preparations we almost are in a position to apply Corollary (4.5). For efficiency the estimator sequence $S_n=\sqrt{n}(\kappa_n-\theta)$ applied to the parameters $\theta_n=\theta+\frac{1}{\sqrt{n}}t$ has to be concentrated as much as possible around t. Therefore the linear function f to be estimated is

$$f(h):=t\quad\text{if }\ h:=(tl_{1\theta}+\psi l_{2\theta})\in T_2(\theta,\Gamma).$$

We only have to check the existence of a function

$$h_{\theta\Gamma}^{**}\in\mathcal{C}_b^*(\mathbb{P}_\theta)\cap\overline{\text{span}}\{V_{\theta\Gamma}\cup T_2(\theta,\Gamma)\}\cap V_{\theta\Gamma}^{\perp}$$

such that

$$\int h_{\theta\Gamma}^{**}(tl_{1\theta}+\psi l_{2\theta})\,dR_{\theta\Gamma}=t\quad\text{for all }t\in\mathbb{R},\ \psi\in L_2(\Gamma).$$

(This implies $T_2(\theta,\Gamma)\cap V_{\theta\Gamma}\subseteq\ker f$.) Then the assertion of Corollary (4.5) gives an asymptotic information bound for the functional model and perturbation symmetric estimator sequences.

But this is not the whole story. The Theorem (3.9) which is our actual goal does not only establish an asymptotic information bound but states that this bound equals the bound valid in the structure model. The following is the key theorem. It is based on our basic Lemmas (2.15) and (2.16).

(5.9) THEOREM *Suppose that $(P_{\theta\eta})$ is regular. Let $h^*_{\theta\Gamma}$ be the uniquely determined tangent vector in $T_1(\theta, \Gamma)$ such that*

$$\int h^*_{\theta\Gamma} E_{\theta\Gamma}(tl_{1\theta} + k|\mathcal{A})\, dR'_{\theta\Gamma} = t \quad \text{for all } t \in \mathbb{R},\ k \in L_2^*(\Gamma).$$

Then $h^*_{\theta\Gamma} \in \overline{\text{span}}\{V_{\theta\Gamma} \cup T_2(\theta, \Gamma)\} \cap V_{\theta\Gamma}^{\perp}$ and satisfies

$$\int h^*_{\theta\Gamma}(tl_{1\theta} + \psi l_{2\theta})\, dR_{\theta\Gamma} = t \quad \text{for all } t \in \mathbb{R},\ \psi \in L_2(\Gamma).$$

Proof: Since $h^*_{\theta\Gamma}$ is $\mathcal{A}$-measurable and $E_{\theta\eta}(h^*_{\theta\Gamma}) = 0$, $\eta \in \Lambda$, it follows by Lemma (2.16) that $h^*_{\theta\Gamma} \in V_{\theta\Gamma}^{\perp}$.

Let $h^*_{\theta\Gamma} = E_{\theta\Gamma}(t^* l_{1\theta} + k^*|\mathcal{A})$ for some $t^* \in \mathbb{R}$, $k^* \in L_2^*(\Gamma)$. By Lemma (2.15) there exists $\psi^* \in L_2(\Gamma)$ such that $E_{\theta\Gamma}(k^*|\mathcal{A}) = E_{\theta\Gamma}(\psi^* l_{2\theta}|\mathcal{A})$. It follows that

$$h^*_{\theta\Gamma} = E_{\theta\Gamma}(t^* l_{1\theta} + \psi^* l_{2\theta}|\mathcal{A}) = t^* l_{1\theta} + \psi^* l_{2\theta} + g^*,$$

with $g^* = (t^* l_{1\theta} + \psi^* l_{2\theta}) - E_{\theta\Gamma}(t^* l_{1\theta} + \psi^* l_{2\theta}|\mathcal{A}) \in V_{\theta\Gamma}$. Thus we have $h^*_{\theta\Gamma} \in \overline{\text{span}}\{T_2(\theta, \Gamma) \cup V_{\theta\Gamma}\}$.

Let $t \in \mathbb{R}$ and $\psi \in L_2(\Gamma)$. By Lemma (2.15) there is some $k \in L_2^*(\Gamma)$ such that $E_{\theta\Gamma}(k|\mathcal{A}) = E_{\theta\Gamma}(\psi l_{2\theta}|\mathcal{A})$. This implies

$$\begin{aligned} t &= \int h^*_{\theta\Gamma} E_{\theta\Gamma}(tl_{1\theta} + k|\mathcal{A})\, dR'_{\theta\Gamma} \\ &= \int h^*_{\theta\Gamma} E_{\theta\Gamma}(tl_{1\theta} + \psi l_{2\theta}|\mathcal{A})\, dR'_{\theta\Gamma} \\ &= \int h^*_{\theta\Gamma}(tl_{1\theta} + \psi l_{2\theta})\, dR_{\theta\Gamma}. \end{aligned}$$

Thus, $h^*_{\theta\Gamma}$ satisfies all claims. □

Proof: (of Theorem (3.9)) Apply Corollary (4.5) with the notation introduced above. □

6 Auxiliary Lemmas and Additional Proofs

Proof: (of Theorem (3.3).)

We apply a basic result on empirical processes which can be found in Shorack and Wellner, [16], chapter 25, section 4, Theorem 1, (p. 810). In view of this result it is sufficient for our proof to consider the limits of the first and second moments of the empirical processes.

Considering the first moments we have

$$\begin{aligned} &\int Z_n(x)\, d \bigotimes_{i=1}^{n} Q_{n\eta_{ni}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\int_{-\infty}^{x} \Big(1 + \frac{1}{\sqrt{n}} g(., \eta_{ni})\Big)\, dF_{\theta\eta_{ni}} \ -\ F_{\theta\eta_{ni}}(x) \Big] \\ &= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{x} g(., \eta_{ni})\, dF_{\theta\eta_{ni}} \\ &\to \int \int_{-\infty}^{x} g(., \eta)\, dF_{\theta\eta}\, \Gamma(d\eta) = \int_{-\infty}^{x} E_{\theta\Gamma}(g|\mathcal{A})\, dR'_{\theta\Gamma}. \end{aligned}$$

Considering the second moments we have

$$\begin{aligned}
&\int Z_n(x)Z_n(y)\, d\bigotimes_{i=1}^{n} Q_{n\eta_{ni}} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[\int_{-\infty}^{x\cap y}\Big(1+\frac{1}{\sqrt{n}}g(.,\eta_{ni})\Big)\,dF_{\theta\eta_{ni}} \;-\; F_{\theta\eta_{ni}}(x)F_{\theta\eta_{ni}}(y)\Big] \\
&\to \int\int_{-\infty}^{x\cap y} dF_{\theta\eta}\,\Gamma(d\eta) - \int F_{\theta\eta}(x)F_{\theta\eta}(y)\,\Gamma(d\eta).
\end{aligned}$$

□

Proof: (of Lemma (3.5) Let $g := g_1 - g_2$ and for $A \in \mathcal{A}$ let

$$f_A(\eta) := \int_A g(.,\eta)p_\theta(.,\eta)\, d\mu.$$

Then $(f_A)_{A\in\mathcal{A}}$ is a family of continuous functions which is uniformly bounded and equicontinuous on compacts. Weak convergence of (η_{ni}) together with the Theorem of Arzelà and Ascoli implies

$$\lim_{n\to\infty}\sup_A \Big|\frac{1}{n}\sum_{i=1}^{n} f_A(\eta_{ni}) - \int f_A\, d\Gamma\Big| = 0.$$

Straightforward reasoning gives

$$\lim_{n\to\infty}\int\Big|\frac{1}{n}\sum_{i=1}^{n} g(.,\eta_{ni})p_\theta(.,\eta_{ni}) \;-\; \int g(.,\eta)p_\theta(.,\eta)\, d\Gamma\Big|\, d\mu = 0.$$

The assertion follows since it is easy to see that

$$\int\Big|\frac{1}{n}\sum_{i=1}^{n} g(.,\eta_{ni})p_\theta(.,\eta_{ni})\Big|\, d\mu = \sqrt{n}\Big|\Big|\frac{1}{n}\sum_{i=1}^{n} Q^1_{n\eta_{ni}} - \frac{1}{n}\sum_{i=1}^{n} Q^2_{n\eta_{ni}}\Big|\Big|$$

and

$$\int\Big|\int g(.,\eta)p_\theta(.,\eta)\, d\Gamma\Big|\, d\mu = \int\Big|E_{\theta\Gamma}(g|\mathcal{A})\Big|\, dR'_{\theta\Gamma}.$$

□

(6.1) LEMMA *Suppose that h is $\mathcal{A}$-measurable, $\eta \mapsto E_{\theta\eta}(h^2)$ is bounded and continuous and $E_{\theta\eta}(h) = 0,\ \eta \in \Lambda$. Then $h \in \mathcal{C}_b^*(\mathbb{P}_\theta)$.*

Proof: The assertion follows since $\eta \mapsto h\sqrt{p_\theta(.,\eta)}$ is μ-continuous and $\eta \mapsto E_{\theta\eta}(h^2) = \int (h\sqrt{p_\theta(.,\eta)})^2\, d\mu$ is continuous. □

(6.2) LEMMA *$\mathcal{C}_b^*(\mathbb{P}_\theta)$ is dense in $L_2^*(\mathbb{P}_\theta,\Gamma)$ for every $\Gamma|\mathcal{B}(\Lambda)$.*

Proof: Let $C_b(\mathbb{P}_\theta)$ be the set of all measurable functions $g : \Omega\times\Lambda \to \mathbb{R}$ such that $\eta \mapsto g(.,\eta)\sqrt{p_\theta(.,\eta)}$ is a continuous and norm-bounded function from Λ to $L_2(\mu)$. It is obvious that $\mathcal{C}_b(\mathbb{P}_\theta)$ is dense in $L_2(R_{\theta\Gamma})$.

Let $h \in L_2^*(\mathbb{P}_\theta,\Gamma)$ and choose $g \in \mathcal{C}_b^*(\mathbb{P}_\theta)$ such that $||h - g||_{\theta\Gamma} < \epsilon/2$. Then it follows that $\int |E_{\theta\eta}(g(.,\eta))|\, d\Gamma < \epsilon/2$ which implies $||h - g^*||_{\theta\Gamma} < \epsilon$ for $g^* = g - E_{\theta\Gamma}(g(.,\eta))$. □

(6.3) LEMMA *Assume that regularity conditions (5.1) and (5.2) are satisfied. Every $g \in \mathcal{C}_b^*(\mathbb{P}_\theta)$ satisfies condition (L) for every tight TA (η_{ni}).*

Proof: By a similar proof as for Lemma (4.7) in [21] one shows that for every compact set $K \subseteq \Lambda$

$$\lim_{a\to\infty} \sup_{\eta\in K} \int_{|g(.,\eta)|>a} g^2(.,\eta)\, dP_{\theta\eta} = 0.$$

This implies the assertion. □

Proof: (of Lemma (5.6)) Let

$$r_n(\eta) := \sqrt{\frac{dP_{\theta+t/\sqrt{n},\eta+\psi(\eta)/\sqrt{n}}}{dP_{\theta\eta}}} - 1 - \frac{1}{2\sqrt{n}}\Big(tl_{1\theta}(.,\eta) + \psi(\eta)l_{2\theta}(.,\eta)\Big).$$

Thus, we have $r_{ni} = r_n(\eta_{ni})$.

By regularity condition (5.1) for every $\epsilon > 0$ there exists $\delta(\theta,\eta,\epsilon) > 0$ such that

$$\int r_n(\eta)^2\, dP_{\theta\eta} < \epsilon^2\Big(\frac{t^2}{n} + \frac{\psi(\eta)^2}{n}\Big) \qquad \text{if } \Big(\frac{t^2}{n} + \frac{\psi(\eta)^2}{n}\Big) < \delta^2(\theta,\eta,\epsilon).$$

Let $K \subseteq \Lambda$ be a compact set such that

$$\frac{1}{n}\sum_{i=1}^n 1_K(\eta_{ni}) > 1-\epsilon \quad \text{for all } n \in \mathbb{N}.$$

For $\eta \in K$ the number $\delta(\theta,\eta,\epsilon)$ may be replaced by a number $\delta_K(\theta,\epsilon)$, independent of $\eta \in K$. Thus, we have

$$\limsup_{n\to\infty} \sup_{\eta\in K} n \int r_n^2(\eta)\, dP_{\theta\eta} = 0.$$

Moreover, regularity condition (5.2) implies a Lipschitz condition for $\eta \mapsto \sqrt{p_\theta(.,\eta)}$ and thus

$$\limsup_{n\to\infty} \sup_{\eta\in\Lambda} n \int r_n^2(\eta)\, dP_{\theta\eta} \le C < \infty.$$

Putting terms together, we obtain

$$\begin{aligned}
&\limsup_{n\to\infty} \sum_{i=1}^n \int r_n^2(\eta_{ni})\, dP_{\theta\eta_{ni}} \\
&\quad\le \limsup_{n\to\infty} \frac{1}{n} \sum_{\eta_{ni}\in K} n \int r_n^2(\eta_{ni})\, dP_{\theta\eta_{ni}} \\
&\quad+ \limsup_{n\to\infty} \frac{1}{n} \sum_{\eta_{ni}\notin K} n \int r_n^2(\eta_{ni})\, dP_{\theta\eta_{ni}} \\
&\quad\le 0 + C\epsilon.
\end{aligned}$$

□

Proof: (of Lemma (2.15)).

Let us denote the sets in question by M_1 and M_2. Both sets are closed linear subspaces of $L_2(R'_{\theta\Gamma})$. We will show that $M_1^\perp = M_2^\perp$ by showing that

$$M_1^\perp = M_2^\perp = \{h \in L_2(R'_{\theta\Gamma}) : \eta \mapsto E_{\theta\eta}(h) \text{ is constant } \Gamma\text{–a.e.}\}$$

The assertion $h \in M_1^\perp$ is equivalent to

$$\int hk\, dR_{\theta\Gamma} = \int\int h(\omega)P_{\theta\eta}(d\omega)k(\eta)\Gamma(d\eta) = 0 \quad \text{if } k \in L_2^*(\Gamma).$$

This is the case iff $\eta \mapsto E_{\theta\eta}(h)$ is constant Γ–a.e.

The assertion $h \in M_2^\perp$ is equivalent to

$$\int h\psi l_{2\theta}\, dR_{\theta\Gamma} = \int \Big(\frac{d}{d\eta} \int h(\omega)\, P_{\theta\eta}(d\omega)\Big)\psi(\eta)\,\Gamma(d\eta) = 0 \quad \text{for all } \psi \in L_2(\Gamma).$$

This is also the case iff $\eta \mapsto E_{\theta\eta}(h)$ is constant Γ–a.e. □

Proof: (of Lemma (2.16))
Let $g \in L_2^*(\mathbb{P}_\theta, \Gamma)$. Then the condition

$$\int g(\omega,\eta)\Big(h(\omega) - E_{\theta\eta}(h)\Big)\, R_{\theta\Gamma}(d\omega, d\eta) = 0 \quad \text{for all } h \in L_2(R'_{\theta\Gamma})$$

is in view of $E_{\theta\eta}(g(.,\eta)) = 0$ equivalent to

$$\int g(\omega,\eta)h(\omega)\, R_{\theta\Gamma}(d\omega, d\eta) = 0 \quad \text{for all } h \in L_2(R'_{\theta\Gamma}).$$

This means $E_{\theta\Gamma}(g|\mathcal{A}) = 0$, thus $g \in V_{\theta\Gamma}$. □

References

[1] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins Univ. Press, 1993.

[2] P.J. Bickel and C.A.L. Klaassen. Empirical Bayes estimation in functional and structural models. *Advances in Applied Mathematics*, pages 55–69, 1986.

[3] C.R. Blyth. On minimax statistical procedures and their admissibility. *Annals of Math. Statistics*, 22:22–42, 1951.

[4] J. Hájek. *Theory of rank tests*. Academic Press, 1967.

[5] J. Hájek. Local asymptotic minimax and admissibility in estimation. In *Proc. 6th Berkeley Symp. Math. Stat. Prob.*, pages 175–194, 1972.

[6] L. LeCam. On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.*, 1:277–330, 1953.

[7] L. LeCam. Limits of experiments. In L.LeCam, editor, *Proc. 6th Berkeley Symp.*, pages 245–261, 1972.

[8] L. LeCam. On a theorem of Hájek. In J. Jurečkova, editor, *J. Hájek memorial volume*, pages 119–137. Akademia Prague, 1979.

[9] L. LeCam. *Asymptotic Methods in Statistical Decision Theory*. Springer, 1986.

[10] L. LeCam and G.L. Yang. *Asymptotics in Statistics*. Springer Series in Statistics. Springer, 1990.

[11] P.W. Millar. The minimax principle in asymptotic theory. In P.L.Hennequin, editor, *Ecole d'été der probabilités de Saint Flour ix–1981*, volume 976 of *Lecture Notes in Mathematics*, pages 76–267. Springer, 1983.

[12] J. Neyman and E. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32, 1948.

[13] J. Pfanzagl. *Estimation is semiparametric models*, volume 63 of *Lecture Notes in Statistics*. Springer–Verlag, 1990.

[14] J. Pfanzagl. Incidental versus random nuisance parameters. *Annals of Statistics*, 21:1663–1691, 1993.

[15] J. Pfanzagl and W. Wefelmeyer. *Contributions to a general asymptotic statistical theory*, volume 13 of *Lecture Notes in Statistics*. Springer, 1982.

[16] G.R. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. John Wiley, 1986.

[17] H. Strasser. *Mathematical theory of statistics: Statistical experiments and asymptotic decision theory*, volume 7 of *De Gruyter Studies in Mathematics*. de Gruyter, 1985.

[18] H. Strasser. Tangent vectors for models with independent but non–identically distributed observations. *Statistics and decisions*, 7:127–152, 1989.

[19] H. Strasser. Asymptotic admissibility and uniqueness of efficient estimates in semiparametric models. In D. Pollard and G. Yang, editors, *Research papers in probability and statistics: a Festschrift for Lucien LeCam*, chapter 20. 1994.

[20] H. Strasser. Asymptotic normality toolbox. Technical report, Department of Statistics, 1994.

[21] H. Strasser. Asymptotic efficiency of estimates for models with incidental nuisance parameters. *Annals of Statistics*, 24, 1996.

[22] A.W. van der Vaart. *Statistical estimation in large parameter spaces*, volume 44 of *C.W.I.Tracts*. Mathematical Centrum, 1988.