WORKING   PAPERS

Hannes Leeb
Benedikt M. Pötscher

Performance Limits for Estimators of the Risk or Distribution of
Shrinkage-Type Estimators, and Some General Lower Risk-Bound Results

October 2002

Working Paper No: 0301

DEPARTMENT OF ECONOMICS

UNIVERSITY   OF   VIENNA

Performance Limits for Estimators of the Risk or Distribution of

Shrinkage-Type Estimators, and Some General Lower Risk-Bound Results[*]

Hannes Leeb
Department of Statistics, Yale University

and

Benedikt M. Pötscher
Department of Statistics, University of Vienna

Preliminary version: January 2002

This version: October 2002

Address for correspondence:

Benedikt M. Pötscher, Department of Statistics, University of Vienna,
Universitätsstrasse 5, A-1010 Vienna, Austria.   Phone: (+43 1)427738640.
E-mail: Benedikt.Poetscher@univie.ac.at.

[*]A preliminary draft of the results in Section 3 of this paper was already
written in 1999.

<u>Summary</u>: We consider the problem of estimating measures of precision of

shrinkage-type estimators like their risk or distribution. The notion of

shrinkage-type estimators here refers to estimators like the James-Stein

estimator or Lasso-type estimators, as well as to "thresholding" estimators

such as, e.g., Hodges' so-called superefficient estimator. While the

precision measures of such estimators typically can be estimated consistently,

we show that they cannot be estimated uniformly consistently (even locally).

This follows as a corollary to (locally) uniform lower bounds on the

performance of estimators of the precision measures that we obtain in the

paper. These lower bounds are typically quite large (e.g., they approach 1/2

or 1 depending on the situation considered). The analysis is based on some

general lower risk bounds and related general results on the (non)existence of

uniformly consistent estimators also obtained in the paper.

1. Introduction and Overview

In virtually any statistical application one is not only interested in a point-estimate of an unknown parameter by itself, but also in a measure of the precision of the estimate. For example, one would like to obtain information on the risk or on the distribution of the estimator. As these quantities often depend on unknown parameters, one frequently has to be content with estimates of the risk or the distribution. In the present paper we are concerned with the estimation of the risk or the distribution of shrinkage-type estimators. Under the umbrella of shrinkage-type estimators we subsume well-known estimators like the James-Stein estimator or Hodges' superefficient estimator, as well as penalized maximum likelihood (least squares) estimators including recent proposals like the Lasso-type and Bridge estimators (cf. Tibshirani (1996), Knight and Fu (2000), Frank and Friedman (1993)) or the SCAD-estimators recently introduced by Fan and Li (2001). Furthermore, any estimator based on some sort of a (soft or hard) thresholding rule falls into this category.

While it is typically not difficult to construct consistent estimators for the risk or the distribution of shrinkage-type estimators, the results we obtain imply that any such estimator of the risk or the distribution necessarily has very poor performance in the sense that the maximum probability of the estimation error exceeding a certain threshold remains large even in large samples. To illustrate, we now sketch a simple special case of the results in Section 2. Let $\hat{\theta}$ be a shrinkage-type estimator (as considered in Sections 2.2-2.3) for a parameter vector $\theta \in \mathbb{R}^k$, and let $F_{n,\theta}(t) = P_{n,\theta}(n^{1/2}(\hat{\theta}-\theta) \leq t)$ be its cdf at sample size n. Then there exists a $\delta > 0$ such that for any consistent estimator $\hat{F}_n(t)$ of $F_{n,\theta}(t)$ the relation

$$\sup_{||\theta||<M} P_{n,\theta}(|\hat{F}_n(t) - F_{n,\theta}(t)|>\delta) \to 1 \qquad (1.1)$$

holds for $n \to \infty$. That is, while the probability $P_{n,\theta}(|\hat{F}_n(t) - F_{n,\theta}(t)|>\delta)$

1

converges to zero for every given θ by consistency, relation (1.1) shows that it does not do so uniformly in θ. It follows that $\hat{F}_n(t)$ can never be uniformly consistent (not even when restricting consideration to compact subsets of the parameter space). Hence, a large sample size does not guarantee a small estimation error with high probability when estimating the risk or distribution of shrinkage-type estimators. As a consequence, reliably assessing the precision of shrinkage-type estimators constitutes an intrinsically hard (if not unsolvable) problem. It is interesting to note that the non-uniformity phenomena like (1.1) arise near the origin which is precisely that region of the parameter space where shrinkage-type estimators typically have the greatest advantage over the maximum likelihood estimator. Apart from results like (1.1), we also provide minimax lower bounds for the performance of arbitrary (not necessarily consistent) estimators of the risk or the distribution of shrinkage-type estimators. For example, we show that there exists a δ>0 such that

$$\liminf_{n \to \infty} \inf_{\hat{F}_n(t)} \sup_{||\theta||<M} P_{n,\theta}(|\hat{F}_n(t) - F_{n,\theta}(t)| > \delta) \geq c > 0 \qquad (1.2)$$

holds, where the lower bound c is typically large and can be computed. In fact, we show that in (1.1)-(1.2) the balls $||\theta||<M$ can be replaced by (suitable) balls shrinking at the rate $n^{-1/2}$. For related results in the context of model selection see Leeb and Pötscher (2002).

The results on shrinkage-type estimators mentioned above share a common mathematical structure which is investigated in a more abstract framework in Section 3. The results in that section are roughly as follows: We consider an (abstract) set B of parameters indexing a set $\{P_{n,\beta} : \beta \in B\}$ of probability measures, and a functional of interest, $\varphi_n : B \to \mathbb{R}$ say, that is to be estimated. Note that we need to allow the functional φ_n to depend on sample size in order to be able to subsume the results relating to shrinkage-type estimators mentioned earlier. (This feature also allows the results of Section 3 to be

applied to the derivation of bounds on convergence rates in general, cf. Remark 3.5(v)). We then provide conditions on the estimand $\varphi_n(\beta)$ and the set of probability measures implying the existence of a $\delta>0$ such that any consistent estimator $\hat{\varphi}_n$ of $\varphi_n(\beta)$ satisfies

$$\sup_{\beta \in B} P_{n,\beta}(|\hat{\varphi}_n - \varphi_n(\beta)| > \delta) \to 1 \qquad (1.3)$$

for $n \to \infty$. This relation implies nonexistence of uniformly consistent estimators for $\varphi_n(\beta)$, but certainly is much stronger. The results we obtain in Section 3 are actually more general than (1.3) as the supremum over B in (1.3) may be replaced by a supremum over suitably "shrinking" subsets B_n, and φ_n may take values in a metric space. Furthermore, we also provide -- analogously to (1.2) -- minimax lower bounds for the performance of arbitrary (not necessarily consistent) estimators.

The remainder of the paper is organized as follows: Estimation of the risk of the James-Stein estimator is discussed in Section 2.1. Sections 2.2-2.3 treat estimation of the finite-sample distribution of Lasso-type estimators and of Hodges' superefficient estimator, respectively; the latter is analyzed as a prototypical yet simple instance of an estimator that is obtained by "hard thresholding" and that has the so-called "oracle property" in the sense of Fan and Li (2001). The techniques used for establishing these results are the subject of Section 3. In this section we study in an abstract framework lower bounds for the performance of estimators and, in particular, the existence/nonexistence of uniformly consistent estimators for sample-size-dependent functionals of interest. All proofs are relegated to appendices.

Some words on notation: The cumulative distribution function (cdf) of a normally distributed random vector with mean zero and covariance matrix Σ is denoted by Φ_Σ, while the cdf of a standard normally distributed random variable is denoted by Φ as usual. The transpose of a matrix A is denoted by A'. The largest eigenvalue of a symmetric matrix A is represented by $\lambda_{max}(A)$.

The symbol $\|.\|$ is used as a generic symbol for a norm on a vector space (e.g., in Section 3), or to denote the Euclidean norm (e.g., in Section 2).

2. Lower Bounds for the Estimation of the Risk or of the Distribution of Shrinkage-Type Estimators

In this section we provide lower bounds on the performance of estimators of the risk or of the distribution of shrinkage-type estimators. In particular, it follows that these objects cannot be estimated uniformly consistently, not even over (suitable) shrinking subsets of the parameter space (and hence a fortiori not over all compacta). This is in stark contrast to the situation where the least squares estimator is employed instead. We furthermore note that all results in this section also hold for randomized estimators, cf. Lemma 3.6. All results in this section are derived for Gaussian models. However, this is not really a restriction, since the results a fortiori also hold for any more general statistical model containing the Gaussian models of Sections 2.1-2.3, respectively.

2.1 Estimating the Risk of the James-Stein Estimator

In order to assess the actual variability of the James-Stein estimator an estimator of its risk is desirable. The problem of estimating the risk of the James-Stein estimator has been considered, e.g., in Jennrich and Oman (1986), Sen (1986), Adkins (1990, 1992), and Venter and Steel (1990). (A related strand of literature treats estimation of the loss; see Lu and Berger (1989), Wan and Zou (2002), and the references in these papers.) For example, Sen (1986) shows that various jackknife estimators for the risk do not work properly near the origin. In the following we establish that this is in fact true for any estimator of the risk of the James-Stein estimator.

Consider observations Y_i, $i \geq 1$, that are independent and identically distributed as $N(\theta, I_k)$, where I_k denotes the k-dimensional identity matrix. Let $P_{n,\theta}$ denote the distribution of the sample $Y_1, \ldots, Y_n$ of size n, and let $E_{n,\theta}$ denote the corresponding expectation. Given an estimator $\hat{\theta}$ for θ we consider the quadratic risk

$$R_n(\hat{\theta}; \theta) = nE_{n,\theta}(\hat{\theta}-\theta)'(\hat{\theta}-\theta).$$

It is well-known that for $k \geq 3$ the James-Stein estimator $\hat{\theta}_{JS}$ given by

$$\hat{\theta}_{JS} = (1-(k-2)/(n\hat{\theta}'_{ML}\hat{\theta}_{ML}))\hat{\theta}_{ML}$$

dominates the maximum likelihood estimator $\hat{\theta}_{ML} = n^{-1}\sum_{i=1}^{n}Y_i$; more precisely, $R_n(\hat{\theta}_{JS}; \theta) < R_n(\hat{\theta}_{ML}; \theta) = k$ holds for every $\theta \in \mathbb{R}^k$ and every $n \geq 1$. This follows immediately from the well-known result

$$R_n(\hat{\theta}_{JS}; \theta) = k-(k-2)^2\exp(-n\theta'\theta)\sum_{j=0}^{\infty}(n\theta'\theta)^j(j!(k-2+2j))^{-1}, \qquad (2.1)$$

cf., e.g., Judge and Bock (1978), eqs. (8.3.6), (8.3.7). The improvement of $\hat{\theta}_{JS}$ over $\hat{\theta}_{ML}$ is substantial when the true parameter θ is close to zero. Indeed, if $n^{1/2}\theta$ is small, then (2.1) shows that $R_n(\hat{\theta}_{JS}; \theta)$ is close to $R_n(\hat{\theta}_{JS}; 0)=2$, which is substantially smaller than $R_n(\hat{\theta}_{ML}; \theta)=k$. For large θ the risk of $\hat{\theta}_{JS}$ is close to the risk of $\hat{\theta}_{ML}$; in fact, $R_n(\hat{\theta}_{JS}; \theta)$ converges to k for every $\theta \neq 0$ and $n \longrightarrow \infty$. (To see this, observe that $\exp(-n\theta'\theta)\sum_{j=m+1}^{\infty}(n\theta'\theta)^j(j!(k-2+2j))^{-1} \leq (k+2m)^{-1}$. This inequality furthermore also shows that that $R_n(\hat{\theta}_{JS}; \theta)$ is continuous in θ.)

Turning to estimation of the risk $R_n(\hat{\theta}_{JS}; \theta)$, it is easy to construct consistent estimators for $R_n(\hat{\theta}_{JS}; \theta)$: E.g., estimate the risk by 2 if $||\hat{\theta}_{ML}||<c_n$ and by k if $||\hat{\theta}_{ML}|| \geq c_n$, where c_n is a sequence of positive numbers satisfying $c_n \longrightarrow \infty$ and $c_n = O(n^{-1/2})$. Obviously this estimator is not very useful and will perform poorly in finite samples when $n^{1/2}\theta$ is close to but different from zero. The poor performance of this estimator is not accidental, but is a genuine feature of the estimation problem and affects <u>any</u> estimator of the risk as the following result shows. For convenience we introduce for $k \geq 3$ the function $\Delta(x) = 2^{-1}(k-2)[1-(k-2)\exp(-x^2)\sum_{j=0}^{\infty}x^{2j}(j!(k-2+2j))^{-1}]$. It is easy to

see that Δ is continuous and strictly increasing for $x \geq 0$, satisfies $\Delta(0) = 0$ and $\lim_{x \to \infty} \Delta(x) = (k-2)/2$.

<u>Theorem 2.1</u>: Suppose $k \geq 3$. Then any consistent estimator $\hat{R}_n$ for the risk $R_n(\hat{\theta}_{JS}; \theta)$ of the James-Stein estimator satisfies

$$\liminf_{n \to \infty} \sup_{||\theta|| < \rho n^{-1/2}} P_{n,\theta}(|\hat{R}_n - R_n(\hat{\theta}_{JS}; \theta)| > \delta) = 1 \qquad (2.2)$$

for every pair (δ, ρ) with $0 < \rho < \infty$ and $\delta < \Delta(\rho)$. Furthermore,

$$\inf_{n \geq 1} \inf_{\hat{R}_n} \sup_{||\theta|| < \rho n^{-1/2}} P_{n,\theta}(|\hat{R}_n - R_n(\hat{\theta}_{JS}; \theta)| > \delta) \geq 1 - \Phi(\Delta^{-1}(\delta)) \geq 1 - \Phi(\rho) > 0$$

$$(2.3)$$

holds for every pair (δ, ρ) with $0 < \rho < \infty$ and $\delta < \Delta(\rho)$, where the inner infimum extends over <u>all</u> estimators $\hat{R}_n$ of the risk $R_n(\hat{\theta}_{JS}; \theta)$. Moreover, for every $\rho > 0$

$$\sup_{\delta > 0} \inf_{n \geq 1} \inf_{\hat{R}_n} \sup_{||\theta|| < \rho n^{-1/2}} P_{n,\theta}(|\hat{R}_n - R_n(\hat{\theta}_{JS}; \theta)| > \delta) \geq 1/2 \qquad (2.4)$$

holds.

The above theorem implies that the risk of the James-Stein estimator is difficult to estimate; in particular, it can not be estimated uniformly consistently over balls around the origin. It is interesting to note that this difficulty arises precisely in that region of the parameter space where the James-Stein estimator has the biggest advantage over the maximum likelihood estimator in terms of risk.

<u>Remark 2.1</u>: The range of δ for which (2.2)-(2.3) hold for <u>some</u> $\rho > 0$ is given by $\delta < (k-2)/2$. This is quite natural, since for $\delta \geq (k-2)/2$ the trivial estimator $\hat{R}_n = (k+2)/2$ satisfies $\sup_{\theta \in \mathbb{R}^k} P_{n,\theta}(|\hat{R}_n - R_n(\hat{\theta}_{JS}; \theta)| > \delta) = 0$ for every $n \geq 1$.

2.2 <u>Estimating</u> <u>the</u> <u>Cdf</u> <u>of</u> <u>Lasso-Type</u> <u>Estimators</u>

Consider the linear model $Y_i = X_i \theta + u_i$ for $i \geq 1$, where X_i is a non-stochastic

1×k vector of regressors (k≥1), θ is an unknown k×1 parameter vector, and the errors u_i are i.i.d. normal with mean zero and unit variance. Furthermore, assume that $n^{-1}\sum_{i=1}^{n} X_i'X_i$ converges to a finite positive-definite matrix Q. For $\theta \in \mathbb{R}^k$, let $P_{n,\theta}$ denote the distribution of the sample $(Y_1,\dots,Y_n)'$. We are interested in the estimator $\hat{\theta}_L$ that is obtained by minimizing the penalized least-squares criterion

$$\sum_{i=1}^{n}(Y_i - X_i\theta)^2 + \lambda_n\sum_{j=1}^{k}|\theta_j| \qquad (2.5)$$

over $\theta \in \mathbb{R}^k$. The constants λ_n are assumed to be positive and to satisfy $n^{-1/2}\lambda_n \to \lambda$ with $0<\lambda<\infty$ as $n \to \infty$. (The solution to the minimization problem (2.5) is unique provided $\sum_{i=1}^{n} X_i'X_i$ is positive-definite as (2.5) is then strictly convex. Hence, except possibly for finitely many n, the estimator $\hat{\theta}_L$ is well-defined; it is also measurable as is easily seen.) This estimator is referred to as a Lasso-type estimator in Knight and Fu (2000), since it is closely related to the Lasso of Tibshirani (1996). It is a member of the class of Bridge estimators introduced by Frank and Friedman (1983). Knight and Fu (2000) also note that in the context of wavelet regression minimizing (2.5) is known as "basis pursuit", cf. Chen, Donoho, and Saunders (1999). In fact, in case $\sum_{i=1}^{n} X_i'X_i$ is diagonal, the Lasso-type estimator reduces to soft-thresholding of the coordinates of the maximum likelihood (least squares) estimator.

Knight and Fu (2000) have studied the asymptotic distribution of the Lasso-type estimator $\hat{\theta}_L$. Let $F_{n,\theta}$ denote the cdf of $n^{1/2}(\hat{\theta}_L - \theta)$, i.e.,

$$F_{n,\theta}(t) = P_{n,\theta}(n^{1/2}(\hat{\theta}_L - \theta) \le t),$$

and consider local perturbations $\theta_n = \theta + \nu n^{-1/2}$ of θ. Then Knight and Fu (2000) show that F_{n,θ_n} converges weakly to a cdf $F_{\infty,\theta,\nu}$ which is given by

$$F_{\infty,\theta,\nu}(t) = P(\text{argmin}_{u \in \mathbb{R}^k} V_{\theta,\nu}(u) \le t). \qquad (2.6)$$

Here the random function $V_{\theta,\nu}(u)$ is given by

$$V_{\theta,\nu}(u) = -2u'W + u'Qu + \lambda\sum_{j:\theta_j \ne 0}u_j\,\text{sgn}(\theta_j) + \lambda\sum_{j:\theta_j = 0}|u_j + \nu_j|, \qquad (2.7)$$

where W is a k-dimensional normally distributed random vector with mean zero and covariance matrix Q. (The argmin of $V_{\theta,\nu}$ is unique by strict convexity; it is clearly also measurable.) This result is established in Theorem 5(a) of Knight and Fu (2000) under assumptions slightly different from ours. Inspection of the proof, however, reveals that the result continues to hold under the assumptions of this subsection.

More explicit expressions for the asymptotic distribution (2.6) can be obtained in special cases and will be useful later. For example, if k=1 simple but tedious calculations show that for $\theta \neq 0$

$$F_{\infty,\theta,\nu}(t) = \Phi_{Q^{-1}}(t-\lambda Q^{-1}\mathrm{sgn}(\theta)/2), \qquad (2.8)$$

where $\Phi_{Q^{-1}}$ denotes the cdf of a $N(0,Q^{-1})$ distribution. For $\theta=0$ one obtains

$$F_{\infty,0,\nu}(t) = \Phi_{Q^{-1}}(t+\lambda Q^{-1}\mathrm{sgn}(t+\nu)/2) \qquad (2.9)$$

if $t \neq -\nu$ and

$$F_{\infty,0,\nu}(-\nu) = \Phi_{Q^{-1}}(-\nu+\lambda Q^{-1}/2). \qquad (2.10)$$

In particular, $F_{\infty,0,\nu}$ puts mass $\Phi_{Q^{-1}}(-\nu+\lambda Q^{-1}/2)-\Phi_{Q^{-1}}(-\nu-\lambda Q^{-1}/2)$ at $t=-\nu$. If k>1 and $Q=\mathrm{diag}(q_{jj})$ is diagonal, the components of the Lasso-type estimators are asymptotically independent, since minimizing (2.7) can be done coordinatewise and the components of W are independent. Explicit formulas for $F_{\infty,\theta,\nu}(t)$ then immediately follow from (2.8)-(2.10):

$$F_{\infty,\theta,\nu}(t) = \prod_{j=1}^{k} F_{\infty,\theta_j,\nu_j}(t_j), \qquad (2.11)$$

where the marginal cdfs $F_{\infty,\theta_j,\nu_j}(t_j)$ are given by (2.8)-(2.10) with q_{jj}^{-1} taking the rôle of Q^{-1}. Note that $F_{\infty,\theta,\nu}(t)$ puts positive mass on the hyperplanes $t_j=-\nu_j$ for which $\theta_j=0$. In the case k>1 and Q not necessarily diagonal, it is also easy to see that $F_{\infty,\theta,\nu}(t)$ is again given by (2.8) provided that θ has only non-zero coordinates and $\mathrm{sgn}(\theta)$ is interpreted as the column vector with i-th coordinate equal to $\mathrm{sgn}(\theta_i)$. However, if some coordinates of θ are zero, the asymptotic distribution is not know explicitly for non-diagonal Q. (We conjecture that, similar to the case of diagonal Q, it puts mass on the above

mentioned hyperplanes and that $F_{\infty,\theta,\nu}(t)$ coincides with the cdf of a normal distribution for t not falling on any of these hyperplanes where the mean of the normal distribution depends on the position of t relative to the hyperplanes). Cf. the discussion following Theorem 3 in Knight and Fu (2000).

Suppose now we are interested in estimating the finite-sample cdf of the Lasso-type estimator $\hat{\theta}_L$. Knight and Fu (2000) propose a bootstrap procedure for this purpose. They show (under their assumptions) that the proposed procedure provides an asymptotically valid approximation to the finite-sample cdf $F_{n,\theta}$ for any θ that has only non-zero coordinates, but fails to work otherwise. Knight and Fu (2000, p.1371) then sketch a modification of the bootstrap procedure that may work asymptotically for all θ, but also express doubts about the usefulness of this modification in finite samples, especially if some coordinates of θ are zero or close to zero. (An alternative estimator for the finite-sample cdf of the Lasso-type estimator $\hat{\theta}_L$ that is consistent is sketched in Remark 2.2 below. The same caveat regarding its finite-sample merits in case some coordinates of θ are zero or close to zero applies also to this estimator.) That such doubts are well-founded not only for the particular procedure suggested in Knight and Fu (2000), but in fact for <u>any</u> estimator of the finite-sample distribution of the Lasso-type estimator $\hat{\theta}_L$, transpires from the following theorems. For simplicity, we shall use in the following -- for given $t \in \mathbb{R}^k$ -- the notation $\hat{F}_n(t)$ to denote an <u>arbitrary</u> estimator of $F_{n,\theta}(t)$. This notation should not be taken as implying that the estimator is obtained by evaluating an estimated cdf at the argument t, or that it is constrained to lie between zero and one. For the remainder of this subsection let $\Delta(t)$ be given by

$$\Delta(t) = 2^{-1}\max\{\Phi_Q-1(t-\lambda Q^{-1}e/2)-\Phi_Q-1(t-\lambda Q^{-1}f/2): e,f\in\{-1,1\}^k\},$$

which is always nonnegative.

<u>Theorem 2.2</u>: Suppose $t \in \mathbb{R}^k$ is given and assume that $\Delta(t)$ is positive. Then

for every $\delta < \Delta(t)$ there exists a $\rho > 0$, which depends only on δ and t, such that any consistent estimator $\hat{F}_n(t)$ of $F_{n,\theta}(t)$ satisfies

$$\liminf_{n \to \infty} \sup_{||\theta|| < \rho n^{-1/2}} P_{n,\theta}(|\hat{F}_n(t) - F_{n,\theta}(t)| > \delta) = 1. \qquad (2.12)$$

Moreover, for every pair (δ, ρ) as above

$$\liminf_{n \to \infty} \inf_{\hat{F}_n(t)} \sup_{||\theta|| < \rho n^{-1/2}} P_{n,\theta}(|\hat{F}_n(t) - F_{n,\theta}(t)| > \delta) \geq 1 - \Phi(\rho \lambda_{max}^{1/2}(Q))$$

$$(2.13)$$

holds, where the infimum extends over <u>all</u> estimators $\hat{F}_n(t)$ of $F_{n,\theta}(t)$.

Furthermore, there exists a $\rho > 0$, which depends only on t, such that

$$\sup_{\delta > 0} \liminf_{n \to \infty} \inf_{\hat{F}_n(t)} \sup_{||\theta|| < \rho n^{-1/2}} P_{n,\theta}(|\hat{F}_n(t) - F_{n,\theta}(t)| > \delta) \geq 1/2$$

$$(2.14)$$

holds, where again the infimum extends over <u>all</u> estimators $\hat{F}_n(t)$ of $F_{n,\theta}(t)$.

The above theorem maintains the assumption that $\Delta(t)$ is positive. For $k=1$ or, more generally, for diagonal Q this assumption is obviously always satisfied. For non-diagonal Q it will be satisfied in most cases. For example, it is satisfied for all $t \in \mathbb{R}^k$ if the matrix Q^{-1} possesses at least one column with only nonnegative entries (to see this, represent this column as $Q^{-1}e$, set $f=-e$, and exploit monotonicity of $\Phi_{Q^{-1}}$). More generally, $\Delta(t)$ is positive for all $t \in \mathbb{R}^k$ if Q^{-1} admits a linear combination of its columns with weights ± 1 such that this linear combination has only nonnegative entries.

For diagonal Q we obtain a result stronger than (2.13) and (2.14) in Theorem 2.3 below. We believe that a similar result could also be established for non-diagonal Q. We also conjecture that for general Q results similar to Theorems 2.2 and 2.3 hold not only over shrinking neighborhoods of zero but of any parameter value that has at least one zero coordinate. This is quite straightforward to prove for diagonal Q, but establishing these results for the non-diagonal case would require a more detailed analysis of the structure of $F_{\infty,\theta,\nu}$, which is beyond the scope of the paper.

<u>Theorem</u> <u>2.3</u>: Suppose Q is diagonal and t∈ℝ is given. Then

$$\liminf_{n\to\infty} \inf_{\hat{F}_n(t)} \sup_{||\theta||<\rho n^{-1/2}} P_{n,\theta}(|\hat{F}_n(t) - F_{n,\theta}(t)|>\delta) \geq 1/2 \quad (2.15)$$

holds for every $\delta<[\Phi_Q^{-1}(t+\lambda Q^{-1}/2)-\Phi_Q^{-1}(t-\lambda Q^{-1}/2)]/2$ and for every $\rho>||t||$.

Here again the infimum in (2.15) extends over <u>all</u> estimators $\hat{F}_n(t)$ of $F_{n,\theta}(t)$.

<u>Remark</u> <u>2.2</u>: An estimator $\hat{F}_n(t)$ that is consistent for $F_{n,\theta}(t)$ at least if t has only non-zero coordinates can be obtained quite easily from the asymptotic distribution. We sketch such a construction in the following. (Of course, this estimator is also subject to the criticism expressed in Theorems 2.2 and 2.3 above.) Apply a (strongly) consistent model selection procedure (e.g., Schwarz' Bayesian information criterion or a collection of hypothesis tests with suitable sample-size dependent significance levels) to determine which coordinates of θ are zero and which are not. For the latter coordinates estimate $\text{sgn}(\theta_j)$ by $\text{sgn}(\tilde{\theta}_j)$, where $\tilde{\theta}$ is a (strongly) consistent estimator for θ (e.g., least-squares). Plug this information into the formula (2.7) for $V_{\theta,0}$, also replacing Q by $n^{-1}\sum_{i=1}^n X_i'X_i$, W by $W_n\sim N(0,n^{-1}\sum_{i=1}^n X_i'X_i)$, and λ by $n^{-1/2}\lambda_n$, with W_n chosen independently of the data. The resulting function $\hat{V}_n$ now converges weakly to $V_{\theta,0}$ for every $u\in\mathbb{R}^k$ and for almost every realization of the data. By convexity it follows that $\text{argmin}_{u\in\mathbb{R}^k}\hat{V}_n(u)$ converges weakly to $\text{argmin}_{u\in\mathbb{R}^k}V_{\theta,0}(u)$ almost surely. Define the estimator $\hat{F}_n(t) = P(\text{argmin}_{u\in\mathbb{R}^k}\hat{V}_n(u)\leq t)$, where P represents the probability measure governing W. It follows that $\hat{F}_n(t)$ converges almost surely to $F_{\infty,\theta,0}(t)$ for all t for which the latter cdf is continuous, and hence that $\hat{F}_n(t)-F_{n,\theta}(t) \to 0$ almost surely for such values of t. Lemma A.3 shows that any t with only non-zero coordinates is a continuity point of $F_{\infty,\theta,0}(t)$ for all $\theta\in\mathbb{R}^k$. (Whether or not $\hat{F}_n(t)$ is consistent for all $t\in\mathbb{R}^k$ is an open (but in light of the above results somewhat moot) question. Note that Lemma A.3 also shows that $\hat{F}_n(t)-F_{n,\theta}(t) \to$

0 almost surely for all $t \in \mathbb{R}^k$ provided θ has only non-zero coordinates.)

2.3 Estimating the Cdf of a Hard Thresholding Estimator

In this subsection we are interested in estimating the finite-sample cdf of Hodges' so-called "superefficient" estimator. (In fact, we shall consider a slightly more general class of estimators.) We focus on Hodges' estimator mainly because it is a simple instance of a "thresholding" estimator, a class of estimators having gained prominence in the wavelet literature in recent years. It should perhaps be noted though that such estimators have already been studied earlier under the heading "pre-test" estimators in the statistics and econometrics literature (e.g., Judge and Bock (1978), Bauer, Pötscher, and Hackl (1988)).

Consider observations Y_i, $i \geq 1$, that are independent and identically distributed as $N(\theta, 1)$. Let $P_{n,\theta}$ denote the distribution of the sample $Y_1, \ldots, Y_n$ of size n. The class of estimators considered is then given by $\hat{\theta}_H = 0$ if $|\bar{Y}| \leq n^{-1/2} c_n$ and by $\hat{\theta}_H = \bar{Y}$ if $|\bar{Y}| > n^{-1/2} c_n$, where $\bar{Y}$ represents the arithmetic mean of the sample and c_n satisfies $c_n > 0$, $c_n \to \infty$ and $n^{-1/2} c_n \to 0$ for $n \to \infty$. Hodges' estimator (in its simplest form) corresponds to the case $c_n = n^{1/4}$. It is well-known and easy to see that $P_{n,\theta}(\hat{\theta}_H = 0) \to 1$ if $\theta = 0$ and that $P_{n,\theta}(\hat{\theta}_H = \bar{Y}) \to 1$ if $\theta \neq 0$ as $n \to \infty$. This immediately implies that the asymptotic distribution of $\hat{\theta}_H$ is pointmass at zero if $\theta = 0$ and coincides with the asymptotic distribution of $\bar{Y}$, the (unrestricted) maximum likelihood estimator, otherwise. This property of the asymptotic distribution is often referred to as "superefficiency" of the estimator.

Let $G_{n,\theta}$ denote the finite-sample distribution of the (scaled and centered) estimator $\hat{\theta}_H$, i.e., $G_{n,\theta}(t) = P_{n,\theta}(n^{1/2}(\hat{\theta}_H - \theta) \leq t)$ for $t \in \mathbb{R}$. It is easy to see that $G_{n,\theta}(t)$ is given by $G_{n,\theta}(t) = \Phi(t)$ if $t < -c_n - n^{1/2}\theta$ or $t > c_n - n^{1/2}\theta$, by $G_{n,\theta}(t) = \Phi(-c_n - n^{1/2}\theta)$ for $-c_n - n^{1/2}\theta \leq t < -n^{1/2}\theta$, and by $G_{n,\theta}(t) = \Phi(c_n - n^{1/2}\theta)$ for

$-n^{1/2}\theta \leq t \leq c_n - n^{1/2}\theta$. From the previous paragraph we know that $G_{n,\theta}$ converges weakly to $G_{\infty,\theta}$ for every $\theta \in \mathbb{R}$, where $G_{\infty,\theta}(t) = \Phi(t)$ if $\theta \neq 0$ and $G_{\infty,\theta}(t) = 1(0 \leq t)$ if $\theta = 0$. The formulas for the finite-sample cdf even show that $G_{n,\theta}(t)$ converges to $G_{\infty,\theta}(t)$ for every $t \in \mathbb{R}$ and $\theta \in \mathbb{R}$. It is now easy to construct a consistent estimator for $G_{n,\theta}(t)$ for any given $t \in \mathbb{R}$: Put $\hat{G}_n(t) = \Phi(t)$ if $\hat{\theta}_H \neq 0$; otherwise define $\hat{G}_n(t) = 1(0 \leq t)$. Alternatively, one can obtain a consistent estimator for $G_{n,\theta}(t)$ by plugging $\hat{\theta}_H$ into the formula for $G_{n,\theta}(t)$ given above. While both estimators are consistent, it is quite obvious that their performance will be poor if θ is close to but different from zero. That this is not only a property of these two particular estimators, but is a genuine feature of the estimation problem, is shown in the following theorem.

<u>Theorem</u> <u>2.4</u>: Suppose $t \in \mathbb{R}$ is given. Then any consistent estimator $\hat{G}_n(t)$ of $G_{n,\theta}(t)$ satisfies

$$\liminf_{n \to \infty} \sup_{|\theta| < \rho n^{-1/2}} P_{n,\theta}(|\hat{G}_n(t) - G_{n,\theta}(t)| > \delta) = 1 \qquad (2.16)$$

for every $\delta < 1/2$ and every $\rho > |t|$. Moreover, for every $n \geq 1$

$$\inf_{\hat{G}_n(t)} \sup_{|\theta| < \rho n^{-1/2}} P_{n,\theta}(|\hat{G}_n(t) - G_{n,\theta}(t)| > \delta) \geq 1/2 \qquad (2.17)$$

holds for every $\delta < [\Phi(c_n + t) - \Phi(-c_n + t)]/2$ and every $\rho > |t|$, where the infimum extends over <u>all</u> estimators $\hat{G}_n(t)$ of $G_{n,\theta}(t)$. In particular,

$$\liminf_{n \to \infty} \inf_{\hat{G}_n(t)} \sup_{|\theta| < \rho n^{-1/2}} P_{n,\theta}(|\hat{G}_n(t) - G_{n,\theta}(t)| > \delta) \geq 1/2 \qquad (2.18)$$

holds for every $\delta < 1/2$ and every $\rho > |t|$.

<u>Remark</u> <u>2.3</u>: (i) The "superefficiency" property of Hodges' estimator, i.e., the fact that its asymptotic distribution coincides with the asymptotic distribution of the restricted or unrestricted maximum likelihood estimator depending on whether the parameter satisfies the restriction $\theta = 0$ or not, is also shared by post-model-selection (pre-test) estimators based on a consistent model selection procedure (cf., Pötscher (1991), Lemma 1). The

recently introduced SCAD estimators (Fan and Li (2001), Theorem 2) exhibit a similar property: The asymptotic distribution of these estimators remains unchanged, whether or not (valid) zero-restrictions are imposed. This property has (somewhat unfortunately) been dubbed "oracle" property in Fan and Li (2001). Also other penalized least-squares estimators like some (but not all) members of the class of Bridge estimators (cf. Knight and Fu (2000), p.1361) have this property for certain choices of the regularization parameter. (It is not shared for example by the Lasso-type estimator as considered in the previous subsection.) It is likely that results similar to Theorem 2.4 can be obtained for SCAD estimators or other estimators satisfying the "oracle" property, although this is beyond the scope of this paper. (ii) Although of no importance for the present paper, we feel compelled to point out that the "oracle" property is only a _pointwise_ asymptotic concept. In view of the lessons learned from Hodges' estimator and its "superefficiency" one should hence be aware that the "oracle" property may give a very misleading impression of the actual performance of an estimator.

3. Lower Bounds on the Performance of Estimators and the (Non)Existence of Uniformly Consistent Estimators: Some General Results

In this section we establish some general lower bounds on the performance of estimators; in particular, we provide conditions for the (non)existence of uniformly consistent estimators. These results form the basis for the results in the preceding section, but are also of independent interest. Consider a sequence of statistical experiments described by a (non-empty) set B, the parameter space, which indexes a set $\mathcal{P}_n = \{P_{n,\beta}: \beta \in B\}$ of probability measures for every $n \geq 1$. Typically (but not necessarily), n represents sample size and $P_{n,\beta}$ describes the distribution of the sample of size n. For each n, the measures $P_{n,\beta}$ are defined on some measurable space $(\Omega_n, \mathcal{F}_n)$. The quantity of

interest to be estimated at sample size n is given by $\varphi_n(\beta)$, where φ_n is a

function on B taking its values in a metric space (M,d). Often M will be the

real line or Euclidean space. An estimator $\hat{\varphi}_n$ for $\varphi_n(\beta)$ is an M-valued

function on Ω_n that is measurable w.r.t. $\mathcal{F}_n$ and the Borel σ-field on M.

A sequence of estimators $\hat{\varphi}_n$ for $\varphi_n(\beta)$ is said to be <u>consistent</u> (on B) if

for every $\beta \in B$ and every $\delta > 0$

$$P_{n,\beta}(d(\hat{\varphi}_n,\varphi_n(\beta)) > \delta) \to 0 \qquad (3.1)$$

holds for $n \to \infty$. Furthermore, $\hat{\varphi}_n$ is said to be <u>uniformly consistent</u> (on B) if

for every $\delta > 0$

$$\sup_{\beta \in B} P_{n,\beta}(d(\hat{\varphi}_n,\varphi_n(\beta)) > \delta) \to 0 \qquad (3.2)$$

holds for $n \to \infty$. For later use we note that (3.2) is equivalent to requiring

that (3.1) holds along sequences of parameters β_n, i.e., that

$P_{n,\beta_n}(d(\hat{\varphi}_n,\varphi_n(\beta_n)) > \delta) \to 0$ holds for every sequence $\beta_n \in B$ and $n \to \infty$.

The classical paper on (non)existence of (uniformly) consistent

estimators is LeCam and Schwartz (1960), which provides a characterization of

existence in the context of an i.i.d. model and in case $\varphi_n \equiv \varphi$ does not depend

on sample size n. Related results can be found in Yatracos (1985), Pfanzagl

(1998), and Pötscher (2002), the latter allowing for dependent and

non-identically distributed data. Roughly speaking, the common theme in these

papers is that the existence of uniformly consistent estimators is tied to

(appropriate) continuity properties of the estimand $\varphi(\beta)$. In the following

we provide conditions for the (non)existence of uniformly consistent

estimators allowing in particular for sample-size-dependent estimands $\varphi_n(\beta)$ as

well as for dependent and non-identically distributed data. (For a discussion

contrasting these results with the results in case $\varphi_n \equiv \varphi$ see Remark 3.5(iv)

below.) In the course of this, we also provide lower bounds for the

performance of estimators, i.e., for the l.h.s. of (3.2).

If B_n is a sequence of (non-empty) subsets of B, we define the

<u>oscillation</u>[1] of φ_n over B_n as

$$\omega(\varphi_n, B_n) = \sup\nolimits_{(\beta^{(1)}, \beta^{(2)}) \in B_n \times B_n} d(\varphi_n(\beta^{(1)}), \varphi_n(\beta^{(2)})).$$

Informally speaking, given that the measures $P_{n,\beta}$ for $\beta \in B_n$ are "close" to one another in an appropriate sense (e.g., picture B_n as suitably "shrinking" sets), the existence of estimators that are "consistent uniformly over B_n" depends crucially on the behaviour of the oscillation $\omega(\varphi_n, B_n)$ as $n \to \infty$. This is formalized in the following two lemmata.

<u>Lemma</u> <u>3.1</u>: Let B_n, $n \ge 1$, be a sequence of subsets of B with non-empty intersection and let α be an element of $\bigcap_{n=1}^{\infty} B_n$. Suppose that the sequence P_{n,β_n} is contiguous w.r.t. $P_{n,\alpha}$ for every sequence $\beta_n \in B_n$.[2]

(a) If $\delta^* = \liminf_{n \to \infty} \omega(\varphi_n, B_n) > 0$, then any estimator $\hat{\varphi}_n$ for $\varphi_n(\beta)$ satisfying $P_{n,\alpha}(d(\hat{\varphi}_n, \varphi_n(\alpha)) > \delta) \to 0$ for every $\delta > 0$ and $n \to \infty$ (e.g., any consistent estimator) satisfies

$$\liminf_{n \to \infty} \sup\nolimits_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n, \varphi_n(\beta)) > \delta) = 1 \qquad (3.3)$$

for every $\delta < \delta^*/2$; furthermore,

$$\liminf_{n \to \infty} \inf\nolimits_{\hat{\varphi}_n} \sup\nolimits_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n, \varphi_n(\beta)) > \delta) > 0 \qquad (3.4)$$

holds for every $\delta < \delta^*/2$, where the infimum in (3.4) extends over <u>all</u> estimators $\hat{\varphi}_n$ for $\varphi_n(\beta)$.[3] In particular, no uniformly consistent estimator

---

[1]   This notion of oscillation is different from the one used in Pötscher (2002).

[2]   Let Q_n and R_n be probability measures defined on $(\Omega_n, \mathscr{F}_n)$. The sequence Q_n is said to be contiguous w.r.t. the sequence R_n if $\lim_{n \to \infty} R_n(F_n) = 0$ implies $\lim_{n \to \infty} Q_n(F_n) = 0$ whenever $F_n \in \mathscr{F}_n$.

[3]   While δ in (3.1)-(3.2) has to be positive, this is obviously not necessary in statements like (3.3), (3.4), or (3.7) below. Moreover, δ^* could be allowed to be zero in Lemma 3.1(a) or Lemma 3.2, although this leads only

for $\varphi_n(\beta)$ exists (neither on B nor on $\bigcup_{n=1}^{\infty} B_n$).

(b) If $\limsup_{n \to \infty} \omega(\varphi_n, B_n) = 0$, then any estimator $\hat{\varphi}_n$ for $\varphi_n(\beta)$ satisfying

$P_{n,\alpha}(d(\hat{\varphi}_n, \varphi_n(\alpha)) > \delta) \to 0$ for every $\delta > 0$ and $n \to \infty$ (e.g., any consistent estimator) satisfies

$$\limsup_{n \to \infty} \sup_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n, \varphi_n(\beta)) > \delta) = 0 \qquad (3.5)$$

for every $\delta > 0$.

Remark 3.1: (Extensions of Lemma 3.1) If $\limsup_{n \to \infty} \omega(\varphi_n, B_n) > 0$ but $\liminf_{n \to \infty} \omega(\varphi_n, B_n) = 0$ holds, applying Lemma 3.1 to appropriate subsequences shows that (3.3)-(3.4) hold for $\delta < \limsup_{n \to \infty} \omega(\varphi_n, B_n)/2$ provided the limit inferior in (3.3)-(3.4) is replaced by a limit superior; similarly, (3.5) holds for every $\delta > 0$ provided the limit superior is replaced by a limit inferior. In particular, it follows that the condition $\limsup_{n \to \infty} \omega(\varphi_n, B_n) = 0$ is not only sufficient but also necessary for (3.5) to hold for every $\delta > 0$.

A result somewhat related to (3.3) above is established in Pfanzagl (1998, Corollary 3.1) for i.i.d. models and for the case $\varphi_n \equiv \varphi$ under a considerably stronger assumption on the functional φ. The next lemma provides a lower bound on the performance of arbitrary estimators that improves upon (3.4). This result is based on the total variation distance as a notion of "closeness" between probability measures, and is closely related to Lemma 1 in Chen (1997). As Chen (1997) notes, the basic inequality underlying this lemma has been used in the literature before (e.g., in Donoho and Liu (1987)), at least implicitly. The lemma is also similar in spirit to ideas used in Le Cam (1973). Recall that the total variation distance between finite measures Q and R defined on a measurable space $(\Omega, \mathscr{F})$ is given by $\|Q-R\|_{TV} =$

---

to completely trivial results.

$\sup_{F\in\mathscr{F}}|Q(F)-R(F)|$.

Lemma 3.2: Let B_n, $n\geq 1$, be a sequence of (non-empty) subsets of B. Suppose that the diameters of the sets $\{P_{n,\beta}: \beta\in B_n\}$ w.r.t. total variation distance satisfy

$$\limsup_{n\to\infty}\sup_{(\beta^{(1)},\beta^{(2)})\in B_n\times B_n}||P_{n,\beta^{(1)}} - P_{n,\beta^{(2)}}||_{TV} \leq \Gamma \qquad (3.6)$$

with $\Gamma<1$. If $\delta^* = \liminf_{n\to\infty}\omega(\varphi_n,B_n) > 0$, then

$$\liminf_{n\to\infty}\inf_{\hat{\varphi}_n}\sup_{\beta\in B_n} P_{n,\beta}(d(\hat{\varphi}_n,\varphi_n(\beta))>\delta) \geq (1-\Gamma)/2 > 0 \qquad (3.7)$$

holds for every $\delta<\delta^*/2$, where the infimum in (3.7) extends over <u>all</u> estimators $\hat{\varphi}_n$ for $\varphi_n(\beta)$. In particular, no uniformly consistent estimator for $\varphi_n(\beta)$ exists (neither on B nor on $\bigcup_{n=1}^{\infty} B_n$).

Remark 3.2: (Finite-sample lower bounds) Lemma 3.2 is at its core actually a finite-sample result: Let A be a non-empty subset of $\mathbb{N}$. (Of primary interest is the case A=$\mathbb{N}$ or the case where A contains only one element.) If the limit superior in (3.6) is replaced by the supremum over A and if the limit inferior in the definition of δ^* is replaced by the infimum over A, then (3.7) holds with the limit inferior replaced by the infimum over A. (This follows either from inspection of the proof of the lemma or from applying the lemma for every $m\in A$ to φ'_n and $\{P'_{n,\beta}: \beta\in B'_n\}$ given by $\varphi'_n=\varphi_m$, $P'_{n,\beta}=P_{m,\beta}$, and $B'_n=B_m$ for all $n\geq 1$.)

Observe that the l.h.s. of (3.7) decreases if the sets B_n are replaced by smaller subsets, whereas the r.h.s. increases. This suggests that for a <u>given</u> sequence B_n and a <u>given</u> $\delta<\delta^*/2$, the lower bound in (3.7) can often be improved by the following strategy: Try to find $C_n\subseteq B_n$ such that the sets $\{P_{n,\beta}: \beta\in C_n\}$ have diameter w.r.t. total variation distance as small as possible, subject to the requirement that $\delta<1/2\liminf_{n\to\infty}\omega(\varphi_n,C_n)$ holds. (Often a natural choice is $C_n=\{P_{n,\beta_n^{(1)}},P_{n,\beta_n^{(2)}}\}$ where $\beta_n^{(1)}$, $\beta_n^{(2)}$ in B_n are such that $P_{n,\beta_n^{(1)}}$ and

$P_{n,\beta_n}^{(2)}$ are as close as possible in the total variation distance, subject to the requirement that $\liminf_{n \to \infty} d(\varphi_n(\beta_n^{(1)}), \varphi_n(\beta_n^{(2)})) > 2\delta$ holds.) A (typically) improved bound for the l.h.s. of (3.7) is then obtained by applying Lemma 3.2 with C_n replacing B_n. A slightly refined version of this strategy is formalized in Lemma B.1 in Appendix B. This strategy gives rise to the following useful results.

Corollary 3.3: Assume B is equipped with a metric π such that the maps $\beta \to P_{n,\beta}$ are asymptotically continuous at $\alpha \in B$, i.e.,

$$\sup_{\pi(\alpha, \beta) < c_n} ||P_{n,\beta} - P_{n,\alpha}||_{TV} \to 0$$

holds for some sequence $c_n \to 0$, $c_n > 0$. Suppose $\varphi_n = \varphi$, and φ is discontinuous at α. Suppose further that the sets $B_n \subseteq B$ contain α as an interior point for every $n \geq 1$. Define $\eta = \inf_U \omega(\varphi, U)$ where the infimum extends over all open balls U in B with center α. Then η is positive and

$$\liminf_{n \to \infty} \inf_{\hat{\varphi}_n} \sup_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n, \varphi(\beta)) > \delta) \geq 1/2 \tag{3.8}$$

holds for every $\delta < \eta/2$, where the infimum in (3.8) extends over all estimators $\hat{\varphi}_n$ for $\varphi_n(\beta)$.

In the next corollary, ζ_n will typically (but not necessarily) be a sequence that converges to zero.

Corollary 3.4: Let B be a subset of a normed vector space $(V, ||.||)$ and let α be an interior point of B. Suppose the sets B_n, $n \geq 1$, are given by the open balls $B(\alpha, \zeta_n) = \{\beta \in V: ||\beta - \alpha|| < \zeta_n\}$ for some sequence $\zeta_n > 0$, and suppose $B_n \subseteq B$ holds. Suppose further that the maps $\gamma \to P_{n, \alpha + \gamma \zeta_n}$ (defined on G = $\{\gamma \in V: ||\gamma|| < 1\}$) are asymptotically uniformly equicontinuous on G, i.e., satisfy

$$\limsup_{n \to \infty} \sup_{\{(\gamma, \delta) \in G \times G: ||\gamma - \delta|| < c\}} ||P_{n, \alpha + \gamma \zeta_n} - P_{n, \alpha + \delta \zeta_n}||_{TV} \to 0 \text{ for } c \to 0.$$

19

(a) If $\delta^* = \liminf_{n\to\infty}\omega(\varphi_n,B_n) > 0$ holds, then

$$\sup_{\delta>0}\liminf_{n\to\infty}\inf_{\hat{\varphi}_n}\sup_{\beta\in B_n}P_{n,\beta}(d(\hat{\varphi}_n,\varphi_n(\beta))>\delta) \geq 1/2 \qquad (3.9)$$

where the infimum in (3.9) extends over <u>all</u> estimators $\hat{\varphi}_n$ for $\varphi_n(\beta)$.

(b) Suppose there exists a $\delta_*>0$ such that for every $\varepsilon>0$ there exist $\beta_{n,1}(\varepsilon)$

and $\beta_{n,2}(\varepsilon)$ in B_n satisfying

$$\limsup_{n\to\infty}||\beta_{n,1}(\varepsilon) - \beta_{n,2}(\varepsilon)||/\zeta_n < \varepsilon \qquad (3.10)$$

and

$$\liminf_{n\to\infty}d(\varphi_n(\beta_{n,1}(\varepsilon)),\varphi_n(\beta_{n,2}(\varepsilon))) \geq \delta_*. \qquad (3.11)$$

Then

$$\liminf_{n\to\infty}\inf_{\hat{\varphi}_n}\sup_{\beta\in B_n}P_{n,\beta}(d(\hat{\varphi}_n,\varphi_n(\beta))>\delta) \geq 1/2 \qquad (3.12)$$

holds for $\delta<\delta_*/2$, where the infimum in (3.12) extends over <u>all</u> estimators $\hat{\varphi}_n$

for $\varphi_n(\beta)$. (If one of $\beta_{n,i}(\varepsilon)$ can always be chosen equal to α, then the

asymptotic uniform equicontinuity condition can be weakened to asymptotic

equicontinuity at $\gamma=0$.)

<u>Remark 3.3</u>: (Finite-sample lower bounds) Let A be a non-empty subset of ℕ.

(Of primary interest is the case A=ℕ or the case where A contains only one

element.)

(i) Suppose the asymptotic continuity condition in Corollary 3.3 is replaced

by continuity of the maps $\beta\to P_{n,\beta}$ at α for every $n\in A$. Then (3.8) continues to

hold provided the limit inferior is replaced by the infimum over $n\in A$. (This

follows from Lemma B.1 together with Remark B.1 upon observing that the

assumed continuity property of $\beta\to P_{n,\beta}$ allows one to choose open balls $C_{nk}\subseteq B_n$

centered at α such that $\Gamma_k\to 0$ for $k\to\infty$, and from the fact that $\omega(\varphi,C_{nk})\geq\eta>0$

holds for $n\in A$ and all $k\geq 1$.)

(ii) Suppose the asymptotic uniform equicontinuity condition in Corollary 3.4

is replaced by the condition

$$\sup_{n\in A}\sup\{(\gamma,\delta)\in G\times G: ||\gamma-\delta||<c\}||P_{n,\alpha+\gamma\zeta_n}-P_{n,\alpha+\delta\zeta_n}||_{TV} \to 0 \text{ for } c\to 0.$$

Then (a) and (b) of that corollary continue to hold if at every occurrence a limit inferior (superior) is replaced by an infimum (supremum) over $n \in A$. The proof is similar to the proof of Corollary 3.4 using Remark B.1 instead of Lemma B.1.

Remark 3.4: Suppose the assumptions of Corollary 3.4 are satisfied with $\zeta_n = \rho \xi_n$, $\rho > 0$. For $0 < \rho' \le \rho$ consider the sets $B_n(\rho') = B(\alpha, \rho' \xi_n) \subseteq B_n(\rho) = B_n$. If now not only $\delta^* = \delta^*(\rho) > 0$, but also $\delta^*(\rho') = \liminf_{n \to \infty} \omega(\varphi_n, B_n(\rho')) > 0$ holds for all $0 < \rho' \le \rho$, we may apply Lemma 3.2 to the sequence $B_n(\rho')$. Observe that $\Gamma(\rho') = \limsup_{n \to \infty} \sup_{(\beta^{(1)}, \beta^{(2)}) \in B_n(\rho') \times B_n(\rho')} ||P_{n, \beta^{(1)}} - P_{n, \beta^{(2)}}||_{TV}$ goes to zero for $\rho' \to 0$ in view of the asymptotic equicontinuity assumption in Corollary 3.4. Since $B_n(\rho') \subseteq B_n(\rho) = B_n$, the lower bound (3.9) immediately follows. However, the assumption in Corollary 3.4(a) that $\delta^* = \delta^*(\rho) > 0$ does <u>not</u> in general imply $\delta^*(\rho') > 0$ for all ρ', $0 < \rho' \le \rho$. (Just imagine a situation where the oscillation of φ_n occurs near the boundary of $B_n(\rho)$ and φ_n is "flat" inside of $B_n(\rho/2)$, say.) Corollary 3.4(a) now shows that the condition $\delta^*(\rho') > 0$ for all ρ', $0 < \rho' \le \rho$, is in fact unnecessary for (3.9) to hold. This comes in handy in applications where $\delta^*(\rho')$ can be zero for some $\rho' < \rho$ or where we can only establish the existence of a ρ such that $\delta^*(\rho)$ is positive but where we do not have information on $\delta^*(\rho')$ for arbitrarily small ρ'; cf. the proof of (2.14).

The somewhat abstract conditions imposed on the statistical experiment in the above results are satisfied in a wide variety of situations. For example, if B is an open subset of Euclidean space, the asymptotic uniform equicontinuity condition in Corollary 3.4 can be shown to be satisfied if the experiment satisfies a LAN-condition (and the radii ζ_n of the balls $B(\alpha, \zeta_n)$ are essentially the reciprocal of the rate appearing in the definition of LAN.) Furthermore, under the LAN-condition and if $B_n = B(\alpha, \zeta_n)$, the contiguity requirement in Lemma 3.1 is satisfied, and the l.h.s. of (3.6) can be

evaluated explicitly.

In the important special case where the sets B_n are (typically shrinking) balls with a common center, also simple sufficient conditions for the assumptions in Lemmata 3.1-3.2 and Corollary 3.4 regarding the oscillation of the estimands φ_n can be given.

Lemma 3.5:   Let B be a subset of a normed vector space $(V,||.||)$ and let α be an interior point of B.   Suppose the sets B_n, $n \geq 1$, are given by the open balls $B(\alpha,\zeta_n) = \{\beta \in V: ||\beta-\alpha|| < \zeta_n\}$ for some sequence $\zeta_n > 0$, and suppose $B_n \subseteq B$ holds. Assume that there is a non-empty subset G_0 of G and a function $\varphi_{\infty,\alpha}: G_0 \to M$ such that the functions $\gamma \to \varphi_n(\alpha+\gamma\zeta_n)$ converge to $\varphi_{\infty,\alpha}(\gamma)$ for every $\gamma \in G_0$ and $n \to \infty$.

(a) If $\varphi_{\infty,\alpha}$ is non-constant on G_0, then $\delta^* = \liminf_{n \to \infty} \omega(\varphi_n, B_n)$ is positive. More precisely, $\delta^* \geq \omega(\varphi_{\infty,\alpha}, G_0)$ holds.

(b) Suppose there exists a $\delta_* > 0$ such that for every $\varepsilon > 0$ one can find elements $\gamma_1(\varepsilon)$, $\gamma_2(\varepsilon)$ in G_0 satisfying $||\gamma_1(\varepsilon)-\gamma_2(\varepsilon)|| < \varepsilon$ and $d(\varphi_{\infty,\alpha}(\gamma_1(\varepsilon)), \varphi_{\infty,\alpha}(\gamma_2(\varepsilon))) \geq \delta_*$. Then the conditions on φ_n in Corollary 3.4(b) are satisfied.

(c) If $\varphi_{\infty,\alpha}$ is discontinuous at some $\gamma_0 \in G_0$, then the conditions on φ_n in Corollary 3.4(b) are satisfied (for every $0 < \delta_* < \eta_\infty$, where $\eta_\infty = \inf\{\omega(\varphi_{\infty,\alpha}, U): \gamma_0 \in U \subseteq G_0$, U relatively open in $G_0\}$ is positive).

(d) If $\varphi_{\infty,\alpha}$ is constant on $G_0 = G$ and if $\varphi_n(\alpha+.\zeta_n)$ converges to $\varphi_{\infty,\alpha}$ uniformly on $G_0 = G$, then $\delta^* = \liminf_{n \to \infty} \omega(\varphi_n, B_n)$ equals zero.

The choice of the set G_0 in Lemma 3.5 will depend on the particular situation in which the lemma is being applied.   Often one will choose $G_0 = G$ (if this is feasible), or one will choose G_0 as the largest subset of G on which the maps $\gamma \to \varphi_n(\alpha+\gamma\zeta_n)$ have a limit.   But sometimes other choices are convenient, cf. the proof of Theorem 2.2 or 2.3.

Remark 3.5:   (i)   The lower bound of 1/2 in (3.8), (3.9), and (3.12) can not be

22

improved in general.

(ii) It is easy to see that the condition in Corollary 3.4(b) is equivalent to the condition that there exist $\beta_{n,1}$ and $\beta_{n,2}$ in B_n such that $\limsup_{n\to\infty} ||\beta_{n,1} - \beta_{n,2}||/\zeta_n = 0$ and $\liminf_{n\to\infty} d(\varphi_n(\beta_{n,1}),\varphi_n(\beta_{n,2})) > 0$ hold.

(iii) The results in this section concern the performance measures $P_{n,\beta}(d(\hat{\varphi}_n,\varphi_n(\beta))>\delta)$, which can be viewed as the risk corresponding to the loss-function $1(|x|>\delta)$. Quite similar results can also be obtained for a wide variety of other risk functions.

(iv) In case the sets B_n are shrinking balls with a common center and the estimand φ_n does not depend on n, i.e., $\varphi_n \equiv \varphi$, the cause for a positive lower bound on the performance measure for estimation is discontinuity of φ. The situation is more complex in case the estimand depends on n. Examples can be given where each φ_n is continuous, but a lower bound like (3.7) holds. Conversely, estimation problems can be constructed, such that $\varphi_n(\beta)$ is discontinuous (and even converges to a discontinuous limit), but where a uniformly consistent estimator for $\varphi_n(\beta)$ exists. Lemma 3.5 shows that what matters are not so much continuity properties of $\varphi_n(\beta)$ or of its limit, but properties of the limit $\varphi_{\infty,\alpha}$ obtained after rescaling the parameter.

(v) The results in this section are also useful for obtaining bounds on convergence rates of estimators: Suppose $\psi_n(\beta)$ is to be estimated, with $\hat{\psi}_n$ denoting corresponding estimators. Setting $\varphi_n(\beta)=a_n\psi_n(\beta)$ and $\hat{\varphi}_n=a_n\hat{\psi}_n$ for a sequence a_n, an application of the results in this section can be used to show that no estimator $\hat{\psi}_n$ for $\psi_n(\beta)$ can converge faster than a_n^{-1} in a minimax sense.

(vi) All results in this section with the exception of Lemma 3.5 continue to hold if (M,d) is allowed to depend on n and/or to be a pseudo-metric space. (Lemma 3.5(b)-(d) continues to hold if (M,d) is a pseudo-metric space but does not depend on n.)

The next result shows that Lemmata 3.1 and 3.2 (and hence all results in this paper) also apply to randomized estimators. To this end let $\Omega_n^* = \Omega_n \times \Xi_n$ and $\mathcal{F}_n^* = \mathcal{F}_n \otimes \mathcal{G}_n$ where $(\Xi_n, \mathcal{G}_n)$ is a measurable space for every $n \geq 1$. The randomization mechanism is described by a Markov-kernel K_n defined on $\Omega_n \times \mathcal{G}_n$. For each $\beta \in B$ define $P_{n,\beta}^*$ via

$$P_{n,\beta}^*(F^*) = \iint 1_{F}^*(\omega, \xi) K_n(\omega, d\xi) P_{n,\beta}(d\omega).$$

A randomized estimator for $\varphi_n(\beta)$ is then a function $\varphi_n^* : \Omega_n^* \to M$ that is measurable w.r.t. $\mathcal{F}_n^*$ and the Borel σ-field on M.

Lemma 3.6: (a) For any $\alpha_n \in B$, $\beta_n \in B$, contiguity of the sequence P_{n,β_n} w.r.t. P_{n,α_n} implies contiguity of P_{n,β_n}^* w.r.t. P_{n,α_n}^*.

(b) $||P_{n,\alpha}^* - P_{n,\beta}^*||_{TV} = ||P_{n,\alpha} - P_{n,\beta}||_{TV}$ for any $\alpha \in B$, $\beta \in B$, and $n \geq 1$.

In some applications, cf. Leeb and Pötscher (2002), it is necessary to transfer the results of this section from a given statistical experiment to one that is obtained by conditioning. Results pertaining to this case are given in Appendix C.

Appendix A: Proofs for Section 2

Lemma A.1: Consider the linear model $Y_i = X_i \theta + u_i$ for $i \geq 1$, where X_i is a non-stochastic $l \times k$ regressor matrix ($l \geq 1, k \geq 1$), θ is an unknown $k \times 1$ parameter vector, and the errors u_i are i.i.d. normal with mean zero and covariance matrix I_l, the $l \times l$ identity matrix. For $\theta \in \mathbb{R}^k$, let $P_{n,\theta}$ denote the distribution of the sample $(Y_1', \ldots, Y_n')'$. Then

(a) $||P_{n,\theta} - P_{n,\vartheta}||_{TV} = 2\Phi([(\theta-\vartheta)' \sum_{i=1}^{n} X_i' X_i (\theta-\vartheta)]^{1/2}/2) - 1 \leq 2\Phi(||\theta-\vartheta|| \lambda_{max}^{1/2}(\sum_{i=1}^{n} X_i' X_i)/2) - 1$, where Φ is the standard normal distribution function.

(b) P_{n,ϑ_n} is contiguous w.r.t. P_{n,θ_n} if $\theta_n - \vartheta_n = O(n^{-1/2})$ and $n^{-1}\sum_{i=1}^{n} X_i'X_i$ is

bounded.

Proof: (a) Since $P_{n,\theta}$ and $P_{n,\vartheta}$ are mutually absolutely continuous we can

write the total variation distance as

$$||P_{n,\theta} - P_{n,\vartheta}||_{TV} = P_{n,\theta}(dP_{n,\theta}/dP_{n,\vartheta} > 1) - P_{n,\vartheta}(dP_{n,\theta}/dP_{n,\vartheta} > 1) =$$

$$P_{n,\theta}(\log(dP_{n,\theta}/dP_{n,\vartheta}) > 0) - P_{n,\vartheta}(\log(dP_{n,\theta}/dP_{n,\vartheta}) > 0).$$

Set $A = (\theta-\vartheta)'\sum_{i=1}^{n} X_i'X_i(\theta-\vartheta)$, and assume first that $A>0$. The log-likelihood

ratio $\log(dP_{n,\theta}/dP_{n,\vartheta})$ is then distributed as $N(A/2,A)$ under $P_{n,\theta}$ and as

$N(-A/2,A)$ under $P_{n,\vartheta}$ as is easily seen. Elementary calculations then show

that $P_{n,\theta}(\log(dP_{n,\theta}/dP_{n,\vartheta}) > 0) = 1-\Phi(-A^{1/2}/2)$ and $P_{n,\vartheta}(\log(dP_{n,\theta}/dP_{n,\vartheta}) > 0)$

$= 1-\Phi(A^{1/2}/2)$. The result then follows for $A>0$. If $A=0$, clearly $X_i\theta=X_i\vartheta$ for

$i=1,\ldots,n$ holds, implying $P_{n,\theta}=P_{n,\vartheta}$. Since $\Phi(0)=1/2$, the result again

follows.

(b) Since $\theta_n - \vartheta_n = O(n^{-1/2})$ and $n^{-1}\sum_{i=1}^{n} X_i'X_i$ is bounded, any subsequence contains

a further subsequence such that along this subsequence $n^{1/2}(\theta_n - \vartheta_n)$ and

$n^{-1}\sum_{i=1}^{n} X_i'X_i$ converge. By LeCam's first lemma (cf. van der Vaart (1998), Lemma

6.4) it thus suffices to show that $dP_{n,\theta_n}/dP_{n,\vartheta_n}$ converges in distribution

under P_{n,ϑ_n} to a random variable that is almost surely positive, provided that

$n^{1/2}(\theta_n - \vartheta_n) \rightarrow \chi$ and $n^{-1}\sum_{i=1}^{n} X_i'X_i \rightarrow C$. The same calculations as in the proof

of part (a) show, that $\log(dP_{n,\theta_n}/dP_{n,\vartheta_n})$ is distributed as $N(-A_n/2,A_n)$ under

P_{n,ϑ_n}, where $A_n=(\theta_n - \vartheta_n)'\sum_{i=1}^{n} X_i'X_i(\theta_n - \vartheta_n)$. Here we use the convention that

$N(-A_n/2,A_n)$ denotes point-mass at zero if $A_n=0$. Hence, $\log(dP_{n,\theta_n}/dP_{n,\vartheta_n})$

converges in distribution under P_{n,ϑ_n} to $Z \sim N(-A/2,A)$, where $A=\chi'C\chi$. It

follows from the continuous mapping theorem that $dP_{n,\theta_n}/dP_{n,\vartheta_n}$ converges in

distribution under P_{n,ϑ_n} to $\exp(Z)$ which is always positive. ∎

Proof of Theorem 2.1: We apply the results of Section 3 with $B=\mathbb{R}^k$, $\beta=\theta$,

$\varphi_n(\beta)=R_n(\hat{\theta}_{JS};\theta)$, $\hat{\varphi}_n=\hat{R}_n$, and $B_n=\{\theta\in\mathbb{R}^k: ||\theta||<\rho n^{-1/2}\}$. We first compute the

oscillation of $R_n(\hat{\theta}_{JS};\theta)$ over B_n: It is easy to see that the risk is a

strictly increasing and continuous function of $\theta'\theta$. Hence, the oscillation is

given by $R_n(\hat{\theta}_{JS};\rho n^{-1/2})-R_n(\hat{\theta}_{JS};0)$ which equals $2\Delta(\rho)$, where Δ was defined

prior to Theorem 2.1. Note that $\Delta(\rho)$ is positive. Applying Lemma A.1(b) with

$X_i=I_k$ shows that P_{n,θ_n} is contiguous w.r.t. $P_{n,0}$ for every sequence $\theta_n\in B_n$.

Lemma 3.1(a) then gives (2.2). From Lemma A.1(a) we obtain

$$\sup_{(\theta,\vartheta)\in B_n\times B_n}||P_{n,\theta} - P_{n,\vartheta}||_{TV} = 2\Phi(\rho)-1,$$

since $\sum_{i=1}^{n}X_i'X_i=nI_k$. Lemma 3.2 together with Remark 3.2 now imply that

the l.h.s. of (2.3) is not less than $1-\Phi(\rho)$ for every $0<\rho<\infty$ satisfying $\delta<\Delta(\rho)$,

or equivalently $\Delta^{-1}(\delta)<\rho$. Since the l.h.s. of (2.3) is nonincreasing and the

lower bound $1-\Phi(\rho)$ is increasing when ρ decreases to $\Delta^{-1}(\delta)$, the result (2.3)

follows.   ∎

Lemma A.2: Let F be a cdf on $\mathbb{R}^k$, $k>1$, and let U be an open subset of $\mathbb{R}^k$.

Then F is continuous on U if and only if for every $t=(t_1,\ldots,t_k)'\in U$ and every

i, $1\leq i\leq k$, the map $u_i\to F(t_1,\ldots,t_{i-1},u_i,t_{i+1},\ldots,t_k)$ is continuous at $u_i=t_i$.

Proof: One direction is trivial. To prove the other one, fix $t\in U$ and let

$t(n)\in\mathbb{R}^k$ denote a sequence converging to t. For sufficiently small $\varepsilon_i>0$ we

have $\prod_{i=1}^{k}[t_i-\varepsilon_i,t_i+\varepsilon_i] \subseteq U$ and

$$F(t_1-\varepsilon_1,\ldots,t_k-\varepsilon_k)\leq F(t(n))\leq F(t_1+\varepsilon_1,\ldots,t_k+\varepsilon_k)$$

for $n\geq N(\varepsilon_1,\ldots,\varepsilon_k)$. Hence also $\limsup_n F(t(n))$ and $\liminf_n F(t(n))$ are

sandwiched between these bounds. Since $(t_1-\varepsilon_1,\ldots,t_k-\varepsilon_k)\in U$, it follows that

$F(t_1-\varepsilon_1,\ldots,t_{k-1}-\varepsilon_{k-1},t_k-\varepsilon_k)$ converges to $F(t_1-\varepsilon_1,\ldots,t_{k-1}-\varepsilon_{k-1},t_k)$ for $\varepsilon_k\to 0$.

Similarly, $F(t_1+\varepsilon_1,\ldots,t_{k-1}+\varepsilon_{k-1},t_k+\varepsilon_k)$ converges to $F(t_1+\varepsilon_1,\ldots,t_{k-1}+\varepsilon_{k-1},t_k)$

for $\varepsilon_k\to 0$. This shows that $\limsup_n F(t(n))$ and $\liminf_n F(t(n))$ are both

sandwiched between $F(t_1-\varepsilon_1,\ldots,t_{k-1}-\varepsilon_{k-1},t_k)$ and $F(t_1+\varepsilon_1,\ldots,t_{k-1}+\varepsilon_{k-1},t_k)$. Observing that the arguments of the latter two functions are elements of U, we may repeat the argument until we arrive at the conclusion that $F(t) \leq$ $\liminf_n F(t(n)) \leq \limsup_n F(t(n)) \leq F(t)$. ∎

Lemma A.3: Let θ, ν, and t be elements of $\mathbb{R}^k$ and let $I(\theta)$ denote the set {i: $\theta_i=0$}. Then the cdf $F_{\infty,\theta,\nu}$ given by (2.6) is continuous at t provided that $t_i \neq -\nu_i$ holds for all $i \in I(\theta)$. In particular, $F_{\infty,\theta,\nu}$ is continuous on all of $\mathbb{R}^k$ if θ has only non-zero components; furthermore, $F_{\infty,\theta,\nu}$ is continuous at t for all θ if $t_i \neq -\nu_i$ holds for all i, $1 \leq i \leq k$.

Proof: We only need to prove the first claim. The case k=1 as well as the case k>1 and $I(\theta)=\varnothing$ follow immediately from (2.8)-(2.10) and the attending discussion. Hence assume k>1 and $I(\theta) \neq \varnothing$ in the sequel. Choose an open neighborhood U of t, such that any $s \in U$ also satisfies $s_i \neq -\nu_i$ for $i \in I(\theta)$. Assume now that $F_{\infty,\theta,\nu}$ is not continuous at t. Lemma A.2 then implies the existence of an element $s^* \in U$ and of an index m, $1 \leq m \leq k$, such that the function taking s_m into $F_{\infty,\theta,\nu}(s_1^*,\ldots,s_{m-1}^*,s_m,s_{m+1}^*,\ldots,s_k^*)$ is not continuous at $s_m=s_m^*$. Since this function is cadlag, it follows that it has a jump of positive height at $s_m=s_m^*$. Hence, the event A, that T_m equals s_m^*, has positive probability, where T_m denotes the m-th coordinate of $T=\text{argmin}_{u \in \mathbb{R}^k} V_{\theta,\nu}(u)$. We can now find a subset $J \subseteq I(\theta)$, J possibly empty, such that the event B=B(J) defined as the intersection of A with the event {$T_i=-\nu_i$ for $i \in J$ and $T_i \neq -\nu_i$ for $i \in I(\theta) \backslash J$} has positive probability. By possibly reducing B further, we can even achieve that the sign of $T_i+\nu_i$ is constant over B for all $i \in I(\theta) \backslash J$, without loosing the property that B has positive probability. Let u[J] denote the vector obtained from u by deleting the coordinates for $i \in J$. Note that $m \notin J$ since $s_m^* \neq -\nu_m$ by construction and since s_m^* equals T_m on the event B. Let $V_{\theta,\nu}^0(u[J])$ denote the function resulting from $V_{\theta,\nu}(u)$ after replacing u_i with

27

$-\nu_i$ for $i \in J$. Observe that on the event B the function $V^0_{\theta,\nu}(u[J])$ is a differentiable function in a neighborhood of its argmin, the coordinates of which coincide with the coordinates of $\operatorname{argmin}_{u \in \mathbb{R}^k} V_{\theta,\nu}(u)$ for $i \notin J$ by construction. Hence, the first derivative of $V^0_{\theta,\nu}(u[J])$ must be zero at its argmin. This leads to a system of equations of the form

$$-2W_i + 2\sum_{j \notin J} q_{ij} T_j = c_i$$

for $i \notin J$, where W_i is the i-th coordinate of the random vector W appearing in (2.7) and c_i is a constant on B depending only on Q, ν, and θ. Now on B we have $m \notin J$ and $T_m = s_m^*$, a fixed constant, implying that the distribution of $(T_j : j \notin J)$ puts positive mass on an hyperplane of dimension k-card(J)-1. Together with the above system of equations, this contradicts the fact that the distribution of $(W_i : i \notin J)$ is a non-singular multivariate normal. ∎

Lemma A.4: Suppose $\theta \in \mathbb{R}^k$ is given and $\gamma \in G$ satisfies $\gamma_i \neq 0$ for all $i \in I(\theta)$, where $G = \{\gamma \in \mathbb{R}^k: ||\gamma|| < 1\}$ and $I(\theta) = \{i: \theta_i = 0\}$. Then for any $t \in \mathbb{R}^k$ we have

$$\lim_{\rho \to \infty} F_{\infty,\theta,\rho\gamma}(t) = F_{\infty,\chi,0}(t) = \Phi_Q^{-1}(t - \lambda Q^{-1} \operatorname{sgn}(\chi)/2)$$

where $\chi \in \mathbb{R}^k$ is given by $\chi_i = \theta_i$ for $i \notin I(\theta)$ and $\chi_i = \gamma_i$ for $i \in I(\theta)$ and $\operatorname{sgn}(\chi)$ denotes the vector with i-th coordinate equal to $\operatorname{sgn}(\chi_i)$.

Proof: From (2.6)-(2.7) one sees that $F_{\infty,\theta,\rho\gamma}(t)$ can equivalently be expressed as $P(\operatorname{argmin}_{u \in \mathbb{R}^k} V^*_{\theta,\rho\gamma}(u) \leq t)$, where

$$V^*_{\theta,\rho\gamma}(u) = -2u'W + u'Qu + \lambda\sum_{j \notin I(\theta)} u_j \operatorname{sgn}(\theta_j) + \lambda\sum_{j \in I(\theta)} (|u_j + \rho\gamma_j| - |\rho\gamma_j|).$$

For $\rho \to \infty$ the random function $V^*_{\theta,\rho\gamma}(u)$ converges for every $u \in \mathbb{R}^k$ and every value of W to

$$-2u'W + u'Qu + \lambda\sum_{j \notin I(\theta)} u_j \operatorname{sgn}(\theta_j) + \lambda\sum_{j \in I(\theta)} u_j \operatorname{sgn}(\gamma_j)$$

which is nothing else than $V_{\chi,0}(u)$. In view of Geyer (1996), it follows that $F_{\infty,\theta,\rho\gamma}(.)$ converges weakly to the cdf of $\operatorname{argmin}_{u \in \mathbb{R}^k} V_{\chi,0}(u)$ for $\rho \to \infty$, which is given by $\Phi_Q^{-1}(. - \lambda Q^{-1} \operatorname{sgn}(\chi)/2)$. Since any $t \in \mathbb{R}^k$ is a continuity point of the latter distribution, the lemma follows. ∎

<u>Proof of Theorem 2.2</u>: We again apply the results of Section 3 and use the following identifications: $B=V=\mathbb{R}^k$, $\beta=\theta$, $\varphi_n(\beta)=F_{n,\theta}(t)$, $\hat{\varphi}_n=\hat{F}_n(t)$, $B_n=\{\theta\in\mathbb{R}^k: \|\theta\|<\rho n^{-1/2}\}$, where $\rho>0$ will be chosen later, and $\varphi_{\infty,0}(\gamma)=F_{\infty,0,\rho\gamma}(t)$ with $\gamma\in G$, the open unit ball in $\mathbb{R}^k$. To obtain information on the oscillation of $F_{n,\theta}(t)$ over B_n, we wish to apply Lemma 3.5(a) with $\alpha=0$, $\zeta_n=\rho n^{-1/2}$ and $G_0=\{e/(2k^{1/2}): e\in\{-1,1\}^k\}$. Observe that $\lim_{n\to\infty}F_{n,\rho\gamma n^{-1/2}}(t) = F_{\infty,0,\rho\gamma}(t)$ for all $\gamma\in G_0$ and all $\rho>(2k^{1/2})\max\{|t_i|:1\leq i\leq k\}$. This follows from the weak convergence of the finite-sample cdf and the fact that $F_{\infty,0,\rho\gamma}$ is continuous at t by Lemma A.3 (applied with $\theta=0$) whenever ρ satisfies $\pm\rho/(2k^{1/2})\neq t_i$ for all i. Hence the general assumptions of Lemma 3.5 are satisfied for such a choice of ρ. The oscillation of $F_{\infty,0,\rho\gamma}(t)$ over G_0 now satisfies

$$\lim_{\rho\to\infty}\omega(F_{\infty,0,\rho\gamma}(t),G_0)=\lim_{\rho\to\infty}\max\{|F_{\infty,0,\rho\gamma}(t)-F_{\infty,0,\rho\gamma'}(t)|: \gamma,\gamma'\in G_0\}=2\Delta(t)$$

by Lemma A.4. Lemma 3.5(a) now implies that

$$\liminf_{\rho\to\infty}\liminf_{n\to\infty}\omega(F_{n,\theta}(t),B_n) \geq 2\Delta(t) \qquad (A.1)$$

holds. This further implies that for every $\delta<\Delta(t)$ we can find a ρ large enough (depending on t and δ), such that $\delta<\liminf_{n\to\infty}\omega(F_{n,\theta}(t),B_n)/2$ holds and that the r.h.s. is positive (because $\Delta(t)$ is so by assumption). Now fix this δ and ρ (and hence B_n). Since Lemma A.1(b) with l=1 implies that P_{n,θ_n} is contiguous w.r.t. $P_{n,0}$ for any sequence $\theta_n\in B_n$, we may apply Lemma 3.1(a) to obtain (2.12). From Lemma A.1(a) we also obtain

$$\limsup_{n\to\infty}\sup_{(\theta,\vartheta)\in B_n\times B_n}\|P_{n,\theta} - P_{n,\vartheta}\|_{TV} \leq 2\Phi(\rho\lambda_{max}^{1/2}(Q))-1 < 1.$$

Lemma 3.2 now implies (2.13). Finally, (A.1) shows that we can find a ρ large enough such that $\liminf_{n\to\infty}\omega(F_{n,\theta}(t),B_n)$ is positive. Lemma A.1(a) shows that the asymptotic uniform equicontinuity condition of Corollary 3.4 is satisfied for B_n. Corollary 3.4(a) now implies (2.14).   ∎

Proof of Theorem 2.3: We again apply the results of Section 3 using the same identifications as in the proof of Theorem 2.2, except that now ρ satisfies $\rho > ||t||$. We now wish to apply Lemma 3.5(b) with $\alpha=0$, $\zeta_n = \rho n^{-1/2}$ and $G_0 = G\setminus\{-t/\rho\}$. Observe that $\lim_{n\to\infty} F_{n,\rho\gamma n^{-1/2}}(t) = F_{\infty,0,\rho\gamma}(t)$ for all $\gamma \in G_0$ by weak convergence and the fact that $F_{\infty,0,\rho\gamma}$ is continuous at t as long as $t \ne -\rho\gamma$, cf. (2.9)-(2.11). This shows that $\varphi_n(\alpha+\gamma\zeta_n)=F_{n,\rho\gamma n^{-1/2}}(t)$ converges to $\varphi_{\infty,0}(\gamma)=F_{\infty,0,\rho\gamma}(t)$ for all $\gamma \in G_0$. Now (2.9)-(2.11) furthermore show that $F_{\infty,0,\rho\gamma}(t)$ equals $\Phi_Q^{-1}(t+\lambda Q^{-1}/2)$ for $\gamma > -t/\rho$ and equals $\Phi_Q^{-1}(t-\lambda Q^{-1}/2)$ for $\gamma < -t/\rho$, where the relation $<$ is to interpreted coordinatewise. Define $\delta_* = \Phi_Q^{-1}(t+\lambda Q^{-1}/2) - \Phi_Q^{-1}(t-\lambda Q^{-1}/2)$ and note that δ_* is positive. Since $-t/\rho \in G$, which is open, it now follows that for every $\varepsilon>0$ we can find $\gamma_i(\varepsilon) \in G_0$, $i=1,2$, such that $|\gamma_1(\varepsilon)-\gamma_2(\varepsilon)|<\varepsilon$ and $|\varphi_{\infty,0}(\gamma_1(\varepsilon))-\varphi_{\infty,0}(\gamma_2(\varepsilon))| \ge \delta_*$ hold, hence the assumptions of Lemma 3.5(b) are satisfied. Applying Corollary 3.4(b) and observing that Lemma A.1(a) implies the asymptotic uniform equicontinuity condition in Corollary 3.4 completes the proof. ∎

Proof of Theorem 2.4: We again apply the results of Section 3 and use the identifications $B=\mathbb{R}$, $\beta=\theta$, $\varphi_n(\beta)=G_{n,\theta}(t)$, and $\hat\varphi_n=\hat G_n(t)$. We first prove (2.17). For given $\rho>|t|$ and $n\ge1$, the sets $C_n(\varepsilon)=\{-n^{-1/2}t, -n^{-1/2}(t+\varepsilon)\}$ are subsets of $B_n=\{\theta: |\theta|<\rho n^{-1/2}\}$ provided, e.g., $0<\varepsilon<(\rho-|t|)/2$ holds. From the formula for $G_{n,\theta}$ given in Section 2.3 we obtain that the oscillation satisfies

$$\omega(G_{n,\theta}(t),C_n(\varepsilon)) = |\Phi(c_n+t) - \Phi(-c_n+t+\varepsilon)| \qquad (A.2)$$

whenever ε also satisfies $\varepsilon<c_n$. Furthermore, Lemma A.1 implies that the total variation distance between the measures $P_{n,\theta}$, $\theta \in C_n(\varepsilon)$, is $2\Phi(\varepsilon/2)-1$. Applying the finite-sample version of Lemma 3.2 given in Remark 3.2 with $A=\{n\}$ for fixed n, gives

$$\inf_{\hat G_n(t)} \sup_{\theta \in C_n(\varepsilon)} P_{n,\theta}(|\hat G_n(t) - G_{n,\theta}(t)|>\delta) \ge 1-\Phi(\varepsilon/2)$$

for all $\delta<|\Phi(c_n+t) - \Phi(-c_n+t+\varepsilon)|/2$ and for all $0<\varepsilon<\min\{c_n, (\rho-|t|)/2\}$.

Observing that $C_n(\varepsilon) \subseteq B_n$ holds and letting ε go to zero establishes (2.17), from which we immediately obtain (2.18) by letting n go to infinity.

To prove (2.16), we apply Lemma 3.1 with $B_n = \{\theta \in \mathbb{R} : |\theta| < \rho n^{-1/2}\}$. The contiguity condition in that lemma follows immediately from Lemma A.1(b). Choose a sequence ε_n satisfying $0 < \varepsilon_n < \min\{c_n, (\rho - |t|)/2\}$. Then using (A.2) we obtain

$$\delta^* = \liminf_{n \to \infty} \omega(G_{n,\theta}(t), B_n) \geq \liminf_{n \to \infty} \omega(G_{n,\theta}(t), C_n(\varepsilon_n))$$

$$= \liminf_{n \to \infty} |\Phi(c_n + t) - \Phi(-c_n + t + \varepsilon_n)| = 1,$$

which completes the proof of (2.16). ∎

Appendix B: Proofs for Section 3

Proof of Lemma 3.1: (a) By definition of δ^*, we can find sequences $\beta_n^{(1)} \in B_n$ and $\beta_n^{(2)} \in B_n$ such that $\liminf_{n \to \infty} d(\varphi_n(\beta_n^{(1)}), \varphi_n(\beta_n^{(2)})) \geq \delta^*$. Fix a $\delta < \delta^*/2$. Since

$$d(\varphi_n(\beta_n^{(1)}), \varphi_n(\beta_n^{(2)})) \leq d(\varphi_n(\beta_n^{(1)}), \varphi_n(\alpha)) + d(\varphi_n(\beta_n^{(2)}), \varphi_n(\alpha)), \qquad (B.1)$$

it follows for any δ' satisfying $\delta < \delta' < \delta^*/2$ that at least one of the terms on the r.h.s. of (B.1) exceeds δ' for sufficiently large n. Thus we can find a sequence $\beta_n \in B_n$ (in fact $\beta_n \in \{\beta_n^{(1)}, \beta_n^{(2)}\}$) such that $\liminf_{n \to \infty} d(\varphi_n(\beta_n), \varphi_n(\alpha)) \geq \delta'$. For any δ'' satisfying $\delta < \delta'' < \delta'$ we have

$$P_{n,\beta_n}(d(\varphi_n(\beta_n), \varphi_n(\alpha)) > \delta'') \leq P_{n,\beta_n}(d(\hat{\varphi}_n, \varphi_n(\beta_n)) > \delta) + P_{n,\beta_n}(d(\hat{\varphi}_n, \varphi_n(\alpha)) > \delta'' - \delta).$$

$$(B.2)$$

By construction, the l.h.s. of (B.2) converges to one. By contiguity, by the assumption on $\hat{\varphi}_n$, and because $\delta'' - \delta > 0$, the second term on the r.h.s. of (B.2) converges to zero. Hence, the first term on the r.h.s. of (B.2) converges to one, which completes the proof of (3.3). Assume now that (3.4) is not true. Then we can find a sequence of estimators $\hat{\varphi}_n$ and a subsequence $n(i) \to \infty$, such that

$$\sup_{\beta \in B_{n(i)}} P_{n(i),\beta}(d(\hat{\varphi}_{n(i)}, \varphi_{n(i)}(\beta)) > \delta) \to 0.$$

It follows that $P_{n(i),\alpha}(d(\hat{\varphi}_{n(i)},\varphi_{n(i)}(\alpha))>\delta) \to 0$ along this subsequence.

Observe that the statistical experiment corresponding to the subsequence $n(i)$

and the estimators $\hat{\varphi}_{n(i)}$ satisfy the requirements of Lemma 3.1(a). Applying

the already established result (3.3) to this experiment we conclude that

$\sup_{\beta \in B_{n(i)}} P_{n(i),\beta}(d(\hat{\varphi}_{n(i)},\varphi_{n(i)}(\beta))>\delta) \to 1$, which leads to a contradiction.

(b) It suffices to show that $P_{n,\beta_n}(d(\hat{\varphi}_n,\varphi_n(\beta_n))\geq\delta) \to 0$ for every $\delta>0$ and

every sequence $\beta_n \in B_n$. Since $d(\varphi_n(\beta_n),\varphi_n(\alpha)) \leq \omega(\varphi_n,B_n)$ holds, this follows

from the assumption on $\hat{\varphi}_n$, from contiguity, and from

$P_{n,\beta_n}(d(\hat{\varphi}_n,\varphi_n(\beta_n))>\delta) \leq P_{n,\beta_n}(d(\hat{\varphi}_n,\varphi_n(\alpha))>\delta/2) + P_{n,\beta_n}(d(\varphi_n(\alpha),\varphi_n(\beta_n))>\delta/2).$ ∎

Inspection of the proof shows that Lemma 3.1 continues to hold if α is

replaced by $\alpha_n \in B_n$ at every occurrence in the formulation of the lemma.

<u>Proof</u> <u>of</u> <u>Lemma</u> <u>3.2</u>: Again we can find sequences $\beta_n^{(1)} \in B_n$ and $\beta_n^{(2)} \in B_n$ such that

$\liminf_{n \to \infty} d(\varphi_n(\beta_n^{(1)}),\varphi_n(\beta_n^{(2)})) \geq \delta^*$. For every given $\delta<\delta^*/2$ and for every

estimator $\hat{\varphi}_n$ we then have

$P_{n,\beta_n^{(1)}}(d(\varphi_n(\beta_n^{(1)}),\varphi_n(\beta_n^{(2)}))>2\delta) \leq P_{n,\beta_n^{(1)}}(d(\hat{\varphi}_n,\varphi_n(\beta_n^{(1)}))>\delta) +$

$P_{n,\beta_n^{(1)}}(d(\hat{\varphi}_n,\varphi_n(\beta_n^{(2)}))>\delta) \leq P_{n,\beta_n^{(1)}}(d(\hat{\varphi}_n,\varphi_n(\beta_n^{(1)}))>\delta) +$

$P_{n,\beta_n^{(2)}}(d(\hat{\varphi}_n,\varphi_n(\beta_n^{(2)}))>\delta) + ||P_{n,\beta_n^{(1)}} - P_{n,\beta_n^{(2)}}||_{TV} \leq$

$2P_{n,\beta_n}(d(\hat{\varphi}_n,\varphi_n(\beta_n))>\delta) + ||P_{n,\beta_n^{(1)}} - P_{n,\beta_n^{(2)}}||_{TV},$ (B.3)

for an appropriate choice $\beta_n \in \{\beta_n^{(1)},\beta_n^{(2)}\} \subseteq B_n$, the choice possibly depending on

δ and $\hat{\varphi}_n$. Now for every $\varepsilon>0$ there exists $n_0=n_0(\delta,\varepsilon)$ such that for $n\geq n_0$ the

left-most side of (B.3) equals one and $||P_{n,\beta_n^{(1)}}-P_{n,\beta_n^{(2)}}||_{TV}<\Gamma+\varepsilon$. Hence,

$P_{n,\beta_n}(d(\hat{\varphi}_n,\varphi_n(\beta_n))>\delta) \geq (1-\Gamma-\varepsilon)/2$ holds for $n\geq n_0$ and every estimator $\hat{\varphi}_n$. It

follows that

$\inf_{\hat{\varphi}_n} \sup_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n,\varphi_n(\beta))>\delta) \geq \inf_{\hat{\varphi}_n} P_{n,\beta_n}(d(\hat{\varphi}_n,\varphi_n(\beta_n))>\delta) \geq (1-\Gamma-\varepsilon)/2,$

holds for $n \geq n_0$, which implies (3.7). $\blacksquare$

<u>Lemma</u> <u>B.1</u>: Let B_n, $n \geq 1$, be a sequence of (non-empty) subsets of B. For every $k \geq 1$, $n \geq 1$, let C_{nk} be a (non-empty) subset of B_n. Define

$$\Gamma_k = \text{limsup}_{n \to \infty} \sup_{(\beta^{(1)}, \beta^{(2)}) \in C_{nk} \times C_{nk}} ||P_{n,\beta}{}^{(1)} - P_{n,\beta}{}^{(2)}||_{TV} \qquad (B.4)$$

and assume that $\Gamma_k < 1$ holds for every k. Suppose $\delta_k^* = \text{liminf}_{n \to \infty} \omega(\varphi_n, C_{nk})$ is positive for every k. Then

$$\sup_{\delta > 0} \text{liminf}_{n \to \infty} \inf_{\hat{\varphi}_n} \sup_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n, \varphi_n(\beta)) > \delta) \geq (1 - \inf_k \Gamma_k)/2 > 0 \qquad (B.5)$$

holds. If, additionally, $\delta^{**} = \inf_k \delta_k^*$ is positive, then even

$$\text{liminf}_{n \to \infty} \inf_{\hat{\varphi}_n} \sup_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n, \varphi_n(\beta)) > \delta) \geq (1 - \inf_k \Gamma_k)/2 > 0 \qquad (B.6)$$

holds for every $\delta < \delta^{**}/2$. In particular, if $\inf_k \Gamma_k = 0$, then the lower bound in (B.5)-(B.6) reduces to 1/2.

<u>Proof</u>: Applying Lemma 3.2 to $C_{n,k}$ and noting that $C_{n,k} \subseteq B_n$ we obtain

$$\text{liminf}_{n \to \infty} \inf_{\hat{\varphi}_n} \sup_{\beta \in B_n} P_{n,\beta}(d(\hat{\varphi}_n, \varphi_n(\beta)) > \delta) \geq (1 - \Gamma_k)/2$$

for every $\delta < \delta_k^*/2$ and for every $k \geq 1$. Inequality (B.6) is then an immediate consequence, and (B.5) follows since the l.h.s. of the above inequality is nonincreasing in δ. $\blacksquare$

<u>Remark</u> <u>B.1</u>: (Finite-Sample Version) Let A be a non-empty subset of ℕ. If the limit superior in (B.4) is replaced by the supremum over A and if the limit inferior in the definition of δ_k^* is replaced by the infimum over A, then (B.5)-(B.6) hold with the limit inferior replaced by the infimum over A.

<u>Proof</u> <u>of</u> <u>Corollary</u> <u>3.3</u>: Since α is interior to B_n, we may without loss of generality set $B_n = \{\beta \in B: \pi(\alpha, \beta) < \zeta_n\}$ for sufficiently small ζ_n that are positive and decrease to zero. By further reducing ζ_n if necessary, we may -- in view of the asymptotic continuity of the maps $\beta \to P_{n,\beta}$ at α and the triangle

inequality -- conclude that (3.6) holds with $\Gamma=0$. Since $\omega(\varphi,B_n) \geq \eta$ clearly

holds, δ^* defined in Lemma 3.2 is not less than η, which is positive due to

discontinuity of φ at α. An application of Lemma 3.2 then completes the

proof.   ■

Proof of Corollary 3.4: (a) We first construct, for every $k \geq 1$, sequences $\beta_{n,1}^{(k)}$

and $\beta_{n,2}^{(k)}$ in B_n satisfying

$$||\beta_{n,1}^{(k)} - \beta_{n,2}^{(k)}|| \leq 2\zeta_n/k \tag{B.7}$$

for every $n \geq 1$ and

$$\liminf_{n \to \infty} d(\varphi_n(\beta_{n,1}^{(k)}), \varphi_n(\beta_{n,2}^{(k)})) > \delta^*/(2k). \tag{B.8}$$

Since $\delta^*>0$ is assumed, we can obviously find $\beta_{n,1}^{(1)}$ and $\beta_{n,2}^{(1)}$ in B_n such that

(B.8) holds for $k=1$. The triangle inequality applied to $\beta_{n,1}^{(1)}$, $\beta_{n,2}^{(1)}$, and α

shows that also (B.7) is satisfied for $k=1$. For $k>1$ define $\chi_n^{(k)}(i) = \beta_{n,1}^{(1)} +$

$(i/k)(\beta_{n,2}^{(1)}-\beta_{n,1}^{(1)})$, $0 \leq i < k$, and observe that $\chi_n^{(k)}(i) \in B_n$ by convexity of B_n.

Since

$$d(\varphi_n(\beta_{n,1}^{(1)}), \varphi_n(\beta_{n,2}^{(1)})) \leq \sum_{i=0}^{k-1} d(\varphi_n(\chi_n^{(k)}(i)), \varphi_n(\chi_n^{(k)}(i+1))),$$

there exists an index i_n, $0 \leq i_n < k$, such that

$$d(\varphi_n(\chi_n^{(k)}(i_n)), \varphi_n(\chi_n^{(k)}(i_n+1))) \geq d(\varphi_n(\beta_{n,1}^{(1)}), \varphi_n(\beta_{n,2}^{(1)}))/k.$$

Define $\beta_{n,1}^{(k)}=\chi_n^{(k)}(i_n)$ and $\beta_{n,2}^{(k)}=\chi_n^{(k)}(i_n+1)$. Relation (B.8) immediately

follows, and (B.7) follows since $||\beta_{n,1}^{(k)} - \beta_{n,2}^{(k)}|| \leq ||\beta_{n,1}^{(1)} - \beta_{n,2}^{(1)}||/k$ holds by

construction.

Define the sets $C_{nk}=\{\beta_{n,1}^{(k)},\beta_{n,2}^{(k)}\}$ and apply Lemma B.1: Clearly, $\delta_k^* =$

$\liminf_{n \to \infty} \omega(\varphi_n,C_{nk})$ is not less than $\delta^*/(2k)$ and hence is positive for every

$k \geq 1$. Furthermore, Γ_k reduces to $\limsup_{n \to \infty} ||P_{n,\beta_{n,1}^{(k)}}-P_{n,\beta_{n,2}^{(k)}}||_{TV}$ which

converges to zero for $k \to \infty$ in view of (B.7) and the asymptotic uniform

equicontinuity assumption. Hence $\Gamma_k<1$ holds at least for $k \geq k_0$. The result

now follows from (B.5).

(b) Define $C_{nk}=\{\beta_{n,1}(k^{-1}),\beta_{n,2}(k^{-1})\}$ and apply Lemma B.1: Observe that

$\delta_k^* \geq \delta_*>0$, and hence $\delta^{**} \geq \delta_*$, holds in view of (3.11). Furthermore, Γ_k converges

to zero for $k \to \infty$ in view of (3.10) and the asymptotic uniform equicontinuity assumption. The result then follows from (B.6).   ∎

Proof of Lemma 3.5: (a) For every $\varepsilon > 0$ we can find γ_1, γ_2 in G_0 such that $d(\varphi_{\infty,\alpha}(\gamma_1), \varphi_{\infty,\alpha}(\gamma_2)) > \omega(\varphi_{\infty,\alpha}, G_0) - \varepsilon$. From pointwise convergence it follows that $\delta^* \geq \omega(\varphi_{\infty,\alpha}, G_0) - \varepsilon$, and hence $\delta^* \geq \omega(\varphi_{\infty,\alpha}, G_0)$ because ε was arbitrary. The oscillation $\omega(\varphi_{\infty,\alpha}, G_0)$ is positive, since $\varphi_{\infty,\alpha}$ is non-constant on G_0.

(b) Define $\beta_{n,i}(\varepsilon) = \alpha + \gamma_i(\varepsilon)\zeta_n$. Then (3.10)-(3.11) are obviously satisfied.

(c) Discontinuity implies that η_∞ is positive. Now, for every $0 < \delta_* < \eta_\infty$ we can find $\gamma_1(\varepsilon)$, $\gamma_2(\varepsilon)$ in G_0 (near γ_0) satisfying the assumptions of part (b).

(d) Follows trivially from uniform convergence.   ∎

Proof of Lemma 3.6: (a) Assume $P^*_{n,\alpha_n}(F^*_n) \to 0$. This can be rewritten as $E^*_{n,\alpha_n}(f_n) \to 0$, where f_n is given by $f_n(\omega) = \int 1_{F^*_n}(\omega,\xi)K_n(\omega,d\xi)$ and E^*_{n,α_n} denotes expectation w.r.t. P^*_{n,α_n}. Clearly, $0 \leq f_n \leq 1$ holds. For every $\varepsilon > 0$ we have $E^*_{n,\alpha_n}(f_n) \geq \varepsilon P^*_{n,\alpha_n}(f_n \geq \varepsilon)$, implying $P^*_{n,\alpha_n}(f_n \geq \varepsilon) \to 0$. By contiguity also $P^*_{n,\beta_n}(f_n \geq \varepsilon) \to 0$. Because of $0 \leq E^*_{n,\beta_n}(f_n) \leq \varepsilon + P^*_{n,\beta_n}(f_n \geq \varepsilon)$, we obtain $\limsup_{n \to \infty} E^*_{n,\beta_n}(f_n) \leq \varepsilon$. But this proves $P^*_{n,\beta_n}(F^*_n) = E^*_{n,\beta_n}(f_n) \to 0$.

(b) By specializing to the events $F \times \Xi_n$ with $F \in \mathscr{F}_n$, one immediately sees that the l.h.s. in (b) is not less than the r.h.s. The reverse inequality is proved, e.g., in Pötscher (2002), Remark 2.1.   ∎

Appendix C

In this appendix we show how the results in Section 3 can be easily transferred from a given statistical experiment to a derived experiment obtained by conditioning. We start with the following lemma.

Lemma C.1: (a) Let Q and R be probability measures defined on some measurable space $(\Omega, \mathcal{F})$, and let $E \in \mathcal{F}$ satisfy $Q(E) > 0$ and $R(E) > 0$. Then $||Q(.|E) - R(.|E)||_{TV}$

$\leq 2||Q-R||_{TV}/(\max(Q(E), R(E)))$.

(b) For $n \geq 1$, let Q_n and R_n be probability measures defined on measurable spaces $(\Omega_n, \mathcal{F}_n)$ and let $E_n \in \mathcal{F}_n$ satisfy $Q_n(E_n) > 0$ and $R_n(E_n) > 0$ for each $n \geq 1$. Suppose $\liminf_{n \to \infty} R_n(E_n) > 0$ holds. If R_n is contiguous w.r.t. Q_n, then $R_n(.|E_n)$ is contiguous w.r.t. $Q_n(.|E_n)$.

Proof: (a) For arbitrary $A \in \mathcal{F}$ we have

$|Q(A|E) - R(A|E)| = |R(E)Q(A \cap E) - Q(E)Q(A \cap E) + Q(E)Q(A \cap E) - Q(E)R(A \cap E)|/(Q(E)R(E)) \leq$

$|Q(E) - R(E)|Q(A \cap E)/(Q(E)R(E)) + |Q(A \cap E) - R(A \cap E)|/R(E) \leq 2||Q-R||_{TV}/R(E)$.

Reversing the rôles of Q and R we obtain $|Q(A|E) - R(A|E)| \leq 2||Q-R||_{TV}/Q(E)$.

(b) If $Q_n(F_n|E_n) \to 0$ then $Q_n(F_n \cap E_n) \to 0$. Contiguity gives $R_n(F_n \cap E_n) \to 0$. Since $\liminf_{n \to \infty} R_n(E_n) > 0$ holds, $R_n(F_n|E_n) \to 0$ follows. ∎

For the rest of this appendix we use the notation of Section 3. In particular, let B_n, $n \geq 1$, denote a sequence of (non-empty) subsets of B, and let $E_n \in \mathcal{F}_n$ satisfy $P_{n,\beta}(E_n) > 0$ for every $\beta_n \in B_n$ and $n \geq 1$. The first corollary follows immediately from Lemma C.1.

Corollary C.2: Suppose

$$M = \liminf_{n \to \infty} \inf_{\beta \in B_n} P_{n,\beta}(E_n) > 0 \qquad (C.1)$$

holds.

(a) If the sequence P_{n,β_n} is contiguous w.r.t. P_{n,α_n} for sequences $\alpha_n \in B_n$ and $\beta_n \in B_n$, then $P_{n,\beta_n}(.|E_n)$ is contiguous w.r.t. $P_{n,\alpha_n}(.|E_n)$.

(b) If $\{P_{n,\beta}: \beta \in B_n\}$ satisfies (3.6), then $\{P_{n,\beta}(.|E_n): \beta \in B_n\}$ satisfies (3.6) with Γ replaced by $2\Gamma/M$.

A simple sufficient condition for (C.1) is that $\liminf_{n\to\infty} P_{n,\alpha_n}(E_n)>0$ for some sequence $\alpha_n\in B_n$, and that P_{n,α_n} is contiguous w.r.t. P_{n,β_n} for all sequences $\beta_n\in B_n$.

Corollary C.3: Let $\alpha_n\in B_n$, $n\geq 1$, be a given sequence and assume that $m=\liminf_{n\to\infty} P_{n,\alpha_n}(E_n)>0$ holds. If $\{P_{n,\beta}: \beta\in B_n\}$ satisfies (3.6), then $\{P_{n,\beta}(.\,|E_n): \beta\in B_n\}$ satisfies (3.6) with Γ replaced by $4\Gamma/m$.

Proof: Applying the triangle inequality and Lemma C.1 we obtain for $\beta,\gamma\in B_n$

$$||P_{n,\beta}(.\,|E_n)-P_{n,\gamma}(.\,|E_n)||_{TV} \leq 2[||P_{n,\beta}-P_{n,\alpha_n}||_{TV} +$$

$$||P_{n,\alpha_n}-P_{n,\gamma}||_{TV}]/P_{n,\alpha_n}(E_n).$$ Taking the supremum over $B_n\times B_n$ followed by the limsup, leads to the bound $4\Gamma/m$. ∎

Remark C.1: The asymptotic continuity condition in Corollary 3.3 immediately carries over from the measures $P_{n,\beta}$ to the conditional measures $P_{n,\beta}(.\,|E_n)$ provided $\liminf_{n\to\infty} P_{n,\alpha}(E_n)>0$ holds. The same is true for the asymptotic uniform equicontinuity condition in Corollary 3.4 provided condition (C.1) is satisfied.

References

Adkins, L.C. (1990): "Small sample performance of jackknife confidence intervals for the James-Stein estimator," Commun. Statist. Simula., B19, 401-418.

Adkins, L.C. (1992): "Finite sample moments of a bootstrap estimator of the James-Stein rule," Econometric Reviews, 11, 173-193.

Bauer, P., B.M. Pötscher, and P. Hackl (1988): "Model selection by multiple test procedures," Statistics, 19, 39-44.

Chen, J. (1997): "A general lower bound of minimax risk for absolute-error loss," Canadian J. Statist., 25, 545-558.

Chen, S.S., D.L. Donoho, and M.A. Saunders (1999): "Atomic decomposition by basis pursuit," SIAM J. Sci. Statist. Comput., 20, 33-61.

Donoho, D.L., and R.C. Liu (1987): "Geometrizing rates of convergence I," Technical Report, Dept. of Statistics, Univ. of California, Berkeley.

Fan, J. and R. Li (2001): "Variable selection via nonconcave penalized likelihood and its oracle properties," J. Amer. Statist. Assoc., 96, 1348-1360.

Frank, I.E. and J. Friedman (1993): "A statistical view of some chemometrics regression tools (with discussion)," Technometrics, 35, 109-148.

Geyer, C.J. (1996): "On the asymptotics of convex stochastic optimization," unpublished manuscript, University of Minnesota.

Jennrich, R.I., and S.D. Oman (1986): "How much does Stein estimation help in multiple linear regression?," Technometrics, 28, 113-121.

Judge, G.G., and M.E. Bock (1978): The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics. North-Holland/Elsevier, Amsterdam.

Knight, K., and W. Fu (2000): "Asymptotics of Lasso-type estimators," Ann. Statist., 28,1356-1378.

Le Cam, L. (1973): "Convergence of estimates under dimensionality restrictions," Ann. Statist., 1, 38-53.

LeCam, L., and L. Schwartz (1960): "A necessary and sufficient condition for the existence of consistent estimators," Ann. Math. Statist., 31, 140-150.

Leeb, H., and B.M. Pötscher (2002): "Can one estimate the distribution of post-model-selection estimators?," manuscript, Department of Statistics, University of Vienna.

Lu, K.L. and J.O. Berger (1989): "Estimation of normal means: frequentist

estimation of loss," Ann. Statist., 17, 890-906.

Pfanzagl, J. (1998): "The nonexistence of confidence sets for discontinuous functionals," J. Statist. Plan. Inf., 75, 9-20.

Pötscher, B.M. (1991): "Effects of model selection on inference," Econometric Theory, 7, 163-185.

Pötscher, B.M. (2002): "Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters," Econometrica, 70, 1035-1065.

Sen, P.K. (1986): "Whither jackknifing in Stein-rule estimation?," Commun. Statist. Theor. Meth., A15, 2245-2266.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," J. Roy. Statist. Soc. Ser. B, 88, 267-288.

van der Vaart, A.W. (1998): Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

Venter, J.H., and S.J. Steel (1990): "Estimating risk reduction in Stein estimation," Canadian J. Statist., 18, 221-232.

Yatracos, Y.G. (1985): "On the existence of uniformly consistent estimators," Proc. Amer. Math. Soc., 94, 479-486.

Wan, A.T.K., and G. Zou (2002): "On estimating the losses of improved estimators," manuscript, Department of Management Sciences, City University of Hongkong.