COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 659

ROBUST AND ASYMPTOTICALLY EFFICIENT ESTIMATION OF LOCATION IN A STATIONARY STRONG MIXING GAUSSIAN PARAMETRIC MODEL

Donald W. K. Andrews

November 1982

ROBUST AND ASYMPTOTICALLY EFFICIENT ESTIMATION OF LOCATION IN A STATIONARY STRONG MIXING GAUSSIAN PARAMETRIC MODEL[1]

Donald W. K. Andrews

Yale University[2]

November 1982

Short title: Robust and Efficient Estimation

SUMMARY

ROBUST AND ASYMPTOTICALLY EFFICIENT ESTIMATION OF LOCATION IN A STATIONARY STRONG MIXING GAUSSIAN PARAMETRIC MODEL

This paper considers the problem of robust estimation of location in a model with stationary strong mixing Gaussian parametric distributions. An estimator is found that is within ε of being asymptotically efficient at the Gaussian parametric distributions <u>and</u> is within ε of being optimally robust! For the robustness results a Huber-type minimax criterion is used, where minimaxing takes place over neighborhoods of the parametric Gaussian distributions. The neighborhood system considered includes distributions of strong mixing processes and allows for deviations from the normal univariate parametric distributions within a Hellinger metric neighborhood, as well as deviations from stationarity and from the Guassian structure of dependence.

1. Introduction

Most of the literature on robust estimation considers models where the random variables are independent and identically distributed (i.i.d.). In this paper we consider robust estimation of the location parameter θ in the model

$$X_j = \theta + U_j \ , \quad j = 1, \ldots, n \ , \tag{1}$$

where the X_j's are non-i.i.d. random variables. Parametric distributions of the infinite process $\langle X_j \rangle_{j \geq 1} \equiv (X_1, X_2, \ldots)$ are specified to be mean θ stationary strong mixing Gaussian distributions. Neighborhoods of these parametric distributions are constructed. The neighborhoods consist of distributions of strong mixing processes which are "close" to the parametric distributions in terms of the Hellinger metric on the univariate marginal distributions and in terms of some weak-convergence-inducing metric on the higher order finite dimensional distributions. (The neighborhoods are defined precisely in Section 3 below.) The true distribution of $\langle X_j \rangle_{j \geq 1}$ is assumed to lie in one of the neighborhoods, and the distribution of $X_1, \ldots, X_n$ is just the distribution of the first n random variables of the infinite process $\langle X_j \rangle_{j \geq 1}$. The use of these neighborhoods allows for deviations from normality in "shape" of the univariate distributions of X_j, $j = 1, 2, \ldots,$ and for deviations from stationarity and from the Gaussian structure of dependence of the distribution of $\langle X_j \rangle_{j \geq 1}$.

As is common for i.i.d. robust models, an asymptotic framework is adopted for the present dependent random variable robust model. Asymptotics are used because the small sample properties of the estimators considered are too difficult to work with. An optimal estimator for this robust model is determined, following Huber (1964), using a minimax asymptotic risk principle. That is, for a given loss function the asymptotic risk of an estimator is calculated for each distribution in the neighborhood, and the optimal estimator is taken to be the one which minimizes the maximum asymptotic risk of the estimator over all distributions in the neighborhood. To avoid the calculation of an overly conservative estimator and to maintain the parametric nature of the model, the neighborhoods are designed to be reasonably small. In fact, the "size" of the neighborhoods is made to shrink to zero as the sample size, n, goes to infinity. That is, in the asymptotic framework (in which a sequence of statistical problems are considered, one for each sample size) the parametric distributions are taken to be the same for all sample sizes, but the neighborhoods of these parametric distributions depend on the sample size and decrease in "size" as $n \to \infty$. Shrinking (or infinitesimal) neighborhoods for i.i.d. robust models were first introduced by Huber-Carol (1973) and Jaeckel (1971), and have been used extensively by Beran (1977a, b, 1980), Rieder (1979, 1980, 1981a,b), and Bickel (1978, 1981).

The use of shrinking neighborhoods is justified as follows: In order to use asymptotic theory one must construct a sequence of statistical problems for increasing sample sizes. Obviously, there are numerous ways of constructing such sequences. The objective is to embed a fixed sample size problem in a sequence of problems so that the asymptotic properties of estimators approximate those of the finite sample size as closely as possible. If the "size" of the neighborhood is not a function of the sample size, then the asymptotic bias dominates the asymptotic variance, provided asymmetric distributions are allowed in the neighborhoods. If only symmetric neighborhood distributions are considered, then the asymptotic bias is zero (for reasonable estimators) and the asymptotic variance is the sole criterion of performance. Since the symmetry assumption is artificial, we are left with the unsatisfactory result that bias and variance are of different orders of magnitude. On the other hand, if the "size" of the neighborhood is made to shrink at rate $O(1/\sqrt{n})$ as $n \to \infty$, then bias and variance are of the same order of magnitude. In this case both the bias and the variance appear as criteria of performance asymptotically, which reflects the finite sample situation.

The class of estimators considered is that of weighted M-estimators. This is a generalization of the well-known class of M-estimators to include estimators whose defining equation

$$\hat{\theta} \text{ solves } \sum_{j=1}^{n} d_{nj}\psi(\theta, x_j) = 0$$

may include weights d_{nj} which depend on parameters of dependence between the random variables (such as correlations). For example, the finite sample uniformly minimum variance unbiased (UMVU) estimator (i.e., the generalized least squares estimator) for the given parametric model depends on such parameters. Within this class, an estimator is found which is within ε of being optimally robust and is within ε of being asymptotically efficient for the parametric model. Thus, robustness and efficiency are reconciled in this dependent model, as they are in the i.i.d. models of Beran (1977a) and Bickel (1981). The ε-optimal estimator has defining equation with equal weights and ψ function given by the normalized score function, viz. $x-\theta$, truncated at a <u>very</u> high point. This estimator gives the same estimate as the sample mean does, except for samples with <u>extremely</u> large outliers.

The proposed model is an extension of that of Bickel (1981) which, in turn, is closely related to the work of Beran, Rieder, and Hampel (1968, 1974, 1978). Other papers which consider robustness in models with dependent random variables, or which consider the behavior of standard i.i.d. robust estimators on dependent data include Hoyland (1968), Gastwirth and Rubin (1975), Koul (1977), Portnoy (1977, 1979), Martin and Zeh (1978), Bickel and Herzburg (1979), Kleiner, Martin, and Thompson (1979), and Martin (1979, 1980, 1981).

The paper is organized as follows: Section 2 gives definitions, regularity conditions, and an explicit statement of the problem whose solution determines an optimal robust estimator. Section 3 shows that the neighborhoods, as defined, contain a rich complement of members. Section 4 contains the results of the paper including those referred to above regarding the reconciliation of robustness and efficiency. Results from Andrews (1982) for models with general (i.e. maximum likelihood-type regular) scalar-parameter parametric univariate distributions and strong mixing dependence are discussed briefly in Section 5. The Appendix gives some proofs which are omitted in Section 4.

2. Definitions, Statement of the Problem, and Regularity Conditions

2.a. Definitions

2.a.1. The model under consideration is given by (1). The X_j are observed, and θ is an unknown parameter to be estimated. Parametric distributions, $\Phi_{\theta\Omega}$, of the infinite process $\langle X_j \rangle_{j\geq 1}$ are specified to be mean θ , (known) variance σ^2 , stationary, strong mixing, Gaussian distributions with (infinite dimensional) correlation matrix Ω . $\Omega \in S$, a specified set of correlation matrices which is such that $\Phi_{\theta\Omega}$ satisfies the mixing condition MIX (see below) for all $\Omega \in S$. The parameter space Θ is an open subset of R . Under $\Phi_{\theta\Omega}$, the univariate marginal distribution of X_j (for any j) is denoted by Φ_θ , and the bivariate marginal distribution of random variables X_i , X_j of distance r apart (i.e. $|i-j| = r$) is denoted $\Phi_{\theta\Omega}^{(2)r}$. General s^{th} order finite-dimensional distributions of $X_j, X_k, \ldots, X_\ell$, some collection of s distinct X's , is denoted by $\Phi_{\theta\Omega}^{(s)j,k,\ldots,\ell}$, for $s > 2$. The densities of Φ_θ and $\Phi_{\theta\Omega}^{(2)r}$, with respect to Lebesgue measure, are denoted by φ_θ and $\varphi_\theta^{(2)}(\cdot,\cdot,\rho_r)$, respectively, where $\underset{\sim}{\rho} \equiv (\rho_1, \rho_2, \ldots)'$ is the correlation vector corresponding to Ω .

(While it is assumed above that σ^2 is known, the ε-optimal estimator found below does not depend on σ^2 provided it is known that $\sigma^2 \leq M$, where M is some arbitrary fixed bound; see Section 4.)

2.a.2. A sequence of "neighborhoods," denoted by $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$, $n = 1, 2, \ldots$, of each parametric distribution $\Phi_{\theta\Omega}$ is constructed. Each neighborhood consists of distributions $F_{\theta\Omega n}$ of a real-valued process $\langle X_j \rangle_{j\geq 1}$ with the properties specified below. Let $F_{\theta n}^{(1)j}$ be the univariate marginal distribution of X_j under $F_{\theta\Omega n}$, for $j = 1, 2, \ldots$. Let $F_{\theta\Omega n}^{(s)j,k,\ldots,\ell}$, $s > 1$, be the general s^{th} order finite-dimensional distribution of $X_j, X_k, \ldots, X_\ell$, some collection of s distinct X's , under $F_{\theta\Omega n}$. The density of $F_{\theta n}^{(1)j}$, with respect to a specified measure, is denoted by $f_{\theta n}^{(1)j}$, $\forall j = 1, 2, \ldots$.

A distribution $F_{\theta\Omega n}$ is in the n^{th} neighborhood of $\Phi_{\theta\Omega}$, i.e., $F_{\theta\Omega n} \in \{\mathcal{F}^H_{t/\sqrt{n}}, \Omega\}_n$, if:

F1: $F_{\theta\Omega n}$ is the distribution of a strong mixing, real-valued process on the positive integers with mixing numbers less than or equal to $\tilde{\alpha}(j)$, $j = 1, 2, \ldots$, where $\tilde{\alpha}(j)$ satisfy condition MIX defined below (see Section 2.c).

F2: $F_{\theta n}^{(1)j} \in \mathcal{F}^{H}_{t/\sqrt{n}}(\theta)$, $\forall j = 1, 2, \ldots$, where $\mathcal{F}^{H}_{t/\sqrt{n}}(\theta)$ is a Hellinger metric neighborhood of radius $t/\sqrt{n}$ of the univariate, parametric distribution Φ_θ . That is,

$\mathcal{F}^{H}_{t/\sqrt{n}}(\theta) = \{F^{(1)} ; F^{(1)}$ is a distribution on R, and $\Delta_H(F^{(1)}, \Phi_\theta) \le t/\sqrt{n}\}$

where $\Delta_H(\cdot,\cdot)$ is the Hellinger metric on the space of distributions on R defined by

$$\Delta_H(F^{(1)}, \Phi_\theta) = \left[\int (\sqrt{f^{(1)}} - \sqrt{\varphi_\theta})^2 d\mu\right]^{1/2} ,$$

where $f^{(1)}(x)$ is the density of $F^{(1)}$ with respect to some measure μ , and $\mu >> F^{(1)}, \Phi_\theta$.

F3: Each s^{th} order finite-dimensional distribution of $F_{\theta\Omega n}$, for $s \ge 2$, lies in a Δ_W^s-metric neighborhood of the s^{th} order finite-dimensional distribution of $\Phi_{\theta\Omega}$, where Δ_W^s is any fixed but arbitrary weak-convergence inducing metric on R^s , and the size of each such neighborhood is $o(1)$ as $n \to \infty$. That is, for some $d_s(n)$

$$\Delta_W^s(F_{\theta\Omega n}^{(s)j,k,\ldots,\ell}, \Phi_{\theta\Omega}^{(s)j,k,\ldots,\ell}) \le d_s(n)$$

for all s-vectors $(j, k, \ldots, \ell)$ and $d_s(n) \xrightarrow{n\to\infty} 0$.

Comments: (i) For the results that follow F3 is more restrictive than necessary. Actually, only the bivariate marginal distributions of each neighborhood distribution, $F_{\theta\Omega n}$, need be in a Δ_W^s-metric neighborhood of size $o(1)$ of those of $\Phi_{\theta\Omega}$. The assumption that all s^{th} order finite-dimensional distributions lie in such Δ_W^s-metric neighborhoods is made because we then have a "natural" infinitesimal neighborhood system for which $F_{\theta\Omega n} \overset{\mathcal{L}}{\Rightarrow} \Phi_{\theta\Omega}$ as $n \to \infty$, $\forall F_{\theta\Omega n} \in \{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$, $n = 1, 2, \ldots$ and $\forall \Omega \in S$. That is, with the restrictions of F1-F3 each neighborhood distribution is "close" to $\Phi_{\theta\Omega}$ in that each univariate marginal distribution is close in Hellinger distance, and the distribution of the whole process is close in the sense of weak convergence.

(ii) t is fixed, and indexes the size of the univariate marginal distribution neighborhoods. In the independent model, Bickel (1981) finds [.5, 2.0] to be a reasonable range of values for t . However, the value of t turns out not to affect the optimal estimating procedure derived below, provided it is known that $t \leq t_0 < \infty$, where t_0 is some arbitrarily large bound.

(iii) The above univariate marginal distribution neighborhoods include asymmetric distributions. This is a particularly attractive feature of the model in comparison to various other robust models which must arbitrarily assume symmetry of the neighborhood distributions.

(iv) The neighborhoods are infinitesimal since they shrink $(t/\sqrt{n} \to 0,\ d_s(n) \to 0,\ s = 2, 3, \ldots)$ as $n \to \infty$.

Let $\langle F_{\theta\Omega n}\rangle_{n\geq 1}$ denote an arbitrary sequence of neighborhood distributions $F_{\theta\Omega n}$, $n = 1, 2, \ldots$. That is, $F_{\theta\Omega n} \in \{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$, $\forall n$, and $\Omega \in S$. Let u.f.θ,F,Ω abbrebiate "uniformly for $\theta \in C$, any compact set in Θ , uniformly for $\langle F_{\theta\Omega n}\rangle_{n\geq 1}$, and uniformly for $\Omega \in S$" .

2.a.3. Define $T = \langle T_n\rangle_{n\geq 1}$ to be a <u>weighted</u> <u>M-estimator</u> of θ if

(i) $\exists$ some estimator $\tilde{T} = \langle \tilde{T}_n\rangle_{n\geq 1}$ of θ which is consistent for θ u.f.θ,F,Ω . That is, $\forall J > 0$

$$\sup_{\theta\in\Theta}\ \sup_{\Omega\in S}\ \sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} F_{\theta\Omega n}(|\tilde{T}_n - \theta| > J) \xrightarrow{n\to\infty} 0 .$$

(ii) For some Lebesgue measurable function $\psi : \Theta \times R \to R$, and for some triangular array of weights $\{d_{nj}\}$

$$T_n \begin{cases} \text{The closest root of } \sum_{j=1}^{n} d_{nj}\psi(\theta, x_j) = 0 \text{ to } \tilde{T}_n & \text{if a root exists;} \\ \tilde{T}_n & \text{if no root exists.} \end{cases}$$

Let $\mathcal{M}$ denote the class of all weighted M-estimators whose ψ functions satisfy conditions A1, A2, and A3 specified below, and whose triangular array of weights satisfy D1 and D2 specified below.

Comments: (i) The class of equally weighted M-estimators is well known. It can be viewed as the class of estimators which are solutions to the likelihood equations for different i.i.d. parametric distributions. Alternatively, one can view M-estimators as the solution to

$$\min_{\theta} \sum_{j=1}^{n} D(x_j, \theta) \quad \text{where} \quad \frac{d}{d\theta}D(x_j, \theta) = \psi(\theta, x_j) ,$$

for different functions $D(x,\theta)$. Then, if $D(x,\theta)$ is considered to be the distance between x and θ , an (equally weighted) M-estimator minimizes the sum of the distances between θ and x_j , $\forall j = 1, \ldots, n$. Weighted M-estimators minimize the weighted sum of distances between θ and x_j , $\forall j = 1, \ldots, n$.

(ii) When the observations are i.i.d. the order statistics are sufficient, and the Rao-Blackwell theorem implies that it suffices to consider only estimators which are invariant under permutations of the data, Equally weighted M-estimators have this property. In the dependent case, however, the order statistics are not sufficient, and it is possible that estimators which are not invariant have advantageous properties. On the other hand, for the mean θ, stationary, strong mixing, Gaussian location model, $\overline{X}$ is asymptotically efficient for any Ω (with positive, continuous spectral density). This suggests that in the corresponding shape and dependence robust location model the optimal estimator may be invariant under permutations of the data. However, we do not want to impose this requirement. In fact, for finite sample sizes the generalized least squares estimator is the UMVU estimator, but is not invariant under permutations of the data and is not an M-estimator. It is a weighted M-estimator, however, with weights that are a function of the correlation matrix Ω. This estimator motivates the use of the class of weighted M-estimators.

2.a.4. Let $L : R \to R^{+} \cup \{0\}$ be a given loss function which satisfies conditions L1 and L2 specified below.

2.b. Statement of the Problem

The problem can now be stated as follows: Find an estimator $T = \langle T_n \rangle_{n \geq 1} \in \mathcal{M}$ of θ which inf's

$$\sup_{a>0} \overline{\lim_{n\to\infty}} \sup_{\Omega \in S} \sup_{\langle F_{\theta\Omega n} \rangle_{n \geq 1}} E_{F_{\theta\Omega n}} L_a(\sqrt{n}(T_n - \theta)) \tag{2}$$

where $L_a(x) \equiv L(x) \wedge a$. That is, we want to find an estimator in $\mathcal{M}$ which yields asymptotic, minimax risk over the distributions in the neighborhoods.

By taking the supremum over $F_{\theta\Omega n}$ before the limit supremum as $n \to \infty$ we are requiring uniform convergence over the neighborhoods. This is desirable for interpretation of the results in terms of finite sample sizes. By truncating L at height a, and then letting a increase to infinity, we are requiring the loss function to be bounded, but with arbitrarily large bound. This may appear to be innocuous, but can actually be of great significance (see Beran (1980)). The justification for the bounded loss function, L_a , is that for many situations the loss attributed to any error of great magnitude is constant. That is, if your estimate is "wild," the exact extend of its "wildness" is inconsequential.

Define $R_H(\psi, d, \theta, t, S)$ to be the minimand given in (2) where ψ is the function and d is the triangular array of weights which define the weighted M-estimator $T = \langle T_n \rangle_{n \geq 1}$.

2.c. Regularity Conditions

2.c.1. The following mixing condition is assumed to hold (see Philipp (1969) for a definition of strong mixing):

MIX: $\Phi_{\theta\Omega}$ and $F_{\theta\Omega n} \in \{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$, $\forall n$, $\forall \theta \in \Theta$, $\forall \Omega \in S$ are strong mixing with mixing numbers generically denoted by $\tilde{\alpha}(m)$, $m = 1, 2, \ldots$, and there exist non-negative, non-increasing numbers $\alpha(m)$, $m = 1, 2, \ldots$, such that for all $\Phi_{\theta\Omega}$ and $F_{\theta\Omega n}$

$$\tilde{\alpha}(m) \leq \alpha(m) , \quad m = 1, 2, \ldots ,$$

and $\alpha(m)$, $m = 1, 2, \ldots$ satisfy

$$\sum_{m=1}^{\infty} m^2 \alpha(m)^{\tau} < \infty , \quad \text{for some} \quad \tau \in (0,1) . \tag{3}$$

Comments: (i) If the dominating mixing numbers, $\alpha(m)$, $m = 1, 2, \ldots$, are eventually dominated by an exponentially decreasing sequence, then condition (3) of MIX clearly holds. In Lemma 1 below we show that this is indeed the case whenever $\alpha(m)$, $m = 1, 2, \ldots$ are the mixing numbers of any autoregressive-moving average (ARMA) stationary, Gaussian process.

(ii) MIX is needed to prove the weak convergence of the empirical process under $F_{\theta\Omega n} \in \{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$, $n = 1, 2, \ldots$, which is used in the proof of the existence of a consistent estimator (Lemma 3). The strong mixing central limit theorem used below requires the weaker condition

$$\sum_{m=1}^{\infty} m^2\alpha(m) < \infty .$$

Below we use the following inequality due to Ibragimov (1962) (see also, Philipp (1969)). It will be referred to as the mixing inequality.

Theorem. Suppose P is the distribution of a strong mixing process $Y_1, Y_2, \ldots$ with mixing numbers $\tilde{\alpha}(m)$, $m = 1, 2, \ldots$. Let $\mathcal{A}_{a,b}$ be the σ-algebra generated by $Y_a, Y_{a+1}, \ldots, Y_b$, for $1 \leq a \leq b \leq \infty$. If W and Z are bounded rv's measurable with respect to $\mathcal{A}_{1,s}$ and $\mathcal{A}_{s+m,\infty}$, respectively, then

$$|\mathrm{cov}_P(W,Z)| \leq 4 \text{ ess sup}|W| \cdot \text{ess sup}|Z| \cdot \tilde{\alpha}(m) . \qquad (4)$$

The set of admissable correlation matrices S is defined to include all infinite-dimensional correlation matrices which correspond to stationary Gaussian processes which satisfy condition MIX.

Comment: The stationarity condition implies that matrices $\Omega \in S$ are positive definite, symmetric, Toeplitz matrices with unit diagonal elements.

It is of interest to know more specifically which correlation matrices Ω are in S , and hence, which parametric Gaussian processes $\Phi_{\theta\Omega}$ are under consideration in the model. S depends on the condition

MIX and, in particular, upon the specific choice of the dominating mixing numbers $\alpha(m)$, $m = 1, 2, \ldots$. These numbers certainly can be chosen such that any M-dependent distribution $\Phi_{\theta\Omega}$ of interest has $\Omega \in S$. Further, more general conditions can be established which ensure that $\Phi_{\theta\Omega}$ is strong mixing and $\Omega \in S$:

Lemma 1. A real-valued stationary Gaussian process on the positive integers which has a continuous and positive spectral density, $f(\lambda)$, is strong mixing. Further, if $f(\lambda)$ is the ratio of polynomials in $e^{i\lambda}$, then the mixing numbers of the process, $\tilde{\alpha}(m)$, decrease exponentially, and, hence, satisfy

$$\sum_{m=1}^{\infty} m^2 \tilde{\alpha}(m)^{\tau} < \infty \quad \text{for some} \quad \tau \in (0,1) . \tag{5}$$

In particular, any stationary (finite-parameter) ARMA Gaussian process on the positive integers satisfies (5).

Comment: Hence, by specifying the dominating mixing numbers $\alpha(m)$, $m = 1, 2, \ldots$ appropriately, S can be made large enough to include any (finite-parameter) ARMA Gaussian process of interest, and more generally, any Gaussian process whose spectral density is continuous and positive.

Proof of Lemma 1. This lemma follows from Theorems 1, 2 and 4 of Kolmogorov and Rozanov (1960), plus the observation that the spectral density of any ARMA process is continuous, positive, and the ratio of polynomials in $e^{i\lambda}$. □

2.c.2. The following assumptions, concerning the function ψ of a weighted M-estimator, hold when specified:

A1: $\psi \in \Psi(\theta) = \left\{ \tilde{\psi} : \Theta \times R \to R \mid E_{\Phi_\theta} \tilde{\psi}(\theta, X) = 0,\ E_{\Phi_\theta} \tilde{\psi}^2(\theta, X) < \infty, \right.$

$$\left. \text{and}\quad E_{\Phi_\theta} \tilde{\psi}(\theta, X) \cdot \left(\frac{X-\theta}{\sigma^2} \right) = 1 \right\} .$$

A2: $\sup_{x \in R} |\psi(\theta, x)| \leq C_1(\theta)$, where $C_1(\theta)$ is bounded for $\theta \in C$, any compact set in Θ .

A3: $\sup_{x \in R} |\psi(\theta + h_1, x) - \psi(\theta + h_2, x)| \leq C_2(\theta) \cdot |h_2 - h_1|$ for all $|h_1|, |h_2| \leq \varepsilon$, where $\varepsilon > 0$ is fixed, and where $C_2(\theta)$ is bounded for $\theta \in C$, any compact set in Θ .

Comments: (i) The first and third conditions of A1 are merely normalization conditions. The second condition must hold if T is to have finite asymptotic variance under the parametric model, and hence, is not restrictive.

(ii) A2, the boundedness condition is restrictive, although any unbounded ψ function may be approximated by a sequence of bounded ψ functions.

(iii) A3, the Lipschitz condition, is only mildly restrictive since it corresponds to a desirable property for an estimator to have, viz. bounded local shift-sensitivity, as defined by Hampel (1974). It is a property of almost all robust M-estimators which have been suggested in the literature, and is weaker, or no stronger, than assumptions made by others when considering classes of M-estimators.

2.c.3. The following assumptions, concerning the triangular array of weights $\{d_{nj}\}$, hold when specified:

D1: $|d_{nj}| \leq B_d < \infty$, $\forall j = 1, \ldots, n$; $n = 1, 2, \ldots$;

D2: The following limit exists: $\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} d_{nj} = \eta \neq 0$.

Without loss of generality we can take $\eta = 1$, since the defining equation of a weighted M-estimator is unaffected by the scaling of the weights.

Also, without loss of generality we can assume the following limits exist:

$$\tilde{\eta} \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} |d_{nj}| \tag{6}$$

and

$$\omega_r \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} (d_{nj} - \bar{d}_n)(d_{nj+r} - \bar{d}_n) \quad \forall r = 0, 1, 2, \ldots \tag{7}$$

where $\bar{d}_n = \frac{1}{n}\sum_{j=1}^{n} d_{nj}$.

To see this, let

$$a_n = \frac{1}{n}\sum_{j=1}^{n} |d_{nj}| \quad \text{and} \quad w_{nr} = \frac{1}{n}\sum_{j=1}^{n-r} (d_{nj} - \bar{d}_n)(d_{nj+r} - \bar{d}_n) ,$$

$\forall r = 1, 2, \ldots$; $\forall n = 1, 2, \ldots$. a_n and w_{nr} take values from compact sets, viz., $a_n \in [0, B_d]$, $\forall n$ and $w_{nr} \in [-2B_d^2, 2B_d^2]$, $\forall r = 1, 2, \ldots$, $\forall n$. By the diagonal process we can find a subsequence $\langle n_m \rangle_{m\geq 1}$ of $n = 1, 2, \ldots$ such that $\tilde{\eta} \equiv \lim_{m\to\infty} a_{n_m}$, and $\omega_r \equiv \lim_{m\to\infty} w_{n_m r}$ $\forall r$, exist. The asymptotic risk of T calculated along the subsequence $\langle n_m \rangle_{m\geq 1}$ is less than or equal to that calculated along $n = 1, 2, \ldots$. Theorem 1 shows that an estimator T' , which has the same ψ function as that of T and which has weights $\{d_{nj}\}$ such that $\tilde{\eta}' = \lim_{n\to\infty} a'_n$ and $\omega'_r \equiv \lim_{n\to\infty} w'_{nr}$, $\forall r$ exist and equal $\tilde{\eta}$ and ω_r , $\forall r$, respectively, has the same asymptotic risk as the risk of T calculated along the subsequence $\langle n_m \rangle_{m\geq 1}$.

This implies that only weights, $\{d_{nj}\}$, for which $\lim_{n\to\infty} a_n$ and $\lim_{n\to\infty} w_{nr}$, $\forall r$ exist need be considered.

N.B. Assumption D2 can be replaced by the assumption that d_{nj} are bounded away from zero.

Comment: Triangular arrays which satisfy D1 and D2 exist in abundance. In fact, with probability one, the observed sequence from an Erodic process of bounded random variables will satisfy the conditions.

2.c.4. The loss function $L : R \to R^+ \cup \{0\}$ satisfies either or both of the following conditions, when specified:

L1: L is symmetric about 0 and increasing on R^+.

L2: L is convex.

3. Richness of the Neighborhoods

We can demonstrate that the neighborhoods as defined are quite rich. Suppose the infinite sequence $\langle Y_j \rangle_{j\geq 1}$ has parametric distribution $\Phi_{\theta\Omega}$, for some $\theta \in \Theta$ and $\Omega \in S$. Consider co-ordinate-wise transformations of $\langle Y_j \rangle_{j\geq 1}$:

$$Z_{nj} = V_{nj}^{-1}(Y_j) \quad \text{for} \quad j = 1, 2, \ldots$$

where $V_{nj}^{-1}(\cdot)$ is any function which yields a univariate marginal distribution for Z_{nj} which is in $\mathcal{F}^H_{t/\sqrt{n}}(\theta)$ (the univariate neighborhood of Φ_θ). All distributions in $\mathcal{F}^H_{t/\sqrt{n}}(\theta)$ can be attained in this manner. $\langle Z_{nj} \rangle_{j\geq 1}$ has mixing numbers less than or equal to those of $\Phi_{\theta\Omega}$, since the σ-algebra generated by a co-ordinate-wise transformation of a sequence of rv's is contained in that generated by the sequence of rv's. Also, $\langle Z_{nj} \rangle_{j\geq 1}$ has finite-dimensional distributions of order greater than one that are of variational distance $o(1)$ as $n \to \infty$ from those of $\Phi_{\theta\Omega}$. Hence, the distributions of all such processes $\langle Z_{nj}\rangle_{j\geq 1}$ are in the neighborhoods $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$, $\forall n$. Since the univariate Hellinger neighborhood $\mathcal{F}^H_{t/\sqrt{n}}(\theta)$ of $\Phi^{(1)}_\theta$ is quite rich (see Bickel (1981)), the above shows that the process neighborhood $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$ also contains a rich complement of members.

Particular examples of point transformations, which generate a set of distributions which are "dense" (in a certain sense) in the neighborhoods, are as follows: Let

$$V_{nj}(y) = y + \frac{1}{\sqrt{n}}\bar{\Delta}(y)$$

where $$\bar{\Delta}(y) \equiv \varphi_\theta(y)^{-1} \int_{-\infty}^{y} h(w)\varphi_\theta(w)dw ,$$

and where $$h \in \mathcal{H}_H(\theta) \equiv \left\{ h \in L_\infty : \int_R h\varphi_\theta = 0, \int_R h^2\varphi_\theta < 4t^2 \right\} .$$

<u>Lemma 2</u>. Suppose $\langle Y_j\rangle_{j\geq 1}$ has distribution given by $\Phi_{\theta\Omega}$ for some $\theta \in \Theta$ and $\Omega \in S$ and let $V_{nj}(\cdot)$ be defined as above. Then, the distribution of $\langle V^{-1}_{nj}(Y_j)\rangle_{j\geq 1}$ is in the neighborhood $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$ of $\Phi_{\theta\Omega}$.

Comment: These distributions (generated by the functions $h \in \mathcal{H}_H(\theta)$) are "dense" in the neighborhoods in the sense that the asymptotic bias of an estimator under <u>any</u> given sequence of distributions in the neighborhoods can be attained (or approached) by that of a sequence of these distributions. In particular, the asymptotic <u>maximum</u> bias, $b_H(\psi,d,\theta)$ (defined below), can be attained (or approached) by sequences of these distributions (see Bickel (1981), Theorem 2).

<u>Proof of Lemma 2</u>. $\langle V_{nj}^{-1}(Y_j)\rangle_{j\geq 1}$, being a co-ordinate-wise transformation of $\langle Y_j\rangle_{j\geq 1}$, has the same or smaller mixing numbers as $\langle Y_j\rangle_{j\geq 1}$, so F1 is satisfied.

F3 is shown by writing the density of the s^{th}-order finite-dimensional distribution of $V_{nj}^{-1}(Y_j)$, $V_{nk}^{-1}(Y_k)$, ..., $V_{n\ell}^{-1}(Y_\ell)$ (denoted by $f_{\theta\Omega n}^{h(s)j,k,\ldots,\ell}(y_j, y_k, \ldots, y_\ell)$ as a function of $y_j, y_k, \ldots, y_\ell$. That is, by change of variables,

$$f_{\theta\Omega n}^{h(s)j,k,\ldots,\ell}(y_j, y_k, \ldots, y_\ell) = \varphi_{\theta\Omega}^{(s)j,k,\ldots,\ell}(V_{nj}(y_j), V_{nk}(y_k), \ldots, V_{n\ell}(y_\ell)) \cdot \prod_{q=j}^{\ell} \frac{d}{dy_q} V_{nq}(y_q) .$$

A Taylor expansion of $\varphi_{\theta\Omega}^{(s)j,k,\ldots,\ell}(V_{nj}(y_j), V_{nk}(y_k), \ldots, V_{n\ell}(y_\ell))$, calculation of $\prod_{q=j}^{\ell} \frac{d}{dy_q} V_{nq}(y_q)$, and simplification of terms shows that $f_{\theta\Omega n}^{h(s)j,k,\ldots,\ell}$ and $\varphi_{\theta\Omega}^{h(s)j,k,\ldots,\ell}$ differ only by terms of order $O(1/\sqrt{n})$. Thus, letting $F_{\theta\Omega n}^{h(s)j,k,\ldots,\ell}$ denote the distribution of the s rv's $V_{nj}^{-1}(Y_j)$, $V_{nk}^{-1}(Y_k)$, ..., $V_{n\ell}^{-1}(Y_\ell)$, we have $\Delta_V(F_{\theta\Omega n}^{h(s)j,k,\ldots,\ell}, \varphi_{\theta\Omega}^{(s)j,k,\ldots,\ell}) < d_s(n)$, where $d_s(n) = o(1)$ as $n \to \infty$, and where $\Delta_V(\cdot,\cdot)$ is the (weak-convergence inducing) variational metric. So, F3 is satisfied.

F2 is proved in Bickel (1981, Lemma 2). □

4. Results

First, the existence of a consistent estimator $\tilde{T}$ of θ u.f.θ,F,Ω needs to be shown. This result implies that the weighted M-estimators are well-defined.

<u>Lemma 3</u>. $\exists$ an estimator $\tilde{T} = \{T_n\}_{n\geq 1}$ such that $T_n - \theta = o_{F_{\theta\Omega n}}(1)$ as $n \to \infty$, u.f.θ,F,Ω .

Comment: The proof is in the Appendix. It is an extension to the strong mixing case of Wang's (1979) extension of LeCam's (1966) result.

We now present various results used in the solution of the problem (2).

Lemma 4. If $\overline{\lim_{n\to\infty}} E_{F_{\theta\Omega n}} L_a(\sqrt{n}(T_n - \theta))$ converges uniformly for $\langle F_{\theta\Omega n}\rangle_{n\geq 1}$ and for $\Omega \in S$, then

$$R_H(\psi,d,\theta,t,S) = \sup_{a>0} \sup_{\Omega\in S} \sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} \overline{\lim_{n\to\infty}} E_{F_{\theta\Omega n}} L_a(\sqrt{n}(T_n - \theta)) .$$

Proof of Lemma 4. By the uniform convergence assumption, the limit as $n \to \infty$ and the supremum over $\Omega \in S$ and $\langle F_{\theta\Omega n}\rangle_{n\geq 1}$ can be exchanged. □

Theorem 1. For $T \in \mathcal{M}$,

$$\mathcal{L}_{F_{\theta\Omega n}}\left(\sqrt{n}\left(T_n - \theta - \frac{1}{n}\sum_{j=1}^{n} \frac{d_{nj}}{\eta}\lambda(\psi,\ \theta,\ F_{\theta n}^{(1)j}\right)\right)$$

$$\to N\left(0,\ \frac{1}{\eta^2}\sum_{j=-\infty}^{\infty}(\eta^2 + \omega_{|j|})E_{\Phi_{\theta\Omega}}\psi(\theta,\ X_0)\psi(\theta,\ X_{|j|})\right)$$

as $n \to \infty$, u.f. θ, F, Ω, where $\lambda(\psi,\theta,F) \equiv E_F\psi(\theta,X)$.

N.B.: $\mathcal{L}_F(Y)$ denotes the distribution (or law) of Y under the probability measure F.

Comments: (i) The theorem is proved uniformly for $F_{\theta\Omega n} \in \{\mathcal{F}^H_{t/\sqrt{n}}(\theta),\ \Omega\}_n$, $\forall n$, and uniformly for $\Omega \in S$ because the minimand $R_H(\psi,d,\theta,t,S)$ takes the supremum over $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta),\ \Omega\}_n$ and over S, before taking the limit supremum as $n \to \infty$. This ordering (of sup's) is necessary to give a meaningful finite sample interpretation to the results.

(ii) The theorem is proved uniformly for θ in compact sets, again for the reason of finite sample interpretation. Uniformity over θ is needed so that the sample size required for the asymptotic results to be applicable to a finite sample situation is not a function of the unknown parameter.

(iii) The theorem holds for the well-known class of (unweighted) M-estimators (since this class is contained in $\mathcal{M}$).

(iv) This theorem can be generalized to models where the parametric distributions are general, stationary, strong mixing, single parameter distributions whose univariate marginal distributions satisfy certain maximum likelihood-type regularity conditions (see Section 5 and Andrews (1982)).

The proof of Theorem 1 uses the following lemmas. For each lemma assume $T \in \mathcal{M}$.

Lemma 5. $\lambda(\psi, s, \Phi_\theta)$ is differentiable in s, and $\frac{\partial}{\partial s}\lambda(\psi, s, \Phi_\theta)\big|_{s=\theta} = -1$.

Lemma 6.

(a)
$$\left|\mathrm{Var}_{F_{\theta\Omega n}}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\psi(\theta,X_j)\right) - \sum_{j=-\infty}^{\infty}(\eta^2+\omega_{|j|})E_{\Phi_{\theta\Omega}}\psi(\theta,X_0)\psi(\theta,X_{|j|})\right| = o(1) \quad (8)$$

as $n \to \infty$, u.f. θ,F,Ω;

and (b) $\sum_{j=-\infty}^{\infty}(\eta^2+\omega_{|j|})E_{\Phi_{\theta\Omega}}\psi(\theta,X_0)\psi(\theta,X_{|j|}) \in (0,\infty)$.

Lemma 7. (a) $T_n \xrightarrow{F_{\theta\Omega n}} \theta$ u.f. θ,F,Ω;

and (b) $F_{\theta\Omega n}(T_n \text{ is a root of } \sum_{j=1}^{n} d_{nj}\psi(\theta,X_j) = 0) \xrightarrow{n\to\infty} 1$ u.f. θ,F,Ω. (9)

Lemma 8. $\int[\psi(T_n,\ \cdot) - \psi(\theta,\cdot)]d\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} d_{nj}(F_n^{(1)j} - \Phi_\theta)\right) = o_{F_{\theta\Omega n}}(1)$ (10)

u.f. θ, F, Ω .

Lemma 9. Suppose $\alpha(m)$, $m = 1, 2, \ldots,$ are as in MIX. Then, for any given $\varepsilon > 0$ and $\delta \in [1/2, 1)$, $\exists$ a constant $\bar{k} \in (0,\infty)$ which does not depend on ε or δ such that $\sum_{m=[(\bar{k}\varepsilon)^{-\delta}]+1}^{\infty} \alpha(m) < \varepsilon$.

Define $Z_n(s) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}(\psi(s,\ X_j) - \lambda(\psi,\ s,\ F_{\theta n}^{(1)j}))$. (11)

Lemma 10. For any sequence $\varepsilon_n > 0$, $n = 1, 2, \ldots,$ such that $\varepsilon_n \xrightarrow{n\to\infty} 0$,

$$\sup_{s:|s-\theta|\leq\varepsilon_n} |Z_n(s) - Z_n(\theta)| \xrightarrow{F_{\theta\Omega n}} 0 \quad \text{u.f.}\ \theta, F, \Omega\ . \tag{12}$$

Lemma 11. $|Z_n(T_n) - Z_n(\theta)| \xrightarrow{F_{\theta\Omega n}} 0$ u.f. θ, F, Ω . (13)

The proofs of Lemmas 5-11 follow that of the theorem.

Proof of Theorem 1. The theorem is an extension of Bickel's (1981) Theorem 5 from i.i.d. parametric and neighborhood distributions to stationary strong mixing parametric distributions and (not necessarily stationary) strong mixing neighborhood distributions, and from M-estimators to weighted M-estimators. (See also, Rieder (1980).)

By a Taylor series expansion,

$$\lambda(\psi,\ T_n,\ \Phi_\theta) = \lambda(\psi,\ \theta,\ \Phi_\theta) + \frac{\partial}{\partial s}\lambda(\psi,\ s,\ \Phi_\theta)\Big|_{s=\theta}\cdot(T_n - \theta) + o_{F_{\theta\Omega n}}(1) \quad \text{u.f.}\ \theta, F, \Omega\ , \tag{14}$$

using Lemmas 5 and 7. So, by Lemma 5,

$$\lambda(\psi,\ T_n,\ \Phi_\theta) = (-1 + o_{F_{\theta\Omega n}}(1))\cdot(T_n - \theta) \quad \text{u.f.}\theta,F,\Omega\ . \tag{15}$$

Next,

$$-\sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n} d_{nj}\lambda(\psi,\ T_n,\ F_{\theta n}^{(1)j})\right\}$$

$$= -\sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n} d_{nj}\lambda(\psi,\ T_n,\ \Phi_\theta) + \frac{1}{n}\sum_{j=1}^{n} d_{nj}\lambda(\psi,\ \theta,\ F_{\theta n}^{(1)j})\right\}$$

$$- \int[\psi(T_n,\cdot) - \psi(\theta,\cdot)]d\sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n} d_{nj}(F_{\theta n}^{(1)j} - \Phi_\theta)\right\} \quad \text{by algebra,}$$

$$= \sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n} d_{nj}[(T_n - \theta)(1 + o_{F_{\theta\Omega n}}(1))] - \frac{1}{n}\sum_{j=1}^{n} d_{nj}\lambda(\psi,\ \theta,\ F_{\theta n}^{(1)j})\right\}$$

$$+ o_{F_{\theta\Omega n}}(1)\ , \tag{16}$$

the first term by (15), and the second by Lemma 8. Hence, the asymptotic law of $-\sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n} d_{nj}\lambda(\psi,\ T_n,\ F_{\theta n}^{(1)j})\right\}$ is the same as that of

$$\sqrt{n}\cdot\eta\left\{T_n - \theta - \frac{1}{n}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\ \theta,\ F_{\theta n}^{(1)j})\right\}\ , \tag{17}$$

the random variable of interest (since by D2, $\frac{1}{n}\sum_{j=1}^{n} d_{nj} = \eta + o(1)$).

Now, with $F_{\theta\Omega n}$ probability that approaches 1 u.f.θ,F,Ω as $n \to \infty$, $-\sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n} d_{nj}\lambda(\psi,\ T_n,\ F_{\theta n}^{(1)j})\right\} = Z_n(T_n)$, because $\sum_{j=1}^{n} d_{nj}\ (T_n,\ X_j) = 0$ if $T_n \neq \tilde{T}_n$, and by Lemma 7, $F_{\theta\Omega n}(T_n \neq \tilde{T}_n) \xrightarrow{n\to\infty} 1$ u.f.θ,F,Ω . Letting "with $F_{\theta\Omega n}$ probability $\to 1$" be absorbed in $o_{F_{\theta\Omega n}}(1)$ and using Lemma 11 gives:

$$-\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} d_{nj}\psi(\theta,\ T_n,\ F_{\theta n}^{(1)j})\right) = Z_n(\theta) + o_{F_{\theta\Omega n}}(1) \quad \text{u.f.}\theta,F,\Omega\ . \quad (18)$$

Lastly, the triangular array central limit theorem for bounded strong mixing random variables gives

$$\frac{Z_n(\theta)}{\mathrm{Var}_{F_{\theta\Omega n}}(Z_n(\theta))} = \frac{\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}(\psi(\theta,\ X_j) - E_{F_{\theta n}^{(1)j}}\psi(\theta,\ X_j)\right)}{\mathrm{Var}_{F_{\theta\Omega n}}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\psi(\theta,\ X_j)\right)} \Rightarrow N(0,1)$$

$$\text{as} \quad n \to \infty \quad \text{u.f.}\theta,F,\Omega\ . \qquad (19)$$

(See Theorem 3.1 of Mehra and Rao (1975) which is based on Theorem 3 of Philipp (1969). This theorem applies u.f.θ,F,Ω if MIX holds. Let Y_{Nj} of their notation be $\frac{d_{nj}\psi(\theta,\ X_j)}{B_d \cdot C_1(\theta)}$, where $\langle X_j \rangle_{j\geq 1} \sim F_{\theta\Omega n}$. Then, the following conditions of their theorem are satisfied: $|Y_{Nj}| \leq 1$,

$$\underline{\lim}_{n\to\infty} \frac{\mathrm{Var}_{F_{\theta\Omega n}}\left(\sum_{j=1}^{n} d_{nj}\psi(\theta,\ X_j)\right)}{\sum_{j=1}^{n} d_{nj}^2} = \frac{1}{\eta^2 + \omega_0}\ \underline{\lim}_{n\to\infty}\ \mathrm{Var}_{F_{\theta\Omega n}}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\psi(\theta,\ X_j)\right) > 0$$

u.f.θ,F,Ω by Lemma 6; and

$$\max_{1\leq \ell \leq n} \frac{|d_{n\ell}|}{\left(\sum_{j=1}^{n} d_{nj}^2\right)} \leq \frac{\frac{1}{n}\cdot B_d}{\left(\frac{1}{n}\sum_{j=1}^{n}|d_{nj}|\right)} \text{ by the Cauchy-Schwartz inequality and D1,}$$

$$= o(1) \quad \text{as} \quad n \to \infty \quad \text{since} \quad \frac{1}{n}\sum_{j=1}^{n}|d_{nj}| \xrightarrow{n\to\infty} \tilde{\eta} \quad \text{by (6)}\Big)$$

So,

$$\sqrt{n}\,\eta\left(T_n - \theta - \frac{1}{n}\sum_{j=1}^{n} \frac{d_{nj}}{\eta}\lambda(\psi,\ \theta,\ F_{\theta n}^{(1)j})\right)$$

$$\Rightarrow N\left(0,\ \sum_{j=-\infty}^{\infty} (\eta^2 + \omega_{|j|}) E_{\Phi_{\theta\Omega}} \psi(\theta,\ X_0)\psi(\theta,\ X_{|j|})\right)$$

as $n \to \infty$ u.f.θ,F,Ω, by (17), (19), and Lemma 6. □

Proof of Lemma 5. Define

$$A_1(h) \equiv h^{-1}[\lambda(\psi,\ \theta+h,\ \Phi_\theta) - \lambda(\psi,\ \theta,\ \Phi_\theta)]$$

$$= -\int \psi(\theta,\cdot)\left[\frac{\varphi_{\theta+h} - \varphi_\theta}{h}\right]d\mu - \int\Big(\psi(\theta+h,\cdot) - \psi(\theta,\cdot)\Big)\left[\frac{\varphi_{\theta+h} - \varphi_\theta}{h}\right]d\mu$$

$$\equiv A_2(h)$$

by adding and subtracting $h^{-1}\left[\int\Big(\psi(\theta+h,\cdot) - \psi(\theta,\cdot)\Big)\varphi_{\theta+h}d\mu\right]$. Letting $h \to 0$ we get

$$\lim_{h\to 0} A_1(h) = \frac{\partial}{\partial s}\lambda(\psi,\ s,\ \Phi_\theta)\Big|_{s=\theta}$$

provided $\lambda(\psi,\ s,\ \Phi_\theta)$ is differentiable in s at $s = \theta$. And $\lim_{h\to 0} A_2(h) = W_1 + W_2$ where

$$W_1 = \lim_{h\to 0} -\int \psi(\theta,\cdot)\left[\frac{\varphi_{\theta+h} - \varphi_\theta}{h}\right],$$

and

$$W_2 = \lim_{h\to 0} -\int\Big(\psi(\theta+h,\cdot) - \psi(\theta,\cdot)\Big)\cdot\left[\frac{\varphi_{\theta+h} - \varphi_\theta}{h}\right].$$

To evaluate W_1,

$$\lim_{h\to 0}\int_R \left|\frac{\varphi_{\theta+h}-\varphi_\theta}{h}\right| = \lim_{h\to 0}\int_R\int_0^1 |\dot{\varphi}_{\theta+\lambda h}(x)|\,d\lambda dx \quad \text{where} \quad \dot{\varphi}_\theta(x) \equiv \frac{d}{d\theta}\varphi_\theta(x) ,$$

using the Lebesgue absolute continuity of the function $\theta \mapsto \varphi_\theta(x)$ for all x ,

$$= \lim_{h\to 0}\int_0^1\int_R |\dot{\varphi}_{\theta+\lambda h}(x)|\,dxd\lambda \quad \text{by Fubini's theorem,}$$

$$= \int_0^1 \lim_{h\to 0}\int_R |\dot{\varphi}_{\theta+\lambda h}(x)|\,dxd\lambda \quad \text{by the bounded convergence}$$

theorem,

$$= \int_R |\dot{\varphi}_\theta(x)|\,dx \quad \text{using the continuity at } \eta = \theta \text{ of} \tag{20}$$

the function $\eta \mapsto \int_R |\dot{\varphi}_\eta(x)|\,dx$.

Since $\varphi_\theta(x)$ is differentiable in θ for all x ,

$$\frac{\varphi_{\theta+h}(x) - \varphi_\theta(x)}{h} \xrightarrow{h\to 0} \dot{\varphi}_\theta(x) . \tag{21}$$

Thus (20), (21), and the dominated convergence theorem imply

$$\lim_{h\to 0}\int_R \frac{\varphi_{\theta+h}-\varphi_\theta}{h} = \int_R \dot{\varphi}_\theta .$$

And, by the boundedness of $\psi(\theta,\cdot)$,

$$\lim_{h\to 0}\int_R \psi(\theta,\cdot)\frac{\varphi_{\theta+h}-\varphi_\theta}{h} = \int_R \psi(\theta,\cdot)\dot{\varphi}_\theta = 1 ,$$

the second equality by A1. That is, $W_1 = -1$.

Now,

$$|W_2| \le \lim_{h\to 0} \int_R |\psi(\theta+h,\cdot) - \psi(\theta,\cdot)| \cdot \left|\frac{\varphi_{\theta+h} - \varphi_\theta}{h}\right| d\mu$$

$$\le \lim_{h\to 0} \int_R C_2(\theta)|h| \cdot \left|\frac{\varphi_{\theta+h} - \varphi_\theta}{h}\right| d\mu \quad \text{by A3,}$$

$$= C_2(\theta) \int_R \lim_{h\to 0} |\varphi_{\theta+h} - \varphi_\theta| d\mu \quad \text{by the bounded convergence theorem,}$$

$$= 0$$

since φ_θ is continuous in θ. Hence, $\lim_{h\to 0} A_2(h) = W_1 + W_2$ exists, and equals -1. This implies $\lim_{h\to 0} A_1(h) = \frac{\partial}{\partial s}\lambda(\psi, \theta, \Phi_\theta)\Big|_{s=\theta}$ exists and equals -1. □

<u>Proof of Lemma 6</u>. $\sum_{r=0}^{\infty} \alpha(r) < \infty$ by MIX, so given any $\varepsilon > 0$, $\exists\ r_\varepsilon$ such that

$$8B_d^2 C_1^2(\theta) \cdot \sum_{r=r_\varepsilon}^{\infty} \alpha(r) < \varepsilon/2 . \tag{22}$$

Now, by condition F3 of the definition of $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$, $F^{(2)j,k}_{\theta\Omega n} \to \Phi^r_{\theta\Omega}$ uniformly for $j, k = 1, 2, \ldots$ such that $|j-k| = r$ and u.f.θ,F,Ω, where $F_{\theta\Omega n}$ is from some sequence $\langle F_{\theta\Omega n}\rangle_{n\ge 1}$. This, plus A2 (which implies $\psi(\theta, X_j)$ is bounded), gives

$$E_{F^{(2)j,k}_{\theta\Omega n}} \psi(\theta, X_j)\psi(\theta, X_k) - E_{\Phi^{(2)r}_{\theta\Omega}} \psi(\theta, X_j)\psi(\theta, X_k) = o(1)$$

uniformly for $j, k = 1, 2, \ldots$ such that $|j-k| = r$, uniformly for all $r = 1, 2, \ldots$, and u.f.θ,F,Ω. Also, $E_{F^{(1)j}_{\theta n}} \psi(\theta, X_j) = o(1)$ uniformly for $j = 1, 2, \ldots$, and u.f.θ,F,Ω (since $E_{\Phi_\theta}\psi(\theta, X_j) = 0$).

This implies that given ε as above, $\exists\, n_\varepsilon$ such that $\forall n \geq n_\varepsilon$

$$q^{(n)}_{|j-k|} \equiv \sup_{\substack{j',k': \\ |j'-k'|=|j-k| \\ \forall j',k'=1,2,\ldots}} \left| E_{F_{\theta\Omega n}}\left[\left(\psi(\theta, X_{j'}) - E_{F_\theta^{(1)j'}}\psi(\theta, X_{j'})\right)\right.\right.$$

$$\left.\left. \times \left(\psi(\theta, X_{k'}) - E_{F_\theta^{(1)k'}}\psi(\theta, X_{k'})\right)\right] - E_{\Phi_{\theta\Omega}}\psi(\theta, X_{j'})\psi(\theta, X_{k'})\right| .$$

$$< \varepsilon/[4B_d^2 r_\varepsilon] , \quad \text{for all } j, k = 1, 2, \ldots . \tag{23}$$

Then,

$$J_{1n} \equiv \left| \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} d_{nj}d_{nk}E_{F_{\theta\Omega n}}\left[\left(\psi(\theta, X_j) - E_{F_{\theta n}^{(1)j}}\psi(\theta, X_j)\right)\right.\right.$$

$$\left. \times \left(\psi(\theta, X_k) - E_{F_{\theta n}^{(1)k}}\psi(\theta, X_k)\right)\right]$$

$$\left. - \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} d_{nj}d_{nk}E_{\Phi_{\theta\Omega}}\psi(\theta, X_j)\psi(\theta, X_k)\right| .$$

$$\leq \frac{B_d^2}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} q^{(n)}_{|j-k|}$$

$$= B_d^2 \sum_{r=-n}^{n}\left(\frac{n-|r|}{n}\right) q^{(n)}_{|r|}$$

$$\leq B_d^2 \sum_{r=-r_\varepsilon+1}^{r_\varepsilon-1} q^{(n)}_{|r|} + 2B_d^2 \sum_{r=r_\varepsilon}^{\infty} 4C_1^2(\theta)\alpha(r)$$

$< \varepsilon$, using A2 and the mixing inequality,

u.f.θ,F,Ω, for n sufficiently large by (22) and (23). Thus, $J_{1n} = o(1)$ as $n \to \infty$ u.f.θ,F,Ω.

Let

$$J_{2n} = \left|\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} d_{nj}d_{nk}E_{\Phi_{\theta\Omega}}\psi(\theta, X_j)\psi(\theta, X_k) - \sum_{r=-\infty}^{\infty}(\eta^2 + \omega_{|r|})E_{\Phi_{\theta\Omega}}\psi(\theta, X_0)\psi(\theta, X_{|r|})\right| .$$

Now

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} d_{nj}d_{nk}E_{\Phi_{\theta\Omega}}\psi(\theta, X_j)\psi(\theta, X_k)$$

$$= \sum_{r=-n}^{n}\left(\frac{1}{n}\sum_{j=1}^{n-|r|} d_{nj}d_{nj+|r|}\right)E_{\Phi_{\theta\Omega}}\psi(\theta, X_0)\psi(\theta, X_{|r|})$$

by stationarity of $\Phi_{\theta\Omega}$,

$$\xrightarrow{n\to\infty} \sum_{r=-\infty}^{\infty}(\eta^2 + \omega_{|r|})E_{\Phi_{\theta\Omega}}\psi(\theta, X_0)\psi(\theta, X_{|r|})$$

by D2, the mixing inequality, MIX, and the bounded convergence theorem. So $J_{2n} = o(1)$ as $n \to \infty$. And, L.H.S. of (8) $\leq J_{1n} + J_{2n} = o(1)$ as $n \to \infty$ u.f.θ,F,Ω, which proves (a).

To show (b),

$$\sum_{r=-\infty}^{\infty}(\eta^2 + \omega_{|r|})E_{\Phi_{\theta\Omega}}\psi(\theta, X_0)\psi(\theta, X_{|r|})$$

$$\leq \sum_{r=-\infty}^{\infty} B_d^2\cdot 4(\text{ess sup}|\psi|)^2\cdot\alpha(r)$$

by D1, D2, and the mixing inequality,

$< \infty$ by MIX.

Let $$Q_1 = \eta^2 \sum_{r=-\infty}^{\infty} E_{\Phi_{\theta\Omega}} \psi(\theta,\ X_0)\psi(\theta,\ X_{|r|}) ,$$

and $$Q_2 = \sum_{r=-\infty}^{\infty} \omega_{|r|} E_{\Phi_{\theta\Omega}} \psi(\theta,\ X_0)\psi(\theta,\ X_{|r|}) .$$

We have $Q_1 > 0$ and we want to show $Q_2 \geq 0$.

$$0 \leq \lim_{n\to\infty} E_{\Phi_{\theta\Omega}} \left[\frac{1}{\sqrt{n}}\sum_{j=1}^{n} (d_{nj} - \bar{d}_n)\psi(\theta,\ X_j)\right]^2$$

$$= \lim_{n\to\infty} E_{\Phi_{\theta\Omega}} \left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\psi(\theta,\ X_j)\right)^2 + \eta^2 \lim_{n\to\infty}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi(\theta,\ X_j)\right)^2$$

$$- 2\eta \lim_{n\to\infty} E_{\Phi_{\theta\Omega}}\left[\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\psi(\theta,\ X_j)\right)\cdot\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\psi(\theta,\ X_k)\right)\right]$$

$$= (Q_1 + Q_2) + Q_1 - 2\eta \lim_{n\to\infty} \sum_{r=-n}^{n}\left(\frac{1}{n}\sum_{j=1}^{n-|r|} d_{nj}\right) E_{\Phi_{\theta\Omega}} \psi(\theta,\ X_0)\psi(\theta,\ X_{|r|})$$

by the stationarity of $\Phi_{\theta\Omega}$,

$$= Q_2 + 2Q_1 - 2\eta^2 \sum_{r=-\infty}^{\infty} E_{\Phi_{\theta\Omega}} \psi(\theta,\ X_0)\psi(\theta,\ X_{|r|})$$

by D2, the mixing inequality, MIX, and the bounded convergence theorem,

$$= Q_2 . \quad \square$$

<u>Proof of Lemma 7</u>. Define

$$\mathcal{F}^{V}_{t/\sqrt{n}}(\theta) = \{F^{(1)} : F^{(1)} \text{ is a distribution on } R, \text{ and } \Delta_V(F^{(1)}, \Phi_\theta) < t/\sqrt{n}\}$$

where
$$\Delta_V(F^{(1)}, \Phi_\theta) \equiv \sup\{F^{(1)}(A) - \Phi_\theta(A) : A \text{ is a Borel set in } R\}$$

$$= \frac{1}{2}\int_R |f^{(1)} - \varphi_\theta| \quad \text{where} \quad f^{(1)}(x) \equiv \frac{dF^{(1)}(x)}{dx} .$$

Note, $\mathcal{F}^{H}_{t/\sqrt{n}}(\theta) \subset \mathcal{F}^{V}_{t/\sqrt{n}}(\theta)$, since $\Delta_H(F^{(1)}, \Phi_\theta) \leq \Delta_V(F^{(1)}, \Phi_\theta)$.

Next, we claim

$$E_n \equiv \sup_{F^{(1)}_{\theta n} \in \mathcal{F}^{H}_{t/\sqrt{n}}(\theta)} |\lambda(\psi, s, F^{(1)}_{\theta n}) - \lambda(\psi, s, \Phi_\theta)| \xrightarrow{n\to\infty} 0 ,$$

uniformly for $s \in C$, any compact set in Θ , and uniformly for all θ in Θ . Proof:

$$E_n \leq \sup_{F^{(1)}_{\theta n} \in \mathcal{F}^{V}_{t/\sqrt{n}}(\theta)} \int |\psi(s,\cdot)|\, d|F^{(1)}_{\theta n} - \Phi_\theta|$$

$$< \operatorname*{ess\,sup}_{x\in R} |\psi(s,x)| \cdot \frac{t}{\sqrt{n}} \quad \text{since} \quad \Delta_V(F^{(1)}_n, \Phi_\theta) < \frac{t}{\sqrt{n}}$$

$$\xrightarrow{n\to\infty} 0$$

uniformly for $s \in C$, since $\operatorname*{ess\,sup}_{x\in R} |\psi(s,x)|$ is bounded for $s \in C$.

Now, by the triangular array central limit theorem for strong mixing bounded random variables (see Mehra and Rao (1975) and Philipp (1969)),

$$\frac{1}{\sqrt{n}} \cdot \frac{1}{\mathrm{Var}_{F_{\theta\Omega n}}\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\psi(\theta,\ X_j)\right]} \cdot \sum_{j=1}^{n} d_{nj}\left(\psi(\theta,\ X_j) - \lambda(\psi,\ \theta,\ F_{\theta n}^{(1)j})\right)$$

$$\xrightarrow{\mathcal{L}} N(0,1) \ ,$$

u.f.θ,F,Ω. Thus, using Lemma 6,

$$\frac{1}{n}\sum_{j=1}^{n} d_{nj}\left(\psi(\theta,\ X_j) - \lambda(\psi,\ \theta,\ F_{\theta n}^{(1)j})\right) \xrightarrow{F_{\theta\Omega n}} 0 \quad \text{u.f.}\theta,F,\Omega.$$

And for any $\varepsilon_1 > 0$,

$$F_{\theta\Omega n}\left(\left|\frac{1}{n}\sum_{j=1}^{n} d_{nj}[\psi(\theta \pm \delta,\ X_j) - \lambda(\psi,\ \theta\pm\delta, F_{\theta n}^{(1)j})]\right| < \varepsilon_1\right) \xrightarrow{n\to\infty} 1 \quad \text{u.f.}\theta,F,\Omega \qquad (24)$$

(where the choice of "+" or "-" is constant throughout the expression).

By Lemma 5, $\lambda(\psi,\ s,\ \Phi_\theta)$ is differentiable in s and has a negative derivative at $s = \theta$. This, plus $\lambda(\psi,\ \theta,\ \Phi_\theta) = 0$, imply for any given $\varepsilon > 0$, there exists δ with $0 < \delta < \varepsilon$ such that $\lambda_- \equiv \lambda(\psi,\ \theta-\delta,\ \Phi_\theta) > 0$ and $\lambda_+ \equiv \lambda(\psi,\ \theta+\delta,\ \Phi_\theta) < 0$. Take the δ in (24) above to be the present δ , and take ε_1 to be $\eta \cdot \frac{\lambda_- \wedge |\lambda_+|}{3}$. By the claim above, $|\lambda_\pm \cdot \eta - \frac{1}{n}\sum_{j=1}^{n} d_{nj}\lambda(\psi,\ \theta\pm\delta,\ F_{\theta n}^{(1)j})| < \varepsilon_1$ for all $n \geq N_{\varepsilon_1}$, for some N_{ε_1} , and for all $\theta \in C$. So,

$$F_{\theta\Omega n}(D_n) \equiv F_{\theta\Omega n}\left(\left|\frac{1}{n}\sum_{j=1}^{n} d_{nj}\psi(\theta-\delta,\ X_j) - \lambda_-\eta\right| < 2\varepsilon_1\right) \xrightarrow{n\to\infty} 1$$

and

$$F_{\theta\Omega n}(D_n') \equiv F_{\theta\Omega n}\left(\left|\frac{1}{n}\sum_{j=1}^{n} d_{nj}\psi(\theta+\delta,\ X_j) - \lambda_+\eta\right| < 2\varepsilon_1\right) \xrightarrow{n\to\infty} 1$$

u.f.θ,F,Ω. Note, ε_1 has been chosen so small that on $D_n \cap D_n'$, $\frac{1}{n}\sum_{j=1}^{n} d_{nj}\psi(\theta-\delta, X_j) > 0$ and $\frac{1}{n}\sum_{j=1}^{n} d_{nj}\psi(\theta+\delta, X_j) < 0$. So, by continuity of $\psi(s, X_j)$ in s (which follows from A3), $\forall\omega \in D_n \cap D_n'$, $\exists\ t^*(\omega) \in [\theta-\delta, \theta+\delta]$ such that $\frac{1}{n}\sum_{j=1}^{n} d_{nj}\psi(t^*(\omega), X_j) = 0$. Hence $F_{\theta\Omega n}(\sum_{j=1}^{n} d_{nj}\psi(\theta, X_j) = 0 \text{ has a root}) \geq F_{\theta\Omega n}(D_n \cap D_n') \xrightarrow{n\to\infty} 1$, u.f.$\theta$,F,$\Omega$, and (b) is proved.

Now, since $\tilde{T}_n \xrightarrow{F_{\theta\Omega n}} \theta$, $F_{\theta\Omega n}(\tilde{T}_n \in [\theta-\varepsilon_2, \theta+\varepsilon_2]) \xrightarrow{n\to\infty} 1$ u.f.θ,F,Ω, $\forall\varepsilon_2 > 0$. For the $\varepsilon > 0$ given above, take $\varepsilon_2 = \frac{\varepsilon-\delta}{2}$. Then, on $D_n'' \equiv (D_n \cap D_n') \cap [\tilde{T}_n \in [\theta-\varepsilon_2, \theta+\varepsilon_2]]$,

$$|\tilde{T}_n - (\text{closest root of } \sum_{j=1}^{n} d_{nj}\psi(\theta, X_j) = 0 \text{ to } \tilde{T}_n)| \leq \delta + \varepsilon_2 \qquad (25)$$

So, $$F_{\theta\Omega n}(D_n'') \leq F_{\theta\Omega n}\Big(|\tilde{T}_n - T_n| \leq \delta+\varepsilon_2 \text{ and } T_n \in [\theta-\varepsilon_2, \theta+\varepsilon_2]\Big) \text{ by (25)}$$
$$\leq F_{\theta\Omega n}(|T_n - \theta| \leq \varepsilon) .$$

Since $F_{\theta\Omega n}(D_n'') \xrightarrow{n\to\infty} 1$, we get

$$F_{\theta\Omega n}(|T_n - \theta| \leq \varepsilon) \xrightarrow{n\to\infty} 1 \quad \text{u.f.}\theta\text{,F,}\Omega,$$

and (a) is proved. □

Proof of Lemma 8. By A3 and Lemma 7,

$$\sup_{x\in R}|\psi(T_n, x) - \psi(\theta,x)| \leq C_2(\theta)\cdot|T_n - \theta| = o_{F_{\theta\Omega n}}(1) \quad \text{u.f.}\theta\text{,F,}\Omega .$$

Hence,

$$\int[\psi(T_n,\cdot)-\psi(\theta,\cdot)]d\sqrt{n}\left[\frac{1}{n}\sum_{j=1}^{n}d_{nj}(F_{\theta n}^{(1)j}-\Phi_\theta)\right]$$

$$\leq o_{F_{\theta\Omega n}}(1)\max_{1\leq j\leq n}\int d\sqrt{n}|F_{\theta n}^{(1)j}-\Phi_\theta|$$

$$= o_{F_{\theta\Omega n}}(1)\quad \text{u.f.}\theta,F,\Omega$$

since $F_{\theta n}^{(1)j}\in\mathcal{F}_{t/\sqrt{n}}^{H}(\theta)\subset\mathcal{F}_{t/\sqrt{n}}^{V}(\theta)$, $\forall j=1,\ldots,n$. □

<u>Proof of Lemma 9</u>. Let $\beta(m)=\dfrac{\alpha(m)}{\sum_{\ell=0}^{\infty}\alpha(\ell)}$, $\forall m=0,1,2,\ldots$.

($\sum_{\ell=1}^{\infty}\alpha(\ell)\leq\sum_{m=0}^{\infty}m^2\alpha(\ell)^\tau<\infty$, since $\alpha(m)$, $\tau\in(0,1)$.) $\{\beta(m):m=0,1,2,\ldots\}$ is a discrete probability distribution. Markov's inequality implies

$$\sum_{m=w}^{\infty}\beta(m)=P_\beta(V\geq w)\leq\frac{E_\beta V^{1/\delta}}{w^{1/\delta}}\quad\text{where } V\ (\geq 0) \text{ has the discrete}$$

distribution given by β ,

$$\leq\frac{1}{w^{1/\delta}}E_\beta V^2\beta(V)^{\tau-1}$$ since $\delta\in[1/2,1)$, and $\beta(v)$, imply $\beta(v)^{\tau-1}\geq 1$, $\forall v$,

$$=\frac{1}{w^{1/\delta}}\sum_{m=1}^{\infty}m^2\beta(m)^{\tau-1}\beta(m)$$

$$\equiv\frac{1}{w^{1/\delta}}\cdot C_\alpha\quad\text{where}\quad C_\alpha\equiv\sum_{m=1}^{\infty}m^2\beta(m)^\tau<\infty\quad\text{by MIX.}\qquad(26)$$

Let $\bar{k}=\left(C_\alpha\cdot\sum_{\ell=0}^{\infty}\alpha(\ell)\right)^{-1}$, and take $w=[(\bar{k}\varepsilon)^{-\delta}]$ in (26), to get

$$\sum_{m=[(\bar{k}\varepsilon)^{-\delta}]+1}^{\infty} \alpha(m) < C_\alpha \cdot \sum_{\ell=0}^{\infty} \alpha(\ell) \cdot \frac{1}{[(\bar{k}\varepsilon)^{-\delta}]^{1/\delta}}$$

$$< \varepsilon . \quad \square$$

<u>Proof of Lemma 10</u>. Given a sequence $<\varepsilon_n>_{n\geq 1}$ such that $\varepsilon_n \in (0, 1/2)$ and $\varepsilon_n \xrightarrow{n\to\infty} 0$, define

$$W_n(s) = \begin{cases} Z_n(\theta - \varepsilon_n) - Z_n(\theta) & \text{for} \quad s \in [\theta - 1/2, \theta - \varepsilon_n] \\ Z_n(s) - Z_n(\theta) & \text{for} \quad s \in [\theta - \varepsilon_n, \theta + \varepsilon_n] \\ Z_n(\theta + \varepsilon_n) - Z_n(\theta) & \text{for} \quad s \in [\theta + \varepsilon_n, \theta + 1/2] . \end{cases}$$

We show, using the inequality of Kolmogorov-Chentsov, that the process $W_n(\cdot)$ converges weakly to the zero function on $[\theta - 1/2, \theta + 1/2]$ under the probability measure $F_{\theta\Omega n}$, u.f.θ,F,Ω. Then, by Durbin (1973, p. 18), $g(x) = \sup_{s\in[\theta-1/2,\theta+1/2]} |x(s)|$ is continuous in the Skorokhod metric on $D[\theta - 1/2, \theta + 1/2]$. So,

$$g(W_n) = \sup_{s:|s-\theta|\leq\varepsilon_n} |Z_n(s) - Z_n(\theta)| \xrightarrow{\mathcal{L}_{F_{\theta\Omega n}}} g(\underline{0}) = 0$$

u.f.θ,F,Ω, and where $\underline{0}$ is the zero function on $[\theta - 1/2, \theta + 1/2]$, which is the desired result.

Now,

$$E_{F_{\theta\Omega n}}(Z_n(s) - Z_n(\theta))^2 = \mathrm{Var}_{F_{\theta\Omega n}}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\Big(\psi(s,X_j) - \psi(\theta,X_j)\Big)\right)$$

$$\leq \frac{B_d^2}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left|\mathrm{Cov}_{F_{\theta\Omega n}}\Big(\psi(s,X_j) - \psi(\theta,X_j),\ \psi(s,X_k) - \psi(\theta,X_k)\Big)\right| \quad \text{by D1,}$$

$$\leq B_d^2 \sum_{r=-\infty}^{\infty}\left(1 - \frac{|r|}{n}\right) \sup_{\substack{j,k:|j-k|=r,\\ \forall j,k=1,2,\ldots}} \left|\mathrm{Cov}_{F_{\theta\Omega n}}(\psi(s,X_j) - \psi(\theta,X_j),\ \psi(s,X_k) - \psi(\theta,X_k))\right|$$

$$\leq B_d^2 \cdot \sum_{r=-[(\bar{k}\varepsilon)^{-\delta}]}^{[(\bar{k}\varepsilon)^{-\delta}]} (s-\theta)^2 + 2B_d^2 \cdot \sum_{r=[(\bar{k}\varepsilon)^{-\delta}]}^{\infty} 4C_1(\theta)\alpha(r)$$

where $\varepsilon > 0$, $\delta \in [1/2, 1)$ are as in Lemma 9, by A3, A2 and the mixing inequality

$$\leq B_d^2 \cdot (2[(\bar{k}\varepsilon)^{-\delta}] + 1) \cdot (s-\theta)^2 + 8 \cdot B_d^2 \cdot C_1^2(\theta) \cdot \varepsilon \quad \text{by Lemma 9}$$

$$\leq B_d^2 \cdot (3(\bar{k})^{-\delta} + 8C_1^2(\theta)) \cdot |s-\theta|^{2-\delta}$$

u.f.θ,F,Ω where $\delta \in [1/2, 1)$ by taking $\varepsilon \doteq (s-\theta)^2$ for s sufficiently close to θ ,

$$= A \cdot |s-\theta|^{2-\delta} \tag{27}$$

u.f.θ,F,Ω where $A \equiv B_d^2 \cdot \Big(3(\bar{k})^{-\delta} + 8C_1^2(\theta)\Big)$, and A does not depend on $|s-\theta|$.

Hence, by Chebyshev's inequality,

$$F_{\theta\Omega n}(|Z_n(s) - Z_n(\theta)| > \tilde{\varepsilon}) \leq \frac{1}{\tilde{\varepsilon}^2} E_{F_{\theta\Omega n}}(Z_n(s) - Z_n(\theta))^2$$

$$\leq \frac{A}{\tilde{\varepsilon}^2}|s-\theta|^{2-\delta} \tag{28}$$

for $\theta \in C$, any compact set in Θ, $\forall F_{\theta\Omega n} \in \{\mathcal{F}^{H}_{t/\sqrt{n}}(\theta), S\}_n$, by (27). So, if $s_n \in [\theta - \varepsilon_n, \theta + \varepsilon_n]$, $\forall n$, then $s_n \xrightarrow{n\to\infty} \theta$, and $F_{\theta\Omega n}(|W_n(s_n)| > \varepsilon) = F_{\theta\Omega n}(|Z_n(s_n) - Z_n(\theta)| > \varepsilon) \xrightarrow{n\to\infty} 0$ u.f.θ,F,Ω. That is, $W_n(s_n) \xrightarrow{F_{\theta\Omega n}} 0$, u.f.$\theta$,F,$\Omega$. Similarly, by (28), if $s_n \in [\theta - 1/2, \theta - \varepsilon_n]$, $\forall n$ or $s_n \in [\theta + \varepsilon_n, \theta + 1/2]$, $\forall n$, $W_n(s_n) \xrightarrow{F_{\theta\Omega n}} 0$ u.f.θ,F,Ω. Combining these results gives: If $s_n \in [\theta - 1/2, \theta + 1/2]$, $\forall n$, then $W_n(s_n) \xrightarrow{F_{\theta\Omega n}} 0$ u.f.θ,F,Ω.

It remains to show tightness of $\langle W_n \rangle_{n\geq 1}$. For any t_1, t_2 in $[\theta - 1/2, \theta + 1/2]$ such that $t_1 \leq t_2$, (27) yields

$$E_{F_{\theta\Omega n}}(W_n(t_1) - W_n(t_2))^2 \leq A\cdot(t_1 - t_2)^{2-\delta}$$

$\forall \theta \in C$, any compact set in Θ, $\forall F_{\theta\Omega n} \in \{\mathcal{F}^{H}_{t/\sqrt{n}}(\theta), S\}_n$. This, plus the result that $\{W_n(\theta - 1/2)\}_{n\geq 1}$ are tight, implies that $\langle W_n \rangle_{n\geq 1}$ are tight, by the inequality of Kolmogorov-Chentsov (see Billingsley (1968, p. 95)). $\{W_n(\theta - 1/2)\}_{n\geq 1}$ are tight since

$$\begin{aligned}
\lim_{b\uparrow\infty} \sup_n F_{\theta\Omega n}(|W_n(\theta - 1/2)| > b) &\leq \lim_{b\uparrow\infty} \sup_n \frac{1}{b^2} E_{F_{\theta\Omega n}} |W_n(\theta - 1/2)|^2 \\
&= \lim_{b\uparrow\infty} \frac{1}{b^2} \sup_n E_{F_{\theta\Omega n}}\Big(Z_n(\theta - \varepsilon_n) - Z_n(\theta)\Big)^2 \\
&\leq \lim_{b\uparrow\infty} \frac{1}{b^2}\cdot A\cdot \sup_n |\varepsilon_n|^{2-\delta} \quad \text{for all } \theta \in C, \text{ any compact set in } \Theta, \ \forall F_{\theta\Omega n} \in \{\mathcal{F}^{H}_{t/\sqrt{n}}(\theta), S\}_n, \text{ by (27)}, \\
&= 0 \quad \text{since } 0 < \varepsilon_n < 1/2 .
\end{aligned}$$

Whence the result. □

Proof of Lemma 11. By Lemma 7, $T_n \xrightarrow{F_{\theta\Omega n}} \theta$ u.f.θ,F,Ω. And by Lemma 10,

$\sup_{s:|s-\theta|<\varepsilon_n} |Z_n(s) - Z_n(\theta)| \xrightarrow{\mathcal{L}_{F_{\theta\Omega n}}} 0$, for any fixed sequence $\varepsilon_n > 0$ such that $\varepsilon_n \xrightarrow{n\to\infty} 0$, u.f.θ,F,Ω. Hence, by a standard argument we get the desired result. □

Corollary 1. For $T \in \mathcal{M}$, and for L satisfying L1 and L2,

$$R_H(\psi,d,\theta,t,S) = E_N L\left(b_H(\psi,d,\theta) + \sup_{\Omega\in S} \sigma(\psi,d,\theta,\Omega)\cdot Z\right) , \tag{29}$$

where $Z \sim N$, the standard normal distribution, and where

$$b_H(\psi,d,\theta) = \frac{\tilde{n}}{\eta} 2t\cdot\left(\int\psi^2(\theta,x)d\Phi_\theta(x)\right)^{1/2} \tag{30}$$

and

$$\sigma^2(\psi,d,\theta,\Omega) = \sum_{j=-\infty}^{\infty}\left(\frac{\eta^2+\omega_{|j|}}{\eta^2}\right) E_{\Phi_{\theta\Omega}} \psi(\theta, X_0)\psi(\theta, X_{|j|}) . \tag{31}$$

Comments: (i) $b_H(\psi,d,\theta)$ is the asymptotic maximum bias over the neighborhoods of the weighted M-estimator T . It is independent of the correlation matrix Ω .

(ii) $\sigma^2(\psi,d,\theta,\Omega)$ is the asymptotic variance of T over the neighborhoods of $\Phi_{\theta\Omega}$. It is a function of Ω .

Proof of Corollary 1. By Lemma 4,

$$R_H(\psi,d,\theta,t,S) = \sup_{a>0} \sup_{\Omega\in S} \sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} \overline{\lim_{n\to\infty}} E_{F_{\theta\Omega n}} L_a\left(\sqrt{n}\left(T_n - \theta - \frac{1}{n}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right)\right.$$

$$\left. + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right)$$

$$= \sup_{a>0} \sup_{\Omega\in S} \sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} E_N L_a\left(\sigma(\psi,d,\theta,\Omega)\cdot Z + \lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right)$$

by Theorem 1, the existence of $\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})$ (shown below), and continuity and boundedness of $L_a(\cdot)$, where $Z \sim N(0,1)$,

$$= \sup_{a>0} \sup_{\Omega\in S} E_N L_a\left(\sigma(\psi,d,\theta,\Omega)\cdot Z + \sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} \lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right)$$

since $|b+\sigma Z| \overset{ST}{\leq} |c+\sigma Z|$ iff $|b| \leq |c|$, for constants b, c, and σ, so if L satisfies L1, we get $L_a(|\cdot|)$ is non-decreasing on R^+, and Lehmann (1959, p. 112) implies $E_N L_a(b+\sigma Z) \leq E_N L_a(c+\sigma Z)$ (see also Bickel (1981, p. 19)).

Further,

$$R_H(\psi,d,\theta,t,S) = E_N L\left(\sup_{\Omega\in S}\sigma(\psi,d,\theta,\Omega)\cdot Z + \sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}}\left|\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right|\right)$$

since (i) the second term in parentheses is independent of Ω, (ii) the function $V(\sigma) = E_N L_a(\sigma Z + b)$, for " b " a constant, is increasing in σ for L convex (because L convex implies $V(\sigma)$ is convex, and Z symmetrically distributed about zero implies $V(\sigma)$ is an even function), and (iii) the monotone convergence theorem implies.

We only need to show $\sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} \left| \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{d_{nj}}{\eta} \lambda(\psi,\theta,F_{\theta n}^{(1)j}) \right|$ exists and equals $b_H(\psi,d,\theta)$. This is proven in Lemma 12. □

Lemma 12. If $\{\psi,d\}$ corresponds to some $T \in \mathcal{M}$, then

$$\sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} \left| \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{d_{nj}}{\eta} \lambda(\psi,\theta,F_{\theta n}^{(1)j}) \right| = b_H(\psi,d,\theta) \ . \tag{32}$$

Comment: The original problem (2) can now be written

$$\inf_{\substack{\psi:A1,A2 \\ \text{and} \\ A3 \text{ hold}}} \ \inf_{\substack{\{d_{nj}\}:D1,D2 \\ \text{and} \\ D3 \text{ hold}}} R_H(\psi,d,\theta,t,S) = E_N L\left(b_H(\psi,d,\theta) + \sup_{\Omega\in S} \sigma(\psi,d,\theta,\Omega)\cdot Z\right) \tag{33}$$

where $Z \sim N(0,1)$.

Proof of Lemma 12. First we show

$$\overline{\lim_{n\to\infty}} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{d_{nj}}{\eta} \lambda(\psi,\theta,F_{\theta n}^{(1)j}) \leq b_H(\psi,d,\theta)$$

for all sequences $\langle F_{\theta\Omega n}\rangle_{n\geq 1}$ in the neighborhoods. For all $j = 1, 2, \ldots$ we have

$$|\lambda(\psi,\theta,F_{\theta n}^{(1)j})| = \left|\int\psi(\theta,x)(f_{\theta n}^{(1)j}(x) - \varphi_\theta(x))dx\right| \quad \text{since} \quad E_{\Phi_\theta}\psi(\theta,X_j) = 0 , \quad \forall j ,$$

$$\leq \int\left|\psi(\theta,\cdot)(\sqrt{f_{\theta n}^{(1)j}} - \sqrt{\varphi_\theta})(\sqrt{f_{\theta n}^{(1)j}} + \sqrt{\varphi_\theta})\right|$$

$$\leq \left(\int[\sqrt{f_{\theta n}^{(1)j}} - \sqrt{\varphi_\theta}]^2\right)^{1/2} \cdot \left(\int\psi(\theta,\cdot)^2[\sqrt{f_{\theta n}^{(1)j}} + \sqrt{\varphi_\theta}]^2\right)^{1/2}$$

by the Cauchy-Schwartz inequality,

$$< \frac{t}{\sqrt{n}}\left[\left(\int\psi^2[f_{\theta n}^{(1)j} + \varphi_\theta + 2f_{\theta n}^{(1)j}]\right)^{1/2} \vee \left(\int\psi^2[f_{\theta n}^{(1)j} + \varphi_\theta + 2\varphi_\theta]\right)^{1/2}\right] \qquad (34)$$

since $\int\psi^2\sqrt{f_{\theta n}^{(1)j}}\sqrt{\varphi_\theta} \leq \left(\int\psi^2 f_{\theta n}^{(1)j}\right)^{1/2} \cdot (\int\psi^2\varphi_\theta)^{1/2} \leq \int\psi^2 f_{\theta n}^{(1)j} \vee \int\psi^2\varphi_\theta$. Now, $\lim_{n\to\infty} \int\psi^2 f_{\theta n}^{(1)j} = \int\psi^2\varphi_\theta$, since ψ is a bounded continuous function, and $F_{\theta n}^{(1)j} \in \mathcal{F}^H_{t/\sqrt{n}}(\theta)$, $\forall n$ implies $F_{\theta n}^{(1)j} \Rightarrow \Phi_\theta$. This gives, for any $\varepsilon > 0$ and for n sufficiently large, $\int\psi^2 f_{\theta n}^{(1)j} \leq \int\psi^2\varphi_\theta + \varepsilon$. Hence, continuing from (34) above, for n sufficiently large and $j = 1, 2, \ldots$,

$$\left|\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right| < \frac{t}{\sqrt{n}}(3\varepsilon + 4\int\psi^2\varphi_\theta)^{1/2} . \qquad (35)$$

Thus,

$$\left|\overline{\lim_{n\to\infty}} \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right| \leq \overline{\lim_{n\to\infty}} \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \frac{|d_{nj}|}{\eta}\left|\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right| \quad \text{since without loss of generality } \eta > 0 ,$$

$$\leq \overline{\lim_{n\to\infty}} \frac{1}{n}\sum_{j=1}^{n} |d_{nj}| \cdot \frac{1}{\eta}2t \cdot (\int\psi^2\varphi_\theta)^{1/2}$$

$$= b_H(\psi,d,\theta) ,$$

for all sequences $\langle F_{\theta\Omega n}\rangle_{n\geq 1}$ in the neighborhoods.

Next, we show $\sup_{<F_{\theta\Omega n}>_{n\geq 1}} \left| \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{d_{nj}}{\eta} \lambda(\psi,\theta,F_{\theta n}^{(1)j}) \right| \geq b_H(\psi,d,\theta)$.

Consider sequences $<F_{\theta\Omega n}^{h}>_{n\geq 1}$ of distributions in $\{\mathcal{F}_{t/\sqrt{n}}^{H}(\theta), \Omega\}_n$, $\forall n$ sufficiently large, and any $\Omega \in S$, whose univectorate marginal distributions with respect to Lebesgue measure are given by $f_{\theta n}^{h(1)j}(x) = \varphi_\theta(x) e^{\frac{1}{\sqrt{n}} h(x) \operatorname{sgn} d_{nj} - c_n}$, $j = 1, 2, \ldots$ (take $d_{nj} = 0$ for $j > n$) where c_n is a constant defined to make this a density, and where $h \in \mathcal{H}_H(\theta)$. ($\mathcal{H}_H(\theta)$ is defined in Section 3.) Such distributions exist since they can be created by point transformations of random variables with distribution $\Phi_{\theta\Omega}$ (see Section 3). Bickel (1981, Lemma 2) shows that for n sufficiently large the univectorate distributions $F_{\theta n}^{h(1)j}$ are in $\mathcal{F}_{t/\sqrt{n}}^{H}(\theta)$, $\forall h \in \mathcal{H}_H(\theta)$. This implies $F_{\theta\Omega n}^{h} \in \{\mathcal{F}_{t/\sqrt{n}}^{H}(\theta), \Omega\}_n$, for n sufficiently large, $\forall h \in \mathcal{H}_H(\theta)$. And by Bickel (1981, Theorem 2),

$$\sup_{h \in \mathcal{H}_H(\theta)} \int \psi(\theta,x) h(x) \varphi_\theta(x) dx = 2t \cdot \left(\int \psi^2(\theta,x) \varphi_\theta(x) dx \right)^{1/2} .$$

Now,

$$e^{c_n} = \int \varphi_\theta e^{\frac{1}{\sqrt{n}} h \operatorname{sgn} d_{nj}} = \int \varphi_\theta \left(1 + \frac{1}{\sqrt{n}} h \operatorname{sgn} d_{nj} + 0\left(\frac{1}{n}\right) \right) = 1 + 0\left(\frac{1}{n}\right) ,$$

and so,

$$e^{\frac{1}{\sqrt{n}} h \operatorname{sgn} d_{nj} - c_n} = \left(1 + \frac{1}{\sqrt{n}} h \operatorname{sgn} d_{nj} + 0\left(\frac{1}{n}\right) \right) \cdot \left(1 + 0\left(\frac{1}{n}\right) \right)$$

$$= 1 + \frac{1}{\sqrt{n}} h \operatorname{sgn} d_{nj} + 0\left(\frac{1}{n}\right) ,$$

where $0\left(\frac{1}{n}\right)$ holds uniformly for $x \in R$, since h is bounded. Hence,

$$\sqrt{n}\ \lambda(\psi,\theta,F_{\theta n}^{h(1)j}) = \sqrt{n}\int\psi\varphi_\theta\left[1+\frac{1}{\sqrt{n}}h\ \text{sgn}\ d_{nj}+0\left(\frac{1}{n}\right)\right]$$

$$= \text{sgn}\ d_{nj}\cdot\int\psi h\varphi_\theta + 0\left(\frac{1}{n}\right)\ ,\quad \forall j = 1,\ 2,\ \ldots\ ,$$

using the uniformity of $0\left(\frac{1}{n}\right)$ over $x \in R$. Thus,

$$\sup_{<F_{\theta\Omega n}>_{n\geq 1}}\left|\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{(1)j})\right|$$

$$\leq \sup_{<F_{\theta\Omega n}^{h}>_{n\geq 1}}\left|\underline{\lim}_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\lambda(\psi,\theta,F_{\theta n}^{h(1)j})\right|$$

$$= \sup_{<F_{\theta\Omega n}^{h}>_{n\geq 1}}\left[\underline{\lim}_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\frac{|d_{nj}|}{\eta}\right]\cdot\left|\int\psi h\varphi_\theta\right| + \left[\underline{\lim}_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\frac{d_{nj}}{\eta}\cdot 0\left(\frac{1}{n}\right)\right]$$

$$= \sup_{h\in\mathcal{H}_H(\theta)}\frac{\tilde{\eta}}{\eta}\cdot\left|\int\psi h\varphi_\theta\right|$$

$$= \frac{\tilde{\eta}}{\eta}2t\cdot\left(\int\psi^2\varphi_\theta\right)^{1/2}$$

$$= b_H(\psi,d,\theta)\ .\quad \square$$

We now calculate the infimum of $R_H(\psi,d,\theta,t,S)$ (as simplified in (33)) over the triangular arrays of weights $\{d_{nj}\}$.

<u>Theorem 2</u>. For $T \in \mathcal{M}$, and for L satisfying L1 and L2

$$\inf_{\substack{\{d_{nj}\}:D1,D2 \\ \text{and D3 hold}}} R_H(\psi,d,\theta,t,S) = R_H(\psi,d^*,\theta,t,S) \tag{36}$$

where $d^* = \{d_{nj}^*\}$ is any triangular array of identical nonzero constants.

Without loss of generality we can take $d^*_{nj} = 1$, $\forall j$, $\forall n$. Then, define

$$R_H(\psi,\theta,t,S) = R_H(\psi,d^*,\theta,t,S) \ , \tag{37}$$

$$b_H(\psi,\theta) = b_H(\psi,d^*,\theta) \ , \tag{38}$$

and
$$\sigma^2(\psi,\theta,\Omega) = \sigma^2(\psi,d^*,\theta,\Omega) \ . \tag{39}$$

Comments: (i) The theorem shows equally weighted M-estimators are optimally minimax over the Hellinger neighborhoods defined above. For the parametric Gaussian distributions in these neighborhoods this is not unexpected, since the sample mean (which has equal weights) is asymptotically efficient, even though the UMVU estimator for finite sample size n has unequal weights (which depend on $(\rho_1, \rho_2, \ldots)$).

(ii) The asymptotic maximum bias and the asymptotic variance simplify to:

$$b_H(\psi,\theta) = 2t\cdot\left\{\int\psi^2(\theta,x)d\Phi_\theta(x)\right\}^{1/2} \tag{40}$$

and

$$\sigma^2(\psi,\theta,\Omega) = \sum_{j=-\infty}^{\infty} E_{\Phi_{\theta\Omega}}\psi(\theta, X_0)\psi(\theta, X_{|j|}) \ . \tag{41}$$

<u>Proof of Theorem 2</u>. Since $|\tilde{\eta}/\eta| \geq 1$, $|b_H(\psi,d,\theta)|$ is minimized by taking all d_{nj} of the same sign. The asymptotic maximum variance is

$$\sup_{\Omega\in S}\sigma^2(\psi,d,\theta,\Omega) = \sup_{\Omega\in S}\sum_{j=-\infty}^{\infty} E_{\Phi_{\theta\Omega}}\psi(\theta,X_0)\psi(\theta,X_{|j|}) + \sum_{j=-\infty}^{\infty}\frac{|\omega_j|}{\eta^2}E_{\Phi_{\theta\Omega}}\psi(\theta,X_0)\psi(\theta,X_{|j|}) \ .$$

By the proof of Lemma 6b, $\sum_{j=-\infty}^{\infty}\frac{|\omega_j|}{\eta^2}E_{\Phi_{\theta\Omega}}\psi(\theta, X_0)\psi(\theta, X_{|j|}) \geq 0$. Hence, for any $\Omega \in S$, $\sigma^2(\psi,d,\theta,\Omega)$ is minimized by taking $\omega_j = 0$, $\forall j = 1, 2, \ldots$.

This is the case if all d_{nj} are equal. Thus, taking all d_{nj} equal minimizes $|b_H(\psi,d,\theta)|$, and $\sup_{\Omega\in S} \sigma^2(\psi,d,\theta,\Omega)$. By the same argument as in the proof of Corollary 1, all d_{nj} equal minimizes the given function of these two expressions. □

Define

$$H(\psi,\theta,t,S) = E_N L\left(b_H(\psi,\theta) + \sup_{\Omega\in S} \sigma(\psi,\theta,\Omega)\cdot Z\right) .$$

Under the assumptions of Theorem 2 (which, in particular, require $\psi(\theta,x)$ to be bounded and satisfy a Lipschitz condition),

$$H(\psi,\theta,t,S) = R_H(\psi,\theta,t,S) . \tag{42}$$

We proceed by minimizing $H(\cdot,\theta,t,S)$ without imposing the boundedness and Lipschitz conditions on ψ, and then by adjusting the solution to get an ε-optimal solution which satisfies these conditions. (42) implies that such a solution also is within ε of minimizing $R_H(\cdot,\theta,t,S)$ and is an ε-optimal solution to the problem of (2).

We have the following result:

Theorem 3. For L satisfying L1 and L2, $i(\theta,x) \equiv x-\theta$ minimizes

$H(\psi,\theta,t,S)$ over ψ subject to

1) $E_{\Phi_\theta} \psi(\theta,X)\cdot\left(\frac{X-\theta}{\sigma^2}\right) = 1$,

and 2) $E_{\Phi_\theta} \psi(\theta,X) = 0$.

Comments: (i) Conditions 1) and 2) of Theorem 3 are the normalization conditions (on ψ) of A1. (ii) Notice $i(\theta,x) = I^{-1}(\theta)\dot{\ell}_\theta(x)$, where $\dot{\ell}_\theta(x) \equiv \frac{d}{d\theta}\ln \varphi_\theta(x)$ and $I(\theta) \equiv E_{\Phi_\theta} \dot{\ell}_\theta^2(X)$ is the Fisher information of Φ_θ.

The estimator based on $\psi(\theta,x) = I^{-1}(\theta)\dot{\ell}_\theta(x)$ is the maximum likelihood estimator for the i.i.d. Φ_θ model (in particular, it is the sample mean). Thus, Theorem 3 implies that the (i.i.d. Φ_θ) maximum likelihood estimator minimizes $H(\psi,\theta,t,S)$! This estimator is asymptotically efficient for the i.i.d. Φ_θ model. In fact, this estimator is asymptotically efficient for <u>all</u> parametric distributions $\Phi_{\theta\Omega}$ (provided Ω is such that $\Phi_{\theta\Omega}$ has a positive, continuous spectral density (see Grenander and Rosenblatt (1957, Chapter 7))!

<u>Proof of Theorem 3</u>. Any function ψ satisfying 1) and 2) can be written as $\psi(\theta,x) = (x-\theta) + g(x)$, where $\langle g,1\rangle \equiv E_{\Phi_\theta} g(X)\cdot 1 = 0$, and $\langle g, X-\theta\rangle \equiv E_{\Phi_\theta} g(X)\cdot(X-\theta) = 0$. We want to show $\sigma^2(X-\theta+g,\theta,\Omega) \geq \sigma^2(X-\theta,\theta,\Omega)$, for all functions g above and all $\Omega \in S$, since this implies $X-\theta$ minimizes $\sup_{\Omega\in S} \sigma(\cdot,\theta,\Omega)$ subject to 1) and 2).

The bivariate normal density $\varphi^{(2)}_{\theta\Omega}(x, y, \rho_j)$ can be expanded in terms of Hermite polynomials as:

$$\varphi^{(2)}_{\theta\Omega}(x, y, \rho_j) = \sum_{\ell=0}^{\infty} h_{\theta\ell}(x)h_{\theta\ell}(y)\frac{(\rho_j)^\ell}{\ell!}\varphi_\theta(x)\varphi_\theta(y)$$

where $h_{\theta\ell}(x) = h_\ell\left(\frac{x-\theta}{\sigma}\right)$ and $h_\ell(x)$ is the ℓ^{th} Hermite polynomial, for $\ell = 0, 1, 2, \ldots$ (see, e.g., Cramér (1946, p. 133)).

Also, without loss of generality we can assume $g \in L^2[\Phi_\theta]$, since otherwise $H(X-\theta+g,\theta,t,S) = \infty$.

Now,

$$\sigma^2(X-\theta+g,\theta,\Omega) - \sigma^2(X-\theta,\theta,\Omega) = 2\sum_{j=-\infty}^{\infty} E_{\Phi_{\theta\Omega}}(X_0-\theta)g(X_{|j|}) + \sum_{j=-\infty}^{\infty} E_{\Phi_{\theta\Omega}} g(X_0)g(X_{|j|})$$

$$\geq 2\cdot E_{\Phi_\theta}(X-\theta)g(X) + 4\sum_{j=1}^{\infty} E_{\Phi_{\theta\Omega}}(X_0-\theta)g(X_j)$$

since $\sum_{j=-\infty}^{\infty} E_{\Phi_{\theta\Omega}} g(X_0)g(X_{|j|})$ is the limit of a sequence of variances, and hence, is non-negative,

$$= 4\sum_{j=1}^{\infty} \int_{R^2} (x-\theta)g(y)\varphi_\theta^{(2)}(x, y, \rho_j)dxdy$$

since $\langle g,X-\theta\rangle = 0$ by condition 2),

$$= 4\sum_{j=1}^{\infty}\sum_{\ell=0}^{\infty} \langle X-\theta,h_{\theta\ell}\rangle\cdot\langle g,h_{\theta\ell}\rangle\cdot\frac{(\rho_j)^\ell}{\ell!}$$ by a Fourier

series result (see, e.g., Andrews (1982), Lemma 1), which uses the bivariate expansion of $\varphi_\theta^{(2)}(x,y,\rho_j)$ in terms of Hermite polynomials and the fact that $x-\theta,\ g \in L^2[\Phi_\theta]$,

$$= 0$$

since $x-\theta \propto h_{\theta 1}(x)$, and $h_{\theta\ell}$, $\ell = 0, 1, 2, \ldots$ orthogonal, imply $\langle X-\theta,h_{\theta\ell}\rangle = 0$, $\forall\ell \neq 1$. For $\ell = 1$, $x-\theta \propto h_{\theta 1}(x)$ and $\langle X-\theta,g\rangle = 0$ imply $\langle g,h_{\theta 1}\rangle = 0$. Thus, $i(\theta,x) = x-\theta$ minimizes $\sup_{\Omega\in S} \sigma(\cdot,\theta,\Omega)$.

For all functions ψ satisfying conditions 1) and 2), the Cauchy-Schwartz inequality implies

$$1 = E_{\Phi_\theta}\psi(\theta,X)\cdot\left(\frac{X-\theta}{\sigma^2}\right) \leq (E_{\Phi_\theta}\psi^2(\theta,X))^{1/2}\cdot\left(E_{\Phi_\theta}\left(\frac{X-\theta}{\sigma^2}\right)^2\right)^{1/2} ,$$

so

$$b_H(\psi,\theta) = 2t\cdot(E_{\Phi_\theta}\psi^2(\theta,X))^{1/2}$$

$$\geq 2t\sigma \; .$$

Since $b_H(X-\theta,\theta) = 2t\sigma$, $i(\theta,x) = x-\theta$ minimizes $b_H(\psi,\theta)$ subject to 1) and 2).

Thus, $i(\theta,x) = x-\theta$ minimizes $b_H(\psi,\theta)$ and $\sup_{\Omega\in S}\sigma(\psi,\theta,\Omega)$, and by the argument used in the proof of Corollary 1, for L satisfying L1 and L2, it minimizes the function $H(\psi,\theta,t,S) = E_N L\left(b_H(\psi,\theta) + \sup_{\Omega\in S}\sigma(\psi,\theta,\Omega)\cdot Z\right)$ of the two. □

Unfortunately, the function $i(\theta,x)$ which minimizes $H(\cdot,\theta,t,S)$ is unbounded, and hence, does not satisfy A2. The following lemma is useful for deriving an estimator whose ψ function is <u>bounded</u> and satisfies

1) $E_{\Phi_\theta}\psi(\theta,X)\cdot\left(\frac{X-\theta}{\sigma^2}\right) = 1$,

and

2) $E_{\Phi_\theta}\psi(\theta,X) = 0$,

and has $H(\cdot,\theta,t,S)$ value arbitrarily close to $H(i,\theta,t,S)$.

Denote

$$[h(x)]_{-\gamma}^{\gamma} = \begin{cases} \gamma & \text{if} \quad h(x) > \gamma \\ h(x) & \text{if} \quad |h(x)| \leq \gamma \\ -\gamma & \text{if} \quad h(x) < -\gamma \; , \end{cases}$$

where h is any function from R to R . Now define

$$h_\gamma(x) = \frac{1}{\alpha_\gamma}([h(x)]_{-\gamma}^{\gamma} - \beta_\gamma)$$

where $\qquad \beta_\gamma \equiv E_{\Phi_\theta}[h(X)]_{-\gamma}^{\gamma} \quad (< \infty)$, and $$\alpha_\gamma \equiv E_{\Phi_\theta}([h(X)]_{-\gamma}^{\gamma} - \beta_\gamma)\cdot\left(\frac{X-\theta}{\sigma^2}\right) .$$

h_γ is a truncated, shifted, and scaled transform of h such that h_γ is bounded and satisfies 1) and 2).

Lemma 13. Suppose $h : R \to R$, $h \in L^{\frac{2}{1-\tau}}[\Phi_\theta]$ where $\tau \in (0,1)$ is as in MIX, h satisfies $E_{\Phi_\theta} h(X)\cdot\left(\frac{X-\theta}{\sigma^2}\right) = 1$ and $E_{\Phi_\theta} h(X) = 0$, and L satisfies L1 and L2, then for h_γ as defined above

$$\lim_{\gamma\to\infty} H(h_\gamma,\theta,t,S) = H(h,\theta,t,S) .$$

Proof of Lemma 13. As $\gamma \to \infty$, we have $\beta_\gamma \to 0$ and $\alpha_\gamma \to 1$. Hence for all x , $h_\gamma(x) \to h(x)$ as $\gamma \to \infty$. For γ sufficiently large, $|\beta_\gamma| < 1$, and $|1/\alpha_\gamma| \le 2$. Thus

$$\begin{aligned} |h_\gamma(x)| &\le 2\cdot(|h(x)| + |\beta_\gamma|) \quad \text{for } \gamma \text{ sufficiently large,} \\ &\le 2\cdot|h(x)| + 2 \\ &\equiv \overline{h}(x) . \end{aligned} \tag{43}$$

Hence, for γ sufficiently large $\overline{h}(x)\overline{h}(y)$ is a dominating function of $h_\gamma(x)h_\gamma(y)$. Also, for all $j = 1, 2, \ldots$,

$$E_{\Phi_{\theta\Omega}}|\overline{h}(X_0)\overline{h}(X_{|j|})| < \infty \quad \text{since } h \in L^2[\Phi_\theta] .$$

Thus, the dominated convergence theorem implies

$$\lim_{\gamma\to\infty} E_{\Phi_{\theta\Omega}} h_\gamma(X_0)h_\gamma(X_{|j|}) = E_{\Phi_{\theta\Omega}} h(X_0)h(X_{|j|}) \quad \forall j = 0, 1, 2, \ldots .$$

For $j = 0$, this gives

$$\lim_{\gamma\to\infty} b_H(h_\gamma,\theta) = b_H(h,\theta) .$$

To get the corresponding result for $\lim_{\gamma\to\infty} \sup_{\Omega\in S} \sigma^2(h_\gamma,\theta,\Omega)$ we need the following:

$$\left|2\sum_{j=J}^{\infty} E_{\Phi_{\theta\Omega}} h_\gamma(X_0)h_\gamma(X_{|j|})\right| \le 20\cdot \sum_{j=J}^{\infty} \alpha^\tau(j)\cdot \left(E_{\Phi_\theta} |h_\gamma(X)|^{\frac{2}{1-\tau}}\right)^{1-\tau}$$

for all $\gamma > 0$ and all $\Omega \in S$, by Lemma 1 of Deo (1973), since $h \in L^{\frac{2}{1-\tau}}[\Phi_\theta]$ and $|h_\gamma| \le |h|$,

$$\le 20\cdot\left(\sum_{j=J}^{\infty} \alpha^\tau(j)\right)\cdot \left(E_{\Phi_\theta} |\overline{h}(X)|^{\frac{2}{1-\tau}}\right)^{1-\tau}$$

for all γ, by (43),

$< \varepsilon$ for all γ, for J sufficiently large, since $\sum_{j=1}^{\infty} \alpha^\tau(j) < \infty$ by MIX.

Similarly,

$$\left|2\sum_{j=J}^{\infty} E_{\Phi_{\theta\Omega}} h(X_0)h(X_{|j|})\right| < \varepsilon \quad \text{for } J \text{ sufficiently large.}$$

Hence, the infinite sum over j, the supremum over Ω, and the limit as $\gamma \to \infty$ can be interchanged in the following:

$$\lim_{\gamma\to\infty} \sup_{\Omega\in S} \sigma^2(h_\gamma,\theta,t,\Omega) = \sup_{\Omega\in S} \sum_{j=-\infty}^{\infty} \lim_{\gamma\to\infty} E_{\Phi_{\theta\Omega}} h_\gamma(X_0)h_\gamma(X_{|j|})$$

$$= \sup_{\Omega\in S} \sigma^2(h,\theta,t,\Omega) .$$

Now, by the properties of L (viz., symmetry about zero and convexity), and the argument of the proof of Corollary 1,

$$\lim_{\gamma\to\infty} H(h_\gamma,\theta,t,S) = E_N L\left(\lim_{\gamma\to\infty} b_H(h_\gamma,\theta) + \lim_{\gamma\to\infty} \sup_{\Omega\in S} \sigma(h_\gamma,\theta,\Omega)\cdot Z\right)$$

$$= H(h,\theta,t,S) \text{ using the above results.} \quad \square$$

Applying Lemma 13 to the function $i(\theta,x)$ (which is in $L^{\frac{2}{1-\tau}}[\Phi_\theta]$, $\forall\tau \in (0,1)$), we get a bounded function, viz., $i_\gamma(\theta,x)$, which is within ε of minimizing $H(\cdot,\theta,t,S)$ subject to 1) and 2), for γ sufficiently large. In particular,

$$i_\gamma(\theta,x) = \frac{1}{\alpha_\gamma}\cdot[x-\theta]_{-\gamma}^{\gamma}$$

where $\alpha_\gamma = E_{\Phi_0}[X]_{-\gamma}^{\gamma}\cdot\frac{X}{\sigma^2}$ (note, $\theta = 0$ in this expectation), and $\beta_\gamma = 0$ by symmetry of Φ_θ and anti-symmetry of $i(\theta,x)$ about θ. $i_\gamma(\theta,x)$ is just a truncated version of the (normalized) score function of Φ_θ.

<u>Theorem 4</u>. For L satisfying L1 and L2, and given $\varepsilon > 0$, the estimator T corresponding to $(\psi,d) = (i_\gamma,d^*)$ (where non-zero constants) satisfies

$$R_H(i_\gamma,d^*,\theta,t,S) \le \sup_{\substack{(\psi,d):(\psi,d)\\ \text{correspond to}\\ \text{some } T\in\mathcal{M}}} R_H(\psi,d,\theta,t,S) + \varepsilon ,$$

for γ sufficiently large.

Comments: (i) The result of this Theorem is striking! It is well known that the sample mean, i.e., the M-estimator with $\psi(\theta,x) = x-\theta$ $(\equiv i(\theta,x))$, is asymptotically efficient for all parametric distributions $\Phi_{\theta\Omega}$ with positive, continuous spectral density (see, e.g., Grenander and Rosenblatt

(1957, Chapter 7). The sample mean has asymptotic risk at $\Phi_{\theta\Omega}$ given by $H(i,\tau,\tilde{t},S_\Omega)$ where $S_\Omega \equiv \{\Omega\}$ and $\underline{\tilde{t} = 0}$. By the proof of Theorem 4, the M-estimator with $\psi(\theta,x) = i_\gamma(\theta,x)$ has asymptotic risk at $\Phi_{\theta\Omega}$ which is less than or equal to $H(i,\theta,\tilde{t},S_\Omega) + \varepsilon$, where $S_\Omega \equiv \{\Omega\}$ and $\tilde{t} = 0$ for γ sufficiently large. Hence, the estimator corresponding to (i_γ,d^*) is not only ε-optimal for the robust model, but is asymptotically ε-efficient at each parametric distribution $\Phi_{\theta\Omega}$ for $\Omega \in S$, provided Ω has a finite spectral density! Thus, asymptotic ε-efficiency and ε-robustness against distributional shape and dependence over time are attained simultaneously in this model. Bickel (1981) and Beran (1977a) obtain analogous results (with respect to distributional shape only) for the i.i.d. Hellinger neighborhood model.

(ii) Note, $i(\theta,x)$ is independent of t, the "radius" of the neighborhoods, of σ^2, the variance of the parametric distributions, and of Ω, the correlation matrix of the parametric distributions. $i_\gamma(\theta,x)$ is also independent of Ω, for $\Omega \in S$, and depends on t and σ^2 only for the determination of the magnitude of γ necessary to make $R(i_\gamma,d^*,\theta,t,S)$ within ε of the optimal value. If it is known that t and σ^2 are bounded above by some values t_0 and M, respectively, then $i_\gamma(\theta,x)$ is also independent of t and σ^2.

(iii) Calculation of specific estimates using the estimator with (i_γ,d^*) does not require evaluation of α_γ $(= E_{\Phi_0}[X]^{\gamma}_{-\gamma}\cdot\frac{X}{\sigma^2})$, since the estimate based on $\alpha_\gamma \cdot i_\gamma$ is identical to that based on i_γ. In fact, α_γ is not even needed for calculation of the asymptotic bias, variance, and risk of the estimator if the ε-approximations given by $b_H(i,\theta)$, $\sigma^2(i,\theta,\Omega)$, and $H(i,\theta,t,S)$ are sufficient.

Proof of Theorem 4. d^* satisfies D1 and D2, and by choice of β_γ and α_γ, i_γ satisfies A1 and A2. Further, $i_\gamma(\theta,x)$ satisfies the Lipschitz condition A3 since it is linear in θ for $\theta \in [x-\gamma,x+\gamma]$, is constant in θ for $\theta > x+\gamma$ and $\theta < x-\gamma$, and has smaller Lipschitz constants at the truncation points $x-\gamma$ and $x+\gamma$ than if no truncation occurred (and the function remained linear). The Lipschitz constants for $\theta \in (x-\gamma,x+\gamma)$ exceed those for $\theta \in [x-\gamma,x+\gamma]^c$ and those at the truncation points, and hence, the Lipschitz constants are independent of x. So, the estwmator T corresponding to (i_γ,d^*) is in $\mathcal{M}$.

Also,

$$\lim_{\gamma\to\infty} R_H(i_\gamma,d^*,\theta,t,S) = \lim_{\gamma\to\infty} H(i_\gamma,\theta,t,S) \qquad \text{by Corollary 1,}$$

$$= H(i,\theta,t,S) \qquad \text{by Lemma 13,}$$

$$\leq \inf_{\substack{\psi:\psi \text{ satisfies} \\ \text{A1, A2 and A3}}} H(\psi,\theta,t,S) \qquad \text{by Theorem 3,}$$

$$= \inf_{\substack{\psi:\psi \text{ satisfies} \\ \text{A1, A2 and A3}}} R_H(\psi,\theta,t,S) \qquad \text{by Corollary 1,}$$

$$\leq \inf_{\substack{(\psi,d)\,:\,(\psi,d) \\ \text{corresponds to} \\ \text{some } T \in \mathcal{M}}} R_H(\psi,d^*,\theta,t,S)$$

by Theorem 2. $\square$

5. Models with More General Parametric Univariate Distributions

In this paper we concentrate on the estimation of a location parameter θ for a robust model with stationary, strong mixing, Gaussian, parametric distributions. Analogous robust models can be considered where the parametric distributions (denoted by $P_{\theta\Omega}$) are stationary, strong mixing distributions of processes on the positive integers, whose univariate distributions depend on a scalar parameter θ and satisfy certain maximum likelihood-type regularity conditions (see Andrews (1982)). Neighborhoods of such parametric distributions can be defined in the same manner as done above for Gaussian distributions. The results of Lemmas 2-13, Theorems 1 and 2, and Corollary 1 all hold for these more general parametric distributions (with $P_{\theta\Omega}$ replacing $\Phi_{\theta\Omega}$). Theorems 3 and 4 hold with $i(\theta,x) = I^{-1}(\theta)\dot{\ell}_\theta(x)$ (i.e., with $i(\theta,x)$ equal to the normalized score function of the univariate parametric distribution, see Comment (ii) to Theorem 3) for the case where the bivariate parametric distributions have expansions in terms of orthogonal functions (see Lancaster (1958, 1963) for details of such expansions), some one of which is orthogonal to $I^{-1}(\theta)\dot{\ell}_\theta(x)$ for each bivariate parametric distribution. For parametric distributions whose bivariate distributions do not have an expansion with some orthogonal function which is orthogonal to the score function, Theorems 3 and 4 can be shown to hold for $i(\theta,x)$ given by an infinite series expansion in terms of the (assumed) common orthogonal functions of the expansions of the bivariate parametric distributions (see Andrews (1982)).

Unfortunately, though $i_\gamma(\theta,x)$ (based on $i(\theta,x)$ referred to above) yields an ε-robust estimator for the two cases discussed, it does not yield (in general) an estimator which is also ε-efficient for the parametric distributions. The reason is the lack of a single estimator which is asymptotically efficient for all parametric distributions $P_{\theta\Omega}$ with $\Omega \in S$. (In the former case, however, it does yield an estimator which is ε-efficient for the i.i.d. processes whose univariate distributions correspond to those of the dependent, parametric distributions under consideration.) Thus, the optimality results for general parametric distributions are considerably less satisfactory than for Gaussian distributions, and for this reason the results are not presented in detail here.

6. Acknowledgments

I would like to thank Peter J. Bickel and Thomas J. Rothenberg for their numerous suggestions and comments, and their continual encouragement.

7. Appendix

Proof of Lemma 3. Let

$$\mathcal{F}^K_{t/\sqrt{n}}(\theta) = \left\{F^{(1)} : F^{(1)} \text{ is a distribution on R, and } \Delta_K(F^{(1)}, \Phi_\theta) < \frac{t}{\sqrt{n}}\right\}$$

where $\Delta_K(\cdot,\cdot)$ is the Kolmogorov metric on the space of distributions on R defined by

$$\Delta_K(F^{(1)}, \Phi_\theta) = \sup_{x \in R} |F^{(1)}(x) - \Phi_\theta(x)|$$

where $F^{(1)}(\cdot)$ and $\Phi_\theta(\cdot)$ are the distribution functions of $F^{(1)}$ and Φ_θ, respectively. Define $\{\mathcal{F}^K_{t/\sqrt{n}}(\theta), \Omega\}_n$ exactly as $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$ is defined except with $\mathcal{F}^K_{t/\sqrt{n}}(\theta)$ replacing $\mathcal{F}^H_{t/\sqrt{n}}(\theta)$. Since $\Delta_H(F^{(1)}, \Phi_\theta) \geq \Delta_K(F^{(1)}, \Phi_\theta)$, $\mathcal{F}^H_{t/\sqrt{n}}(\theta) \subset \mathcal{F}^K_{t/\sqrt{n}}(\theta)$, and $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n \subset \{\mathcal{F}^K_{t/\sqrt{n}}(\theta), \Omega\}_n$. We show the result of Lemma 3 holds uniformly over the neighborhoods $\{\mathcal{F}^K_{t/\sqrt{n}}(\theta), \Omega\}_n$, $\forall\theta \in C$, any compact set in Θ , and $\forall\Omega \in S$, which implies the same for the smaller neighborhoods $\{\mathcal{F}^H_{t/\sqrt{n}}(\theta), \Omega\}_n$.

Wang (1979) has proven the following result: Under the hypothesis of Lemma 3, $\exists\ \tilde{T} = \{\tilde{T}_n\}_{n\geq 1}$ such that $\tilde{T}_n - \theta = o_{F_{\theta In}}(1)$ uniformly for $\theta \in C$, any compact set, where $F_{\theta In}$ is the distribution of an i.i.d. process with univariate distribution in $\mathcal{F}^K_{t/\sqrt{n}}(\theta)$. Lemma 3 extends this result to the case of strong mixing processes where $F_{\theta\Omega n} \in \{\mathcal{F}^K_{t/\sqrt{n}}(\theta), S\}_n$, $\forall n$.

$\tilde{T}$ is constructed as follows. Θ , the parameter space, is open and separable. Thus Θ can be represented as the countable union of an increasing sequence of compact sets. That is, $\Theta = \bigcup_{k=1}^{\infty} \Theta_k$ where Θ_k is compact and $\Theta_k \subseteq \Theta_\ell$, for $k \leq \ell$. Let $K(\theta)$ be the first k such that θ (the true θ) is in Θ_k . Let $\hat{F}_n$ be the empirical distribution based on $X_1, \ldots, X_n$. We now define $\tilde{T}_n$ so that with $F_{\theta\Omega n}$-probability $\xrightarrow{n\to\infty} 1$, $\tilde{T}_n \in \Theta_{K(\theta)}$. Let

$$\tilde{T}_{nk} = \text{any } \theta' \in \Theta_k \text{ s.t. } \Delta_K(\hat{F}_n, \Phi_{\theta'}) \leq \inf\{\Delta_K(\hat{F}_n, \Phi_{\tilde{\theta}}) : \tilde{\theta} \in \Theta_k\} + \frac{1}{n} .$$

(For measurability concerns see Wang (1979).) Define $\tilde{T}_n$ as $\tilde{T}_{nK_n}$, where

$$K_n = \begin{cases} \text{the first } k \text{ s.t. } \Delta_K(\hat{F}_n, \Phi_{\tilde{T}_{nk}}) \leq n^{-1/4} , & \text{if such a } k \text{ exists,} \\ 1 , & \text{if no such } k \text{ exists.} \end{cases}$$

Now we claim: $\Delta_K\left(\hat{F}_n, \frac{1}{n}\sum_{j=1}^{n} F^{(1)j}_{\theta n}\right) = O_{F_{\theta\Omega n}}(n^{-1/2})$ u.f.θ,F,Ω (the proof is given below). Then,

$$\Delta_K(\hat{F}_n,\ \Phi_\theta) \le \Delta_K\left(\hat{F}_n,\ \frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}\right) + \Delta_K\left(\frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j},\ \Phi_\theta\right)$$

$$\le O_{F_{\theta\Omega n}}(n^{-1/2}) + \frac{t}{\sqrt{n}} \quad \text{since} \quad \forall j\ ,\quad F_{\theta n}^{(1)j} \in \mathcal{F}_{t/\sqrt{n}}^{K}(\theta)\ ,\quad \forall\theta \in C\ ,$$

$$= o_{F_{\theta\Omega n}}(n^{-1/4}) \quad \text{u.f.}\theta,F,\Omega. \tag{44}$$

So, for n large, with $F_{\theta\Omega n}$-probability $\xrightarrow{n\to\infty} 1$, and for $\theta \in \Theta_k$, i.e., for $k \ge K(\Theta)$, $\inf\{\Delta_K(\hat{F}_n,\ \Phi_{\tilde{\theta}}) : \tilde{\theta} \in \Theta_k\} + n^{-1} \le n^{-1/4}$. In this case, $\tilde{T}_{nk}$ satisfies $\Delta_K(\hat{F}_n,\ \Phi_{\tilde{T}_{nk}}) \le n^{-1/4}$. Hence, for large n and with $F_{\theta\Omega n}$-probability $\xrightarrow{n\to\infty} 1$, $K_n \le K(\theta)$.

Now for $k < K(\theta)$, $\theta \notin \Theta_k$. Since Θ_k is compact, $\exists$ a positive distance ε_k such that $\inf_{\tilde{\theta}\in\Theta} \Delta_K(\Phi_{\tilde{\theta}},\ \Phi_\theta) = \varepsilon_k$. For n sufficiently large, with $F_{\theta\Omega n}$-probability $\xrightarrow{n\to\infty} 1$, $\Delta_K(\hat{F}_n,\ \Phi_\theta) < \varepsilon_k/2$ by (44). So,

$$\Delta_K(\hat{F}_n,\ \Phi_{\tilde{\theta}}) \ge \Delta_K(\Phi_{\tilde{\theta}},\ \Phi_\theta) - \Delta_K(\Phi_\theta,\ \hat{F}_n) \quad \text{for all} \quad \tilde{\theta} \in \Theta_k\ ,$$

$$> \varepsilon_k/2 \quad \text{u.f.}\theta,F,\Omega.$$

Thus, $\inf_{\tilde{\theta}\in\Theta_k} \Delta_K(\hat{F}_n,\ \Phi_{\tilde{\theta}}) \ge \varepsilon_k/2 > 0$. Hence, for n sufficiently large that $n^{-1/4} < \varepsilon_k/2$, with $F_{\theta\Omega n}$-probability $\xrightarrow{n\to\infty} 1$,

$$\inf\{\Delta_K(\hat{F}_n,\ \Phi_{\tilde{\theta}}) : \tilde{\theta} \in \Theta_k\} + n^{-1} > n^{-1/4}\ .$$

So, with $F_{\theta\Omega n}$-probability $\xrightarrow{n\to\infty} 1$, $K_n \not< K(\theta)$. That is,

$$F_{\theta\Omega n}(K_n = K(\theta)) \xrightarrow{n\to\infty} 1 \quad \text{and} \quad F_{\theta\Omega n}(T_n \in \Theta_{K(\theta)}) \xrightarrow{n\to\infty} 1 \quad \text{u.f.}\theta,F,\Omega.$$

This, plus (44), gives:

$$\Delta_K(\Phi_\theta, \Phi_{\tilde{T}_n}) \leq \Delta_K(\Phi_\theta, \hat{F}_n) + \Delta_K(\hat{F}_n, \Phi_{\tilde{T}_n}) = o_{F_{\theta\Omega n}}(1) \quad \text{u.f.}\theta,F,\Omega. \tag{45}$$

The function $g : \theta \mapsto \Phi_\theta$ on $\Theta_{K(\theta)}$ is continuous, is onto the image of $\Theta_{K(\theta)}$, is defined on a compact set, and is to a Hausdorff space, hence by a well known theorem from analysis (see Royden (1968)) this function is a homeomorphism. That is, the map $g^{-1} : \Phi_\theta \longrightarrow \theta$ on the image of $\Theta_{K(\theta)}$ onto $\Theta_{K(\theta)}$ is continuous and one-to-one. Then, for $\tilde{T}_n \in \Theta_{K(\theta)}$,

$$\Delta_K(\Phi_\theta, \Phi_{\tilde{T}_n}) \xrightarrow{n\to\infty} 0 \Longleftrightarrow |g^{-1}(\Phi_\theta) - g^{-1}(\Phi_{\tilde{T}_n})| = |\theta - T_n| \xrightarrow{n\to\infty} 0 .$$

By (45) and since $F_{\theta\Omega n}(\tilde{T}_n \in \Theta_{K(\theta)}) \xrightarrow{n\to\infty} 1$, $|\tilde{T}_n - \theta| = o_{F_{\theta\Omega n}}(1)$ u.f.θ,F,Ω, as desired. □

It just remains to prove the claim. The claim is

$$\Delta_K\left(\hat{F}_n, \frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}\right) = O_{F_{\theta\Omega n}}(n^{-1/2}) \quad \text{u.f.}\theta,F,\Omega.$$

We use the following two lemmas in the proof of the claim.

<u>Lemma 3b</u>. For any sequence, $\langle X_j \rangle_{j\geq 1}$, of real-valued rv's with distribution function $F^{(1)}$, there exists a sequence $\langle U_j \rangle_{j\geq 1}$, of uniform [0,1] rv's such that the joint distribution of $\langle F^{(j)^{-1}}(U_j) \rangle_{j\geq 1}$ is identical to that of $\langle X_j \rangle_{j\geq 1}$.

Lemma 3c. Let $\langle U_{nj} \rangle_{j \geq 1}$ be a sequence of uniform [0,1] random variables with distribution $G_{\theta\Omega n}$. Let $\hat{G}_n$ and $G^{(1)}$ denote the empirical distribution function of $U_{n1}, \ldots, U_{nn}$, and the distribution function of U_{nj}, $\forall j$, respectively. Then, $\Delta_K(\hat{G}_n, G^{(1)}) = O_{G_{\theta\Omega n}}(n^{-1/2})$ uniformly for all distributions $G_{\theta\Omega n}$ provided $G_{\theta\Omega n}$ is strong mixing with mixing numbers less than or equal to $\alpha(m)$, $\forall m$, where $\langle \alpha(m) \rangle_{m \geq 0}$ satisfies (3) of MIX, and provided $E_{G_{\theta\Omega n}}(\sqrt{n}(\hat{G}_n(t) - t)) \cdot (\sqrt{n}(\hat{G}_n(s) - s))$ converges uniformly (over $G_{\theta\Omega n}$) to some $\sigma(t,s)$ as $n \to \infty$.

Lemma 3b is proved in Andrews (1981). Lemma 3c is proved below, following the proof of the claim.

Proof of Claim. We will show

$$B \equiv \sqrt{n}\Delta_K\left(\hat{F}_n, \frac{1}{n}\sum_{j=1}^{n} F_n^{(1)j}\right) \equiv \sup_{x \in R}\left|\sqrt{n}\left(\hat{F}_n(x) - \frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}(x)\right)\right|$$

$$= O_{F_{\theta\Omega n}}(1) \quad \text{u.f.}\theta,F,\Omega,$$

by showing that B converges weakly u.f.θ,F,Ω. By Lemma 3b, for all distributions $F_{\theta\Omega n}$ in the neighborhoods, there exists a process $\langle U_{nj} \rangle_{j \geq 1}$, of uniform [0,1] random variables such that $\langle F_{\theta n}^{(1)j^{-1}}(U_{nj}) \rangle_{j \geq 1} \sim F_{\theta\Omega n}$. Denote the distribution of $\langle U_{nj} \rangle_{j \geq 1}$ by $G_{\theta\Omega n}$. The mixing numbers of $G_{\theta\Omega n}$ are less than or equal to those of $F_{\theta\Omega n}$ and hence, are less than or equal to $\alpha(m)$, $\forall m$, where $\alpha(m)$, $m = 1, 2, \ldots$ satisfy (3) of MIX. Let $\hat{G}_n(t)$ be the empirical d.f. of $U_{n1}, \ldots, U_{nn}$, i.e., $\hat{G}_n(t) = \frac{1}{n}\sum_{j=1}^{n} 1_{[U_{nj} \leq t]}$, $\forall t \in [0,1]$. Let $\hat{F}_n(x)$ be the empirical d.f. of $F_{\theta n}^{(1)1^{-1}}(U_{n1}), \ldots, F_{\theta n}^{(1)n^{-1}}(U_{nn})$, i.e., $\hat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{[F_{\theta n}^{(1)j^{-1}}(U_{nj}) \leq x]}$,

$\forall x \in R$. Define $\hat{GF}_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{[U_{nj} \leq F_{\theta n}^{(1)j}(x)]}$, $\forall x \in R$.

$$\text{Now,} \quad \hat{G}_n(\max_{1 \leq k \leq n} F_{\theta n}^{(1)k}(x)) = \frac{1}{n}\sum_{j=1}^{n} 1_{[U_{nj} \leq \max_{1 \leq k \leq n} F_{\theta n}^{(1)k}(x)]}$$

$$\geq \frac{1}{n}\sum_{j=1}^{n} 1_{[U_{nj} \leq F_{\theta n}^{(1)j}(x)]}$$

$$= \hat{GF}_n(x) ,$$

and
$$\hat{G}_n(\min_{1 \leq k \leq n} F_{\theta n}^{(1)k}(x)) \leq \hat{GF}_n(x) .$$

We claim $\hat{GF}_n(x) = \hat{F}_n(x)$. $\hat{F}_n(x)$ is right continuous. $F_{\theta n}^{(1)j}(x)$ is right continuous, $\forall j = 1, \ldots, n$. And $\frac{1}{n}\sum_{j=1}^{n} 1_{[U_{nj} \leq t_j]}$ is right continuous as a function of t_j , $j = 1, \ldots, n$. Thus, $\hat{GF}_n(x)$ is right continuous. At all points x where $F_{\theta n}^{(1)j}(x)$ is continuous $\forall j = 1, \ldots, n$ we have $\hat{GF}_n(x) = \hat{F}_n(x)$, since

$$\hat{GF}_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{[F_{\theta n}^{(1)j^{-1}}(U_{nj}) \leq F_{\theta n}^{(1)j^{-1}}(F_{\theta n}^{(1)j}(x))]}$$

$$= \hat{F}_n(x) .$$

So, we have equality except at a countable number of points. Two right continuous functions which agree except on a countable set agree on all of R . Thus,

$$\hat{G}_n(\max_{1\le k\le n} F_{\theta n}^{(1)k}(x)) \ge \hat{F}_n(x)$$

and

$$\hat{G}_n(\min_{1\le k\le n} F_{\theta n}^{(1)k}(x)) \le \hat{F}_n(x) \ .$$

So,

$$B \le \sup_{x\in R}\left|\sqrt{n}\left(\hat{G}_n(\max_{1\le k\le n} F_{\theta n}^{(1)k}(x)) - \frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}(x)\right)\right|$$

$$+ \sup_{x\in R}\left|\sqrt{n}\left(\hat{G}_n(\min_{1\le k\le n} F_{\theta n}^{(1)k}(x)) - \frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}(x)\right)\right|$$

$$\equiv B_1 + B_2$$

Since $F_{\theta n}^{(1)\ell} \in \mathcal{F}_{t/\sqrt{n}}^{K}(\theta)$, $\ell = 1, \ldots, n$,

$$\sup_{x\in R}\left|\max_{1\le \ell\le n} F_{\theta n}^{(1)\ell}(x) - \frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}(x)\right|$$

$$\le \sup_{x\in R}\left|\max_{1\le \ell\le n} F_{\theta n}^{(1)\ell}(x) - \Phi_\theta(x)\right| + \sup_{x\in R}\left|\frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}(x) - \Phi_\theta(x)\right|$$

$$< \frac{t}{\sqrt{n}} + \frac{t}{\sqrt{n}}$$

$$= O(1) \ .$$

Hence,

$$B_1 \leq \sup_{x\in R}\left|\sqrt{n}\left(\hat{G}_n(\max_{1\leq k\leq n} F_{\theta n}^{(1)k}(x)) - \max_{1\leq k\leq n} F_{\theta n}^{(1)k}(x)\right)\right|$$

$$+ \sup_{x\in R}\left|\sqrt{n}\left(\max_{1\leq k\leq n} F_{\theta n}^{(1)k}(x) - \frac{1}{n}\sum_{j=1}^{n} F_{\theta n}^{(1)j}(x)\right)\right|$$

$$\leq \sup_{s\in[1,0]} |\sqrt{n}(\hat{G}_n(s) - s)| + O(1) \ .$$

Similarly, $B_2 \leq \sup_{s\in[0,1]} |\sqrt{n}(\hat{G}_n(s) - s)| + O(1)$ u.f.θ,F,Ω.

So, $B \leq 2\cdot \sup_{s\in[0,1]} |\sqrt{n}(\hat{G}_n(s) - s)| + O(1)$ u.f.θ,F,Ω.

Now by Lemma 3c, $\sqrt{n}(\hat{G}_n - G^{(1)}) \xrightarrow{\mathcal{L}} G_0$, uniformly for $G_{\theta\Omega n}$ with α-mixing numbers less than or equal to $\alpha(m)$, $\forall m$, provided the convergence of $n\cdot E_{G_{\theta\Omega n}}(\hat{G}_n(t) - t)(\hat{G}_n(s) - s)$ is uniform over $G_{\theta\Omega n}$, where G_0 is a Brownian bridge on [0,1]. By Durbin's (1973) result this implies $w(\sqrt{n}(\hat{G}_n - G^{(1)})) \equiv \sup_{s\in[0,1]} |\sqrt{n}(\hat{G}_n(s) - G^{(1)}(s))|$ converges weakly, uniformly over distributions $G_{\theta\Omega n}$. That is, $\sup_{s\in[0,1]} \sqrt{n}|\hat{G}_n(s) - s| = O_{G_{\theta\Omega n}}(1)$ uniformly over $G_{\theta\Omega n}$. This implies $B = O_{F_{\theta\Omega n}}(1)$ u.f.θ,F,Ω, as desired. It just remains to show $n\cdot E_{G_{\theta\Omega n}}(\hat{G}_n(t) - t)(\hat{G}_n(s) - s)$ converges uniformly over all $G_{\theta\Omega n}$ which correspond to $F_{\theta\Omega n} \in \{\mathcal{F}^K_{t/\sqrt{n}}(\theta),\ \Omega\}_n$, $\forall n$, $\forall\theta \in C$, $\forall\Omega \in S$. But this follows by an argument analogous to that used in the proof of Lemma 6, where $\mathrm{Var}_{F_{\theta\Omega n}}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} d_{nj}\psi(\theta,\ X_j)\right)$ is shown to converge u.f.θ,F,Ω. □

Proof of Lemma 3c. We will show that Mehra and Rao's (1975) Theorem 3.2, which extends Billingsley's (1968) Theorem 22.1 from ϕ-mixing processes to strong (or α-) mixing processes, holds uniformly for the distributions $G_{\theta\Omega n}$. In Mehra and Rao's theorem we take C_{Nj} , $\forall j$, $\forall N$. This implies

$\psi_N = 1$, $\forall N$, and $a_{Nj} = 1$, $\forall j$, $\forall N$. Define $\sigma(s,t) = \lim_{n\to\infty} n \cdot E_{G_{\theta\Omega n}} (\hat{G}_n(t) - t)(\hat{G}_n(s) - s)$. Their theorem states: If $\{U_{nj} : j \geq 1\}$, $n = 1, 2, \ldots$ are strong mixing with mixing numbers which satisfy (3) of MIX, then $\sigma(s,t)$ exists, $|\sigma(s,t)| < \infty$, and if $\sigma(s,t) > 0$ $\forall 0 < s,t < 1$, then $\sqrt{n}(\hat{G}_n - G^{(1)}) \xrightarrow{\mathcal{L}_{G_{\theta\Omega n}}} G_0$, where $G_0 = \{G_0(t) : 0 \leq t \leq 1\}$ is a Gaussian process tied down at 0 and 1 and defined by $EG_0(t) = 0$ and $EG_0(s)G_0(t) = \sigma(s,t)$, $\forall 0 \leq s,t \leq 0$. The proof of this theorem consists of showing tightness and of showing convergence of all finite dimensional distributions. Their proof of tightness (which relies on their Lemmas 2.2, 2.4i), 2.5i), and 2.6i), and on Billingsley's (1968) proof of Theorem 22.1) depends on the structure of dependence of $G_{\theta\Omega n}$ only through the mixing numbers of $G_{\theta\Omega n}$. Thus the tightness result holds uniformly for all distributions with mixing numbers less than or equal to $\alpha(m)$, $\forall m$.

The convergence of the finite dimensional distributions is shown using their Theorem 3.1, which is an application of Philipp's (1969) Theorem 3. The theorem states: Suppose $\{Y_{nj}, j \geq 1\}$, $n = 1, 2, \ldots$ is strong mixing with $|Y_{nj}| \leq 1$ and that $\sum_{j=1}^{\infty} j^2\alpha(j) < \infty$. Assume $\lim_{n\to\infty} \frac{\tau_n^2}{n} > 0$, where $\tau_n^2 = \mathrm{Var}(\sum_{j=1}^{n} Y_{nj})$. Then $\frac{1}{\tau_n} \sum_{j=1}^{n} Y_{nj} \xrightarrow{\mathcal{L}} N(0,1)$ as $n \to \infty$. The proof of this theorem (and of Philipp's (1969) Theorem 3) also depends on the structure of dependence of the random variables only through their mixing numbers. Thus, the theorem holds uniformly for all triangular arrays with mixing numbers less than or equal to $\alpha(m)$, $\forall m$, where $\alpha(m)$, $m = 0, 1, \ldots$ satisfy (3) of MIX. However, the result actually used to show convergence of the finite dimensional distributions is $\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Y_{nj} \to N\left(0, \lim_{n\to\infty} \frac{\tau_n^2}{n}\right)$. Hence we must have the convergence of τ_n^2/n uniform. Hence, we must the convergence of τ_n^2/n uniform. That is, we must have $nE_{G_{\theta\Omega n}}(\hat{G}_n(t)-t)(\hat{G}_n(s)-s)$ converge uniformly over distributions $G_{\theta\Omega n}$ to $\sigma(s,t) > 0$, $\forall 0 < s,t < 1$ as $n \to \infty$. With this proviso, the finite dimensional distributions converge to Gaussian random variables uniformly over distributions $G_{\theta\Omega n}$ with appropriate mixing numbers. Since tightness and convergence of the finite dimensional distributions both hold uniformly, we are done. (If $\sigma(s,t) = 0$, $\forall 0 \leq s,t \leq 1$, the finite dimensional distributions converge weakly (and in probability) to zero vectors. In this case the conclusion of Lemma 3c still holds.) $\square$

FOOTNOTES

1. AMS 1980 subject classifications. Primary 62F35, 62E20; Secondary 62F10. Key works and phrases. Robust and asymptotically efficient estimation, strong mixing Gaussian processes, Hellinger metric, shrinking neighborhoods, minimax criterion.

2. This paper is based on work done by the author at the University of California, Berkeley for his Ph.D. thesis.

REFERENCES

Andrews, D. W. K. (1981). "Generalized empirical processes of dependent random variables with discontinuous marginals," manuscript.

Andrews, D. W. K. (1982). A Model for Robustness against Distributional Shape and Dependence over Time. Ph.D. Thesis, University of California, Berkeley.

Beran, R. (1977a). "Robust location estimates," Ann. Statist. 5, 431-444.

Beran, R. (1977b). "Minimum Hellinger distance estimates for parametric models," Ann. Statist. 5, 445-463.

Beran, R. (1980). "Asymptotic lower bounds for risk in robust estimation," Ann. Statist. 8, 1252-1264.

Bickel, P. J. (1978). "Some recent developments in robust statistics," manuscript.

Bickel, P. J. (1981). "Quelque aspects de la statistique robuste," Lecture Notes in Mathematics—876: Ecole d'Ete de Probabilites de Saint-Flour IX-1979. Springer-Verlag, Heidelberg.

Bickel, P. J. and A. M. Herzburg (1979). "Robustness of design against autocorrelation in time: I," Ann. Statist. 7, 77-95.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton.

Deo, C. M. (1973). "A note on empirical processes of strong-mixing sequences," Ann. Probability 1, 870-875.

Durbin, J. (1973). Distribution Theory for Tests Based on the Sample Distribution Function. S.I.A.M., Philadelphia.

Gastwirth, J. L. and H. Rubin (1975). "The behavior of robust estimators on dependent data," Ann. Statist. 3, 1070-1100.

Grenander, U. and M. Rosenblatt (1957). The Statistical Analysis of Stationary Time Series. Wiley, New York.

Hajek, J. (1971). "Limiting properties of likelihoods and inference," Foundations of Statistical Inference," D. A. Sprott and V. Godanbe (eds.). Hold, Rinehart, and Winston, Toronto.

Hajek, J. (1972). "Local asymptotic minimaxity and admissibility in estimation," Proceedings Sixth Berkeley Symp. Math. Stat. Prob., University of California, Berkeley.

Hampel, F. R. (1968). Contributions to the Theory of Robust Estimation, Ph.D. thesis, University of California, Berkeley.

Hampel, F. R. (1974). "The influence curve and its role in robust estimation," J.A.S.A. 69, 383-393.

Hampel, F. R. (1978). "Optimally bounding the gross-error-sensitivity and the influence of position in factor space," 1978 Proceedings of the A.S.A. Statistical Computing Section, American Statistical Association, 59-64.

Hoyland, A. (1968). "Robustness of the Wilcoxin estimate of location against a certain dependence," Ann. Math. Stat. 39, 1196-1201.

Huber, P. J. (1964). "Robust estimation of a location parameter," Ann. Math. Statist. 35, 73-101.

Huber-Carol, C. (1970). Etude Asymptotique de Test Robustes, Ph.D. Thesis. Eidgen. Technische Hochschule, Zurich.

Ibragimov, I. A. (1962). "Some limit theorems for stationary processes," Theor. Probability Appl. 7, 349-382.

Jaeckel, L. A. (1971). "Robust estimates of location: Symmetry and asymmetric contamination," Ann. Math. Statist. 42, 1020-1034.

Kleiner, B., R. D. Martin, and D. J. Thomson (1979). "Robust estimation of power spectra," J. R. Statist. Soc. B 41, No. 3, 313-351.

Koul, H. L. (1977). "Behavior of robust estimators in the regression model with dependent errors," Ann. Statist. 5, 681-699.

Lancaster, H. O. (1958). "The structure of bivariate distributions," Ann. Math. Statist. 29, 719-736.

LeCam, L. (1966). "Likelihood functions for large numbers of independent observations," Research Papers in Statistics, F. N. Davis (ed.), Wiley, New York.

Martin, R. D. (1979). "Robust estimation of location with autoregressive errors," unpublished manuscript, Dept. of Electrical Engineering, University of Washington, Seattle.

Martin, R. D. (1980). "Robust estimation of autoregressive models," Directions in Time Series, Brillinger, Cleveland, Davis and Tiao (eds.), Institute of Mathematical Statistics, Hayward, California.

Martin, R. D. (1981). "The Cramer-Rao bound and robust M-estimates for autoregressions," Technical Report 9, Dept. of Statistics, University of Washington, Seattle.

Martin, R. D. and J. E. Zeh (1978). "Generalized M-estimates for autoregressions, including small-sample efficiency robustness," Technical Report 214, Dept. of Electrical Engineering, University of Washington, Seattle.

Mehra, K. L. and M. S. Rao (1975). "Weak convergence of generalized empirical processes relative to d_q under strong mixing," Ann. Probability 3, 979-991.

Papantoni-Kazakos, P. and R. M. Gray (1979). "Robustness of estimators on stationary observations," Ann. Probability 7, 989-1002.

Philipp, W. (1969). "The central limit problem for mixing sequences of random variables, Z. Wahrscheinlichkeitstheorie verw. Geb. 12, 155-171.

Portnoy, S. L. (1977). "Robust estimation in dependent situations," Ann. Statist. 5, 22-43.

Portnoy, S. L. (1979). "Further remarks on robust estimation in dependent situations," Ann. Statist. 7, 224-231.

Reed, M. and B. Simon (1981). Methods of Modern Mathematical Physics, I: Functional Analysis (revised and enlarged edition). Academic Press, New York.

Rieder, H. (1979). "A robust asymptotic testing model," Ann. Statist. 6, 1086-1094.

Rieder, H. (1980). "Estimates derived from robust tests," Ann. Statist. 8, 106-115.

Rieder, H. (1981a). "Robustness of one- and two-sample rank tests against gross errors," Ann. Statist. 9, 245-265.

Rieder, H. (1981b). "On local asymptotic minimaxity and admissibility in robust estimation," Ann. Statist. 9, 266-277.

Royden, H. L. (1968). Real Analysis. Macmillan, New York.

Wang, P. C. (1979). Asymptotic Robust Tests in the Presence of Nuisance Parameters, Ph.D. thesis, University of California, Berkeley.

Wang, P. C. (1981). Robust asymptotic tests of statistical hypotheses involving nuisance parameters, Ann. Statist. 9, 1096-1106.