COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 697

ROBUST ESTIMATION OF LOCATION IN A GAUSSIAN PARAMETRIC MODEL: II[1]

Donald W. K. Andrews

March 1984

ROBUST ESTIMATION OF LOCATION IN A GAUSSIAN PARAMETRIC MODEL: II[1]

Donald W. K. Andrews

Yale University

February 1984

Short title: Robust Estimation: II

SUMMARY

ROBUST ESTIMATION OF LOCATION IN A GAUSSIAN PARAMETRIC MODEL: II

This paper extends the results of Andrews (1984) which considers the problem of robust estimation of location in a model with stationary strong mixing Gaussian parametric distributions. Three neighbourhood systems are considered, each of which contains the Hellinger neighbourhoods used in Andrews (1984). Optimal robust estimators for this dependent random variable model are found to be bounded influence estimators with optimal ψ functions which are very nearly of Huber shape. These estimators are quite robust against different "amounts" of dependence, and against lack of dependence. To generate the optimal estimators a minimax asymptotic risk criterion is used, where minimaxing is done over neighbourhoods of the parametric Gaussian distributions. The neighbourhood systems include distributions of strong mixing processes. They allow for deviations from stationarity and from the Gaussian structure of dependence. In addition, deviations from the normal univariate parametric distributions are allowed within neighbourhoods defined by (i) ε_n-contamination, (ii) variational metric distance, and (iii) Kolmogorov metric distance.

1. Introduction

In this paper we consider the dependent random variable (rv) robustness model of Andrews (1984) with larger systems of neighbourhood distributions. Qualitatively distinct results are obtained. The models considered here are based on the independent identically distributed (i.i.d.) models of Bickel (1981). The problem is to estimate 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 dependent rv's. Parametric distributions of the infinite process $\langle X_j \rangle_{j \geq 1} \equiv (X_1, X_2, \ldots)$ are specified, and neighbourhoods of these parametric distributions are constructed. In fact, three different neighbourhood systems are considered--each corresponding to a distinct model. As in Andrews (1984), the parametric distributions are mean θ stationary strong mixing Gaussian distributions. The neighbourhoods contain strong mixing distributions which are "close" to the Gaussian distributions in the sense that their univariate marginal distributions are close in terms of 1) ε-contamination, 2) variational metric distance, or 3) Kolmogorov metric distance, and their bivariate distributions are close in terms of some weak-convergence-inducing metric. The neighbourhoods of the contamination (c) model are contained in those of the variational (V) model which, in turn, are contained in those of the Kolmogorov (K) model. For any one of the three models, the true distribution of $\langle X_j \rangle_{j \geq 1}$ is assumed to lie in the neighbourhood of some parametric distribution. The true distribution of $X_1, \ldots, X_n$ is then just the distribution of the first n rv's of the infinite process $\langle X_j \rangle_{j \geq 1}$. These neighbourhood systems allow for the relaxation of the univariate parametric assumption of

normality, and of the assumption of independence and stationarity of the rv's.

The univariate contamination and variational neighbourhoods used in this paper can be given gross error interpretations (for the latter a theorem of Strassen (1965) is used). Also, Kolmogorov neighbourhoods lie between variational and Prokhorov neighbourhoods in size, and the Prokhorov neighbourhood (which is not considered in this paper) has a gross error interpretation which is similar to that of the variational neighbourhoods (see Hampel (1968, 1971)). In the present model, however, the neighbourhoods are not intended to be interpreted solely as the consequence of gross errors. The neighbourhoods are designed to reflect the belief that the parametric model is merely a close approximation of the true distribution of the process $\langle X_j \rangle_{j \geq 1}$, due to the complexity of nature. Thus, the parametric model is an over-simplification of reality, of which the existence of gross errors is only one possible cause.

The estimation problem under consideration is cast in an asymptotic framework. An optimal robust estimator is determined using a criterion of minimax asymptotic risk (following Huber (1964)). That is, the optimal estimator minimizes, over a class of estimators, the asymptotic maximum risk over all distributions in the neighbourhood system. In the present models, this includes minimaxing over a prespecified set of correlation matrices of the parametric distributions. The effect of this procedure is discussed below (see Section 4). The use of the minimax principle requires careful design of the size and character of the neighbourhood system. The size of the neighbourhoods considered here is made to shrink to zero as the sample size (n) tends to infinity. Thus, for each n we have a neighbourhood system of the whole process $\langle X_j \rangle_{j \geq 1}$, and as n increases the size of the neighbourhoods shrink. This modelling

construct is somewhat analogous to that of local power used in testing theory. The idea is not that the true distribution of the process $\langle X_j \rangle_{j \geq 1}$ somehow becomes closer to that of a parametric distribution as the sample size increases. Rather, shrinking neighbourhoods are used as a device to generate asymptotic properties of estimators which approximate their properties for some specified finite sample size. Further discussion of the use of shrinking (or infinitesimal) neighbourhoods can be found in Andrews (1984) and Bickel (1981). Such neighbourhoods have been used extensively in i.i.d. robustness models, see Huber-Carol (1970), Jaeckel (1971), Beran (1977a, b, 1980), Rieder (1979, 1980, 1981a, b), Bickel (1978, 1981, 1982, 1983), Wang (1979, 1981), and Holmes (1981).

It is important that the neighbourhoods contain a rich array of distributions which are close to the parametric distributions. In particular, the dependence structure of the neighbourhoods should be such that all distributions in the univariate ε-contamination, variational metric, or Kolmogorov metric neighbourhoods are attained as the univariate distribution(s) of some neighbourhood distribution(s) of the process $\langle X_j \rangle_{j \geq 1}$. This is the case, in fact, as shown in Andrews (1982a, section III.C). Thus the neighbourhoods are not sparse.

An optimal estimator is sought from the class of weighted M-estimators. Such estimators solve for θ an equation of the form

$$\sum_{j=1}^{n} d_{nj} \psi(x_j - \theta) = 0$$

where the weights d_{nj} may depend on parameters of dependence (such as correlations). The use of weights is motivated by the generalized least squares estimator which is a weighted M-estimator and is the uniformly

minimum variance unbiased estimator. (See Andrews (1984) for a discussion of this class of estimators.)

The optimal estimators from this class are found for each of the three different neighbourhood models. In fact, the optimal estimators for any two of the three models are expressible quite simply in terms of that of the third, so the description of the solution for any one of the three models is sufficient. The optimal estimator is shown to have equal weights. Note, the sample mean, which is an asymptotically efficient (non-robust) estimator for the parametric model, also has equal weights. The least favorable dependence structure of the neighbourhood distributions is found to be that which has the largest correlation coefficients. This is in accord with the results of Gastwirth and Rubin (1975) (which concern estimator performance only under the parametric model). The optimal ψ function is of the form

$$\psi(x) \begin{cases} = \gamma & \text{for } x > b \\ \in [-\gamma,\gamma] & \text{for } |x| \leq b, \text{ for a constant } b < \infty \\ = -\gamma & \text{for } x < -b \end{cases}.$$

Thus, the optimal estimator has a bounded influence function. In fact, the dependent robust model considered here is closely related to the bounded influence method of Hampel (1968, 1974, 1978) and extended by others such as Handschin et al. (1975), Rousseeuw (1981), Krasker and Welsch (1982), and Huber (1983). The i.i.d. model of Bickel (1981), upon which the present model is based, can be viewed as a formalization of the bounded influence approach of Hampel. The results of the present paper show that this theoretical justification for bounded influence estimators carries over to the dependent case.

The shape of the optimal ψ function for $x \in [-b,b]$ is given by the solution to a calculus of variations problem. This problem yields a constrained integral equation. The solution of this integral equation is given as an infinite expansion in terms of orthonormal polynomials. The coefficients of this expansion, the truncation height γ, and the constant b are calculated numerically. The appearance of an integral equation solution to the calculus of variations problem, as opposed to a closed-form solution, is directly attributable to dependence in the model. Estimators based on dependent data typically have asymptotic variances given by a doubly infinite sum of bivariate expectations involving the ψ function. These cross-product expectations prohibit pointwise minimization, which is possible in the independent case, and lead to the solution being given as an integral equation.

The optimal estimator is found to be nearly linear for $x \in [-b,b]$ and is very well approximated by a Huber estimator (see Huber (1964)) whose truncation height depends upon the "size" of the univariate neighbourhoods and upon the least-favourable dependence structure. (Note, Huber estimators are strictly optimal for the i.i.d. version of the model considered here.) Huber ψ functions are of the form

$$\psi(x) = \left[\frac{\gamma}{b}x\right]_{-\gamma}^{\gamma} \equiv \begin{cases} \gamma & x > b \\ \frac{\gamma}{b}x & x \in [-b,b], \text{ for a constant } b < \infty . \\ -\gamma & x < -b \end{cases}$$

This is a useful approximation to the optimal ψ function because the determination of the optimal Huber estimator (which only requires finding the optimal truncation height) is much easier than the determination of the strictly optimal estimator (which requires calculation of the infinite

expansion referred to above).

It is shown that the truncation height of the optimal estimator (or of the optimal Huber estimator) is relatively insensitive to the least favourable dependence structure, unless considerable positive dependence is allowed. Thus, the optimal ψ function for dependent rv's is remarkably similar to that for independent rv's. This result is most convenient.

A clear dichotomy exists between the optimal ψ function for the Hellinger model considered in Andrews (1984) and those of the models considered here. This stems from the small size of the Hellinger neighbourhoods as compared to the neighbourhoods considered here. Optimal Hellinger ψ functions are Huber ψ functions with arbitrarily large truncation height γ. Consequently, the bound on their influence function is arbitrarily large. On the other hand, the corresponding estimator is arbitrarily close to being asymptotically efficient at the parametric model. For contamination, variational, and Kolmogorov models, such estimators are far from optimal. In fact, their asymptotic maximum risk over c, V, and K neighbourhoods is arbitrarily large. Whether or not the Hellinger neighbourhoods are too small depends, of course, on the unknown true distribution. However, if there is a possibility that the true distribution lies farther away from the Gaussian distribution than the Hellinger neighbourhoods allow, then the c, V, or K neighbourhoods are more appropriate.

This paper is organized as follows: Section 2 sets up the problem to be solved. It contains definitions, regularity conditions, and a formal statement of the problem. Section 3 gives the theoretical results of the paper including a solution to the problem in terms of a constrained integral equation, and a solution to the integral equation in terms of

an infinite expansion using orthonormal polynomials. Section 4 then describes the results of numerically solving the integral equation via calculation of the coefficients of the expansion. It also contains comparisons of the performance of the optimal estimators with the simpler Huber estimators, and with the optimal non-robust estimator, under a variety of scenarios. Section 5 contains some proofs which are omitted from Section 3 for purposes of continuity and ease of reading.

The results presented in this paper concern the case where the parametric distributions are mean θ stationary Gaussian distributions. The neighbourhood system and theoretical results are extended in Andrews (1982a) to more general single parameter stationary parametric distributions (whose univariate marginal distributions satisfy maximum-likelihood-type regularity conditions). For these generalizations, numerical calculations are necessary to yield the optimal ψ function in closed form. (Such calculations have not been performed.)

For references to other papers which consider robust inference in the context of dependent rv's see Andrews (1984).

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

Some of the definitions and regularity conditions given here are analogous to those of Andrews (1984). The reader is referred to comments given in that paper regarding the choice and generality of the definitions and assumptions.

2.1. The model under consideration is given by (1). The X_j are observed, and θ is an unknown parameter to be estimated. The parameter space, Θ, is an open subset of R.

2.2. 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 defined in section 2.3 below. $\Phi_{\theta\Omega}$ satisfies the mixing condition MIX defined in section 2.5. 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 rv's X_j, X_k of distance r apart (i.e., $|j-k| = r$) is denoted $\Phi^r_{\theta\Omega}$. The densities of Φ_θ and $\Phi^r_{\theta\Omega}$, with respect to Lebesgue measure, are denoted by φ_θ and $\varphi_\theta(\cdot,\cdot,\rho_r)$, respectively, where $\underset{\sim}{\rho} = [\rho_j]_{j\geq 0} \equiv [\mathrm{Cov}_{\Phi_{\theta\Omega}}(X_1, X_{j+1})/\sigma^2]_{j\geq 0}$ is the correlation vector corresponding to Ω.

2.3. Each correlation matrix Ω of a stationary parametric distribution $\Phi_{\theta\Omega}$ is completely characterized by its correlation vector $\underset{\sim}{\rho}$. We specify restrictions, in terms of the vector $\underset{\sim}{\rho}$, on the matrices Ω over which minimaxing is to take place. Correlation matrices Ω of parametric distributions $\Phi_{\theta\Omega}$ must lie in the set S, where S is defined by

$$S = \Big\{\Omega : \Omega \text{ is a positive definite, semi-infinite, Toeplitz matrix with unit main diagonal elements and } j^{\text{th}} \text{ diagonal elements } \rho_j, \text{ where } |\rho_j| \leq \rho_j^B, \forall j \geq 1\Big\},$$

where $\underset{\sim}{\rho}^B = (1, \rho_1^B, \rho_2^B, \ldots)'$ is some specified positive definite correlation vector (i.e., $\underset{\sim}{\rho}^B$ corresponds to a positive definite Toeplitz matrix Ω^B) with $\rho_j^B \geq 0$, $\forall j$. (By semi-infinite we mean Ω corresponds to a process on the positive integers.) $\underset{\sim}{\rho}^B$ and Ω^B are referred to as the boundary correlation vector and matrix, respectively.

Comments: (i) For example, the boundary correlation vector can be taken as that corresponding to any stationary, finite-parameter autoregressive-moving average (ARMA) process.

(ii) The boundary correlation vector is specified on the basis of experience in the same manner as is the choice of parametric family. However, the results of Section 4 show that the optimal robust estimator is not too sensitive to the choice of $\underset{\sim}{\rho}^B$, unless quite large "amounts" of correlation are considered.

2.4. Three different models are considered, each corresponding to a different sequence of neighbourhoods of the parametric distributions $\Phi_{\theta\Omega}$. The three sequences of neighbourhoods are denoted by $\{\mathcal{F}^i_{t/\sqrt{n}}(\theta),\ \Omega\}_n$, $n = 1, 2, \ldots$, for $i = c$ (contamination), V (variational), and K (Kolmogorov). Each neighbourhood 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 $F^j_{\theta n}$ be the univariate marginal distribution of X_j under $F_{\theta\Omega n}$, for $j = 1, 2, \ldots$. Let $F^{j,k}_{\theta\Omega n}$ denote the bivariate distribution of X_j, X_k under $F_{\theta\Omega n}$. The density of $F^j_{\theta n}$, with respect to a specified measure, is denoted by $f^j_{\theta n}$, $\forall j = 1, 2, \ldots$.

A distribution $F_{\theta\Omega n}$ is in the n^{th} i-type neighbourhood of $\Phi_{\theta\Omega}$, i.e., $F_{\theta\Omega n} \in \{\mathcal{F}^i_{t/\sqrt{n}}(\theta),\ \Omega\}_n$, for $i = c$, V or K, if:

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

F2: $F^{j}_{\theta n} \in \mathcal{F}^{i}_{t/\sqrt{n}}(\theta)$, $\forall j = 1, 2, \ldots$, where $\mathcal{F}^{i}_{t/\sqrt{n}}(\theta)$ is a neighbourhood of the univariate parametric distribution Φ_θ defined as in Bickel (1981) by

a) Contamination $(i = c)$:

$$\mathcal{F}^{c}_{t/\sqrt{n}}(\theta) = \left\{F : F(x) = (1 - t/\sqrt{n})\Phi_\theta(x) + H(x)t/\sqrt{n}, \text{ for } H(\cdot) \text{ arbitrary}\right\} ,$$

where $F(\cdot)$ and $H(\cdot)$ are distribution functions (dfs) on R .

b) Variational and Kolmogorov $(i = V \text{ and } K)$:

$$\mathcal{F}^{i}_{t/\sqrt{n}}(\theta) = \{F : \Delta_i(F, \Phi_\theta) < t/\sqrt{n}\} ,$$

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

$$= \frac{1}{2}\int |f - \varphi_\theta| d\mu ,$$

where $f = dF/d\mu$ and $\mu >> \Phi_\theta$, and

$$\Delta_K(F, \Phi_\theta) \equiv \sup_{x \in R} |F(x) - \Phi_\theta(x)| .$$

F3: The bivariate distributions of $F_{\theta\Omega n}$ lie in Δ_W-metric neighbourhoods of $\Phi_{\theta\Omega}$, where Δ_W is any fixed but arbitrary weak-convergence-inducing metric on R^2 , and the sizes of such neighbourhoods are $o(1)$ as $n \to \infty$. That is, for some $d(n)$ which satisfies $d(n) \xrightarrow{n\to\infty} 0$, $\Delta_W(F^{j,k}_{\theta\Omega n}, \Phi^{r}_{\theta}) \leq d(n)$ for all r , and for all j, k such that $|j-k| = r$.

Comment: The univariate neighbourhoods are related as follows:

$$\left.\begin{array}{l} \mathcal{F}^{c}_{t/\sqrt{n}}(\theta) \subset \\ \\ \mathcal{F}^{H}_{t/\sqrt{n}}(\theta) \subset \end{array}\right\} \mathcal{F}^{V}_{t/\sqrt{n}}(\theta) \subset \mathcal{F}^{K}_{t/\sqrt{n}}(\theta) \ ,$$

where $\mathcal{F}^{H}_{t/\sqrt{n}}(\theta)$ is the Hellinger metric neighbourhood of radius $t/\sqrt{n}$ about Φ_θ .

2.5. The following mixing condition is assumed to hold:

MIX: $\Phi_{\theta\Omega}$ and $F_{\theta\Omega n}$, where $F_{\theta\Omega n} \in \{\mathcal{F}^{i}_{t/\sqrt{n}}(\theta), \Omega\}_n$, $\forall n$, $\forall \theta \in \Theta$, $\forall \Omega \in S$, are strong mixing with mixing numbers generically denoted by $\tilde{\alpha}(m)$, $m = 1, 2, \ldots$. There exist non-negative non-increasing numbers $\alpha(m)$, $m = 1, 2, \ldots$, such that for all $\Phi_{\theta\Omega}$ and $F_{\theta\Omega n}$, $\tilde{\alpha}(m) \leq \alpha(m)$, $m = 1, 2, \ldots$, and $\alpha(m)$, $m = 1, 2, \ldots$, satisfy $\sum_{m=1}^{\infty} m^2 \alpha(m)^{\tau} < \infty$, for some $\tau \in (0,1)$.

2.6. Let $\langle F_{\theta\Omega n} \rangle_{n\geq 1}$ denote an arbitrary sequence of neighbourhood distributions $F_{\theta\Omega n}$, $n = 1, 2, \ldots$. That is, $F_{\theta\Omega n} \in \{\mathcal{F}^{i}_{t/\sqrt{n}}(\theta), \Omega\}_n$, $\forall n$, and $\Omega \in S$ where $i = c$, V, or K. Let $\text{unif}(\theta,F,\Omega)_i$ abbreviate "uniformly for $\theta \in C$, where C is any compact set in Θ , uniformly for $\langle F_{\theta\Omega n} \rangle_{n\geq 1}$, and uniformly for $\Omega \in S$, where the neighbourhoods are of type i ," for $i = c$, V, or K.

2.7. Define $T = \langle T_n \rangle_{n\geq 1}$ to be a <u>weighted M-estimator</u> of θ for some Lebesgue measurable function $\psi : \Theta \times R \to R$, and for some triangular array of weights $\{d_{nj}\}$, if

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

where $\tilde{T} = \langle \tilde{T}_n \rangle_{n\geq 1}$ is some estimator of θ which is consistent unif$(\theta, F, \Omega)_i$. Such an estimator exists by Andrews (1982a, b).

The following assumptions on ψ hold when specified:

A1: $\psi \in \Psi(\theta) = \left\{ \tilde{\psi} : \Theta \times R \to R \,\middle|\, E_{\Phi_\theta} \tilde{\psi}(\theta, X) = 0,\ E_{\Phi_\theta} \tilde{\psi}(\theta, X)\cdot(X-\theta)/\sigma^2 = 1, \right.$
$\left. \text{and } E_{\Phi_\theta} \tilde{\psi}^2(\theta, X) < \infty \right\}$.

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.

A4: $\psi(\theta, x) = \psi(0, x-\theta)$ $(\equiv \psi(x-\theta))$.

Comment: Assumptions A1-A3 are discussed in Andrews (1984). A4 implies that the estimator is location equivariant with probability that converges to one as $n \to \infty$. This condition is not used for the theoretical results of the paper, but it simplifies the numerical calculations of Section 4 considerably.

For the K neighbourhood model the following two stronger assumptions replace A2 and A3:

A2': There exists a finite, signed measure ν_θ and some constant C_3 such that $\psi(\theta,x) = \nu_\theta((-\infty, x]) + C_3$ a.e. and $\|\nu_\theta\|_V \leq C_1(\theta)$, where $C_1(\theta)$ is bounded on compact sets, and $\|\cdot\|_V$ is the variational norm (see condition F2).

A3': $\|\nu_{\theta+h_2} - \nu_{\theta+h_1}\|_V \leq C_2(\theta)|h_2 - h_1|$, for ν_θ is as in A2', where $C_2(\theta)$ is bounded on compact sets.

Comment: The existence of a measure ν_θ in A2' and A3' is not restrictive since the minimand is finite if and only if there exists such a measure (see Bickel (1981)).

The triangular array of weights $\{d_{nj}\}$ are assumed to satisfy:

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

D2: The limit of $\bar{d}_n$ $(\equiv \frac{1}{n}\sum_{j=1}^{n} d_{nj})$, $n = 1, 2, \ldots$, exists and equals η $(\neq 0)$.

Without loss of generality, take $\eta = 1$ and assume the following limits exist:

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

and (2)

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

2.8. Let $\mathcal{M}_i$, i = c, V, or K be classes of weighted M-estimators with weights $\{d_{nj}\}$ which satisfy D1 and D2, and ψ functions which satisfy A1, A2, and A3 for i = c or V , and A1, A2' and A3' for i = K . Let $\mathcal{M}_i^{\ell}$ be the location equivariant subset of $\mathcal{M}_i$. That is, estimators in $\mathcal{M}_i^{\ell}$ also satisfy A4.

2.9. Let $L : R \to [0,\infty)$ be a given loss function which satisfies conditions L1 and L2 when specified:

L1: L is symmetric about 0 and increasing;

L2: L is convex.

2.10. The problem for i = c, V, or K can now be stated as follows: Find an estimator $T = \langle T_n \rangle_{n\geq 1} \in \mathcal{M}_i^{\ell}$ of θ which attains the infimum of the expression

$$R_i(\psi,d,\theta,t,S) \equiv \sup_{a>0} \overline{\lim_{n\to\infty}} \sup_{\Omega\in S} \sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} E_{F_{\theta\Omega n}} L_a\left(\sqrt{n}(T_n - \theta)\right) \quad (3)$$

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

3. Theoretical Results

We now proceed to simplify the minimand $R_i(\psi,d,\theta,t,S)$. Some of the results of this section parallel those of Andrews (1984). In such cases, the reader is referred to that paper for the proofs. Any necessary adjustments of those proofs are given below. Theorems, lemmas, and corollaries of this paper are numbered B1, B2, ... Results referred to by numbers 1, 2, ... are from Andrews (1984).

Theorem B1. For $T \in \mathcal{M}_i$,

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

$$\Rightarrow N\left(0, \sum_{j=-\infty}^{\infty}(1+\omega_{|j|})E_{\Phi_{\theta\Omega}}\psi(\theta, X_1)\psi(\theta, X_{1+|j|})\right) \qquad (4)$$

as $n \to \infty$, $\text{unif}(\theta,F,\Omega)_i$ for $i = c, V,$ or K , where $\lambda(\psi,\theta,F) \equiv E_F\psi(\theta,X)$, and where $\mathcal{L}_F(Y)$ denotes the distribution (or law) of Y under F .

Comment: The proof of Theorem 1 (of Andrews (1984)) holds as stated when Hellinger neighbourhoods are replaced by c or V neighbourhoods, as in Theorem B1. The proof needs to be altered slightly for $i = K$ (see Appendix).

Corollary B1. If $T \in \mathcal{M}_i$ and L satisfies L1 and L2, then

$$R_i(\psi,d,\theta,t,S) = E_N L\left(b_i(\psi,d,\theta) + \sup_{\Omega\in S}\sigma(\psi,d,\theta,\Omega)\cdot Z\right) , \qquad (5)$$

for $i = c, V,$ or K, where $Z \sim N(0,1)$, and where

$$b_c(\psi,d,\theta) \equiv \tilde{\eta}\cdot t\cdot\frac{1}{2}\left(\operatorname*{ess\,sup}_{x\in R}\psi(\theta,x) - \operatorname*{ess\,inf}_{x\in R}\psi(\theta,x)\right)$$

$$+ t\cdot\frac{1}{2}\cdot\left|\operatorname*{ess\,sup}_{x\in R}\psi(\theta,x) + \operatorname*{ess\,inf}_{x\in R}\psi(\theta,x)\right| , \qquad (6)$$

$$b_V(\psi,d,\theta) \equiv \tilde{\eta}\cdot t\cdot\left(\operatorname*{ess\,sup}_{x\in R}\psi(\theta,x) - \operatorname*{ess\,inf}_{x\in R}\psi(\theta,x)\right) , \qquad (7)$$

$$b_K(\psi,d,\theta) \equiv \tilde{\eta}\cdot t\cdot\|\nu_\theta\|_V , \qquad (8)$$

and

$$\sigma^2(\psi,d,\theta,\Omega) = \sum_{j=-\infty}^{\infty} (1+\omega_{|j|}) E_{\Phi_{\theta\Omega}} \psi(\theta, X_1)\psi(\theta, X_{1+|j|}) \ . \qquad (9)$$

Comment: The proof is the same as for Corollary 1 with Lemma 6 replaced by Lemma B1.

Lemma B1. If $\{\psi,d\}$ corresponds to some $T \in \mathcal{M}_i$, then

$$\sup_{<F_{\theta\Omega n}>_{n\geq 1}} \left| \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} d_{nj} \lambda(\psi, \theta, F^j_{\theta n}) \right| = b_i(\psi,d,\theta) \ , \text{ for } i = c, V, \text{ and } K \ . \qquad (10)$$

N.B.: The proof is given in the Appendix.

We now calculate the infinum over the triangular arrays of weights $\{d_{nj}\}$:

Theorem B2. If $T \in \mathcal{M}_i$ and L satisfies L1 and L2, then

$$\inf_{\{d_{nj}\}: D1 \text{ and } D2 \text{ hold}} R_i(\psi,d,\theta,t,S) = R_i(\psi,d^*,\theta,t,S) \ , \qquad (11)$$

where $d^* = \{d^*_{nj}\}$ is any triangular array of identical nonzero constants, for $i = c, V,$ and K.

Without loss of generality we can take $d^*_{nj} = 1$, $\forall j$, $\forall n$. Then, define $R_i(\psi,\theta,t,S) = R_i(\psi,d^*,\theta,t,S)$, $b_i(\psi,\theta) = b_i(\psi,d^*,\theta)$, and $\sigma^2(\psi,\theta,\Omega) = \sigma^2(\psi,d^*,\theta,\Omega)$, for $i = c, V,$ and K. The asymptotic maximum bias and the asymptotic variance simplify to:

$$b_c(\psi,\theta) = t \cdot \operatorname*{ess\,sup}_{x\in R} |\psi(\theta,x)| \ ; \quad b_V(\psi,\theta) = t \cdot (\operatorname*{ess\,sup}_{x\in R} \psi(\theta,x) - \operatorname*{ess\,inf}_{x\in R} \psi(\theta,x)) \ ;$$

$$b_K(\psi,\theta) = t \cdot ||\nu_\theta||_V \ ; \text{ and } \sigma^2(\psi,\theta,\Omega) = \sum_{j=-\infty}^{\infty} E_{\Phi_{\theta\Omega}} \psi(\theta, X_1)\psi(\theta, X_{1+|j|}) \ .$$

Proof of Theorem B2. The proof is the same as for Theorem 2, since $b_i(\psi,d,\theta)$ is minimized by minimizing $\tilde{\eta}$ for $i = c, V,$ and K. □

Next we show that Ω^B is the least favorable correlation matrix.

Theorem B3. For ψ satisfying A2,

$$\sup_{\Omega\in S} \sigma^2(\psi,\theta,\Omega) = \sigma^2(\psi, \theta, \Omega^B) . \tag{12}$$

Comments: (i) This shows that the (perhaps unrealistic) inclusion of Ω matrices in S with large amounts of negative correlation is irrelevant.

(ii) Suppose the true Ω is known, call it Ω^T. Then, the optimal estimator can be calculated by taking $\Omega^B = \Omega^T$.

Proof of Theorem B3. The bivariate normal density $\varphi_\theta(x,y,\rho)$ can be expanded in terms of the Hermite polynomials $H_k(\cdot)$, $k = 0, 1, \ldots,$ as

$$\varphi_\theta(x,y,\rho) = \sum_{k=0}^{\infty} \frac{1}{k!} H_k\left(\frac{x-\theta}{\sigma}\right) H_k\left(\frac{y-\theta}{\sigma}\right) \rho^k \varphi_\theta(x)\varphi_\theta(y) \tag{13}$$

where equality is the sense of mean square convergence. For $j = \pm1, \pm2, \ldots,$

$$\begin{aligned} E_{\Phi_{\theta\Omega}} \psi(\theta, X_1)\psi(\theta, X_{1+|j|}) &= \sum_{k=0}^{\infty} \left[\int_R \psi(\theta,x) H_k\left(\frac{x-\theta}{\sigma}\right)\varphi_\theta(x)dx\right]^2 (\rho_{|j|})^k \\ &\le \sum_{k=0}^{\infty} \left[\int_R \psi(\theta,x) H_k\left(\frac{x-\theta}{\sigma}\right)\varphi_\theta(x)dx\right]^2 (\rho^B_{|j|})^k \\ &= E_{\Phi_{\theta\Omega^B}} \psi(\theta, X_1)\psi(\theta, X_{1+|j|}) , \end{aligned} \tag{14}$$

where the two equalities follow using (13), the boundedness of ψ, and a Fourier series result (see, e.g., Andrews (1982a, Lemma 11)). Summing over $j = 0, \pm 1, \pm 2, \ldots$ gives the desired result. □

The results obtained thus far are summarized by:

Corollary B2. For $T \in \mathcal{M}_i$ and for L satisfying L1 and L2,

$$R_i(\psi,\theta,t,S) = E_N L\left(b_i(\theta,\psi) + \sigma(\psi,\theta,\Omega^B)\cdot Z\right) \quad \text{for} \quad i = c, V, \text{and } K. \tag{15}$$

Under A4, $\psi(\theta,x) = \psi(0, x-\theta)$. Define $\psi(x-\theta) = \psi(0, x-\theta)$. The next result shows that only odd (about θ) $\psi(\cdot)$ functions need to be considered:

Theorem B4. If there exists a $\psi(\cdot)$ function satisfying A1-A4 which minimizes (15), then there exists an odd (about θ) $\psi(\cdot)$ function which also solves the problem, for $i = c, V,$ and K. If there exists only a sequence of functions $\langle\psi_q\rangle_{q\geq 1}$ which approaches the infimum of (15) as $q \to \infty$, then there also exists such a sequence of odd (about θ) $\psi(\cdot)$ functions.

Proof of Theorem B4. For any function $\psi(\theta,x)$ which satisfies A1-A4, consider the odd version of $\psi(\theta,x)$ defined by

$$\tilde{\psi}(\theta,x) \equiv [\psi(\theta,x) - \psi(-\theta, -x)]/2 . \tag{16}$$

$\tilde{\psi}$ satisfies A1-A4, since φ_θ is symmetric about θ. Without loss of generality assume $\theta = 0$. We have

$$J(\tilde{\psi}) \equiv \sigma^2(\tilde{\psi}, 0, \Omega^B) \leq \sigma^2(\psi, 0, \Omega^B) \equiv J(\psi) ,$$

since $J(\psi)$ is a convex function. The latter follows because

$$J\left(c\psi_1 + (1-c)\psi_2\right) = \lim_{n\to\infty} E_{\Phi_{0\Omega^B}}\left(c\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi_1(X_j) + (1-c)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi_2(X_j)\right)^2$$

$$\leq \lim_{n\to\infty} E_{\Phi_{0\Omega^B}}\left[c\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi_1(X_j)\right)^2 + (1-c)\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi_2(X_j)\right)^2\right]$$

$$= cJ(\psi_1) + (1-c)J(\psi_2) \ , \tag{17}$$

where the inequality follows because $g(x) = x^2$ is convex. Also,

$$b_c(\tilde{\psi},0) \leq b_c(\psi,0) \quad \text{by the triangle inequality,}$$

$$b_V(\tilde{\psi},0) \leq t\cdot([\text{ess sup } \psi - \text{ess inf } \psi]/2 - [\text{ess inf } \psi - \text{ess sup } \psi]/2)$$

$$= b_V(\psi,0) \ ,$$

and $$b_K(\tilde{\psi},0) = t\cdot\frac{1}{2}\int_R d\tfrac{1}{2}|\nu_0(x) - \nu_0(-x)| \leq b_K(\psi,0) \ ,$$

by the triangle inequality. Thus, by the argument of the proof of Corollary 1, for L satisfying L1 and L2, and for $i = c$, V, and K ,

$$E_N L\left(b_i(\tilde{\psi},\theta) + \sigma(\tilde{\psi},\ \theta,\ \Omega^B)\cdot Z\right) \leq E_N L\left(b_i(\psi,\theta) + \sigma(\psi,\ \theta,\ \Omega^B)\cdot Z\right) \ .$$

The case for sequences $\langle\psi_q\rangle_{q\geq 1}$ is proved analogously, term by term. □

For odd ψ functions,

$$b_V(\psi,\theta) = 2t\cdot\text{ess sup}|\psi(x-\theta)| = 2b_c(\psi,\theta) \ . \tag{18}$$

Thus the variational model with $t = t'/2$ has the same asymptotic maximum bias as the contamination model with $t = t'$. Since the asymptotic

variance is the same in both models:

$$R_c(\psi,\theta,t',S) = R_V(\psi,\ \theta,\ t'/2,\ S)\ . \tag{19}$$

Further, if the optimal ψ function for the variational (or contamination) model, call it ψ^*, is monotone (as it is for the numerical calculations carried out), then

$$b_K(\psi^*,\theta) = b_V(\psi^*,\theta) \quad (= 2\cdot b_c(\psi^*,\theta))\ , \tag{20}$$

and

$$R_K(\psi^*,\theta,t,S) = R_V(\psi^*,\theta,t,S)\ . \tag{21}$$

Since $\{\mathcal{F}^V_{t/\sqrt{n}}(\theta),\ \Omega\}_n \subset \{\mathcal{F}^K_{t/\sqrt{n}}(\theta),\ \Omega\}_n$, $\forall n$, $\forall \Omega \in S$, (21) implies that ψ^* is also optimal for the Kolmogorov model. Hence, it suffices to consider just the contamination model.

Minimizing (15) is easier if the doubly infinite sum of $\sigma^2(\psi,\theta,\Omega^B)$ is truncated at a finite number of terms. By truncating at a sufficiently high point the optimal ψ function for the "truncated" minimand has risk $R_c(\psi,\theta,t,S)$ which is arbitrarily close to the minimum of (15) over $T \in \mathcal{M}_1^\ell$.

Lemma B2. Suppose $T \in \mathcal{M}_1^\ell$ and L satisfies L1 and L2. Given any $\varepsilon > 0$, there exists an integer $M \in (0,\infty)$ such that

$$\left|\sigma^2_M(\psi,\theta,\Omega^B) - \sigma^2(\psi,\theta,\Omega^B)\right| < \varepsilon \ , \text{ and } \ \left|H_{cM}(\psi) - R_c(\psi,\theta,t,S)\right| < \varepsilon \ , \tag{22}$$

where
$$\sigma^2_M(\psi,\theta,\Omega^B) \equiv \sum_{j=-M}^{M} E_{\Phi_{\theta\Omega}}\psi(\theta,\ X_1)\psi(\theta,\ X_{1+|j|}) \quad (>0)\ , \tag{23}$$

and
$$H_{cM}(\psi) \equiv E_N L\Big(b_c(\psi,\theta) + \sigma_M(\psi,\theta,\Omega^B)\cdot Z\Big)\ . \tag{24}$$

Proof of Lemma B2. Given any $\varepsilon > 0$, for M sufficiently large,

$$2 \cdot \left| \sum_{j=M+1}^{\infty} E_{\Phi_0 \Omega^B} \psi(X_1)\psi(X_{1+|j|}) \right| \leq 8C_1(0)^2 \sum_{j=M+1}^{\infty} \alpha(j) < \varepsilon , \quad (25)$$

using MIX, A2, and a mixing inequality (see Philipp (1969, Lemma 2)). This gives the first inequality of the Lemma. Note, $\sigma_M^2(\psi, 0, \Omega^B)$ is necessarily positive, since equation (14) of the proof of Theorem B3 and $\rho_j^B \geq 0$, $\forall j$, yield

$$\sigma_M^2(\psi, 0, \Omega^B) = \int \psi^2(x)\varphi_0(x)dx + 2 \sum_{j=1}^{M} \sum_{k=1}^{\infty} \left[\int \psi(x) H_k(x/\sigma)\varphi_0(x)dx\right]^2 (\rho_j^B)^k > 0 . \quad (26)$$

The second inequality of the Lemma holds because $E_N L(b_c(\psi,0) + \sigma \cdot Z)$ is continuous in σ. □

By Lemma B2, we can solve

$$\inf_{\substack{\psi: A1, A2, A3, \\ \text{and } A4 \text{ hold}}} H_{cM}(\psi) \quad (27)$$

to get an ε-optimal solution to (3). Of course, if the parametric distributions are M-dependent this solution is strictly optimal. For convenience, and without loss of generality by A4, take $\theta = 0$. As above, $\psi(x) \equiv \psi(0,x)$. The "truncated" problem (27) can be split into two parts. Let

$$\Psi_\gamma = \Big\{ \psi(x) : R \to R \,|\, \psi \text{ satisfies A1, A2, and A3 with } \theta = 0 \text{ and } \operatorname*{ess\,sup}_{x \in R} |\psi(x)| \leq \gamma \Big\} . \quad (28)$$

Then, assuming L1 and L2 hold, (27) can be written as

$$\inf_{\gamma \in R^+} E_N L\left(t \cdot \gamma + \inf_{\psi_\gamma \in \Psi_\gamma} \sigma_M(\psi_\gamma, 0, \Omega^B) \cdot Z\right) , \tag{29}$$

by the argument of the proof of Corollary 1. Thus, we need to solve

$$\inf_{\psi_\gamma \in \Psi_\gamma} \sigma_M^2(\psi, 0, \Omega^B) . \tag{30}$$

Once this is solved for different values of γ, yielding solutions ψ_γ^*, the optimal value of γ, call it γ^*, is simply the smallest value of γ which solves

$$\inf_{\gamma \in R^+} E_N L\left(t \cdot \gamma + \sigma_M(\psi_\gamma^*, 0, \Omega^B) \cdot Z\right) . \tag{31}$$

Equation (31) is relatively easy to solve, at least for simple loss functions. For example, if loss is quadratic, (31) is just

$$\inf_{\gamma \in R^+} \left[t^2 \cdot \gamma^2 + \sigma_M^2(\psi_\gamma^*, 0, \Omega^B)\right] . \tag{32}$$

Note, γ^* must be finite (whether or not quadratic loss is assumed), since $R_c(\psi_\gamma, 0, t, S) \to \infty$ as $\gamma \uparrow \infty$, for all ψ_γ under consideration.

Before attempting to solve problem (30) via standard calculus of variations methods, the question of existence must be addressed, since there may not exist a ψ_γ function which satisfies all the constraints and attains the infimum.

Lemma B3. From the class of ψ_γ functions which satisfy A1, A2, A4, and $\operatorname{ess\,sup}_{x\in R}|\psi_\gamma(x)| \leq \gamma$ (but which do not necessarily satisfy A3), there exists an (a.s.) unique extremal function $\overline{\psi}^*_\gamma$ for the problem

$$\inf_{\psi_\gamma} \sigma^2_M(\psi_\gamma, 0, \Omega^B) \ . \tag{33}$$

Further, $\forall \varepsilon > 0$, $\exists \psi^*_\gamma \in \Psi_\gamma$ (i.e., ψ^*_γ also satisfies A3) such that

$$\sigma^2_M(\psi^*_\gamma, 0, \Omega^B) < \inf_{\psi_\gamma} \sigma^2_M(\psi_\gamma, 0, \Omega^B) + \varepsilon \ . \tag{34}$$

Proof of Lemma B3. The proof is sketched here; details can be found in the proof of Lemma 12 of Andrews (1982a). Define

$$B = \{\psi \in L_2(\Phi_0) : E_{\Phi_0}\psi(X) = 0 \text{ and } E_{\Phi_0}\psi(X)X/\sigma^2 = 1\} \ ,$$

and

$$B_\gamma = \{\psi \in L_2(\Phi_0) : \operatorname{ess\,sup}_{x\in R}|\psi(x)| \leq \gamma\} \ .$$

Consider the weak* topology on the space of $L_2(\Phi_0)$ functions. In this topology, B is closed, and B_γ is closed and compact. Hence $B \cap B_\gamma$ is compact. Using Reed and Simon (1981, Theorem 5.7), it can be shown that $J_M(\cdot) \equiv \sigma^2_M(\cdot, 0, \Omega^B)$ is lower semi-continuous with respect to the weak* topology on $B \cap B_\gamma$. This, coupled with the compactness result, implies $J_M(\cdot)$ attains its infimum on $B \cap B_\gamma$.

If the function $\overline{\psi}^*_\gamma$ which attains the infimum of $J_M(\cdot)$ on $B \cap B_\gamma$ does not satisfy A3, then take

$$\psi_\gamma^*(x) = \int_R \overline{\psi}_\gamma^*(x-z)s_\varepsilon(z)dz \tag{35}$$

where $s_\varepsilon(\cdot)$ is a close (but smooth) approximation of point mass at zero. □

Lemma B3 implies the existence of an (a.s.) unique extremum for the problem (30) provided the Lipschitz condition is ignored. Hence, the Euler-Lagrange necessary conditions for this calculus of variations problem can be applied. (See Gelfand and Fomin (1963) or Akhiezer (1962) for the classical calculus of variations theory.) For points $x \in R$ not on the boundary of the constraint

$$\operatorname*{ess\,sup}_{x\in R} |\psi_\gamma(x)| \leq \gamma \ , \tag{36}$$

a necessary condition for an extremum is

$$\psi_\gamma(x) = -\frac{\lambda x}{2\sigma^2} - 2\int_R \psi_\gamma(y) \sum_{j=1}^{M} \varphi_0(x,y,\rho_j^B)dy/\varphi_0(x) \tag{37}$$

where the multiplier λ is such that $E_{\Phi_0}\psi_\gamma(X)X/\sigma^2 = 1$. This is a <u>constrained</u> integral equation of Fredholm second type. (Note, the truncation of the sum in (37) at M has allowed the change in order of integration and summation which otherwise would not be justified.) Points $x \in R$ on the boundary of (36) are on the upper (lower) boundary if (37) holds with "=" replaced by "$\leq$" ("$\geq$"). Whether a point is on the boundary of (36) or not is determined by the coincident satisfaction of the integral equation (37) and the constraints. Since $\overline{\psi}_\gamma^*$ is odd, (37) can

be rewritten for x in R^+ (the set of positive real numbers) as

$$\psi_\gamma(x) = \frac{-\lambda x}{2\sigma^2} + g_\gamma(x) + \int_{D_\gamma} \psi_\gamma(y) K_M(x,y)dy \ , \tag{38}$$

where $$K_M(x,y) \equiv -2 \sum_{j=1}^{M} [\varphi_0(x,y,\rho_j^B) - \varphi_0(x,-y,\rho_j^B)]/\varphi_0(x) \ , \tag{39}$$

$$g_\gamma(x) \equiv \gamma \int_{D_\gamma^+} K_M(x,y)dy - \gamma \int_{D_\gamma^-} K_M(x,y)dy \ , \tag{40}$$

and D_γ , D_γ^- , and D_γ^+ are the regions in R^+ where $-\gamma < \psi_\gamma(x) < \gamma$, $\psi_\gamma(x) = -\gamma$, and $\psi_\gamma(x) = +\gamma$, respectively. For $x \in R^-$, $\psi(x) = -\psi(-x)$. Manipulation of the integral equation (38) yields the following:

Lemma B4. The optimal ψ_γ function, $\bar{\psi}_\gamma^*$, which solves problem (33) of Lemma B3 satisfies

$$\bar{\psi}_\gamma^*(x) = \begin{cases} \gamma & \text{for all } x \text{ sufficiently large} \\ -\gamma & \text{for all } x \text{ sufficiently small.} \end{cases} \tag{41}$$

Comment: The optimal ψ_γ function, $\bar{\psi}_\gamma^*$, is similar to the ψ function generated via the bounded influence approach (see Hampel (1968, 1974)) --both are constant for large and small values of x .

Proof of Lemma B4. Since $\rho_j^B \geq 0$, $\forall j$, implies $K_M(x,y) < 0$, $\forall x \in R^+$, the right-hand-side of (38), denoted $W_\psi(x)$, satisfies

$$W_\psi(x) \geq -\frac{\lambda x}{2\sigma^2} + \gamma \int_{R^+} K_M(x,y)dy \geq -\frac{\lambda x}{2\sigma^2} - 2M\gamma > \gamma \ , \tag{42}$$

for x sufficiently large, since $\lambda < 0$. Hence, (38) holds with "$=$" replaced by "$\leq$," and $\bar{\psi}_\gamma^*$ must be on the upper boundary, for x sufficiently

large. The corresponding result for x sufficiently small follows since ψ is odd. □

The solution $\bar{\psi}^*_\gamma$ to the constrained integral equation (37) can be expressed in terms of orthonormal functions. See Courant and Hilbert (1937, Chapter III, §5) for a discussion of this method. For notational convenience, let $\langle s_1, s_2 \rangle$ denote $\int_{D_\gamma} s_1(x)s_2(x)\varphi_0(x)dx$, for any two functions $s_1(\cdot)$ and $s_2(\cdot)$ defined on D_γ. Let $\xi_\ell(x)$, for $x \in D_\gamma$, $\ell = 1, 2, \ldots$ be the orthonormal polynomials (with respect to φ_0 over the region D_γ) obtained by applying the Gram-Schmidt process to the polynomials $1, x, x^2, \ldots$. Then, by definition, $\langle \xi_k, \xi_\ell \rangle = \delta_{k\ell}$, for $k, \ell = 1, 2, \ldots$, where $\delta_{k\ell}$ is the Kronecker delta function. By standard Fourier series results, the bivariate normal density $\varphi_0(x, y, \pm\rho_j^B)$ can be written

$$\varphi_0(x, y, \pm\rho_j^B) = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \xi_k(x)\xi_\ell(x)\varphi_0(x)\varphi_0(y)r_{k\ell}(\pm\rho_j^B) \quad , \forall x, y \in D_\gamma \quad , \tag{43}$$

where equality is in the sense of mean square convergence (against the density $\varphi_0(x)\varphi_0(y)$ over the region $(D_\gamma)^2$), and

$$r_{k\ell}(\rho) \equiv \int_{(D_\gamma)^2} \int_2 \xi_k(x)\xi_\ell(x)\varphi_0(x,y,\rho)dxdy \quad . \tag{44}$$

Thus,

$$K_M(x,y) = -2 \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \xi_k(x)\xi_\ell(y)\varphi_0(y)\tau_{k\ell} \tag{45}$$

where $\tau_{k\ell} \equiv \sum_{j=1}^{M} (r_{k\ell}(\rho_j^B) - r_{k\ell}(-\rho_j^B))$, for $k, \ell = 1, 2, \ldots$. Note, $\tau_{k\ell} \geq 0$ because $\rho_j^B \geq 0$ implies $r_{k\ell}(\rho_j^B) \geq r_{k\ell}(-\rho_j^B)$, $\forall j, k, \ell$.

Define the following infinite dimensional vectors:

$\underset{\sim}{\psi} \equiv (\langle\psi, \xi_1\rangle, \langle\psi, \xi_2\rangle, \ldots)'$, $\underset{\sim}{g}_\gamma \equiv (\langle g_\gamma, \xi_1\rangle, \langle g_\gamma, \xi_2\rangle, \ldots)'$,
$\underset{\sim}{X} \equiv (\langle X/\sigma^2, \xi_1\rangle, \langle X/\sigma^2, \xi_2\rangle, \ldots)'$, and $\underset{\sim}{\xi}(x) \equiv (\xi_1(x), \xi_2(x), \ldots)'$.
Let T be the semi-infinite matrix with (k,ℓ) element $\tau_{k\ell}$, for $k, \ell = 1, 2, \ldots$.

Theorem B5. Let D_γ , D_γ^- , and D_γ^+ be the regions in R^+ where $\bar{\psi}_\gamma^*$, the solution to problem (33) of Lemma B3, is unconstrained, constrained below, and constrained above, respectively. Then $\bar{\psi}_\gamma^*$ is given by

$$\bar{\psi}_\gamma^*(x) = \begin{cases} \gamma & \text{for } x \in D_\gamma^+ \\ (-\lambda \underset{\sim}{X}/2 - 2\underset{\sim}{g}_\gamma)'(I+2T)^-\underset{\sim}{\xi}(x) & \text{for } x \in D_\gamma \\ -\gamma & \text{for } x \in D_\gamma^- , \end{cases} \tag{46}$$

where I is the identity matrix, $(I+2T)^-$ is some generalized inverse of $I+2T$, $\lambda = (2\underset{\sim}{g}_\gamma'(I+2T)^-\underset{\sim}{X} + \gamma[\Phi_0(D^+) - \Phi_0(D^-)] - 1)/\underset{\sim}{X}'(I+2T)^-\underset{\sim}{X}$ (where $\Phi_0(D^+)$ and $\Phi_0(D^-)$ are the Φ_0-probabilities of D_γ^+ and D_γ^- , respectively), and the equality in (46) is in the sense of mean square convergence. For $x \in R^-$, $\bar{\psi}_\gamma^*(x) = -\bar{\psi}_\gamma^*(-x)$.

Comments: 1. The proof of this result follows the classical treatment of linear integral equations (see Courant and Hilbert (1937, Chapter III).

2. By Lemma B3 the solution to problem (33) is (a.s.) unique. Thus the sets D_γ , D_γ^- , and D_γ^+ of Theorem B5 are (a.s.) unique. Also, given D_γ , D_γ^- , and D_γ^+ , the integral equation (38) has an (a.s.) unique solution, provided $I+2T$ is non-singular. The latter holds if 1 is not an eigenvalue of the symmetric kernel $K_M(x,y)$. This is shown using Mercer's expansion theorem (see Courant and Hilbert (1937,

p. 138), which guarantees an expansion of the form $\sum_{\ell=1}^{\infty} Z_\ell(x)Z_\ell(y)/\lambda_\ell$ where Z_ℓ, and λ_ℓ, $\ell = 1, 2, \ldots$, are the orthonormal eigenvectors and eigenvalues of $K_M(x,y)$, respectively. For the numerical calculations described below, singularity or near-singularity of $I + 2T$ never occurred.

Proof of Theorem B5. Substituting (45) in (38), multiplying (38) by $\xi_m(x)\varphi_0(x)$, and integrating (38) over D_γ, gives

$$\begin{aligned} \langle \psi_\gamma, \xi_m \rangle &= -\frac{\lambda}{2}\left\langle \frac{X}{\sigma^2}, \xi_m \right\rangle + \langle g_\gamma, \xi_m \rangle - 2\sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} \langle \xi_k, \xi_m \rangle \cdot \langle \psi_\gamma, \xi_\ell \rangle \cdot \tau_{k\ell} \\ &= -\frac{\lambda}{2}\left\langle \frac{X}{\sigma^2}, \xi_m \right\rangle + \langle g_\gamma, \xi_m \rangle - 2\sum_{\ell=1}^{\infty} \langle \psi_\gamma, \xi_\ell \rangle \cdot \tau_{\ell m} \end{aligned} \qquad (47)$$

where the exchange of integral and summation signs necessary for the first equality is justified by a Fourier series result (see, e.g., Andrews (1982a, Lemma 11)) since $\xi_\ell, \xi_k \in L^2[\varphi_0]$, and where the second equality follows by the orthonormality of ξ_ℓ, $\ell = 1, 2, \ldots$. Equation (47) for $m = 1, 2, \ldots$ can be written as

$$\underset{\sim}{\psi}_\gamma = (I + 2T)^{-}(-\lambda X/2 + \underset{\sim}{g}_\gamma) \; . \qquad (48)$$

The vector of Fourier coefficients $\underset{\sim}{\psi}_\gamma$ gives the solution to (38):

$$\psi_\gamma(x) = \sum_{\ell=1}^{\infty} \langle \psi_\gamma, \xi_\ell \rangle \xi_\ell(x) \equiv \underset{\sim}{\psi}_\gamma' \underset{\sim}{\xi}(x) = (-\lambda \underset{\sim}{X}/2 + \underset{\sim}{g}_\gamma)'(I + 2T)^{-}\underset{\sim}{\xi}(x) \; , \qquad (49)$$

where the first equality is in the sense of mean square convergence.

The constraint $E_{\Phi_0}\psi_\gamma(X)X/\sigma^2 = 1$ is satisfied when

$$1 = 2\int_{D_\gamma}\left(-\frac{\lambda}{2}\underset{\sim}{X}+\underset{\sim}{g}_\gamma\right)'(I+2T)^{-}\underset{\sim}{\xi}(x)\frac{x}{\sigma^2}\varphi_0(x)dx + \gamma\int_{D_\gamma^+}\varphi_0(x)dx - \gamma\int_{D_\gamma^-}\varphi_0(x)dx$$

$$= (-\lambda\underset{\sim}{X}+2\underset{\sim}{g}_\gamma)'(I+2T)^{-}\underset{\sim}{X} + \gamma\left(\Phi_0(D_\gamma^+) - \Phi_0(D_\gamma^-)\right) . \qquad (50)$$

This yields λ as defined above. □

If the $\overline{\psi}_\gamma^*$ function of Theorem B5 does not satisfy condition A3, then a smoothed version of it, call it ψ_γ^*, which does satisfy A3, can be used (see Lemma B3 and its proof). Minimization of (31) over γ then yields an ε-optimal solution to problem (3) for contamination neighbourhoods. As discussed above, this immediately gives the solution for variational neighbourhoods, and if the optimal contamination neighbourhood ψ function is monotone, for Kolmogorov neighbourhoods as well. The next section presents the results of carrying out these procedures numerically.

4. Numerical Results

The solution to problem (3) given by (31) and (46) has been calculated numerically for contamination neighbourhoods and quadratic loss. The results are given in Table 1 for a number of different boundary correlation vectors $\underset{\sim}{\rho}^B$. More specifically, first order autoregressive (AR(1)) and moving average (MA(1)) boundary correlation vectors are considered with several different correlation parameters. σ^2 is taken to be 1 (for $\sigma^2 \neq 1$, the ψ function just needs to be rescaled). The degenerate kernel method utilizing orthonormal expansions was used to solve the integral equation (38) numerically. (See Atkinson (1976) for a brief description of this method and references to the literature concerning it. Theoretical error bounds can be calculated for this method

but are often too large for practical use, and hence, are not reported here.) This method consists of truncating the expansion of the solution (46) given in Theorem B5 at a finite number of terms. For the degrees of accuracy given, it was necessary to calculate the first three terms of the expansion. Hence, the optimal ψ function for $x \in D_\gamma$ is given by a polynomial of degree three (without a constant term, since the solution to (38) at $x = 0$ is necessarily 0). The sets D_γ, D_γ^-, and D_γ^+ and the optimal truncation height γ were found by search procedures. D_γ, D_γ^-, and D_γ^+ are of the form $[0,b)$, the null set, and $[b,\infty)$, respectively, where b is given in the tables. The table only gives $\psi(x)$ values for $x \geq 0$, since ψ is odd. The table shows the optimal ψ function corresponding to $\theta = 0$; for $\theta \neq 0$, $\psi(\theta,x) = \psi(x-\theta)$, by assumption A4. Four values of t (the parameter which indexes the size of the neighbourhoods) are considered, viz., $t = 0.5$, 1.0, 1.5, and 2.0. For samples of size 100, these correspond to the "reasonable" contamination levels of 5, 10, 15 and 20%, respectively. Bickel (1981) provides additional arguments which suggest that this is a reasonable range of values for t. As discussed above, the contamination results for $t = t'$ correspond to the variational results for $t = t'/2$, thus the table actually gives solutions for both the contamination and variational neighbourhoods models. Further, since the optimal ψ functions given in Table 1 are monotone, the variational and Kolmogorov models have the same solution, and the results of the table apply to the Kolmogorov model as well (with $t = t'/2$).

Table 1 gives the optimal ψ functions and the corresponding minimum risk values (labelled R1 risk) for the different situations described above. As expected, γ (i.e., the value of $\psi(x)$ for $x \in [b,\infty)$)

TABLE 1. ψ functions for the optimal estimators (E1) in the contamination model with squared error loss.

The ψ functions are odd and on R^+ are of the form

$$\psi(x) = \begin{cases} \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 & \text{for } x \in [0,b) \\ \gamma & \text{for } x \in [b,\infty) . \end{cases}$$

For $\theta \neq 0$, $\psi(\theta,x) = \psi(x-\theta)$.

Boundary Correlation Vector	t	α_1	α_2	α_3	γ	b	R1 Risk
AR(1)ρ_1^B = .0	0.5	1.636	.000	.000	1.410	.862	1.636
	1.0	2.971	.000	.000	1.293	.435	2.964
	1.5	4.953	.000	.000	1.267	.256	5.000
	2.0	7.634	.000	.000	1.259	.165	7.787
AR(1)ρ_1^B = .1	0.5	1.632	.000	.000	1.411	.864	1.859
	1.0	2.971	.000	.001	1.293	.435	3.186
	1.5	4.953	-.002	.006	1.267	.256	5.223
	2.0	7.990	-.002	.007	1.259	.158	8.010
AR(1)ρ_1^B = .3	0.5	1.617	.000	.008	1.414	.871	2.499
	1.0	2.936	-.001	.011	1.294	.441	3.829
	1.5	4.953	-.006	.025	1.267	.256	5.867
	2.0	7.636	-.004	.023	1.259	.165	8.653
AR(1)ρ_1^B = .5	0.5	1.538	.002	.043	1.431	.908	3.666
	1.0	2.837	-.023	.098	1.297	.456	5.010
	1.5	4.631	-.009	.089	1.269	.274	7.050
	2.0	7.634	-.005	.088	1.259	.165	9.839
AR(1)ρ_1^B = .7	0.5	1.223	.416	-.198	1.490	1.030	6.404
	1.0	2.469	.008	.305	1.310	.513	7.809
	1.5	4.345	.116	.022	1.271	.290	9.870
	2.0	7.038	-.003	.396	1.260	.179	12.664
MA(1)ρ_1^B = .1	0.5	1.636	.000	.000	1.410	.862	1.837
	1.0	2.971	.000	.001	1.293	.435	3.164
	1.5	4.953	-.002	.006	1.267	.256	5.201
	2.0	8.402	-.002	.008	1.258	.150	7.987
MA(1)ρ_1^B = .3	0.5	1.617	.000	.007	1.414	.871	2.242
	1.0	2.903	-.001	.011	1.295	.446	3.572
	1.5	4.953	-.005	.022	1.267	.258	5.609
	2.0	7.636	-.003	.021	1.259	.165	8.394
MA(1)ρ_1^B = .499,99	0.5	1.442	.307	-.171	1.427	.903	2.662
	1.0	2.865	-.001	.058	1.296	.451	4.005
	1.5	4.951	-.006	.077	1.267	.256	6.045
	2.0	7.635	-.004	.077	1.259	.165	8.833

and b decrease in t (for any given $\underset{\sim}{\rho}^B$). That is, as the size of the neighbourhoods increases the truncation height decreases. For any given t, γ and b increase as the "amount" of dependence (as measured by the value of ρ_1^B) increases. Increased dependence inflates the asymptotic variance but not the asymptotic maximum bias. Hence, the tradeoff between the two shifts in favor of increased weight given to the asymptotic variance as ρ_1^B increases. However, it is worth noting that the sensitivity of the optimal truncation height to the value of ρ_1^B is slight for ρ_1^B less than .7. A very large value of ρ_1^B is needed to significantly increase the truncation height relative to the $\rho_1^B = 0$ optimal height.

Most of the optimal ψ functions have slopes which are slightly increasing on $[0,b]$, or which are decreasing very near zero and increasing from there to b. This contrasts with the constant slope on $[0,b]$ of the optimal, <u>independent</u>, robust ψ function. The intuition to explain this pattern of change in slope is not clear. However, the change in slope is <u>very</u> slight for all cases except those with quite large ρ_1^B value, and even in those cases it is small. Generally speaking, the optimal ψ functions are remarkably close to the Huber ψ functions (see Huber (1964)) which are continuous, symmetric, linearly increasing in the middle of R, and constant in the tails. This is not surprising, especially for small ρ_1^B values, since Huber ψ functions are optimal in the above robust model with independent data (i.e., the model with $\Omega^B = I$; see Bickel (1981)).

A number of interesting questions can be asked concerning the relative performances of:

TABLE 2. ψ functions for optimal Huber estimators E3 in the contamination model with squared error loss.

The ψ functions are odd, and on R^+ are of the form

$$\psi(x) = \begin{cases} \alpha x & \text{for } x \in [0,b) \\ \tilde{\gamma} & \text{for } x \in [b,\infty) \end{cases}$$

For $\theta \neq 0$, $\psi(\theta,x) = \psi(x-\theta)$.

Boundary Correlation Vector	t	α	$\tilde{\gamma}$	b
AR(1)ρ_1^B = .0	0.5	1.636	1.410	.862
	1.0	2.971	1.293	.435
	1.5	4.953	1.267	.256
	2.0	7.634	1.259	.165
AR(1)ρ_1^B = .1	0.5	1.636	1.410	.862
	1.0	2.971	1.293	.435
	1.5	4.953	1.267	.256
	2.0	7.638	1.259	.165
AR(1)ρ_1^B = .3	0.5	1.624	1.413	.870
	1.0	2.937	1.294	.441
	1.5	4.953	1.267	.256
	2.0	7.638	1.259	.165
AR(1)ρ_1^B = .5	0.5	1.563	1.430	.915
	1.0	2.841	1.297	.457
	1.5	4.785	1.268	.265
	2.0	7.638	1.259	.165
AR(1)ρ_1^B = .7	0.5	1.409	1.490	1.057
	1.0	2.520	1.310	.520
	1.5	4.496	1.270	.282
	2.0	7.048	1.260	.179
AR(1)ρ_1^B = .9	0.5	1.161	1.719	1.481
	1.0	1.535	1.439	.938
	1.5	2.374	1.318	.555
	2.0	4.150	1.273	.307
AR(1)ρ_1^B = .95	0.5	1.093	1.883	1.723
	1.0	1.287	1.568	1.218
	1.5	1.636	1.410	.862
	2.0	2.141	1.335	.624
MA(1)ρ_1^B = .1	0.5	1.636	1.410	.862
	1.0	2.971	1.293	.435
	1.5	4.953	1.267	.256
	2.0	7.638	1.259	.165
MA(1)ρ_1^B = .3	0.5	1.624	1.413	.870
	1.0	2.971	1.293	.435
	1.5	4.953	1.267	.256
	2.0	7.638	1.259	.165
MA(1)ρ_1^B = .499,99	0.5	1.580	1.425	.902
	1.0	2.872	1.296	.451
	1.5	4.953	1.267	.256
	2.0	7.638	1.259	.165

The estimators can be compared for different underlying models. For each such model, a comparison of two estimators can be made by looking at their ratio of risks. The risk function of the underlying models considered are:

(R1) the asymptotic maximum risk over neighbourhoods of dependent processes (i.e., the value of (3)) for given $\underset{\sim}{\rho}^B$ and $t > 0$;

(R2) the asymptotic maximum risk over neighbourhoods of <u>independent</u> processes (i.e., the value of (3) for $\Omega^B = I$ and given $t > 0$);

(R3) the asymptotic risk (equivalently variance) under parametric Gaussian processes with a given boundary correlation vector $\underset{\sim}{\rho}^B$ and <u>no</u> neighbourhood distributions (i.e., the value of (3) for given $\underset{\sim}{\rho}^B$ and <u>$t = 0$</u>); and

(R4) the asymptotic risk (equivalently variance) under <u>independent</u> parametric Gaussian processes with <u>no</u> neighbourhood distributions (i.e., the value of (3) for $\Omega^B = I$ and <u>$t = 0$</u>).

As in Table 1, loss is taken to be squared error, and $\sigma^2 = 1$.

Ratios of R1 risk for the estimators E1 and E3 have been calculated for the cases considered in Table 1. In each case the optimal estimator E1 has smaller R1 risk, but the difference in risks is less than .1%. Hence, the class of Huber ψ functions yields approximately optimal ψ functions for the problem defined in (3) which are virtually as good as the strictly optimal ψ functions. As mentioned above, it is much easier to calculate the optimal truncation point $\tilde{\gamma}$ for the Huber ψ function than it is to calculate the strictly optimal ψ function using an expansion in terms of orthonormal polynomials. Thus, Huber ψ functions

(with truncation points depending on $\underset{\sim}{\rho}^B$ and t) are recommended for use in the dependent robust models developed above.

Column 3 of Table 3 assesses the performance of the optimal robust estimator for dependent processes E1 with respect to the optimal robust estimator for independent processes E2 on neighbourhoods of independent processes (i.e., using the ratio of their R2 risks). Note, when R2 risk is considered, different $\underset{\sim}{\rho}^B$ values in the table do not reflect different underlying neighbourhoods of distributions; they reflect different estimators E1. The table shows that the E1 estimators do remarkably well, especially if $\rho_1^B \leq .7$ for AR(1) boundary correlation vectors. (If $\rho_1^B \leq .7$, the ratio of R2 risks is greater than .99.) Hence, the specification of a large set of covariance matrices S still yields an estimator which is quite good when the data are actually independent. Virtually identical results hold for the performance of the optimal Huber estimators E3 with respect to the estimator E2 using R2 risk values.

Table 3 also considers the obverse situation (see column 4) and assesses the performance of the optimal robust estimator for independent processes E2 with respect to the optimal robust estimator for dependent processes E1 on neighbourhoods of dependent processes (i.e., using the ratio of their R1 risks) for different boundary correlation vectors $\underset{\sim}{\rho}^B$ and different values of t. Column 4 of the table gives the comforting result that the optimal robust estimator for independent processes performs extremely well on dependent data, and hence, is quite robust against positive dependence. (N.B.: Columns 3 and 4 of Table 3 exhibit an asymmetry when very large amounts of dependence are considered, i.e., when $\rho_1^B \geq .9$ for an AR(1) boundary correlation vector. In particular, the optimal robust estimator for independent processes does better (relatively

speaking) on dependent distribution neighbourhoods than the optimal Huber estimator for dependent processes does on independent distribution neighbourhoods. While this asymmetry also exists for $\rho_1^B < .9$, it is not of significant magnitude.)

Suppose the data actually are generated by the parametric distribution, i.e., a stationary Gaussian distribution with boundary correlation vector $\underset{\sim}{\rho}^B$. One might ask: How much is lost by using the optimal robust estimator for dependent processes E1 rather than the asymptotically efficient estimator for Gaussian processes E4, i.e., $\bar{x}_n$? Column 5 of Table 3 answers this question by exhibiting the ratio of the R3 risk of E4 to that of E1. For given $\underset{\sim}{\rho}^B$, the relative efficiency of E1 increases as t (the size of the neighbourhoods) decreases. For given t, the relative efficiency of E1 increases as ρ_1^B increases for MA(1) and AR(1) boundary correlation vectors. This follows because 1) the truncation height of ψ increases as ρ_1^B increases, and the higher the truncation height, the closer is the estimator E1 to $\bar{x}_n$, and 2) even for the same ψ function, the ratio of R3 risk of an M-estimator with given ψ function to that of $\bar{x}_n$ increases as ρ_1^B increases. (This second assertion is proved using the argument of Gastwirth and Rubin (1975, Proposition 4.2, Lemma 4.2, and Corollary 4.1).) For the case where $t = 1.0$, the loss declines from 19% for an AR(1) least-favorable correlation vector with $\rho_1^B = .1$ to 2% for an AR(1) vector with $\rho_1^B = .95$.

The same question as above can be posed for the case where the data is generated by an <u>independent</u> Gaussian distribution. Column 6 of Table 3 gives the relevant risk ratios, viz., the ratio of R4 risk for the estimator E4 to that of estimator E1. (In this case, for different $\underset{\sim}{\rho}^B$ in the table, the underlying distributions used for risk calculations

TABLE 3. Relative Risks of Estimators

Boundary Correlation Vector	t	R2 risk of E2* / R2 risk of E1	R1 risk of E1 / R1 risk of E2	R3 risk of E4 / R3 risk of E1	R4 risk of E4* / R4 risk of E1
AR(1)ρ_1^B = .0	0.5	1.000	1.000	.878	.878
	1.0	1.000	1.000	.774	.774
	1.5	1.000	1.000	.720	.720
	2.0	1.000	1.000	.691	.691
AR(1)ρ_1^B = .1	0.5	1.000	1.000	.898	.878
	1.0	1.000	1.000	.807	.774
	1.5	1.000	1.000	.759	.720
	2.0	1.000	1.000	.730	.689
AR(1)ρ_1^B = .3	0.5	1.000	1.000	.929	.880
	1.0	1.000	1.000	.862	.776
	1.5	1.000	1.000	.824	.720
	2.0	1.000	1.000	.803	.691
AR(1)ρ_1^B = .5	0.5	.999	1.000	.951	.888
	1.0	1.000	1.000	.902	.780
	1.5	1.000	1.000	.875	.726
	2.0	1.000	1.000	.857	.691
AR(1)ρ_1^B = .7	0.5	.991	.998	.969	.912
	1.0	.998	.999	.930	.797
	1.5	1.000	1.000	.909	.731
	2.0	1.000	1.000	.897	.696
MA(1)ρ_1^B = .1	0.5	1.000	1.000	.896	.878
	1.0	1.000	1.000	.804	.774
	1.5	1.000	1.000	.755	.720
	2.0	1.000	1.000	.724	.686
MA(1)ρ_1^B = .3	0.5	1.000	1.000	.918	.880
	1.0	1.000	1.000	.844	.777
	1.5	1.000	1.000	.801	.720
	2.0	1.000	1.000	.779	.692
MA(1)ρ_1^B = .499,99	0.5	1.000	1.000	.929	.886
	1.0	1.000	1.000	.860	.779
	1.5	1.000	1.000	.822	.720
	2.0	1.000	1.000	.802	.691

(continued)

TABLE 3 (continued)

Boundary Correlation Vector	t	R2 risk of E2 / R2 risk of E3	R1 risk of E3 / R1 risk of E2	R3 risk of E4** / R3 risk of E3	R4 risk of E4** / R4 risk of E3
AR(1)$\rho_1^B = .9$	0.5	.921	.985	.988	.959
	1.0	.929	.988	.966	.885
	1.5	.971	.995	.942	.806
	2.0	.993	.999	.926	.735
AR(1)$\rho_1^B = .95$	0.5	.858	.978	.993	.979
	1.0	.840	.974	.979	.935
	1.5	.891	.984	.961	.878
	2.0	.933	.993	.947	.824

N.B.: Estimators E1 and E3 are so similar that the above table is virtually the same with E1 replaced by E3.

*When R2 or R4 risk are considered, the underlying neighbourhoods of distributions are the same for all $\underset{\sim}{\rho}^B$ cases considered; only the estimators E1, E2, and E3 varies with $\underset{\sim}{\rho}^B$.

**These columns give the same risk ratios as the columns above except that the estimator in the denominator is the optimal Huber estimator E3 rather than the strictly optimal robust estimator E1. This change in the table is made because the estimator E1 has not been calculated for the AR(1), $\rho_1^B = .90, .95$ boundary correlation vectors.

remain unchanged, but the estimators E1 vary with $\underset{\sim}{\rho}^B$.) Again, the relative performance of E1 improves as t decreases or as ρ_1^B increases. And as expected, E1 does better for the case of dependent Gaussian data than for independent Gaussian data. (This is a general result which can be proved using the argument of Gastwirth and Rubin (1975, Lemma 4.1 and Theorem 4.1).)

The obverse of the last two cases considered is the assessment of the performances of $\bar{x}_n$ compared to those of E1 or E2 when the distribution generating the data is arbitrary within the neighbourhoods of the dependent or the independent parametric distributions. These are given by the ratios of R1 and R2 risks of the estimators E1 and E4. For any given $\underset{\sim}{\rho}^B$, these ratios are zero. That is, $\bar{x}_n$ performs so poorly for some distributions in the neighbourhoods that its R1 and R2 risks are infinite.

We summarize briefly as follows: For neighbourhoods of contamination, variational, and Kolmogorov type, the optimal robust estimators for the dependent model are found to have ψ functions very close in shape to Huber ψ functions. In fact, for all practical purposes, a Huber ψ function with truncation point adjusted depending on t and $\underset{\sim}{\rho}^B$ is as good as the optimal ψ function. Further, for AR(1) and MA(1) boundary correlation vectors with $\rho_1^B \leq .7$, the truncation point is not very sensitive to the value of ρ_1^B . The optimal robust estimators for the dependent model are robust against lack of dependence, and the optimal robust estimator for the independent model is robust against positive dependence. When the data are truly Gaussian, the optimal robust estimator for the dependent model is less efficient than $\bar{x}_n$, but decreasingly so as the size (t) of the neighbourhoods decreases, and as the amount of dependence (as measured by ρ_1^B) increases.

On the other hand, $\bar{x}_n$ performs disastrously for some distributions in the neighbourhoods, and hence, has infinite maximum risk over the neighbourhoods. The specifications of t and $\underset{\sim}{\rho}^B$ determine how closely the optimal robust estimator for dependent processes resembles $\bar{x}_n$, and how closely it resembles the sample median $\tilde{x}_n$ (which has the minimum possible asymptotic bias value for c, V and K neighbourhoods, and is optimal for these neighbourhoods as $t \to \infty$).

5. Appendix

Proof of Theorem B1. The proof of Theorem 1 of Andrews (1984) is sufficient except that equations (20) and (13) of the proofs of Lemma 3 and Theorem 1, respectively, need to be proven for $i = K$. That is, it suffices to show (a) $E_{1n} \equiv \sup_{F_{\theta n} \in \mathcal{F}^K_{t/\sqrt{n}}(\theta)} |\lambda(\psi,s,F_{\theta n}) - \lambda(\psi,s,\Phi_\theta)| \xrightarrow{n\to\infty} 0$, uniformly for all θ, and uniformly for $s \in C$, any compact set in Θ, and (b) $G_{1n} \equiv \int [\psi(T_n,x) - \psi(\theta,x)] \frac{1}{\sqrt{n}} \sum_{j=1}^{n} d_{nj} \left(dF^j_{\theta n}(x) - d\Phi_\theta(x) \right) = o_{F_{\theta\Omega n}}(1)$ $\text{unif}(\theta,F,\Omega)_K$.

To prove (a), assumption A2' gives

$$0 \le E_{1n} = \sup_{F_{\theta n} \in \mathcal{F}^K_{t/\sqrt{n}}(\theta)} \left| \int\int_{R^2} 1_{(-\infty,x]}(y)\, d\nu_s(y)\, d\left(F_{\theta n}(x) - \Phi_\theta(x) \right) \right|$$

$$= \sup_{F_{\theta n} \in \mathcal{F}^K_{t/\sqrt{n}}(\theta)} \left| \int_R [\Phi_\theta(y) - F_{\theta n}(y)]\, d\nu_s(y) \right|$$

$$\le \int d|\nu_s|(y) \cdot t/\sqrt{n} \le C_1(s) \cdot t/\sqrt{n} \xrightarrow{n\to\infty} 0 , \qquad (51)$$

uniformly for all θ, and uniformly for $s \in C$, where the second equality holds by Fubini's Theorem, the second inequality by the definition

of $\mathcal{F}^K_{t/\sqrt{n}}(\theta)$, and the third inequality by A2'.

To show (b), A2' and Fubini's Theorem (used as above) yield

$$0 \le |G_{1n}| = \left| \int_R \frac{1}{\sqrt{n}} \sum_{j=1}^{n} d_{nj} \left(\Phi_\theta(y) - F^j_{\theta n}(y) \right) d\left(\nu_{T_n}(y) - \nu_\theta(y) \right) \right|$$

$$\le tB_d \int_R d|\nu_{T_n}(y) - \nu_\theta(y)| \le tB_d C_2(\theta)|T_n - \theta| = o_{F_{\theta\Omega n}}(1) \quad \text{unif}(\theta,F,\Omega)_K \ , \quad (52)$$

where the second inequality uses D1 and the definition of $\mathcal{F}^K_{t/\sqrt{n}}(\theta)$, the third inequality uses A3', and the last equality follows from Lemma 3. □

Proof of Lemma B1. The proof follows that of Bickel (1981). Let $G_{2n} = \frac{1}{n}\sum_{i=1}^{n} d_{nj}\lambda(\psi,\theta,F^j_{\theta n})$, and for any function $g(x)$, let $g(x)^+ = g(x) \vee 0$ and $g(x)^- = (-g(x)) \vee 0$. For all sequences $\langle F_{\theta\Omega n}\rangle_{n\ge 1}$ in the variational neighborhoods.

$$|\lambda(\psi,\theta,F^j_{\theta n})| = \left| \int \psi(\theta,x)(f^j_{\theta n}(x) - \varphi_\theta(x))^+ d\mu(x) - \int \psi(\theta,x)(f^j_{\theta n}(x) - \varphi_\theta(x))^- d\mu(x) \right|$$

$$\le (\operatorname{ess\,sup}_{x\in R} \psi(\theta,x) - \operatorname{ess\,inf}_{x\in R} \psi(\theta,x)) \cdot \int |f^j_{\theta n}(x) - \varphi_\theta(x)| d\mu(x)/2$$

$$< (\text{ess sup } \psi - \text{ess inf } \psi) \cdot t/\sqrt{n} \ , \quad (53)$$

and so,

$$\left| \overline{\lim_{n\to\infty}} \sqrt{n}\, G_{2n} \right| \le \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{j=1}^{n} |d_{nj}| \cdot t \cdot (\text{ess sup } \psi - \text{ess inf } \psi) = b_V(\psi,d,\theta) \ . \quad (54)$$

For sequences $\langle F_{\theta\Omega n}\rangle_{n\ge 1}$ in the Kolmogorov neighbourhoods, A2' and Fubini's Theorem give (as in the proof of Theorem B1)

$$|\lambda(\psi,\theta,F^j_{\theta n})| = \left| \int (\Phi_\theta(y) - F^j_{\theta n}(y)) d\nu_\theta(y) \right| \le \|\nu_\theta\|_V \cdot t/\sqrt{n} \ , \quad (55)$$

and so,

$$\left|\overline{\lim_{n\to\infty}}\ \sqrt{n}\ G_{2n}\right| \le \overline{\lim_{n\to\infty}}\ \frac{1}{n}\sum_{j=1}^{n}|d_{nj}|\cdot\|v_\theta\|_V\cdot t = b_K(\psi,d,\theta)\ . \tag{56}$$

For sequences $\langle F_{\theta\Omega n}\rangle_{n\ge 1}$ in the contamination neighbourhoods, each df $F^j_{\theta n}(x)$ can be written as $F^j_{\theta n}(x) = (1 - t/\sqrt{n})\Phi_\theta(x) + Q_j(x)t/\sqrt{n}$, for some df Q_j. Thus,

$$\begin{aligned}&\lambda(\psi,\theta,F^j_{\theta n}) = \int\psi(\theta,x)\,dQ_j(x)\cdot t/\sqrt{n} \le \operatorname*{ess\,sup}_{x\in R}\ \psi(\theta,x)\cdot t/\sqrt{n}\ , \text{ and}\\ &\lambda(\psi,\theta,F^j_{\theta n}) \ge \operatorname*{ess\,inf}_{x\in R}\ \psi(\theta,x)\cdot t/\sqrt{n}\ .\end{aligned} \tag{57}$$

Let $d^+_{nj} = d_{nj} \vee 0$ and $d^-_{nj} = (-d_{nj}) \vee 0$. Using D2, $\frac{1}{n}\sum_{j=1}^{n} d^+_{nj} \xrightarrow{n\to\infty} (\tilde{n}+1)/2 \equiv \eta_1$, and $\frac{1}{n}\sum_{j=1}^{n} d^-_{nj} \xrightarrow{n\to\infty} (\tilde{n}-1)/2 \equiv \eta_2$. So,

$$\begin{aligned}\left|\overline{\lim_{n\to\infty}}\ \sqrt{n}\ G_{2n}\right| &\le \left(\overline{\lim_{n\to\infty}}\ \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(d^+_{nj}\ \operatorname*{ess\,sup}\psi - d^-_{nj}\ \operatorname*{ess\,inf}\psi)t/\sqrt{n}\right)\\ &\qquad \vee \left(\overline{\lim_{n\to\infty}}\ \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(d^-_{nj}\ \operatorname*{ess\,sup}\psi - d^+_{nj}\ \operatorname*{ess\,inf}\psi)t/\sqrt{n}\right)\\ &= t\cdot[(\eta_1\ \operatorname*{ess\,sup}\psi - \eta_2\ \operatorname*{ess\,inf}\psi) \vee (\eta_2\ \operatorname*{ess\,sup}\psi - \eta_1\ \operatorname*{ess\,inf}\psi)\\ &= b_c(\psi,d,\theta)\ .\end{aligned} \tag{58}$$

Next, for $i = V$ and K, consider sequences $\langle F^{(h)}_{\theta\Omega n}\rangle_{n\ge 1}$ of distributions which are in $\{\mathcal{F}^i_{t/\sqrt{n}}(\theta),\ \Omega\}_n$ for all n sufficiently large and all $\Omega \in S$, whose univariate marginal densities with respect to Lebesgue measure are given by

$$f^{(h)j}_{\theta n}(x) = \varphi_\theta(x)\cdot\exp[\operatorname{sgn}(d_{nj})\cdot h(x)/\sqrt{n} - c_n] \quad \text{for} \quad j = 1, 2, \ldots ,$$

where c_n is a constant defined to make this a density, $h \in \mathcal{H}_i(\theta)$,

$$\mathcal{H}_V(\theta) \equiv \{h \in L^\infty : \int h(x)\varphi_\theta(x)dx = 0 \text{ and } \int |h(x)|\varphi_\theta(x)dx < 2t\} , \text{ and}$$

$$\mathcal{H}_K(\theta) \equiv \{h \in L^\infty : \int h(x)\varphi_\theta(x)dx = 0 \text{ and } \sup_{y\in R}\left|\int_{-\infty}^{y} h(x)\varphi_\theta(x)dx\right| < t\} .$$

Such distributions exist since they can be created by point transformations of a sequence of rv's with distribution $\Phi_{\theta\Omega}$ (see Andrews (1982a, b)). Bickel (1981, Lemma 2) shows that for n sufficiently large the univariate distributions $F_{\theta n}^{(h)j}$ are in $\mathcal{F}^i_{t/\sqrt{n}}(\theta)$ for all $h \in \mathcal{H}_i(\theta)$, for $i = V$ and K. This implies $F_{\theta\Omega n}^{(h)} \in \{\mathcal{F}^i_{t/\sqrt{n}}(\theta), \Omega\}_n$, for n sufficiently large, for all $h \in \mathcal{H}_i(\theta)$, for $i = V$ and K. Also, by Bickel (1981, Theorem 2),

$$\sup_{h\in\mathcal{H}_i(\theta)} \int \psi(\theta,x)h(x)\varphi_\theta(x)dx = \begin{cases} t\cdot(\text{ess sup } \psi - \text{ess inf } \psi) & \text{for } i = V \\ t\cdot\|\nu_\theta\|_V & \text{for } i = K . \end{cases} \tag{59}$$

Now, $\exp[c_n] = 1 + O(1/n)$ as $n \to \infty$, and so,

$\exp[\text{sgn}(d_{nj})\cdot h(x)/\sqrt{n} - c_n] = 1 + \text{sgn}(d_{nj})\cdot h(x)/\sqrt{n} + O(1/n)$ as $n \to \infty$,

where $O(1/n)$ holds uniformly for $x \in R$, since h is bounded. Hence,

$$\lim_{n\to\infty}\left|\sqrt{n}\ \lambda(\psi,\theta,F_{\theta n}^{(h)j}) - \text{sgn}(d_{nj})\cdot\int\psi(\theta,x)h(x)\varphi_\theta(x)dx\right| = 0 , \text{ and}$$

$$\sup_{<F_{\theta\Omega n}>_{n\geq 1}} \underline{\lim}_{n\to\infty}\left|\sqrt{n}\ G_{2n}\right| \geq \sup_{<F^{(h)}_{\theta\Omega n}>_{n\geq 1}:h\in\mathcal{H}_i(\theta)} \underline{\lim}_{n\to\infty}\left|\sqrt{n}\ G_{2n}\right|$$

$$= \sup_{h\in\mathcal{H}_i(\theta)} \underline{\lim}_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} |d_{nj}|\cdot\int\psi(\theta,x)h(x)\varphi_\theta(x)dx$$

$$= b_i(\psi,d,\theta) , \tag{60}$$

using (59), for $i = V$ and K.

Contamination neighbourhoods contain sequences of distributions $\langle F^{Q}_{\theta\Omega n}\rangle_{n\geq 1}$, which are created by point transformations of rv's with distribution $\Phi_{\theta\Omega}$ (see Andrews (1982a, b)), and whose univariate marginal df's are given by $(1 - t/\sqrt{n})\Phi_{\theta} + Q_j t/\sqrt{n}$ for <u>arbitrary</u> df's Q_j, $j = 1, 2, \ldots$. Thus, using (57) and (58),

$$\sup_{\langle F_{\theta\Omega n}\rangle_{n\geq 1}} \overline{\lim_{n\to\infty}} \left|\sqrt{n}\, G_{2n}\right| \geq \sup_{\langle F^{Q}_{\theta\Omega n}\rangle_{n\geq 1}} \overline{\lim_{n\to\infty}} \frac{1}{n}\sum_{j=1}^{n} d_{nj} \left| \int \psi(\theta,x)\, dQ_j(x) \cdot t \right| = b_c(\psi,d,\theta) \quad (61)$$

Since the inequalities of (54) and (60), (56) and (60), and (58) and (61) hold in opposite directions, they hold as equalities. □

Acknowledgments

I am greatly indebted to Peter J. Bickel and Thomas J. Rothenberg for their help and encouragement.

FOOTNOTE

1. AMS 1980 subject classifications. Primary 62F35, 62E20; Secondary 62F10. Key words and phrases. Robust estimation for dependent random variables; contamination, variational metric, and Kolmogorov metric neighbourhoods; shrinking neighbourhoods; location model.

REFERENCES

Akhiezer, N. I. (1962). The Calculus of Variations. Translated by A. H. Frink. Blaisdell Publishing Company, New York.

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

Andrews, D. W. K. (1982b). "Robust and asymptotically efficient estimation of location in a stationary strong mixing Gaussian parametric model," Cowles Foundation Discussion Paper No. 659, Yale University, New Haven, CT.

Andrews, D. W. K. (1983). "Robust and efficient estimation of nonlinear regression models with dependent errors," manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K. (1984). "Robust and efficient estimation of location in a Gaussian parametric model: I," manuscript, Cowles Foundation, Yale University.

Atkinson, K. E. (1976). A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind. S.I.A.M., Philadelphia.

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, University of California, Berkeley.

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. (1982). "Robust regression based on infinitesimal neighborhoods," Technical Report No. 16, Dept. of Statistics, University of California, Berkeley.

Bickel, P. J. (1983). "Comment on minimax aspects of bounded-influence regression," J.A.S.A. 78, 75-77.

Courant, R. and D. Hilbert (1937). Methods of Mathematical Physics, Vol. I, first ed. Interscience Publishing Inc., New York.

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

Gelfand, I. M. and S. V. Fomin (1963). Calculus of Variations. McGraw-Hill, New York.

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.

Handschin, E., J. Kohlas, A. Fiechter and F. Schweppe (1975). "Bad data analysis for power system state apparatus," IEEE Transactions on Power Apparatus and Systems PAS-94, Vol. 2, 329-337.

Holmes, R. M. (1981). Contributions to the theory of parametric estimation in randomly censored data. Ph.D. Thesis, University of California, Berkeley.

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

Huber, P. J. (1983). "Minimax aspects of bounded influence regression," J.A.S.A. 78, 66-72.

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

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

Krasker, W. and R. Welsch (1982). "Efficient bounded influence regression estimation," J.A.S.A. 77, 595-604.

LeCam, L. (1960). "Locally asymptotically normal families of distributions," U. of California Publications in Statistics 3:2, 37-98.

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.

Rousseeuw, P. J. (1981). New Infinitesimal Methods in Robust Statistics. Ph.D. Thesis, Free University of Brussels.

Strassen, V. (1965). "The existence of probability measures with given marginals," Ann. Math. Statist. 36, 423-429.

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.