Technical Report 88.21

ON THE CHOICE OF SUPPORT OF RE-DESCENDING ψ-FUNCTIONS IN LINEAR MODELS WITH ASYMMETRIC ERROR DISTRIBUTIONS

by

M. Hlynka, J.N. Sheahan, and D.P. Wiens

ON THE CHOICE OF SUPPORT OF RE-DESCENDING ψ-FUNCTIONS IN LINEAR MODELS WITH ASYMMETRIC ERROR DISTRIBUTIONS

[ABBREVIATED TITLE: LINEAR MODELS WITH ASYMMETRIC ERRORS]

M.Hlynka,* J.N. Sheahan,† and D.P. Wiens**

AMS 1980 subject classifications: Primary 62F35; secondary 62G05, 62J05

Key words and phrases: Linear models, asymmetric errors, robust estimation, re-descending influence functions

ABSTRACT

In M-estimation of the regression parameter vector in the linear model, we discuss the choice of the support of certain re-descending ψ-functions for both cases when the distribution of the i.i.d. errors is partially known and when it is completely functionally unknown.

* Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario N9B 3P4

† ** Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta T6G 2G1

** †* Research supported in part by the Natural Sciences and Engineering Research Council of Canada

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent random variables with X_i having distribution function

$$P(X_i \le x) = F\left(\frac{x - \sum_{j=1}^{p} c_{ij}\theta_j}{\sigma}\right), i = 1, \ldots, n, \tag{1.1}$$

where the c_{ij} are known constants, the θ_j are unknown parameters to be estimated from the observed values $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$, and σ is an unknown scale parameter.

If we define $U_i = X_i - \sum_{j=1}^{p} c_{ij}\theta_j, i = 1, \ldots, n$, we can write

$$\left.\begin{array}{l} X^{(n)} = C^{(n)}\theta + U \\ \text{where } X^{(n)} = (X_1, \ldots, X_n)' \ (' \text{ denotes transpose}), \\ C^{(n)} = ((c_{ij}^{(n)})) = (c_1^{(n)}, \ldots, c_p^{(n)}) \text{ is an } n \times p \text{ design matrix} \\ \text{of known constants,} \quad \theta = (\theta_1, \ldots, \theta_p)' \text{ is a} \\ \text{vector of unknown regression parameters, and} \\ U = (U_1, \ldots, U_n)' \text{ is a vector of independent and} \\ \text{identically distributed (i.i.d.) random errors with} \\ \text{distribution function } F(u/\sigma). \end{array}\right\} \tag{1.2}$$

When F is an unknown member of a class $\mathcal{F}$ of distributions, one method of estimating θ in (1.2) is to solve a system of p equations of the form

$$\sum_{i=1}^{n} c_i \psi(u_i) = 0 \tag{1.3}$$

or, to ensure that the estimator is scale equivariant,

$$\sum_{i=1}^{n} c_i \psi(u_i/\widehat{\sigma}) = 0 \tag{1.4}$$

where $\widehat{\sigma}$ is a scale equivariant estimator of σ. (One could also consider simultaneous estimation of θ and σ, as in Sheahan 1988).

Examples of functions $\psi : \mathbf{R}^1 \to \mathbf{R}^1$ commonly used in (1.3) are:

$\psi(y) = y$, in which case the solution of (1.3) is the least squares estimator, which is optimal (-its components are efficient) if F is normal, but is not robust if F has heavy tails;

$\psi(y) = -f'(y)/f(y)$, if F has a known density f, giving the maximum likelihood estimator;

$\psi(y) = |y|$, resulting in the least absolute deviation estimator whose asymptotic theoretical properties have been examined by e.g. Koenker and Bassett (1978), and which was one of the estimators whose finite-sample performance was investigated by simulation in Sheahan, Lind and Mehra (1989);

and

$\psi(y) = -\psi(-y) = \max(-k, \min(y, k))$, for fixed k, leading to a "Huber" estimator (Huber (1973); see also Huber (1964, 1981)).

In this article, we consider using a "re-descending" ψ-function of the form

$$\psi(y) = \begin{cases} y, & |y| \leq y_0 \\ y_1 \tanh[\frac{1}{2} y_1 (a_0 - \omega - |y|)] sgn(y), & y_0 \leq |y| \leq a_0 - \omega \\ 0, & |y| \geq a_0 - \omega \end{cases} \tag{1.5}$$

where y_0, y_1, ω and a_0 are constants to be specified.

More generally, we investigate a ψ -function of the form

$$\psi(y) = \begin{cases} y, & |y| \leq y_o \\ y_0 \xi(\frac{a_0 - \omega - |y|}{a_0 - \omega - y_0}) sgn(y), & y_0 \leq |y| \leq a_0 - \omega \\ 0, & |y| \leq a_0 - \omega \end{cases} \tag{1.6}$$

where $\xi : [0,1] \to [0,1]$ is any fixed continuously differentiable and strictly increasing function. Note that if we wish ψ to be continuous, ξ will satisfy $\xi(0) = 0$ and $\xi(1) = 1$.

The ψ of (1.5) was discovered independently by each of Collins, Hampel and Huber (see Collins (1976), Hampel (1973) and Huber (1981, sec. 4.8)). See Collins and Wiens (1985) for generalizations. Relying heavily on the results of Collins (1976) and Collins, Sheahan and Zheng (1986), Sheahan, Lind and Mehra (1989) showed that, under regularity conditions, (1.5) is, according to a certain asymptotic minimax criterion, the optimal ψ-function to use in (1.3), if in (1.2) σ is known and F is known to belong to a class $\mathcal{F}$ of distributions given below:

2. ESTIMATION OF a_0 IN (1.7).

In this section, we assume that in the linear model (1.2), F is known to belong to the class $\mathcal{F}$ of (1.7) and that ϵ and σ are known. Now if a_0 were also known, the optimal M-estimator of θ would be the solution of (1.3) with ψ given by (1.5), y_0 and y_1 given in (1.8) and ω as near zero as desired - see Sheahan, Lind and Mehra (1989). When a_0 is unknown an asymptotically optimal procedure is as follows: Find a consistent estimator $\widehat{a}_0$ of a_0; then solve (1.3) with $\widehat{a}_0$ replacing the unknown a_0 in (1.5). The aim of this section is to present some heuristic methods for estimating a_0. The resulting estimators will each lead to shift but not scale equivariant estimators of θ. For scale equivariance one can solve (1.4) with the known σ in place of $\widehat{\sigma}$. (We remark that if σ is unknown one can obtain scale equivariance either by solving (1.4) or by simultaneous estimation of θ and σ, as in Sheahan (1988)).

To simplify the discussion we assume $\epsilon = 0$ and without loss of generality we assume $\sigma = 1$. Then the distribution of the i.i.d. errors is standard normal in an <u>unknown</u> interval $(-a_0, a_0)$ and completely unknown outside $(-a_0, a_0)$. We may assume that the error distribution is non-normal in every interval containing (a_0, a_0), for otherwise we can re-define a_0 by $a_0 = \sup\{a_0^* |$ the distribution is of standard normal form in $(-a_0^*, a_0^*)\}$. Our aim is then to estimate the point a_0 where outside $(-a_0, a_0)$ the distribution is not of standard normal form.

2.1. One heuristic procedure for estimating a_0 is to make a normal probability plot of the ordered residuals $\widehat{U}_1, \leq \ldots \leq \widehat{U}_n$, where

$$\widehat{U}_i = X_i - c_i'\tilde{\theta}, \quad i = 1, \ldots, n \tag{2.1}$$

and $\tilde{\theta}$ is an initial consistent and shift equivariant estimator of θ. Such an estimator can be constructed as follows, in the case where F has a positive and continuous density f. Let $\tilde{\theta}_L$ be the least absolute deviation estimator, defined in the introduction. It was shown by Koenker and Bassett (1978) - see also Bassett and Koenker (1978) and Ruppert and Carroll (1980) - that

$$\tilde{\theta}_L \quad \longrightarrow_{\mathcal{P}} \theta + (F^{-1}(\frac{1}{2}), \ldots, F^{-1}(\frac{1}{2}))'.$$

$$S(a_0) = \sup_{u \epsilon A_n} |F_n(u) - \Phi(u)|$$

where $A_n = (-a_0 - \delta_n, -a_0 + \delta_n) \cup (a_0 - \delta_n, a_0 + \delta_n)$, and δ_n depends only on n and satisfies $\delta_n \to 0$ as $n \to \infty$. An estimator of the true a_0 is then

$$\widehat{a}_3 = \max\{a_0 | \text{ the observed value of } S(a_0) \text{ is not significant}\}.$$

2.5 At the cost of quite extensive analytical and computational inconvenience, we can estimate a_0 using a likelihood ratio approach. (Recall that, for simplicity of presentation, we are supposing that $\epsilon = 0$ in (1.7)). The density of U_i in the model (1.2) can, from (1.7), be written as

$$\phi(u)I(|u| \leq a_0) + g(u)I(|u| > a_0)$$

where g is a (assumed to exist) density of the unknown tail portion of F and I(A) denotes the indicator function of a set A. The likelihood function of the sample $x = (x_1, \ldots, x_n)'$ is then

$$L(\theta, a_0, g|x) = \prod_{i=1}^{n} \{\phi(x_i - c_i'\theta)I(|x_i - c_i'\theta| \leq a_0) + g(x_i - c_i'\theta)I(|x_i - c_i'\theta| > a_0)\}.$$

In this, we propose replacing θ by the estimate $\tilde{\theta}$ of sec. 2.1. In order to use L to perform a likelihood ratio test of the alternatives $H_0 : a_0 = a_0^*, H_1 : a_0 > a_0^*$, we have to contend with the complication that g is unknown. One possibility, which we have not explored theoretically, is to replace g by a certain least favourable density g_0 for the likelihood ratio test of H_0 versus H_1. Specifically, if we define

$$\lambda(X; \tilde{\theta}, a_0^*, g) = \sup_{a_0} L(\tilde{\theta}, a_0, g|X)/L(\tilde{\theta}, a_0^*, g|X),$$

we let g_0 minimize the power of the test based on λ, of H_0 versus H_1, at some specified alternative.

A rather complicated estimate of a_0 is then defined as

$$\widehat{a}_4 = \max\{a_0^* | \lambda(x; \tilde{\theta}, a_0^*, g_0) \text{ is not significant}\}.$$

3. ESTIMATION OF PARAMETERS IN (1.5) AND (1.6) WHEN F IS ARBITRARY

In this section we relax the assumption that F belongs to the class $\mathcal{F}$ of (1.7), and consider again the problem of estimating θ in (1.2) by solving an equation of the form (1.3). If we wish to use a ψ-function of the form (1.5) or (1.6), which as indicated earlier is advisable if we anticipate gross errors in the data, we are led to the problem of choosing suitable values for the parameters in (1.5) and (1.6). Note that if y_1 in (1.5) is known, then (1.6) includes (1.5) as a special case, so we deal exclusively with the choices of y_o and a_0 in (1.6). The numbers y_0 and a_0 are easily interpreted - ψ is linear in $(-y_0, y_0)$ so we are using a least squares prodecure on the residuals in that interval; outside $(-a_0, a_0)$ residuals have no influence on the solution of (1.3) , while residuals are down-weighted in $(-a_0, -y_0)$ and (y_0, a_0) by ξ. Clearly the points y_0 and a_0 must be chosen appropriately: a poor choice of a_0, for example, may lead to the dismissal of "good" observations, or the retention of "poor" ones which may result in inconsistency of the estimator of θ in (1.3). The choice of the functional form of ξ is not crucial; one should however ensure that the resulting ψ does not descend too rapidly, in order that the asymptotic variance functional (-see (3.2) below) does not become inflated.

We commence our analysis by making the following assumptions about $C^{(n)}$ in (1.2).

A1)There exists a positive definite matrix C_0 such that

$$C^{(n)'}C^{(n)}/n \rightarrow C_0 \text{ as } n \rightarrow \infty$$

A2)$\sup\{|c_{ij}^{(n)}| | i = 1, \ldots, n; j = 1, \ldots, p; n \geq 1\} \leq K$ for some constant K.

We assume further that an "initial" shift equivariant and consistent estimator $\tilde{\theta}$ of θ exists: that is

A3)There exists $\tilde{\theta} = \tilde{\theta}^{(n)}$ such that $\tilde{\theta}(X^{(n)} + C^{(n)}t) = \tilde{\theta}(X^{(n)}) + t$ for all $t \in \mathbf{R}^p$,and $\tilde{\theta} \underset{\mathcal{P}}{\longrightarrow} \theta$. Note that, without some conditions on F, one cannot guarantee the existence of such a $\tilde{\theta}$. However, if for example F can be assumed to have a positive and continuous density and $F^{-1}(1/2) = 0$ then, as pointed out in Sec. 2.1, $\tilde{\theta}_L$ is consistent. It is in addition scale equivariant, i.e. $\tilde{\theta}_L(\lambda X^{(n)}) = |\lambda|\tilde{\theta}_L(X^{(n)}), \lambda \neq 0$. Even if $F^{-1}(1/2) \neq 0$, the estimator $\tilde{\theta}$ of sec. 2.1 is, under certain conditions, consistent, in particular if the central part of F is assumed symmetric and strongly unimodal. For other candidates for $\tilde{\theta}$, such as trimmed least squares estimators, see the simulation results in Sheahan, Lind and Mehra (1989) and Lind (1988).

For any given (fixed) y_0 and a_0 in (1.6), and with $\omega = 10^{-6}$ say, define a "final"estimator $\widehat{\theta}_{y_0,a_0} = \widehat{\theta}^{(n)}_{y_0,a_0}$ of θ as follows:

$$\widehat{\theta}_{y_0,a_0} = \begin{cases} \theta^* , & \text{if the equation (1.3) solved by Newton's method with} \\ & \text{initial value } \tilde{\theta} \text{ has an unique solution } \theta^* \\ \tilde{\theta} , & \text{if the Newton iterates do not converge.} \end{cases} \tag{3.1}$$

The following theorem gives the asymptotic behaviour of $\widehat{\theta}_{y_0,a_0}$. We omit the proof, since it is similar to the proofs of Lemma 3.1 and Theorem 3.1 of Collins, Sheahan and Zheng (1986).

Theorem 3.1.

Under assumptions A1), A2) and A3) and the conditions on ξ given in sec. 1, the estimator $\widehat{\theta}_{y_0,a_0}$ is for each fixed y_0 and a_0 a consistent estimator of θ. Further, $n^{\frac{1}{2}}(\widehat{\theta}_{y_0,a_0} - \theta)$ converges in distribution to the multivariate normal distribution with mean 0 and covariance matrix $C_0^{-1}V(\psi_{y_0,a_0}, F)$ where ψ_{y_0,a_0} is the ψ of (1.6) with its dependence on y_0 and a_0 emphasized, and

$$V(\psi_{y_0,a_0}, F) = \frac{\int_{-a_0+\omega}^{a_0-\omega} \psi^2_{y_0,a_0}(u)dF(u)}{[\int_{-a_0+\omega}^{a_0-\omega} \psi'_{y_0,a_0}(u)dF(u)]^2} \tag{3.2}$$

An estimator of $V(\psi_{y_0,a_0}, F)$ by the method of moments is

$$\widehat{V}(_{y_0,a_0}, F) = \frac{\frac{1}{n}\sum_{i=1}^{n} \psi^2_{y_0,a_0}(X_i - c_i'\tilde{\theta})}{[\frac{1}{n}\sum_{i=1}^{n} \psi'_{y_0,a_0}(X_i - c_i'\tilde{\theta})]^2} \tag{3.3}$$

It is tacitly assumed that F is such that the denominator of (3.2) does not vanish, so that with high probability for large n, (3.3) is well-defined.

Now if F were known, a natural choice for the pair of numbers $(y_0, a_0)'$ would be that which minimizes (3.2). We propose choosing $(y_0, a_0)'$ to be the vector minimizing (3.3). This procedure will then be asymptotically efficient if the minimizer of (3.3) is a consistent estimator of the minimizer of (3.2), and this we shall show under certain conditions. We shall agree that whatever choice we make for $(y_0, a_0)'$, y_0 should not be less than $y_0'\sigma$ and a_0 should not exceed $a_0'\sigma$, where y_0' and $a_0', (y_0' < a_0')$ are fixed numbers determined from experience: the experimenter will know not to expect "aberrant" errors close to zero, so wants to use least squares on residuals in some internal $(-y_o'\sigma, y_0'\sigma)$, but also knows that he can expect gross errors outside $(-a_0'\sigma, a_0'\sigma)$.

We now add the following assumption:

A4) There exists $\widehat{\sigma} = \widehat{\sigma}^{(n)}$ such that $\widehat{\sigma}$ is shift equivariant, scale invariant and satisfies $\widehat{\sigma} \underset{\mathcal{P}}{\longrightarrow} \sigma$.

As with $\tilde{\theta}$, the existence of such a scale estimator cannot be guaranteed without imposing further conditions on F. We remark however that if F happened to belong to $\mathcal{F}$ and if the rows of $C^{(n)}$ contained repetitions, such a $\widehat{\sigma}$ can be obtained - see Sheahan (1988). Other estimators of σ are possible, depending on what functional of F one selects to define the scale parameter. While consistency of $\widehat{\sigma}$ is required for the optimality theory we are presenting to be valid, in practice one may be required to use the median absolute deviation or other such robust scale estimate.

We now define the following subsets of $\mathbf{R}^2$:

$$S = \{(y_0, a_0)' | -y_0'\sigma \leq y_0 \leq a_0 \leq a_0'\sigma\} \text{ and}$$

$$\widehat{S} = \{(y_0, a_0)' | -y_0'\widehat{\sigma} \leq y_0 \leq a_0 \leq a_0'\widehat{\sigma}\}.$$

Finally, let $(y_0^*, a_0^*)'$ satisfy

$$V(\psi_{y_0^*, a_0^*}, F) = \inf\{V(\psi_{y_0, a_0}, F) | (y_0, a_0) \in S\} \tag{3.4}$$

and define an estimator $(\widehat{y}_0^*, \widehat{a}_0^*)'$ of $(y_0^*, a_0^*)'$ by letting $(\widehat{y}_0^*, \widehat{a}_0^*)'$ satisfy

$$\widehat{V}(\psi_{\widehat{y}_0^*, \widehat{a}_0^*}, F) = \inf\{\widehat{V}(\psi_{y_0, a_0}, F) | (y_0, a_0) \epsilon \widehat{S}\} \tag{3.5}$$

We now have the following theorem, the proof of which will be given in the appendix.

Theorem 3.2.

Assume that $(y_0^*, a_0^*)'$ is the unique minimizer in S of $V(\psi_{y_0,a_0}, F)$. Then under assumptions A1), A2), A3) and A4), and the conditions on ξ given in sec. 1, the statistic $(\widehat{y}_0^*, \widehat{a}_0^*)'$ satisfies

$$(\widehat{y}_0^*, \widehat{a}_0^*)' \quad \longrightarrow_{\mathcal{P}} (y_0^*, a_0^*)'.$$

We remark that uniqueness of (y_0^*, a_0^*) is assumed only to simplify the proof of Theorem 3.2. If (y_0^*, a_0^*) is not unique then, with positive probability, $\widehat{V}(\psi_{y_o,a_0}, F)$ will have more than one solution even for large n. In such a case, care must be taken in practise to identify an appropriate minimizer of $\widehat{V}(\psi_{y_0,a_0}, F)$ by a specified algorithm, to ensure that this minimizer is a consistent estimate of a minimizer of $V(\psi_{y_0,a_0}, F)$. (Compare with the problem of solving (1.3) in practice - it has infinitely many solutions because ψ vanishes outside an interval, and hence we chose a solution (3.1) which is a consistent estimator of θ, stated in Theorem 3.1.)

From Theorems (3.1) and (3.2) we immediately have:

Theorem 3.3.

Under the same asssumptions as in Theorem 3.2, $n^{\frac{1}{2}}(\widehat{\theta}_{\widehat{y}_0^*, \widehat{a}_0^*} - \theta)$ converges in distribution to the multivariate normal distribution with mean 0 and covariance matrix $C_0^{-1} V(\psi_{y_0^*, a_0^*}, F)$.

We remark that if one wishes to use Theorem 3.3 in practice to obtain confidence regions for, or to perform hypothesis tests about, θ, one can estimate $V(\psi_{y_0^*, a_0^*}, F)$ by $\widehat{V}(\psi_{\widehat{y}_0^*, \widehat{a}_0^*}, F)$, whose value will not depend on F.

We remark further that if in practice one knows nothing about F, one may, at least for slight analytical convenience in the computation of $(\widehat{y}_0^*, \widehat{a}_0^*)'$ from (3.5), consider using a linear ξ. The ψ-function used in solving (1.3) would then be (with $w = 10^{-6}$)

$$\psi(y) = \begin{cases} y, & |y| \le \widehat{y}_0^* \\ \frac{\widehat{y}_0^*}{\widehat{a}_0^* - 10^{-6} - \widehat{y}_0^*}(\widehat{a}_0^* - 10^{-6} - |y|), & \widehat{y}_0^* \le |y| \le \widehat{a}_0^* - 10^{-6} \\ 0, & |y| \ge \widehat{a}_0^* - 10^{-6}. \end{cases}$$

In any case, whatever the choice of ξ, subject to its properties given in sec.1, the resulting estimator $\widehat{\theta}_{\widehat{y_0^*},\widehat{a_0^*}}$ is asymptotically normal, and optimal in the sense that it has minimum asymptotic variance among all solutions of (1.3) that are based on ψ-functions of the form (1.6).

In conclusion, we remark that an alternative procedure to the one of this section is to replace $\tilde{\theta}$ in (3.3) by the unknown θ and then choose a value for $(y_0, a_0)'$ that minimizes (3.3) subject to (1.3) holding. We have not examined the theoretical properties of this procedure, which is an analogue of Huber's "proposal 3" (Huber 1964).

APPENDIX

To prove Theorem 3.2 we first prove the following lemma.

Lemma.

Define $B = \{(y_0, a_0)' | y_0' \leq y_0 \leq a_0 \leq a_0'\}$.

Under assumptions A2) and A3), and the conditions on ξ given in sec. 1, we have

$$\sup\{\widehat{V}(\psi_{y_0,a_0}, F) - V(\psi_{y_0,a_0} F) \big| (y_0, a_0) \in B\} \underset{\mathcal{P}}{\longrightarrow} 0.$$

Proof of the Lemma:

From the definitions of $\widehat{V}(\psi_{y_0,a_0}, F)$ and $V(\psi_{y_0,a_0}, F)$ in (3.3) and (3.2) respectively, it is sufficient to show that

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n}\psi^2_{y_0,a_0}(X_i - c_i'\tilde{\theta}) - \int_{-a_0+\omega}^{a_0-\omega}\psi^2_{y_0,a_0}(u)dF(u)\big|(y_0,a_0) \in B\right\} \underset{\mathcal{P}}{\longrightarrow} 0\Big\} \qquad (4.1).$$

and that

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n}\psi'_{y_0,a_0}(X_i - c_i'\tilde{\theta}) - \int_{-a_0+\omega}^{a_0-\omega}\psi'_{y_0,a_0}(u)dF(u)\big|(y_0,a_0) \in B\right\} \underset{\mathcal{P}}{\longrightarrow} 0 \qquad (4.2).$$

To prove (4.1) it is clearly sufficient, by the triangle inequality, to show that

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n}\psi^2_{y_0,a_0}(X_i - c_i'\tilde{\theta}) - \psi^2_{y_0,a_0}(X_i - c_i'\theta)|\big|(y_0,a_0) \in B\right\} \underset{\mathcal{P}}{\longrightarrow} 0 \qquad (4.3)$$

and that

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n}\psi^2_{y_0,a_0}(X_i - c_i'\theta) - \int_{-a_0+\omega}^{a_0-\omega}\psi^2_{y_0,a_0}(u)dF(u)|\big|(y_0,a_0) \in B\right\} \underset{\mathcal{P}}{\longrightarrow} 0 \qquad (4.4).$$

We first prove (4.4). Let $\epsilon > 0$.

For any fixed $(y_0^*, a_0^*) \in B$, the Weak Law of Large Numbers implies that

$$\frac{1}{n}\sum_{i=1}^{n}\psi^2_{y_0^*,a_0^*}(X_i - c_i'\theta) \underset{\mathcal{P}}{\longrightarrow} \int_{-a_0+\omega}^{a_0-\omega}\psi^2_{y_0^*,a_0^*}(u)dF(u) \qquad (4.5).$$

By compactness of B, and noting that $\psi_{y_o,a_0}(x - c_i'\theta)$ vanishes outside $(-a_0 + \omega, a_0 - \omega)$, there exist neighbourhoods $B_j = B_j[(y_0^j, a_0^j)]$ of a finite number m of points $(y_0^j, a_0^j), j = 1, \ldots, m$, in B such that $\bigcup_{j=1}^{m} B_j = B$ and

$$(y_0, a_0) \in B_j \Rightarrow \sup |\psi^2_{y_0,a_0}(x - c_i'\theta) - \psi^2_{y_0^j,a_0^j}(x - c_i'\theta)| |x - c_i'\theta \in (-a_0 + \omega, a_0 - \omega\} < \epsilon.$$

Hence

$$\sup\left\{ \left|\frac{1}{n}\sum_{i=1}^{n} \psi^2_{y_0,a_0}(X_i - c_i'\theta) - \int_{-a_0+\omega}^{a_0-\omega} \psi^2_{y_0,a_0}(u)dF(u)\right| (y_0, a_0) \in B \right\}$$

$$\leq \max_{j=1,\ldots,m} \sup_{(y_0,a_0)\in B_j} \left\{ \frac{1}{n}\sum_{i=1}^{n} \left|\psi^2_{y_0,a_0}(X_i - c_i'\theta) - \psi^2_{y_0^j,a_0^j}(X_i - c_i'\theta)\right| \right\} +$$

$$\max_{j=1,\ldots,m} \sup_{(y_0,a_0)\in B_j} \left\{ \int_{-a_0+\omega}^{a_0-\omega} \left|(\psi^2_{y_0^j,a_0^j}(u) - \psi^2_{y_0 a_0}(u)\right| dF(u) \right\} +$$

$$\max_{j=1,\ldots,m} \left\{ \left|\frac{1}{n}\sum_{i=1}^{n} \psi^2_{y_0^j,a_0^j}(X_i - c_i'\theta) - \int_{-a_0+\omega}^{a_0-\omega} \psi^2_{y_0^j,a_0^j}(u)dF(u)\right| \right\}$$

$$< \epsilon + \epsilon + \max_{j=1,\ldots,m} \left\{ \left|\frac{1}{n}\sum_{i=1}^{n} \psi^2_{y_0^j,a_0^j}(X_i - c_i'\theta) - \int_{-a_0+\omega}^{a_0-\omega} \psi^2_{j_0^j,a_0^j}(u)dF(u)\right| \right\} \quad (4.6).$$

(4.4) now follows immediately from (4.5) and (4.6).

To prove (4.3), first note that

$$\left| \max_{i=1,\ldots,n} (c_i'\tilde{\theta} - c_i'\theta)\right| \leq \max_{i=1,\ldots,n} \sum_{j=1}^{P} |c_{ij}||\tilde{\theta}_j - \theta| \quad \xrightarrow[\mathcal{P}]{} 0$$

by A2) and A3).

It thus suffices to show (-see also Billingsley 1968, theorem 8.2) that given $\epsilon > 0$ and $\eta > 0$, there exist $\delta > 0$ and an integer n_0 such that

$$P\left(\sup_{(y_0,a_0)\in B} \sup_{||t_1-t_2||<\delta} \frac{1}{n}\sum_{i=1}^{n} |\psi^2_{y_0,a_0}(X_i - t_1) - \psi^2_{y_0,a_0}(X_i - t_2)| > \epsilon \right) \leq \eta \quad (4.7)$$

Using again (uniform) continuity of $\psi_{y_0,a_0}(x - t)$ and compactness of B, we can find $\delta > 0$ such that

$$\sup_{(y_0,a_0)\in B} \sup_{||t_1-t_2||<\delta} \sup_{x\in\mathbf{R}'} |\psi^2_{y_0,a_0}(x - t_1) - \psi^2_{y_0,a_0}(x - t_2)| < \epsilon$$

and hence it follows that (4.7) holds with, in fact, $\eta = 0$ and $n_0 = 1$. The proof of (4.2) is identical with that of (4.1) on replacing ψ^2 by ψ'. The proof of the lemma is thus complete.

Proof of Theorem 3.1

Fix $\epsilon > 0$. Since $(y_0^*, a_0^*)'$ is the unique minimizer of $V(y_0, a_0, F)$ in S, continuity of $V(y_0, a_0, F)$ shows that there exists $\delta > 0$ such that the events

$$\left.\begin{array}{l} ||(\widehat{y}_0^*, \widehat{a}_0^*)' - (y_0^*, a_0^*)'|| > \epsilon \text{ and } (\widehat{y}_0^*, \widehat{a}_0^*)' \in S \\ \text{imply the event} \\ \qquad V(\widehat{y}_0^*, \widehat{a}_0^*, F) > V(y_0^*, a_0^*, F) + \delta. \end{array}\right\} \tag{4.8}$$

Since $S \cap \widehat{S} = S$ with probability tending to one as $n \to \infty$ by A4), it follows from (4.8) and the Lemma that with probability tending to one as $n \to \infty$, the event

$$||(\widehat{y}_0^*, \widehat{a}_0^*)' - (y_0^*, a_0^*)'|| > \epsilon \text{ implies the event } \widehat{V}(\widehat{y}_0^*, \widehat{a}_0^*, F) > V(y_0^*, a_0^*, F) + \delta.$$

Since this last event has probability tending to zero as $n \to \infty$, we have $P(||(\widehat{y}_0^*, \widehat{a}_0^*)' - (y_0^*, a_0^*)'|| > \epsilon) \to 0$ as $n \to \infty$, completing the proof.

REFERENCES

Bassett, G., Jr., and Koenker, R. (1978). Asymptotic Theory of Least Absolute Error Regression. Journal of the American Statistical Association, 73, 618-622.

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

Collins, J.R. (1976). Robust Estimation of a Location Parameter in the Presence of Asymmetry. Annals of Statistics, 4, 68-85.

Collins, J.R. and Wiens, D.P. (1985). Minimax Variance M–estimators in ϵ-contamination Models. Annals of Statistics, 3, 1078-1096.

Collins, J.R., Sheahan, J.N. and Zheng, Z. (1986). Robust Estimation in the Linear Model with Asymmetric Error Distributions. Journal of Multivariate Analysis, 20, 220-243.

Hampel, F.R. (1973). Robust Estimation: A Condensed Partial Survey. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 27, 87-104.

Huber, P.J. (1964). Robust Estimation of a Location Parameter. Annals of Mathematical Statistics, 35,73-101.

Huber, P.J. (1973). Robust Regression. Annals of Statistics, 1, 799-821.

Huber, P.J. (1981). Robust Statistics. John Wiley and Sons, New York.

Jaeckel, L.A. (1971). Some Flexible Estimates of Location. Annals of Mathematical Statistics, 42, 1540-1552.

Koenker, R. and Bassett, G., Jr. (1978). Regression Quantiles.Econometrika, 46, 33-50.

Lind, J.C. (1988). Monte Carlo Results for a Variety of Estimators of Parameters in Linear Models with Asymmetric Error Distributions. Univ. of Alberta, Dept. of Statistics Technical Report No. 88.17, Edmonton, Canada.

Ruppert, D. and Carroll, R.J. (1980). Trimmed Least Squares Estimation in the Linear Model. Journal of the American Statistical Association, 75, 828-138.

Sheahan, J.N. (1988). Robust Estimation of Regression and Scale Parameters in Linear Models with Asymmetric Error Distributions. Mathematische Operationsforschung und Statistik, Series Statistics, 19, 27-37.

Sheahan, J.N., Lind, J.C. and Mehra, K.L. (1989). Assymmetric Errors in Linear Models ...Estimation Theory and Monte Carlo. Univ. of Alberta, Dept. of Statistics Technical Report No. 88.20, Edmonton, Canada.

Yohai, V.J. (1974). Robust Estimation in the Linear Model.Annals of Statistics, 2, 562-567.