On the Choice of Support of Re-Descending ψ-Functions in Linear Models with Asymmetric Error Distributions

By M. Hlynka[1], J. N. Sheahan[2], and D. P. Wiens[3]

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.

1 Introduction

We consider the linear regression model

$$X^{(n)} = C^{(n)}\theta + \sigma U \tag{1.1}$$

where $C^{(n)}$ is an $n \times p$ design matrix with rows $c_i^{(n)'}$, and the components U_i of U are i.i.d., with d.f. $F(u)$.

For an appropriately chosen function ψ, we estimate θ by an *M*-estimator $\hat{\theta}$, defined as a solution to the system of equations

---

[1] M. Hlynka, Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario N9B 3P4.
[2,3] J. N. Sheahan and D. P. Wiens, Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta T6G 2G1.

$$\sum_{i=1}^{n} c_i \psi \left(\frac{X_i - c_i'\hat{\theta}}{\hat{\sigma}} \right) = 0 \; , \tag{1.2}$$

where $\hat{\sigma}$ is a scale-equivariant estimator of σ. Note that, where convenient, we omit the dependence of the model on n. In this article, we consider using a "redescending" ψ-function of the form

$$\psi(y) = \begin{cases} y \; , & |y| \leq y_0 \\ y_1 \tanh \left[\tfrac{1}{2} y_1 (a_0 - \omega - |y|) \right] \operatorname{sgn}(y) \; , & y_0 \leq |y| \leq a_0 - \omega \\ 0 \; , & |y| \geq a_0 - \omega \end{cases} \tag{1.3}$$

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

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

$$\psi(y) = \begin{cases} y \; , & |y| \leq y_0 \\ y_0 \xi \left(\dfrac{a_0 - \omega - |y|}{a_0 - \omega - y_0} \right) \operatorname{sgn}(y) \; , & y_0 \leq |y| \leq a_0 - \omega \\ 0 \; , & |y| \geq a_0 - \omega \end{cases} \tag{1.4}$$

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

Let $\varepsilon, 0 < \varepsilon < 1$ and a_0 be *known* constants. Then $\mathscr{F} = \{F | F$ is a distribution function;

$$\left. \begin{array}{l} y \in (-a_0, a_0) \Rightarrow F \text{ has a density } f \text{ of the form} \\ \qquad f(y) = (1 - \varepsilon) \phi(y) + \varepsilon h(y) \text{ where } \phi \text{ is the standard} \\ \qquad \text{normal density and } h \text{ is an unknown density sym-} \\ \qquad \text{metric about 0;} \\ y \notin (-a_0, a_0) \Rightarrow F \text{ is completely unknown (and hence, in particular,} \\ \qquad \text{possibly asymmetric).} \} \end{array} \right\} \tag{1.5}$$

The optimal choices of y_0, y_1 and ω are then given as follows: y_0 and y_1 are the solutions of the equations

$$y_0 = y_1 \tanh\left[\tfrac{1}{2}y_1(a_0 - \omega - y_0)\right] \quad \text{and}$$

$$\left.\begin{array}{l} \dfrac{\varepsilon}{1-\varepsilon} = \dfrac{\phi(y_0)}{y_1 \cosh^2\left[\tfrac{1}{2}y_1(a_0 - \omega - y_0)\right]}\left[\sinh\left(y_1(a_0 - \omega - y_0)\right) + y_1(a_0 - \omega - y_0)\right] \\[2ex] \qquad\qquad - 2\,\Phi(a_0 - \omega) + 2\,\Phi(y_0) \end{array}\right\} (1.6)$$

where $\Phi(x) = \int_{-\infty}^{x} \phi(y)\,dy$, and ω is any number as close to zero as desired (e.g. $\omega = 10^{-6}$, as in the simulations of Lind (1988)). For reasons of identifiability of θ in (1.1), ε must be "reasonably small" and y_0 "reasonably large" – see Lind, Mehra and Sheahan (1989) for more detail, and Lind (1988) for numerical work.

If then F is known to belong to $\mathscr{F}$ with ε, σ, and a_0 known, the optimal M-estimator of θ in (1.1) is obtained by solving (1.2), with $\hat{\sigma} = \sigma$, ψ given by (1.3) and y_0 and y_1 given by (1.6). Two questions then arise:

(i) If F is known to belong to $\mathscr{F}$ but a_0 is unknown, how should one estimate this parameter?

(ii) If little is known about F, in particular if F is not known to belong to $\mathscr{F}$, but one wishes to use a ψ-function of the form (1.3) or (1.4), how should one choose the parameters of such a ψ-function?

Question (i) is motivated by the following considerations. In many practical situations – see Sheahan (1988) for examples – it can be argued that the family $\mathscr{F}$ is a reasonable model for errors in the linear model (1.1), in particular since it allows for gross errors that may or may not be generated from a symmetric distribution. The resulting optimal estimator of θ relies on the parameter a_0 in (1.5) being known, for the ψ of (1.3) depends on this a_0. Now if one's intuition or background knowledge of the mechanism generating the data does not give one a specification of a_0, one is led to the problem of estimating it from the data. A number of such estimates are proposed in a technical report by Hlynka, Sheahan and Wiens (1989).

Concerning (ii), when F is not known to be a member of F, the a_0 appearing in (1.3) and (1.4) cannot now be thought of as the parameter a_0 of (1.5). Rather we simply interpret $(-a_0, a_0)$, or more precisely $(-a_0 + \omega, a_0 - \omega)$, as the support of a ψ-function that we wish to use in solving (1.2). The use of such a ψ, vanishing outside an interval, is reasonable if one wishes to protect against the influence of gross errors towards the solution of (1.2). See Hampel et al. (1986) for further attractive properties of re-descenders. In particular, note that a tanh form for ψ which is a special case of (1.4), and is very similar to (1.3), is optimal V-robust – it minimizes the supremum (over R) of the normalized change-of-variance function. Furthermore, the breakdown point then has the maximum possible value of $1/2$. In practice it is sometimes recommended that when outliers are suspected in the data one should use such a re-descender as well as a non re-descender, and compare the resulting estimates – see Hogg (1979).

In answering (ii) we will, in Sect. 2, assume that in (1.3) both a_0 and y_0 are unknown, and, without assuming σ known, will choose values for them that minimize an estimator of the asymptotic variance of the estimator $\hat{\theta}$ in (1.2). This procedure is based on an idea of Jaeckel (1971) who estimated the proportion of observations to be deleted in computing a trimmed mean; see also "proposal 3" of Huber (1964), and Yohai (1974), who estimated in a similar manner the tuning constant k of Huber's ψ-function. Some of our derivations are based on those of Yohai (1974).

Since the work contained in Sect. 2 is rather technical and notationally cumbersome, we briefly summarize here its main content:

For each fixed (y_0, a_0), $0 < y_0 < a_0 < \infty$, an estimator $\hat{\theta} = \hat{\theta}_{y_0, a_0}$ is defined in (2.1), and the asymptotic behaviour of this estimator is described in Theorem 2.1. The question that remains then is: What value of (y_0, a_0) should one use in practice in constructing $\hat{\theta}$? A reasonable value (y_0^*, a_0^*), given in (2.23), would be that which minimizes (over all (y_0, a_0), $0 < y_0 < a_0 < \infty$) the asymptotic variance of solutions $\hat{\theta} = \hat{\theta}_{y_0, a_0}$ of (1.2). Since (y_0^*, a_0^*) will depend on the unknown $F \in \mathcal{F}$, it cannot be used in practice. Accordingly, in (2.24) we define an estimator $(\hat{y}_0^*, \hat{a}_0^*)$ of (y_0^*, a_0^*), this estimator being shown in Theorem 2.2 to be consistent. Finally, Theorem 2.3 establishes asymptotic normality of $\hat{\theta} = \hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}$, this estimator having the same asymptotic optimality properties as the solution of (1.2) one would have used had the "correct" (y_0, a_0), that is, (y_0^*, a_0^*), been known in the first place.

2 Estimation of Parameters in (1.3) and (1.4) when F is Arbitrary

In this section we relax the assumption that F belongs to the class $\mathcal{F}$ of (1.5), and consider again the problem of estimating θ in (1.1) by solving an equation of the form (1.2). If we wish to use a ψ-function of the form (1.3) or (1.4), 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 (1.3) and (1.4). Note that if y_1 in (1.3) is known, then (1.4) includes (1.3) as a special case, so we deal exclusively with the choices of y_0 and a_0 in (1.4). 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 procedure on the residuals in that interval; outside $(-a_0, a_0)$ residuals have no influence on the solution of (1.2), 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 (1.2). 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 (2.2) below) does not become inflated.

We commence our analysis by making the following assumptions about $C^{(n)}$ in (1.1):

A1) There exists a positive definite matrix C_0 such that

$$C^{(n)'} C^{(n)}/n \to C_0 \quad \text{as} \quad n \to \infty \ .$$

A2) $\sup \{|c_{ij}^{(n)}| \,|\, i = 1, \ldots, n; \, j = 1, \ldots, p; \, n \ge 1\} \le K$ 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 $t \in R^p$, and $\tilde{\theta} \xrightarrow[\mathscr{P}]{} \theta$. 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 $\tilde{\theta}_L$, the least absolute deviation estimator (i.e. the solution of $\sum_{i=1}^{n} c_i |X_i - c_i'\theta| = 0$), is consistent (see Lind, Mehra, Sheahan 1989 and references cited therein). It is in addition scale-equivariant, i.e. $\tilde{\theta}_L(\lambda X^{(n)}) = |\lambda| \, \tilde{\theta}_L(X^{(n)})$, $\lambda \ne 0$. Even if $F^{-1}(\frac{1}{2}) \ne 0$, $\tilde{\theta}_L$ 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 Lind, Mehra and Sheahan (1989) and Lind (1988).

For any given (fixed) y_0 and a_0 in (1.4), and with $\omega = 10^{-6}$ say, define a "final" estimator $\hat{\theta}_{y_0, a_0} = \tilde{\theta}_{y_0, a_0}^{(n)}$ of θ as follows:

$$\hat{\theta}_{y_0, a_0} = \begin{cases} \theta^* , & \text{if the equation (1.2) solved by Newton's method with} \\ & \text{initial value } \tilde{\theta} \text{ has a unique solution } \theta^* \\ \tilde{\theta} , & \text{if the Newton iterates do not converge.} \end{cases} \quad (2.1)$$

The following theorem gives the asymptotic behaviour of $\hat{\theta}_{y_0, a_0}$. We shall assume, for simplicity, that the distribution F of the i.i.d. errors U_i has a density f with respect to Lebesque measure. Finally, we add the non-stringent assumption that for each ψ of the form (1.4) we have

A4) $\int \psi(x) f'(x) \, dx < 0$

where f' is the derivative of f, which we assume to exist.

Theorem 2.1. Under assumptions A1), A2), A3), A4) and the conditions on ξ given in Sect. 1, the estimator $\hat{\theta}_{y_0, a_0}$ is for each fixed y_0 and a_0 a consistent estimator of θ. Further, $n^{1/2}(\hat{\theta}_{y_0, a_0} - \theta)$ 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 (1.4) with its dependence on y_0 and a_0 emphasized, and

$$V(\psi_{y_0, a_0}, F) = \frac{\int_{-a_0+\omega}^{a_0-\omega} \psi_{y_0, a_0}^2(u)\,dF(u)}{[\int_{-a_0+\omega}^{a_0-\omega} \psi'_{y_0, a_0}(u)\,dF(u)]^2} \ . \tag{2.2}$$

Proof of Theorem 2.1: We shall content ourselves with a sketch (albeit a rather detailed one) of the proof of Theorem 2.1, since the details are similar to those in the proofs of Lemma 3.1, Lemma 3.2 and Theorem 3.1 of Collins, Sheahan and Zheng (1986). For brevity of the argument that follows, we shall assume that in our model (1.1) σ is known, so that without loss of generality we take $\sigma = 1$. Thus, in (1.2), we omit the term $\hat{\sigma}$ (− in the case where σ is unknown, an easy modification of the proof below goes through if $\hat{\sigma}$ is a consistent estimator of σ), and hence we replace (1.2) by

$$\sum_{i=1}^{n} c_i \psi(X_i - c_i'\theta) = 0 \ , \tag{2.3}$$

In proving the consistency and asymptotic normality of the estimator $\hat{\theta} = \hat{\theta}_{y_0, a_0}$ of (2.1) there is no loss of generality in assuming that the true value θ_0 of θ is 0. This follows because $\tilde{\theta}$ is, by A3), a shift-equivariant estimator of θ_0, so that clearly $\hat{\theta}$ is also shift-equivariant. In establishing consistency of $\hat{\theta}$, we note that the standard consistency argument of Huber (1964) must be modified to take into account the fact that the equation (2.3) has multiple roots, due to the fact that ψ vanishes outside an interval.

We define the random process

$$H_n^*(t) = \frac{1}{n} \sum_{i=1}^{n} c_i \psi(X_i - c_i't) \ , \quad t \in R^p \tag{2.4}$$

and the function

$$H_n(t) = \frac{1}{n} \sum_{i=1}^{n} c_i \int \psi(x - c_i't) f(x)\,dx \ , \quad t \in R^p \ .$$

Note that if, for simplicity, we assume that f is symmetric on $(-a_0+\omega, a_0-\omega)$ then since ψ is skew-symmetric equation (2.3) will now read

$$H_n^*(t) = H_n(0) \ . \tag{2.5}$$

The idea underlying the proof of consistency of (2.1) is to show that the process $\{H_n^*(t) | t \in R^p\}$ "behaves like" its expected value $H_n(t)$, for sufficiently large n,

and that this function $H_n(t)$ has a unique root 0 in a neighbourhood of $\theta_0 = 0$, so that it will follow that (2.5) has, with probability, tending to 1 as $n \to \infty$, a root which is a consistent estimator of θ_0. Since by A3) $\bar{\theta}$ is also consistent, it will then follow that the $\hat{\theta}$ of (2.1) is itself consistent, i.e. that $\hat{\theta} = \hat{\theta}^{(n)} \xrightarrow{\mathscr{P}} 0$.

It follows from Lemma 3.2 of Collins, Sheahan and Zheng (1986) that for each $\delta \in (0, \omega)$,

$$\left. \begin{array}{l} \sup\{|H_n^*(t) - H_n(t)| : |t| \leq \delta\} \xrightarrow{\mathscr{P}} 0 \quad \text{as} \quad n \to \infty \\[2mm] \text{and} \\[2mm] \sup\left\{\left|\dfrac{\delta H_n^*(t)}{\delta t} - \dfrac{\delta H_n(t)}{\delta t}\right| : |t| \leq \delta\right\} \xrightarrow{\mathscr{P}} 0 \quad \text{as} \quad n \to \infty \ . \end{array} \right\} \tag{2.6}$$

Here $|\cdot|$ denotes any norm on the indicated spaces, and $\dfrac{\delta G(t)}{\delta t}$ denotes the Frechét derivative of a function $G : R^p \to R$. It follows from (2.6) that given any $c > 0$, there exists $\delta_2(c) > 0$ such that

$$P_F(\sup\{|H_n^*(t) - H_n(t)| : |t| \leq \delta_2(c)\} \leq c) \to 1 \quad \text{as} \quad n \to \infty \ , \tag{2.7}$$

and

$$P_F\left(\sup\left\{\left|\frac{\delta H_n^*(t)}{\delta t} - \frac{\delta H_n(t)}{\delta t}\right| : |t| \leq \delta_2(c)\right\} \leq c\right) \to 1 \quad \text{as} \quad n \to \infty \ , \tag{2.8}$$

where $P_F(A)$ denotes the probability of an event A when the distribution of the U_i is F. Since $\lim\limits_{\substack{|t| \to 0 \\ n \to \infty}} \dfrac{\delta H_n^*(t)}{\delta t} = -C_0 \int \psi(x) f'(x) dx$, it can be seen from A1), A4) and the perturbation lemma (see Ortega and Rheinboldt 1970) that

$$P_F\left(\det \frac{\delta H_n^*(t)}{\delta t} \neq 0 \quad \text{for all } t \text{ satisfying} \quad |t| \leq \delta_2(c)\right) \to 1 \quad \text{as} \quad n \to \infty \ . \tag{2.9}$$

Next, by a one-term Taylor expansion of $H_n(t)$, we see that

$$H_n(t) - H_n(0) = -\frac{1}{n} \sum_{i=1}^{n} c_i c_i' \left[\int \psi(x) f'(x) dx + 0(1)\right] t \quad \text{as} \quad t \to 0 \ ,$$

so that A1) and A4) imply that there exist $c_1 > 0$ and $\delta_3 > 0$ such that

$$|H_n(t) - H_n(0)| \geq c_1 |t| \quad \text{for all } t \text{ satisfying} \quad |t| \leq \delta_3 \ . \tag{2.10}$$

If we now define $c_2 = c_1 \delta_3 / 4$ and $\delta_1 = \min\{\omega, \delta_2(c_2), \delta_3\}$, it follows from (2.7)−(2.10) that

$$P_F (H_n^* \text{ is a one-to-one mapping on the set } \{t: |t| \le \delta_1\}) \to 1 \qquad (2.11)$$

and

$$P_F (H_n(0) \text{ is an inner point of the image of the set } \{t: |t| \le \delta_1\}$$
$$\text{under the mapping } H_n^*) \to 1 \text{ as } n \to \infty . \qquad (2.12)$$

It now immediately follows that

$$P_F ((2.3) \text{ has a unique solution in the set } \{t: |t| \le \delta_1\}) \to 1 \text{ as } n \to \infty . \quad (2.13)$$

If we now let $\delta > 0$ be any number satisfying $\delta \le \delta_1/4$, a repetition of the above argument, with an appropriate change of constants, shows that

$$P_F ((2.3) \text{ has a unique solution in the set } \{t: |t| \le \delta_1\}$$
$$\text{and this solution lies in the set } \{t: |t| \le \delta\}) \to 1 \text{ as } n \to \infty . \qquad (2.14)$$

By (2.14), the consistency of $\tilde{\theta}$ (from A3)), and the definition of $\hat{\theta}$ in (2.1), it follows that

$$P_F (\hat{\theta} \text{ is the unique solution of (2.3) in the set } \{t: |t| < \delta\}) \to 1 \text{ as } n \to \infty .$$
$$\qquad (2.15)$$

Because $\delta > 0$ is arbitrary, (2.15) implies that

$$\hat{\theta} = \hat{\theta}^{(n)} \xrightarrow[\mathscr{P}]{} 0 \qquad (2.16)$$

as was to be seen. It remains to sketch a proof of the asymptotic normality of $\hat{\theta}$, with variance-covariance matrix given by C_0^{-1} times (2.2). For clarity, we now revert to denoting by θ_0 (rather than 0) the true value of θ. Since ψ has, by assumption, a continuous derivative, an application of the mean value theorem yields

$$H_n^*(\theta) - H_n^*(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} c_i [\psi(X_i - c_i'\theta) - \psi(X_i - c_i'\theta_0)]$$

$$= -\frac{1}{n} \sum_{i=1}^{n} c_i c_i' \{\psi'[X_i - c_i'\theta_0 + \alpha_i c_i'(\theta_0 - \theta)]\}(\theta - \theta_0) , \qquad (2.17)$$

where the α_i, $0 \le \alpha_i \le 1$, are constants, $i = 1, \ldots, n$. Setting $\theta = \hat{\theta}$, we obtain from (2.17),

$$n^{1/2}(\hat{\theta} - \theta_0) \left(\frac{1}{n} \sum_{i=1}^{n} c_i c_i' \psi' [X_i - c_i' \theta_0 + \alpha_i c_i'(\theta_0 - \hat{\theta})] \right)$$

$$= -n^{1/2}[H_n^*(\hat{\theta}) - H_n(0)] + n^{1/2}[H_n^*(\theta_0) - H_n(0)] . \tag{2.18}$$

Now in (2.18), first note that

$$-n^{1/2}[H_n^*(\hat{\theta}) - H_n(0)] \xrightarrow{\mathcal{P}} 0 . \tag{2.19}$$

(2.19) follows from the fact that for each n and any fixed $\varepsilon > 0$, the event $\{H_n^*(\hat{\theta}) - H_n(0) = 0\}$ is contained in the event $\{n^{1/2} |H_n^*(\hat{\theta}) - H_n(0)| < \varepsilon\}$ and the fact that the former event has by (2.15) [see also (2.5)] probability tending to 1 as $n \to \infty$.

Next, by the Central Limit Theorem,

$$n^{1/2}[H_n^*(\theta_0) - H_n(0)] \xrightarrow{\mathcal{D}} N(0, C_0 \int \psi^2 f dx) , \tag{2.20}$$

where $\xrightarrow{\mathcal{D}}$ denotes converges in distribution, and finally we see that

$$\frac{1}{n} \sum c_i c_i' \psi' [X_i - c_i' \theta_0 + \alpha_i c_i'(\theta_0 - \hat{\theta})] \xrightarrow{\mathcal{P}} C_0 \int \psi'(x) f(x) dx . \tag{2.21}$$

Note that (2.21) follows from A 1), A 2), (2.16), (with θ_0 denoting the true value of θ) and piecewise uniform continuity of ψ' on $(-a_0 + \omega, a_0 - \omega)$. Since C_0 is positive definite, it follows from (2.18)-(2.21) and Slutsky's Theorem that

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \int \psi^2 f dx / [\int \psi' f dx]^2) ,$$

completing the proof of Theorem 2.1. $\square$

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

$$\hat{V}(\psi_{y_0,a_0},F) = \frac{\dfrac{1}{n}\sum_{i=1}^{n}\psi^2_{y_0,a_0}(X_i-c'_i\tilde{\theta})}{\left[\dfrac{1}{n}\sum_{i=1}^{n}\psi'_{y_0,a_0}(X_i-c'_i\tilde{\theta})\right]^2} \; . \qquad (2.22)$$

It is tacitly assumed that F is such that the denominator of (2.2) does not vanish, so that with high probability for large n, (2.22) 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 (2.2). We propose choosing $(y_0,a_0)'$ to be the vector minimizing (2.22). This procedure will then be asymptotically efficient if the minimizer of (2.22) is a consistent estimator of the minimizer of (2.2), 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 interval $(-y'_0\sigma,y'_0\sigma)$, but also knows that he can expect gross errors outside $(-a'_0\sigma,a'_0\sigma)$.

We now add the following assumption:

A5) There exists $\hat{\sigma} = \hat{\sigma}^{(n)}$ such that $\hat{\sigma}$ is shift-invariant,

scale-equivariant and satisfies $\hat{\sigma}\xrightarrow[\mathscr{P}]{}\sigma$.

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 happens to belong to $\mathscr{F}$ and if the rows of $C^{(n)}$ contain repetitions, such a $\hat{\sigma}$ 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 $\hat{\sigma}$ 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 estimator.

We now define the following subsets of R^2:

$$S = \{(y_0,a_0)' \mid -y'_0\sigma \le y_0 \le a_0 \le a'_0\sigma\} \quad \text{and}$$

$$\hat{S} = \{(y_0,a_0)' \mid -y'_0\hat{\sigma} \le y_0 \le a_0 \le a'_0\hat{\sigma}\} \; .$$

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

$$V(\psi_{y_0^*,a_0^*},F) = \inf\{V(\psi_{y_0,a_0},F) \mid (y_0,a_0)\in S\} \qquad (2.23)$$

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

$$\hat{V}(\psi_{\hat{y}_0^*, \hat{a}_0^*}, F) = \inf \{\hat{V}(\psi_{y_0, a_0}, F) \mid (y_0, a_0) \in \hat{S}\} \ . \tag{2.24}$$

We now have the following theorem.

Theorem 2.2. 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 A5), and the conditions on ξ given in Sect. 1, the statistic $(\hat{y}_0^*, \hat{a}_0^*)'$ satisfies

$$(\hat{y}_0^*, \hat{a}_0^*)' \xrightarrow[\mathscr{P}]{} (y_0^*, a_0^*)' \ .$$

Proof of Theorem 2.2: To prove Theorem 2.2 we first prove the following Lemma.

Lemma. Define $B = \{(y_0, a_0)' \mid y_0' \le y_0 \le a_0 \le a_0'\}$.

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

$$\sup \{\hat{V}(\psi_{y_0, a_0}, F) - V(\psi_{y_0, a_0}, F) \mid (y_0, a_0) \in B\} \xrightarrow[\mathscr{P}]{} 0 \ .$$

Proof of the Lemma: From the definitions of $\hat{V}(\psi_{y_0, a_0}, F)$ and $V(\psi_{y_0, a_0}, F)$ in (2.22) and (2.2) respectively, it is sufficient to show that

$$\sup \left\{ \frac{1}{n} \sum_{i=1}^n \psi_{y_0, a_0}^2 (X_i - c_i'\tilde{\theta}) - \int_{-a_0 + \omega}^{a_0 - \omega} \psi_{y_0, a_0}^2 (u) \, dF(u) \mid (y_0, a_0) \in B \right\} \xrightarrow[\mathscr{P}]{} 0 \tag{2.25}$$

and that

$$\sup \left\{ \frac{1}{n} \sum_{i=1}^n \psi_{y_0, a_0}' (X_i - c_i'\tilde{\theta}) - \int_{-a_0 + \omega}^{a_0 - \omega} \psi_{y_0, a_0}' (u) \, dF(u) \mid (y_0, a_0) \in B \right\} \xrightarrow[\mathscr{P}]{} 0. \tag{2.26}$$

To prove (2.25) it is clearly sufficient, by the triangle inequality, to show that

$$\sup \left\{\frac{1}{n}\sum_{i=1}^{n}|\psi_{y_0,a_0}^2(X_i-c_i'\tilde{\theta})-\psi_{y_0,a_0}^2(X_i-c_i'\theta)|\,|\,(y_0,a_0)\in B\right\} \xrightarrow[\mathscr{P}]{}0 \qquad (2.27)$$

and that

$$\sup \left\{\frac{1}{n}\sum_{i=1}^{n}|\psi_{y_0,a_0}^2(X_i-c_i'\theta)-\int_{-a_0+\omega}^{a_0-\omega}\psi_{y_0,a_0}^2(u)\,dF(u)|\,|\,(y_0,a_0)\in B\right\} \xrightarrow[\mathscr{P}]{}0 . \qquad (2.28)$$

We first prove (2.28). Let $\varepsilon > 0$.
For any fixed $(y_0^*,a_0^*)\in B$, the Weak Law of Large Numbers implies that

$$\frac{1}{n}\sum_{i=1}^{n}\psi_{y_0^*,a_0^*}^2(X_i-c_i'\theta) \xrightarrow[\mathscr{P}]{}\int_{-a_0+\omega}^{a_0-\omega}\psi_{y_0^*,a_0^*}^2(u)\,dF(u) . \qquad (2.29)$$

By compactness of B, and noting that $\psi_{y_0,a_0}(x-c_i'\theta)$ 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

$$(y_0,a_0)\in B_j \Rightarrow \sup\,|\psi_{y_0,a_0}^2(x-c_i'\theta)-\psi_{y_0^j,a_0^j}^2(x-c_i'\theta)|\,|\,(x-c_i'\theta)$$

$$\in(-a_0+\omega, a_0-\omega)\} < \varepsilon .$$

Hence

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

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

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

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

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

$$(2.30)$$

(2.28) now follows immediately from (2.29) and (2.30).
To prove (2.27), first note that

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

by A2) and A3) .

We now observe that since ψ is continuous and vanishes outside the compact set $[-a_0 + \omega, a_0 - \omega]$, it attains its supremum over that interval. Hence if we define $b = \sup\{|\psi(y)| : y \in (-a_0 + \omega, a_0 - \omega)\}$, we have $\left| \dfrac{1}{n} \sum_{i=1}^{n} \psi^2(X_i) \right| \leq b^2$ for all $n \geq 1$.
Consequently, if we decide to use Theorem 8.2, p. 55, of Billingsley (1968) to prove (2.27), we see that condition (i) of that Theorem is satisfied with his $a = b^2$ and his $\eta = 0$. By the above-mentioned Theorem of Billingsley (1968), or by a direct argument, (2.27) follows if we can show that given $\varepsilon > 0$ and $\eta > 0$, there exist $\delta > 0$ and an integer n_0 such that

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

for all $n \geq n_0$. \hfill (2.31)

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

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

and hence it follows that (2.31) holds with, in fact, $\eta = 0$ and $n_0 = 1$. The proof of (2.26) is identical with that of (2.25) on replacing ψ^2 by ψ'. The proof of the Lemma is thus complete. $\quad \Box$

Proof of Theorem 2.2: Fix $\varepsilon > 0$. Write $V(y_0, a_0, F)$ for $V(\psi_{y_0, a_0}, F)$. 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

$$\left. \begin{array}{l} |(\hat{y}_0^*, \hat{a}_0^*)' - (y_0^*, a_0^*)'| > \varepsilon \quad \text{and} \quad (\hat{y}_0^*, \hat{a}_0^*)' \in S \\[2ex] \text{imply the event} \\[2ex] V(\hat{y}_0^*, \hat{a}_0^*, F) > V(y_0^*, a_0^*, F) + \delta \ . \end{array} \right\} \qquad (2.32)$$

Since $S \cap \hat{S} = S$ with probability tending to one as $n \to \infty$ by A4), it follows from (2.32) that with probability tending to one as $n \to \infty$, the event

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

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

We remark that uniqueness of (y_0^*, a_0^*) is assumed only to simplify the proof of Theorem 2.2. If (y_0^*, a_0^*) is not unique then, with positive probability, $\hat{V}(\psi_{y_0, a_0}, F)$ will have more than one minimizer even for large n. In such a case, care must be taken in practice to identify an appropriate minimizer of $\hat{V}(\psi_{y_0, a_0}, F)$ by a specified algorithm, to ensure that this minimizer is a consistent estimator of a minimizer of $V(\psi_{y_0, a_0}, F)$. (Compare with the problem of solving (1.2) in practice — it has infinitely many solutions because ψ vanishes outside an interval, and hence we chose a solution (2.1) which is a consistent estimator of θ, stated in Theorem 2.1.)

Our last Theorem gives the asymptotic properties of $\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}$, the estimator (2.1) with the estimator $(\hat{y}_0^*, \hat{a}_0^*)'$, satisfying (2.24), as our choice for the vector $(y_0, a_0)'$.

Theorem 2.3. Under the same assumptions as in Theorem 2.2, and (A4), $n^{1/2}(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*} - \theta)$ converges in distribution to the multivariate normal distribution with mean 0 and covariance matrix $C_0^{-1} V(\psi_{y_0^*, a_0^*}, F)$.

Proof of Theorem 2.3: By Theorem 2.1, with the fixed (but arbitrary) (y_0, a_0) of (2.1) replaced by the fixed (y_0^*, a_0^*) of (2.23), we have

$$\hat{\theta}_{y_0^*, a_0^*} \xrightarrow[\mathscr{P}]{} \theta_0 \quad \text{(the true value of } \theta) \tag{2.33}$$

and by Theorem 2.2 we have

$$(\hat{y}_0^*, \hat{a}_0^*)' \xrightarrow[\mathscr{P}]{} (y_0^*, a_0^*)' \ . \tag{2.34}$$

By continuity of ψ on the compact set $[-a_0 + \omega, a_0 - \omega]$, and assumption A2), it follows from (2.34) that

$$|\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*} - \hat{\theta}_{y_0^*, a_0^*}| \xrightarrow[\mathscr{P}]{} 0 \ . \tag{2.35}$$

As in (2.18) we write, for some numbers γ_i, $0 \leq \gamma_i \leq 1$, $i = 1, \ldots, n$,

$$n^{1/2}(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*} - \theta_0) \left(\frac{1}{n} \sum c_i c_i' \psi' \left[X_i - c_i' \theta_0 + \gamma_i c_i'(\theta_0 - \hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) \right] \right)$$

$$= -n^{1/2}[H_n^*(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) - H_n(0)] + n^{1/2}[H_n^*(\theta_0) - H_n(0)] \ . \quad (2.36)$$

Now in (2.36), we claim first that

$$-n^{1/2}[H_n^*(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) - H_n(0)] \xrightarrow[\mathscr{P}]{} 0 \ . \quad (2.37)$$

To prove (2.37), fix $\varepsilon > 0$ and, in analogy with the proof of (2.19), observe that for each n,

$$P_F(n^{1/2}[H_n^*(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) - H_n(0)] < \varepsilon) \geq P_F(H_n^*(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) - H_n(0) = 0)$$

$$= P_F([H_n^*(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) - H_n^*(\hat{\theta}_{y_0^*, a_0^*})] + [H_n^*(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) - H_n(0)] = 0)$$

$$\to 1 \ , \quad \text{as} \quad n \to \infty \ , \quad \text{because}$$

$$|H_n^*(\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}) - H_n^*(\hat{\theta}_{y_0^*, a_0^*})| \xrightarrow[\mathscr{P}]{} 0 \quad (2.38)$$

by (2.35) and uniform continuity of H_n^* on compact subsets of R^p, while

$$|H_n^*(\hat{\theta}_{y_0^*, a_0^*}) - H_n(0)| \xrightarrow[\mathscr{P}]{} 0 \quad (2.39)$$

by (2.19) with the fixed (but arbitrary) (y_0, a_0) of (2.1) replaced by the fixed (y_0^*, a_0^*) of (2.23). This proves (2.37).

Next, in (2.36) we know from (2.20) that

$$n^{1/2}[H_n^*(\theta_0) - H_n(0)] \xrightarrow[\mathscr{D}]{} N(0, C_0 \int \psi^2 f dx) \ . \quad (2.40)$$

Finally, in (2.36) we see, in analogy with (2.21), that

$$\frac{1}{n} \sum c_i c_i' \psi' [X_i - c_i' \theta_0 + \gamma_i c_i'(\theta_0 - \hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*})] \xrightarrow[\mathscr{P}]{} C_0 \int \psi'(x) f(x) dx \quad (2.41)$$

by A 1), A 2), piecewise uniform continuity of ψ' on $(-a_0 + \omega, a_0 - \omega)$, (2.35) and (2.33).

The proof of Theorem 2.3 is now completed upon inserting the results (2.37), (2.40) and (2.41) into (2.36) and applying Slutsky's Theorem. $\quad\square$

We remark that if one wishes to use Theorem 2.3 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 $\hat{V}(\psi_{\hat{y}_0^*, \hat{a}_0^*}, F)$, which can be computed without knowledge of 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 $(\hat{y}_0^*, \hat{a}_0^*)'$ from (2.24), consider using a linear ξ. The ψ-function used in solving (1.2) would then be (with $\omega = 10^{-6}$)

$$
\psi(y) = \begin{cases}
y\,, & |y| \leq \hat{y}_0^* \\[2mm]
\dfrac{\hat{y}_0^*}{\hat{a}_0^* - 10^{-6} - \hat{y}_0^*}(\hat{a}_0^* - 10^{-6} - |y|)\,, & \hat{y}_0^* \leq |y| \leq \hat{a}_0^* - 10^{-6} \\[2mm]
0\,, & |y| \geq \hat{a}_0^* - 10^{-6}\,.
\end{cases}
$$

In any case, whatever the choice of ξ, subject to its properties given in Sect. 1, the resulting estimator $\hat{\theta}_{\hat{y}_0^*, \hat{a}_0^*}$ is asymptotically normal, and optimal in the sense that it has minimum asymptotic variance among all solutions of (1.2) that are based on ψ-functions of the form (1.4).

In conclusion, we remark that an alternative procedure to the one of this section is to replace $\bar{\theta}$ in (2.22) by the unknown θ and then choose a value for $(y_0, a_0)'$ that minimizes (2.22) subject to (1.2) holding. We have not examined the theoretical properties of this procedure, which is an analogue of Huber's "proposal 3" (Huber 1964).

References

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

Collins JR (1976) Robust Estimation of a Location Parameter in the Presence of Asymmetry. Annals of Statistics 4:68–85

Collins JR, Wiens DP (1985) Minimax variance M-estimators in ε-contamination models. Annals of Statistics 3:1078–1096

Collins JR, Sheahan JN, Zheng Z (1986) Robust Estimation in the Linear Model with Asymmetric Error Distributions. Journal of Multivariate Analysis 20:220–243

Hampel FR (1973) Robust Estimation: A Condensed Partial Survey. Z Wahrscheinlichkeitstheorie und Verw Gebiete 27:87–104

Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust Statistics. John Wiley and Sons, New York

Hlynka M, Sheahan JN, Wiens DP (1989) On Deleting Influential Responses Under a Model of Asymmetric Errors in Regression. University of Alberta, Dept of Statistics, Technical Report No 89.25, Edmonton, Canada

Hogg RV (1979) An Introduction to Robust Estimation. In: Launer R, Wilkinson G (eds) Robustness in Statistics. Academic Press, New York, pp 1–29

Huber PJ (1964) Robust Estimation of a Location Parameter. Annals of Mathematical Statistics 35:73–101

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

Jaeckel LA (1971) Some Flexible Estimates of Location. Annals of Mathematical Statistics 42:1540–1552

Lind JC (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

Lind JC, Mehra KL, Sheahan JN (1989) Asymmetric Errors in Linear Models... Estimation-Theory and Monte Carlo. Univ of Alberta, Dept of Statistics, Technical Report No 88.20, Edmonton, Canada

Ortega JM, Rheinboldt WC (1970) Iterative Solution of Non-Linear Equations in Several Variables. Academic Press, New York

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

Yohai VJ (1974) Robust Estimation in the Linear Model. Annals of Statistics 2:562–567