*The Annals of Statistics*
1985, Vol. 13, No. 4, 1490–1497

# ASYMPTOTIC BEHAVIOR OF ROBUST ESTIMATORS OF REGRESSION AND SCALE PARAMETERS WITH FIXED CARRIERS

BY MERVYN J. SILVAPULLÉ

*University of Melbourne and Bureau of Agricultural Economics, Canberra*

For the linear regression model, $y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$ with fixed $\mathbf{x}_i$s, the asymptotic normality of $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$ which minimizes the Huber–Dutter loss function, $\sum \sigma\rho\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\} + A_n\sigma$, is established under rather general conditions.

**1. Introduction.** Consider the linear regression model $y_i = \mathbf{x}_i\boldsymbol{\beta}_0 + \varepsilon_i$, where the $\mathbf{x}_i = (1, x_{i2}, \dots, x_{ip})$, $i = 1, \dots, n$ are known, $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{0p})$ is a vector of $p$ unknown parameters to be estimated and $\mathbf{x}_i\boldsymbol{\beta}_0$ is the usual inner product. The errors, $\varepsilon_i$, $i = 1, \dots, n$ are assumed to be independent and identically distributed with common distribution function, $G$.

A number of robust estimators of $\boldsymbol{\beta}_0$ have been proposed and investigated over the last decade. Perhaps the best-known class of these are the so called $M$ estimators. They are obtained by minimizing a loss function of the form $\sum\rho\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/s\}$ where $\rho$ is some suitably chosen function and $s$ is some estimate of the scale of the errors $\varepsilon_i$. In most practical applications of $M$ estimation [e.g., Andrews (1974)], the scale estimate $s$ is chosen so that it is insensitive to large errors in the data. The loss function is then minimized with $s$ held fixed. This is usually called *fixed-scale estimation.*

An alternative, perhaps more elegant, approach has been suggested by Huber and Dutter (1974) and Huber (1977). In this approach, $\boldsymbol{\beta}_0$ and a scale for the errors are estimated simultaneously by minimizing

$$Q(\boldsymbol{\beta}, \sigma) = \sum \rho\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\}\sigma + A_n\sigma,$$

where $\rho \geq 0$ is convex, $\rho(0) = 0$, $|t|^{-1}\rho(t) \to k$ as $|t| \to \infty$ for some $k > 0$, and $\{A_n\}$ is a suitably chosen sequence of constants. We shall call a point $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$ at which $Q(\boldsymbol{\beta}, \sigma)$ attains its minimum, an HD estimator (HD for Huber–Dutter).

The purpose of this paper is to establish that $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$ is asymptotically normal under rather general conditions. Similar results for the case of random carriers are established in Maronna and Yohai (1981).

**2. Preliminaries.** Let us first introduce some notation: $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$; $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}})$; $\Theta = \{\boldsymbol{\theta}\colon\ \boldsymbol{\beta} \in R^p, \sigma > 0\}$ where $R^p$ is $p$-dimensional Euclidean space; $x = (\mathbf{x}_1^t \dots, \mathbf{x}_n^t)^t$ with the superscript $t$ denoting the transpose; for any set $C$, $\overline{C}$ is

Received August 1983; revised October 1984.

*AMS* 1980 *subject classifications*. Primary 62J05, 62F35; secondary 62F11, 62F12.

*Key words and phrases*. Linear regression, robust estimation, $M$ estimator, simultaneous scale estimation.

the closure of $C$; for an arbitrary function $f$, $f'$ and $f''$ are the first and second derivatives of $f$, respectively. $\|\mathbf{x}\|$ is the Euclidean norm of $\mathbf{x}$.

The estimating equations for $(\boldsymbol{\beta}, \sigma)$ may be written as

$$\sum \psi\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\}\mathbf{x}_i = 0, \qquad \sum \chi\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\} = A_n, \tag{2.1}$$

where $\psi = \rho'$ and $\chi(t) = t\psi(t) - \rho(t) = \int_0^t x\, d\psi(x)$.

In what follows, it will be assumed that the errors are nondegenerate, $n \geq p + 1$, $X$ is of full rank, $A_n > 0$ and $\psi''$ is continuous.

Since we are interested only on estimators having some robustness properties, it will be assumed that $\lim |t|^{-1}\rho(t) = k < \infty$, since $k = \infty$ leads to an unbounded influence curve for $\hat{\boldsymbol{\beta}}$. Without loss of generality, we may assume $k = 1$. *Therefore $\psi$ is bounded and increases from $-1$ to $+1$. It will also be assumed that $\chi$ is bounded*; otherwise the influence curve of $\hat{\sigma}$ is unbounded. Although nonrobustness of $\hat{\sigma}$ itself may not be of much independent interest, we have to ensure that it is robust since $\hat{\boldsymbol{\beta}}$ depends on $\hat{\sigma}$.

In fixed-scale estimation, one usually centers the error by $E\psi(\varepsilon/s_0) = 0$, where $s_0$ is the asymptotic value of the estimator of scale, $s$, under some ideal model, for example the Gaussian errors (see Huber, 1973 and Bickel, 1975). Here we shall center the errors by $E\psi(\varepsilon/\sigma^*) = 0$, where $\sigma^*$ is defined in the proposition below, which may be deduced from the results on pages 138 and 139 of Huber (1981).

PROPOSITION 1. *Suppose that $\eta < 1 - A\nu^{-1}$ and $A > 0$, where $\eta$ is the largest jump in the error distribution, $\nu = \min\{\chi(\infty), \chi(-\infty)\}$ and $A = \lim n^{-1}A_n$. Then the equation, $E[\psi\{(\varepsilon - \mu)/\sigma\}, \chi\{(\varepsilon - \mu)/\sigma\} - A] = 0$, has a solution at $(\mu^*, \sigma^*)$ with $\sigma^* > 0$.*

We close this section with a result which gives a lower bound on sample size to ensure that $\hat{\sigma} > 0$. This result is essentially the same as that of Huber [see Huber (1981), page 189] except that we do not require $\chi$ to be symmetric.

PROPOSITION 2. *Let $p'$ be the maximum number of residuals that may be made simultaneously zero. Then $\hat{\sigma} > 0$ whenever $(n - p') > \nu^{-1}A_n$ where $\nu = \min\{\chi(\infty), \chi(-\infty)\}$.*

PROOF.

$$\begin{aligned}\lim_{\sigma \to 0+} (\partial/\partial\sigma)Q(\hat{\boldsymbol{\beta}}, \sigma) &= n^{-1}A_n - n^{-1}\sum \lim_{\sigma \to 0+} \chi\{(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})/\sigma\} \\ &= n^{-1}A_n - n^{-1}\chi(\infty)\#(y_i > \mathbf{x}_i\hat{\boldsymbol{\beta}}) \\ &\quad - n^{-1}\chi(-\infty)\#(y_i < \mathbf{x}_i\hat{\boldsymbol{\beta}}),\end{aligned}$$

where # is the abbreviation for the number of elements. The rest of the proof is rather easy. □

**3. Asymptotic normality of $\hat{\boldsymbol{\theta}}$.** In this section we prove that $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$ is approximately normal for large $n$. The presentation of the proof is facilitated by a linearization result similar to that in Brown and Kildea (1979). Let us first introduce some notation: $\boldsymbol{\theta}^* = (\boldsymbol{\beta}_0, \sigma^*)$;

$$\mathbf{U}_i(\boldsymbol{\theta}) = -(\partial/\partial\boldsymbol{\theta})\left[\sigma\rho\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\} + n^{-1}A_n\sigma\right]$$
$$= \left[\mathbf{x}_i\psi\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\}, \chi\{(y_i - \mathbf{x}_i\boldsymbol{\beta})/\sigma\} - n^{-1}A_n\right]$$
$$\mathbf{B}_n(\boldsymbol{\theta}) = n^{-1}E\{(\partial^2/\partial\boldsymbol{\theta}^t\partial\boldsymbol{\theta})Q(\boldsymbol{\theta})\}; \qquad \mathbf{T}_n(\boldsymbol{\theta}) = -n^{-1/2}\sum\mathbf{U}_i(\boldsymbol{\theta});$$
$$\mathbf{V}_n(\boldsymbol{\theta}) = \operatorname{var}\{\mathbf{T}_n(\boldsymbol{\theta})\};$$

*For the rest of this paper we will assume that the following conditions are satisfied*: (i) $\lim(n^{-1}A_n - A) = 0$ for some $0 < n^{-1}A_n$, $A < \nu$ where $\nu = \min\{\chi(\infty), \chi(-\infty)\}$, and $\sigma^* > 0$; (ii) $x^2\psi'(x)$ and $x^3\psi''(x)$ are bounded; (iii) the eigenvalues of $n^{-1}X^tX$ are bounded above and away from zero, and $n^{-1}\max\{\|\mathbf{x}_i\|^2\colon 1 \le i \le n\} \to 0$; and (iv) $a, ab - c^2, u, uv - w^2 > 0$, where $a = E\{\psi'(\eta)\}$, $b = E\{\eta^2\psi'(\eta)\}$, $c = E\{\eta\psi'(\eta)\}$, $u = \operatorname{var}\{\psi(\eta)\}$, $v = \operatorname{var}\{\chi(\eta)\}$, and $w = \operatorname{cov}\{\psi(\eta), \chi(\eta)\}$ with $\eta = (\varepsilon/\sigma^*)$.

Now, the linearization result is the following:

THEOREM 1. $\mathbf{T}_n(\boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{\gamma}) = \mathbf{T}_n(\boldsymbol{\theta}^*) + \boldsymbol{\gamma}\mathbf{B}_n(\boldsymbol{\theta}^*) + p_n(\boldsymbol{\gamma})$, *where* $\sup\{\|p_n(\boldsymbol{\gamma})\|\colon \|\boldsymbol{\gamma}\| < K\} \to 0$ *a.s. as* $n \to \infty$ *for any* $K > 0$.

PROOF. Note that $\mathbf{T}_n(\boldsymbol{\theta}) = n^{-1/2}(\partial/\partial\boldsymbol{\theta})Q(\boldsymbol{\theta})$. Expanding $(\partial/\partial\theta_l)Q(\boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{\gamma})$ about $\boldsymbol{\gamma} = \mathbf{0}$, we have

$$(\partial/\partial\theta_l)Q(\boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{\gamma})$$
$$= (\partial/\partial\theta_l)Q(\boldsymbol{\theta}^*) + n^{-1/2}\boldsymbol{\gamma}(\partial/\partial\boldsymbol{\theta}^t)(\partial/\partial\theta_l)Q(\boldsymbol{\theta}^*) + n^{-1}R_{nl}(\boldsymbol{\gamma}, \lambda_l),$$

where $\mathbf{R}_{nl}(\boldsymbol{\gamma}, \lambda) = \boldsymbol{\gamma}(\partial^2/\partial\boldsymbol{\theta}^t\partial\boldsymbol{\theta})(\partial/\partial\theta_l)Q(\boldsymbol{\theta}^* + n^{-1/2}\lambda\boldsymbol{\gamma})\boldsymbol{\gamma}^t$, and $0 < \lambda_l < 1$. This expansion may be rewritten as

$$\mathbf{T}_n(\boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{\gamma}) = \mathbf{T}_n(\boldsymbol{\theta}^*) + \boldsymbol{\gamma}n^{-1}(\partial^2/\partial\boldsymbol{\theta}^t\partial\boldsymbol{\theta})Q(\boldsymbol{\theta}^*) + n^{-3/2}\mathbf{R}_n(\boldsymbol{\gamma}, \boldsymbol{\lambda}).$$

The rest of the proof follows from the next two lemmas, the proofs of which appear in Appendix B.

LEMMA 1. $\{n^{-1}(\partial^2/\partial\boldsymbol{\theta}^t\partial\boldsymbol{\theta})Q(\boldsymbol{\theta}^*) - \mathbf{B}_n(\boldsymbol{\theta}^*)\} \to_P 0$ *as* $n \to \infty$.

LEMMA 2. *For any* $K > 0$ *and* $l = 1, \dots, p + 1$

$$\sup\left\{n^{-3/2}|R_{nl}(\boldsymbol{\gamma}, \lambda)|\colon \|\boldsymbol{\gamma}\| \le K \text{ and } 0 < \lambda < 1\right\} \to 0 \text{ a.s. as } n \to \infty.$$

We now use this theorem to establish that $\hat{\boldsymbol{\theta}}$ is asymptotically normal.

THEOREM 2. *Writing* $\mathbf{D}_n = \{V_n(\boldsymbol{\theta}^*)\}^{-1/2}\mathbf{B}_n(\boldsymbol{\theta}^*)$, *we have* $\mathbf{D}_n n^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \to N(0, \mathbf{I})$ *as* $n \to \infty$. *Furthermore, the eigenvalues of* $\mathbf{D}_n$ *are bounded above and away from zero.*

PROOF. To prove this theorem, we need the following two lemmas; the proofs are given in Appendix B.

LEMMA 3. *The eigenvalues of* $\mathbf{B}_n(\boldsymbol{\theta}^*)$ *and* $\mathbf{V}_n(\boldsymbol{\theta}^*)$ *are bounded above and away from zero.*

LEMMA 4. $\{\mathbf{V}_n(\boldsymbol{\theta}^*)\}^{-1/2}\mathbf{T}_n(\boldsymbol{\theta}^*) \to_d N(\mathbf{0}, \mathbf{I})$.

Now, integrating the expansion in Theorem 2, we have

$$n^{1/2}\{Q(\boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{\gamma}) - Q(\boldsymbol{\theta}^*)\} = \mathbf{T}_n(\boldsymbol{\theta}^*)\boldsymbol{\gamma}^t + 2^{-1}\boldsymbol{\gamma}\mathbf{B}_n(\boldsymbol{\theta}^*)\boldsymbol{\gamma}^t + \varepsilon_n(\boldsymbol{\gamma}), \tag{3.1}$$

where $\varepsilon_n(\boldsymbol{\gamma}) = \|\boldsymbol{\gamma}\| o_p(1)$ on bounded sets. Let the right hand side of (3.1) be denoted by $S_n(\boldsymbol{\gamma})$. For any $K > 0$, define $\mathbf{d}_n = \mathbf{T}_n(\boldsymbol{\theta}^*)I\{\|\mathbf{T}_n(\boldsymbol{\theta}^*)\| \leq K\}$, where $I$ is the indicator function. Let $s_n(\boldsymbol{\gamma}) = \mathbf{d}_n\boldsymbol{\gamma}^t + 2^{-1}\boldsymbol{\gamma}\mathbf{B}_n(\boldsymbol{\theta}^*)\boldsymbol{\gamma}^t$.

For a given $\delta > 0$, let $C_n = \{\boldsymbol{\gamma}: s_n(\boldsymbol{\gamma}) \leq 2\delta\}$. Since the eigenvalues of $\mathbf{B}_n(\boldsymbol{\theta}^*)$ are bounded above and away from zero and $\mathbf{d}_n$ is bounded, there exists $r = r(K, \delta)$ such that $\sup\{\|\boldsymbol{\gamma}\|: \boldsymbol{\gamma} \in C_n\} \leq r$ for every $n$. Clearly, $s_n(\boldsymbol{\gamma})$ is minimized at $\boldsymbol{\gamma}_{0n} = -\mathbf{d}_n\{\mathbf{B}_n(\boldsymbol{\theta}^*)\}^{-1}$; further, $\boldsymbol{\gamma}_{0n} \in C_n$ since $s_n(\boldsymbol{\gamma}_{0n}) = -2\mathbf{d}_n\mathbf{B}_n(\boldsymbol{\theta}^*)\mathbf{d}_n^t < 0$.

The following arguments hold with probability arbitrarily close to 1 for sufficiently large $K$ and $n$:

Since $\mathbf{T}_n(\boldsymbol{\theta}^*) = \mathbf{d}_n$ and $|\varepsilon_n(\boldsymbol{\gamma})| < \delta/2$, we have

$$\sup\{|S_n(\boldsymbol{\gamma}) - s_n(\boldsymbol{\gamma})|: \|\boldsymbol{\gamma}\| \leq r + 1\} < \delta/2. \tag{3.2}$$

Now,

$$\begin{aligned}
\inf\{S_n(\boldsymbol{\gamma}): \|\boldsymbol{\gamma}\| = r + 1\} &> \inf\{s_n(\boldsymbol{\gamma}): \|\boldsymbol{\gamma}\| = r + 1\} - \delta/2 \quad \text{by (3.2)} \\
&> 3\delta/2 \quad \text{since } C_n \cap \{\boldsymbol{\gamma}: \|\boldsymbol{\gamma}\| = r + 1\} \text{ is empty} \\
&> s_n(\boldsymbol{\gamma}_{0n}) + 3\delta/2 \quad \text{since } s_n(\boldsymbol{\gamma}_{0n}) < 0 \\
&> S_n(\boldsymbol{\gamma}_{0n}) + \delta \quad \text{by (3.2)}.
\end{aligned}$$

So, for the convex function, $S_n(\boldsymbol{\gamma})$, the infimum on the circle $\|\boldsymbol{\gamma}\| = r + 1$ is larger than its value at a point inside the circle. Therefore, $S_n(\boldsymbol{\gamma})$ attains its global minimum on $\|\boldsymbol{\gamma}\| < r + 1$.

Hence, we conclude that $\hat{\boldsymbol{\gamma}} = O_p(1)$. The desired result follows by substituting $\hat{\boldsymbol{\gamma}}$ into the expansion in Theorem 2. □

**4. Some results on the asymptotic behavior of $\hat{\boldsymbol{\theta}}$.** A number of useful asymptotic properties of the HD estimators may be inferred from their asymptotic covariance matrices. Let us write the regression model as $y_i = \alpha + \boldsymbol{\beta}_{(2)}\mathbf{z}_i + \varepsilon_i$, where $\boldsymbol{\beta}_{(2)} = (\beta_{02}, \dots, \beta_{0p})$, $\alpha = \beta_{01} + \mathbf{z}_i\boldsymbol{\beta}_{(2)}$ and $z_{ij} = x_{ij} - \bar{x}_{.j}$ for $j = 2, \dots, p$ and $i = 1, \dots, n$. Then, the matrix representation of the regression equation takes the form $\mathbf{Y} = \alpha\mathbf{1} + Z\boldsymbol{\beta}_{(2)} + \boldsymbol{\varepsilon}$ where $\mathbf{1}$ is a column of 1s and $\mathbf{Z} = (z_1^t, \dots, z_n^t)^t$.

Now, with $a$, $b$, $c$, $u$, $v$ and $w$ defined as in condition (iv) at the outset of Section 3, we have

PROPOSITION 3. *$n^{1/2}\{(\hat{\alpha}, \hat{\boldsymbol{\beta}}_{(2)}, \hat{\sigma}) - (\alpha, \boldsymbol{\beta}_{(2)}\sigma)\}$ is asymptotically normal with mean zero and covariance*

$$(\sigma^*)^2 \begin{bmatrix} h_1 & 0 & h_4 \\ 0 & nh_2(\mathbf{Z}'\mathbf{Z})^{-1} & 0 \\ h_4 & 0 & h_3 \end{bmatrix},$$

*where* $h_1 = (ab - c^2)^{-2}(b^2u - 2bcw + c^2v)$, $h_2 = (u/a^2)$, $h_3 = (ab - c^2)^{-2}$ $(a^2v - 2acw + c^2u)$, *and* $h_4 = (ab - c^2)^{-2}(abw - bcu - acv + c^2w)$.

PROOF. (An outline only.) The proof of the asymptotic normality is the same as that of Theorem 2 with only minor modification; hence it is not given here. The following matrix result is easily verified:

$$\begin{bmatrix} p_1 & 0 & p_4 \\ 0 & p_2 & 0 \\ p_4 & 0 & p_3 \end{bmatrix}^{-1} = \begin{bmatrix} p_3(p_1p_3 - p_4^2)^{-1} & 0 & -p_4(p_1p_3 - p_4^2)^{-1} \\ 0 & p_2^{-1} & 0 \\ -p_4(p_1p_3 - p_4^2)^{-1} & 0 & p_1(p_1p_3 - p_4^2)^{-1} \end{bmatrix}.$$

The product of two matrices of the above form is also of the same form. The $B$ and $V$ matrices corresponding to the $(\alpha, \boldsymbol{\beta}_{(2)})$ parametrization are of the above form. Now, it is only a simple matter to evaluate the covariance matrix of $(\hat{\alpha}, \hat{\boldsymbol{\beta}}_{(2)}, \hat{\sigma})$. □

It follows from the above proposition that (i) $\hat{\beta}_2, \ldots, \hat{\beta}_p$ are asymptotically independent of $(\hat{\alpha}, \hat{\sigma})$ and (ii) the asymptotic behavior of $(\hat{\alpha}, \hat{\sigma})$ and that of $(\hat{\mu}, \hat{\sigma})$, the HD estimator of $(\mu, \sigma)$ in the location/scale model $y_i = \mu + \varepsilon_i$, are the same. Some properties of $h_1$ and $h_3$, when the error distribution is asymmetrically contaminated normal, are discussed in Heathcote and Silvapulle (1981).

Although the results established here hold under rather general conditions, the immediate applications of this method are likely to be when the errors are symmetric, in which case, it is natural to choose $\rho$ symmetric. So, let us assume that $\rho$ and errors are symmetric. Then $c = w = 0$, and the asymptotic covariance of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is

$$\begin{bmatrix} h_2 n(\mathbf{X}'\mathbf{X})^{-1} & 0 \\ 0 & h_5 \end{bmatrix},$$

where $h_5 = (v/b^2)$. Thus, it follows that the robustness properties of the HD estimators of location and scale, with respect to least-squares estimators, carry over to the regression model.

**5. Discussion.** In fixed-scale estimation, the choice of scale $s$ represents some minor problems which translate to the choice of $n^{-1}A_n$ in HD estimation. As $s$ ranges from 0 to $\infty$ the fixed-scale estimator of $\beta$ covers the wide spectrum from the extremely robust minimum absolute deviation estimator to the extremely nonrobust least-squares estimator. Similarly, as $n^{-1}A_n$ range from $\max\{\chi(\infty), \chi(-\infty)\}$ to 0, the HD estimator of $\beta$ also covers the same spectrum. Of course, this does not mean that for a given estimator of scale $s$ we can choose $n^{-1}A_n$ so that the two estimators are equivalent. It seems that fixed-scale and HD estimators are unlikely to be equivalent except in very special cases.

Intuitively, a logical choice for $A_n$ seems to be $(n-p)E(\chi)$, where the expectation is taken with respect to a distribution function "close" to the error distribution. Huber (1977) favors this when the error is approximately normal. On the other hand, the result in Heathcote and Silvapulle (1981) may be useful. These authors obtained the HD-estimating equation for location and scale with $A_n = nE\psi(\chi)$ by minimizing a loss function similar to the well known Cramér–von Mises criterion, where $E_\psi$ is the expectation with respect to the distribution $(\psi+1)/2$. Thus, $A_n = nE_\psi(\chi)$ may be a potential choice for the regression model as well. Our limited experience with different data sets and the choices $\psi = \{2(\text{normal})-1\}$, $\{2(\text{logistic})-1\}$ and $\{2(\text{uniform})-1\}$ with $A_n = nE_\psi(\chi)$ has been that the fixed-scale and HD methods lead to rather close results. This is not surprising since, under symmetry, the asymptotic variance of the fixed-scale estimator of the regression parameter is the same as that of the HD estimator with $p\lim s = \sigma^*$.

A point worth noting is on the computation of $(\hat{\boldsymbol\beta}, \hat\sigma)$. The algorithm of Huber and Dutter (1974) performed very well in a number of different situations. We found that (although not unexpected) whenever the starting value is far away from $(\hat{\boldsymbol\beta}, \hat\sigma)$ the iterates moved towards $(\hat{\boldsymbol\beta}, \hat\sigma)$ very quickly, thus the choice of starting values was not all that crucial for the convergence of the algorithm.

## Appendix: Proofs of the lemmas in Section 3

PROOF OF LEMMA 1. For sufficiently large $n$ and $1 \le k, l, m \le p$, we have

$$\begin{aligned}
&E\{(\partial^2/\partial\theta_k\partial\theta_l)Q(\boldsymbol\theta) - E(\partial^2/\partial\theta_k\partial\theta_l)Q(\boldsymbol\theta)\}^2\\
&\quad= n^{-2}\sigma^{-2}\sum x_{ik}^2x_{il}^2E\left[\psi\{(y_i-\mathbf{x}_i\boldsymbol\beta)/\sigma\} - E\psi\{(y_i-\mathbf{x}_i\boldsymbol\beta)/\sigma\}\right]^2\\
&\quad\le M\left\{n^{-1}\sum\|\mathbf{x}_i\|^2\right\}n^{-1}\max\left\{\|\mathbf{x}_i\|^2: 1\le i\le n\right\}, \quad \text{for some } M>0\\
&\quad\to 0 \quad \text{as } n\to\infty.
\end{aligned}$$

Similarly, it can be shown that the same holds for $1 \le k, l, m \le p+1$.

PROOF OF LEMMA 2. Let $\bar{\boldsymbol\theta} = \boldsymbol\theta^* + n^{-1/2}\lambda\boldsymbol\gamma$, where $0 \le \lambda \le 1$ and $1 \le k, l, m \le p$. Then

$$\begin{aligned}
&\sup\left\{n^{-3/2}|\gamma_k(\partial^3/\partial\theta_k\partial\theta_l\partial\theta_m)Q(\bar{\boldsymbol\theta})\gamma_l|: \|\boldsymbol\gamma\| \le K, 0\le\lambda\le 1\right\}\\
&\quad\le Mn^{-1/2}\max\{\|\mathbf{x}_i\|: 1\le i\le n\}n^{-1}\sum\|\mathbf{x}_i\|^2, \quad \text{for some } M>0\\
&\quad\to 0 \quad \text{as } n\to\infty.
\end{aligned}$$

Similarly, it can be shown that the same holds for $1 \le k, l, m \le p+1$.

PROOF OF LEMMA 3. Since the first column of $(X^tX)$ is $\sum \mathbf{x}_i$,

$$B_n(\boldsymbol{\theta}^*) = \begin{bmatrix} an^{-1}X^tX & \vdots & \\ \cdots\cdots & \vdots & \cdots \\ cn^{-1}\sum \mathbf{x}_i & \vdots & b \end{bmatrix} = P^t \begin{bmatrix} an^{-1}X^tX & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots\cdots \\ 0 & \vdots & a^{-1}(ab - c^2) \end{bmatrix} P$$

where $P$ is the same as the identity matrix except that the $(1, p+1)$th element is $-ca^{-1}$.

Since $ab - c^2 > 0$, the determinant of $B_n(\boldsymbol{\theta}^*)$ is bounded above and away from zero, hence so are its eigenvalues.

It follows from the definition that $V_n(\theta^*)$ is the same as $B_n(\theta^*)$ with $u$, $v_n$, and $w_n$ in place of $a$, $b$, and $c$, respectively. Now, by arguments similar to those above, it may be verified that the eigenvalues of $V_n(\boldsymbol{\theta}^*)$ are also bounded above and away from zero.

PROOF OF LEMMA 4. Let $\boldsymbol{\lambda} \in R^{p+1}$ and $\sigma > 0$. Since $\psi$, $\chi$, and $n^{-1}A_n$ are bounded, there exists $M > 0$ such that

$$\|U_i(\boldsymbol{\theta}^*) - E\{U_i(\boldsymbol{\theta}^*)\}\|^2 \le M\{1 + \|\mathbf{x}_i\|^2\}, \qquad 1 \le i \le n.$$

Since the eigenvalues of $V_n(\boldsymbol{\theta}^*)$ are bounded away from zero, there exists $K > 0$ such that, as $n \to \infty$

$$|\boldsymbol{\lambda}^t\{V_n(\boldsymbol{\theta}^*)\}^{-1/2}[U_i(\boldsymbol{\theta}^*) - E\{U_i(\boldsymbol{\theta}^*)\}]|^2 \le KM\|\boldsymbol{\lambda}\|^2\{1 + \|\mathbf{x}_i\|^2\} \to 0$$

uniformly on $1 \le i \le n$ since $n^{-1}\max\{\|\mathbf{x}_i\|^2 : 1 \le i \le n\} \to 0$. Therefore,

$$\text{prob}\left[|\boldsymbol{\lambda}^t\{V_n(\boldsymbol{\theta}^*)\}^{-1/2}[U_i(\boldsymbol{\theta}^*) - E\{U_i(\boldsymbol{\theta}^*)\}]| > \sigma n^{1/2}\right] = 0, \qquad 1 \le i \le n$$

for sufficiently large $n$, and hence

$$\{V_n(\boldsymbol{\theta}^*)\}^{-1/2}[T_n(\boldsymbol{\theta}^*) - E\{T_n(\boldsymbol{\theta}^*)\}] \to_d N(\mathbf{0}, I).$$

Now, it suffices to establish that $E\{T_n(\boldsymbol{\theta}^*)\} \to 0$. To prove this, let $(\mu_n, \sigma_n)$ be the unique solution of

$$E\left[\psi\{(\varepsilon - \mu)/\sigma\}, \chi\{(\varepsilon - \mu)/\sigma\} - n^{-1}A_n\right] = 0, \sigma \ge 0.$$

The existence and uniqueness of $(\mu_n, \sigma_n)$ follow by replacing $A$ by $n^{-1}A_n$ in Proposition 1. Applying the mean value theorem to the l.h.s. in the above equation, it may be verified that $n^{1/2}\mu_n$ and $n^{1/2}(\sigma_n - \sigma^*)$ converge to zero. This may be written as $n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*) \to 0$, where $\boldsymbol{\theta}_n = (\beta_{01} + \mu_n, \beta_{02}, \dots, \beta_{0p}, \sigma_n)$.

Since the equations $\mathbf{x}_i(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \mu_n$, $1 \le i \le n$ have a unique solution at $\boldsymbol{\beta} = (\beta_{01} + \mu_n, \beta_{02}, \dots, \beta_{0p})$, $E\{U_i(\boldsymbol{\theta}_n)\} = 0$ for $1 \le i \le n$ and hence $E\{T_n(\boldsymbol{\theta}_n)\} = 0$.

Now, expanding $(\partial/\partial\theta_l)Q(\boldsymbol{\theta}_n)$ about $\boldsymbol{\theta}^*$ as in the proof of Theorem 2 and taking expectation, we have

$$0 = E\{T_n(\boldsymbol{\theta}_n)\} = E\{T_n(\boldsymbol{\theta}^*)\} + n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)B_n(\boldsymbol{\theta}^*)$$
$$+ n^{-3/2}E\left[R_n\{n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*), \boldsymbol{\lambda}\}\right].$$

By Lemmas 2 and 3 we have $E\{T_n(\boldsymbol{\theta}^*)\} \to 0$.

**Acknowledgment.** I am grateful to Professor C. R. Heathcote for his assistance in the preparation of this article. Professor C. C. Heyde, Drs. R. L. Chambers, T. K. Donaldson, P. Hall, and T. J. O'Neill have also helped me at various stages. Most of the research leading to this article was carried out at the Australian National University.

## REFERENCES

ANDREWS, D. F. (1974). A robust method for multiple regression, *Technometrics* **16** 523–531.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–343.

BROWN, B. M. and KILDEA, D. G. (1979). Outlier detection tests and robust estimators based on signs of residuals. *Comm. Statist. A*—Theory Methods **9** 257–269.

HEATHCOTE, C. R. and SILVAPULLE, M. J. (1981). Minimum mean squared estimation of location and scale parameters under misspecification of the model. *Biometrika* **68** 501–514.

HUBER, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

HUBER, P. J. (1977). *Robust Statistical Procedures.* Regional Conference Series in Applied Mathematics No. 27, Soc. Industr. Appl. Math. Philadelphia, Penn.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

HUBER, P. J. and DUTTER, R. (1974). Numerical solutions of robust regression problems. In: *COMPSTAT 1974* (G. Brickmann, ed.) 165–172. Physika Verlag, Wein.

MARONNA, R. A. and YOHAI, V. J. (1981). Asymptotic behaviour of general $M$-estimators for regression and scale with random carriers. *Z. Wahrsch. verw. Gebiete* **58** 7–20.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MELBOURNE
PARKVILLE, VICTORIA 3052
AUSTRALIA