The Annals of Statistics
1998, Vol. 26, No. 3, 1147–1169

FUNCTIONAL STABILITY OF ONE-STEP *GM*-ESTIMATORS IN APPROXIMATELY LINEAR REGRESSION[1]

BY DOUGLAS G. SIMPSON AND VICTOR J. YOHAI

University of Illinois, Urbana–Champaign, University of Buenos Aires and Conicet

This paper provides a comparative sensitivity analysis of one-step Newton–Raphson estimators for linear regression. Such estimators have been proposed as a way to combine the global stability of high breakdown estimators with the local stability of generalized maximum likelihood estimators. We analyze this strategy, obtaining upper bounds for the maximum bias induced by ε-contamination of the model. These bounds yield breakdown points and local rates of convergence of the bias as ε decreases to zero. We treat a unified class of Newton–Raphson estimators, including one-step versions of the well-known Schweppe, Mallows and Hill–Ryan *GM* estimators. Of the three well-known types, the Hill–Ryan form emerges as the most stable in terms of one-step estimation. The Schweppe form is susceptible to a breakdown of the Hessian matrix. For this reason it fails to improve on the local stability of the initial estimator, and it may lead to falsely optimistic estimates of precision.

1. Introduction. Statistical models are seldom believed to be complete descriptions of how real data are generated; rather, the model is an approximation that is useful if it captures essential features of the data. Good robust methods perform well even if the data deviate from the theoretical model. The best-known example of this behavior is the outlier resistance and transformation invariance of the median. There have been considerable challenges in developing good robust methods for more general problems. A major challenge in regression modeling has been to combine gross stability, that is, a breakdown point near 50%, with local stability as indicated by a bounded influence function; see, for example, Ruppert and Simpson (1990) for discussion and further references.

Several authors have proposed one-step and k-step estimators to combine local and global stability, as well as a degree of efficiency under the target linear model. This approach follows from seminal work of Bickel (1975), who showed that one-step M-estimators have the same first order asymptotic properties as fully iterated estimators. Citing a suggestion by Bickel, Jurečková and Portnoy (1987) proved that the initial estimator need only converge at a rate $o_p(n^{-1/4})$. They used this fact to provide estimators that retain the high breakdown point yet improve on the efficiency of the least median of

Received November 1996; revised December 1997.
[1]Supported in part by NSF Grant DMS 95-05290, ONR Grant N0014-84-K0265, UBACYT Grant EX187 and CONICET Grant PIP 4186.
AMS 1991 *subject classifications*. Primary 62F35, 62J02; secondary 62F10, 62J05.
Key words and phrases. Breakdown point, maximum bias function, Newton–Raphson, robust statistics, weighted least squares.

squares estimator, which was shown by Rousseeuw (1984) to have breakdown point equal to 1/2, and for which cube root-n convergence was established under general conditions by Kim and Pollard (1990) and Davies (1990). Simpson, Ruppert and Carroll (1992) and Davies (1993) presented one-step *GM*-estimators of the Mallows type to combine local and global stability. Coakley and Hettmansperger (1993) considered one-step Schweppe-type estimators for design independent efficiency under the target model, assuming the initial estimator to be root-n consistent. In each case the idea was to do one iteration, or a fixed number of iterations, of a quadratically converging algorithm for *GM* estimation. *GM* estimation is a generalization of maximum likehood estimation and is closely related to the method of generalized estimating equations, which is finding many applications in biostatistics, econometrics and other fields.

Recently, Simpson and Chang (1997) studied a unified class of k-step Newton–Raphson *GM*-estimators including versions of the preceding proposals as special cases. They established root-n consistency and asymptotic normality under regularity conditions that depend on the the score function and regression weights. We provide a complementary sensitivity analysis of these one-step *GM*-estimators, comparing stability properties within the unified class. Based on the theoretical results and numerical computations, we develop specific recommendations concerning one-step *GM*-estimators. We find much to recommend the Hill–Ryan type of estimator, which is essentially a maximum likelihood estimator for a heteroscedastic linear regression model. In contrast to the well-known Schweppe form, and, to a lesser degree, the Mallows form, the Hill–Ryan form requires fewer constraints on the regression design and error distribution.

An important drawback of the Schweppe form, and with an inadequately weighted Mallows form, is that the inverse Hessian easily breaks down to a subspace. This causes the one-step estimator to have the same local stability (or instability) as the preliminary regression estimator. It also leads to a breakdown of the Huber (1967) "sandwich" covariance functional, so that inferences based on the estimator would lead to falsely optimistic estimates of precision. Simpson, Ruppert and Carroll (1992) advocated stronger downweighting in the Mallows case for similar reasons.

The rest of the paper is organized as follows. Section 2 reviews the general approach to the sensitivity analysis and some previous results on optimal sensitivity. Section 3 defines the class of estimators under investigation and provides results about the regression weights. Problems with the Hessian matrix are treated in Section 4, where we develop and justify an extended definition of the one-step functional. Section 5 provides the main results on the global and local stability of one-step *GM* functionals, while Section 6 presents a solution to an optimal sensitivity problem. Section 7 provides numerical comparisons of the bias functions of one- and two-step Hill–Ryan estimators, as well as the least median of squares functional, the minimax *GM* functional, and certain S- and P-estimators. Section 8 provides some further discussion and recommendations. Proofs for the main results are given in Section 9.

2. Bias under the contaminated linear model: general sensitivity analysis. We consider the effects of deviations from the linear regression model, $y = \beta' x + e$, where $x \in R^p$ and e are stochastically independent. Under this model (x, y) has a distribution of the form

$$F_\beta(x_0, y_0) = \int_{-\infty}^{x_{01}} \cdots \int_{-\infty}^{x_{0p}} G_0(y_0 - \beta' x)\, dH_0(x).$$

The parameter β describes the dependence of y on the different components of x. Given a random sample $(x_1, y_1), \dots, (x_n, y_n)$ from F_β, many estimates of β can be expressed as functionals of the empirical distribution, that is,

$$\hat\beta = T(F_n), \qquad F_n(x_0, y_0) = n^{-1} \sum_{i=1}^{n} 1_{[0,\infty)}(y_0 - y_i) \prod_{j=1}^{p} 1_{[0,\infty)}(x_{0j} - x_{ij}). \tag{2.1}$$

If the data come from the distribution F, then $T(F_n)$ is typically a consistent estimate of $T(F)$. Following Huber (1981), Hampel, Ronchetti, Rousseeuw and Stahel (1986) and others, we study the behavior of $T(F)$ as F departs from the class $\{F_\beta\}$ along ε-contaminations of the form $F_{\beta,\varepsilon} := (1-\varepsilon)F_\beta + \varepsilon F^*$. The measure of the stability of the functional is provided by the maximum bias,

$$b_T(\varepsilon) = \sup_{F^*} \| T(F_{\beta,\varepsilon}) - \beta \|, \tag{2.2}$$

where $\|\cdot\|$ is a suitable norm; see Huber (1981). The bias function has been used by various authors as a way to combine local and global aspects of the stability of an estimator.

Global stability is often measured by the breakdown point,

$$\varepsilon_T^* = \inf\{\varepsilon > 0 : b_T(\varepsilon) = \infty\}. \tag{2.3}$$

It indicates how far from the model the estimator becomes completely uninformative. A positive breakdown point, equivalently a finite maximum bias for some $\varepsilon > 0$, implies a global stability of the estimator. Equivariant functionals always have $\varepsilon_T^* \le 1/2$ if the bias norm is invariant; see, for example, He and Simpson (1993), Remark 2.3. For further background see Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986).

EXAMPLE. The least squares estimate corresponds to $T(F) = \{E_F(xx')\}^{-1} \times E_F(xy)$. Because $T\{(1-\varepsilon)F_\beta + \varepsilon F_{\beta+\delta}\} = \beta + \varepsilon\delta$ for each $\varepsilon > 0$, the breakdown point, ε_T^*, equals 0, as is well known.

To measure the local stability of T we define the *sensitivity of order* r,

$$\gamma_T^{(r)} := \limsup_{\varepsilon \downarrow 0} b_T(\varepsilon)/\varepsilon^r, \tag{2.4}$$

where $0 \le r \le 1$. If $\gamma_T^{(r)} < \infty$, then T is *locally stable of order* r. He and Simpson (1993) showed that order 1 stability is the best possible for Fisher consistent functionals.

Hampel's (1974) influence function is a widely used heuristic tool for assessing local stability. It uses a slightly narrower class of deviations than contamination bias, restricting to point mass contamination. In our experience, the main technical challenge in going from an influence analysis to a local analysis in terms of the contamination bias is not the size of the neighborhood, but, rather, the interchange of the limit in ε and the supremum over the neighborhood; compare Huber [(1981), page 15]. For instance, the existence problem treated in Section 4 is revealed by the bias analysis, but fails to emerge in an influence analysis.

Local bias and influence analysis can be compared if we slightly generalize Hampel's notion of gross-error sensitivity. He and Simpson (1992) defined the generalized gross-error sensitivity,

$$\gamma_T^* := \sup_{x,\,y} \limsup_{\varepsilon \downarrow 0} \frac{\|T\{(1-\varepsilon)F_\beta + \varepsilon\Delta_{x,\,y}\} - T(F_\beta)\|}{\varepsilon}, \tag{2.5}$$

avoiding the assumption that the influence function exists. They observed that $\gamma_T^* \le \gamma_T^{(1)}$ for any functional T. Thus, an unbounded influence function implies local instability of order 1, whereas local stability of order 1 implies a bounded gross error sensitivity. Martin, Yohai and Zamar (1989) showed that bounded influence *GM*-estimators are also locally stable of order 1. Yohai and Zamar (1992) showed that the least median of squares functional is locally stable of order 1/2.

Few functionals are known to achieve both the optimal $O(\varepsilon)$ rate locally and the optimal global breakdown point with respect to contamination bias. He and Simpson (1993) and Maronna and Yohai (1993) presented equivariant functionals having contamination bias within a dimension-free factor of the smallest possible bias for equivariant functionals. He and Simpson (1993) showed that such functionals forfeit a property they called *local linearity*. Roughly, local linearity means a functional has an influence function with finite second moments under the model. Here we aim to develop reasonable locally linear functionals with good local stability and optimal global stability, as measured by the contamination bias. Davies (1993) established an even stronger form of stability for a locally linear k-step functional. Our goal is to compare stability properties across a unified class of estimators.

Definition (2.2) extends to other types of deviations from the model, and a variety of neighborhoods have been used in the literature. Huber (1981) and Bickel (1981) discussed several possibilities. Donoho and Liu (1988) considered various distance based neighborhoods. Davies (1993) argued for "weak" metric neighborhoods. Since the Davies neighborhoods are larger than contamination neighborhoods, stability with respect to those neighborhoods would be a stronger result than stability with respect to contamination neighborhoods. Conversely, lack of stability for contamination would imply lack of stability for the Davies neighborhoods. Here we use the smaller contamination neighborhoods for screening. Contamination neighborhoods are tractible and intuitive, if somewhat heuristic, but they are sufficient to reveal existence problems with poorly chosen one-step estimators.

3. *K*-step *GM*-estimators. *GM*-estimators include many well-known regression estimators, for instance, ordinary least squares, weighted least squares, minimum L_1 distance, Huber's (1973) regression M-estimate, Mallows' (1975) generalized M-estimate, the efficient bounded–influence estimate of Krasker and Welsch (1982) and the locally linear regression estimates studied by Cleveland (1979), Fan (1992) and others.

A unified class of *GM*-estimators consists of functionals of the form $\hat{\beta} = T(F_n)$ with

$$T(F) := \underset{t \in R^p}{\operatorname{argmin}}\ E_F\left[\rho\left\{\frac{y - t'x}{S(F)w(x;H)^\alpha}\right\}w(x;H)^{1+\alpha}\right], \tag{3.1}$$

where ρ is an even function such that $\rho(|r|)$ is nondecreasing in $|r|$, $S(F)$ is a residual scale functional and the weights w depend only on the marginal distribution of x. If $w(x;H) \equiv 1$ we obtain the Huber (1973) M-estimates, which have unbounded influence functions. Weighted estimates with $\alpha = -1, 0, 1$ correspond to the well-known Hill–Ryan, Mallows and Schweppe forms, respectively; see Hampel, Ronchetti, Rousseeuw and Stahel (1986).

If ρ has a bounded derivative and the weights decrease to zero at least as fast as $\|x\|^{-1}$ as $\|x\|$ increases, then the *GM*-estimator has a bounded influence function. Despite good local stability, *GM*-estimators have breakdown points that decrease to zero as the dimension p becomes large [Maronna, Bustos and Yohai (1979)]. They cannot achieve the optimal global stability. To address this deficiency we consider one-step *GM* functionals.

We need a preliminary regression functional, $T_0(F)$, and residual scale functional $S(F)$ which are both equivariant, that is, if (y, x) has distribution F and F^* is the distribution of (y^*, x^*), where $y^* = ay + b'x$, $x^* = Ax$, where $a \in R$, $b \in R^p$ and A is a $p \times p$ matrix, then $T_0(F^*) = (A')^{-1}(aT_0(F) + b)$ and $S(F^*) = |a|S(F)$. Further conditions on S and T_0 are given below. In the numerical calculations of Section 7, we take $T_0(F)$ to be the LMS functional, and we take $S(F)$ to be the median absolute deviation (MAD) of residuals from $T_0(F)$, scaled to be consistent under the normal model.

Assume $\rho(x)$ is twice differentiable with first and second derivatives $\psi(x)$ and $\dot{\psi}(x)$. Then the Newton–Raphson and Scoring algorithms iterate according to $T_1(F) := T_0(F) - \dot{\Psi}_0(F)^{-1}\Psi_0(F)$, where

$$\Psi_0(F) = S(F)\ E_F\left[\psi\left\{\frac{y - T_0(F)'x}{S(F)w(x;H)^\alpha}\right\}w(x;H)x\right], \tag{3.2}$$

where H is the marginal distribution of x under F, and where $\dot{\Psi}_0(F)$ is one of the following:

$$\begin{aligned} \dot{\Psi}_0^{\mathrm{nr}}(F) &= -E_F\left[\dot{\psi}\left\{\frac{y - T_0(F)'x}{S(F)w(x;H)^\alpha}\right\}w(x;H)^{1-\alpha}xx'\right], \\ \dot{\Psi}_0^{\mathrm{sc}}(F) &= -E_F\left[\dot{\psi}\left\{\frac{y - T_0(F)'x}{S(F)w(x;H)^\alpha}\right\}\right]E_H[w(x;H)^{1-\alpha}xx']. \end{aligned} \tag{3.3}$$

Iterating a fixed number of times leads to k-step GM-estimators.

REMARK 1. In order for the scoring algorithm to have the accelerated convergence of Newton–Raphson, it is essentially necessary that $w(x; H)^{-\alpha}(y - \beta' x)$ be stochastically independent of x. If $\alpha = 0$, this condition implies a homoscedastic linear regression model, but if $\alpha \neq 0$, it implies a nonstandard heteroscedastic model. Simpson, Ruppert and Carroll (1992) observed that Newton–Raphson with a monotone score function appears more robust to heteroscedasticity. In the sequel we restrict to Newton–Raphson and monotone ψ.

Under sufficient regularity conditions, the k-step *GM*-estimators have the same first-order asymptotic theory as fully iterated *GM*-estimators. Results of this kind have been known at least since the work of Bickel (1975). Simpson and Chang (1997) provided an asymptotic result of this type for the general class of estimators described in (3.2) and (3.3). From Huber's (1967) theory of M-estimation, they noted that we may estimate the covariance of $\hat{\beta} = T(F_n)$ by the sandwich functional, $V_0(F_n)$, where

$$V_0(F) = \dot{\Psi}_0(F)^{-1}\Gamma_0(F)\dot{\Psi}_0(F)^{-1}, \tag{3.4}$$

and

$$\Gamma_0(F) = S^2(F) E_F\left[\psi^2\left\{\frac{y - T_0(F)'x}{S(F)w(x; H)^\alpha}\right\} w^2(x; H)xx'\right];$$

see also Simpson, Ruppert and Carroll (1992) and Coakley and Hettmansperger (1993).

We require the weights to have certain invariance properties so that the regression estimator will be equivariant. Thus, we consider weights of the form

$$w(x; H) = w\{\|x - m(H)\|_{C(H)}\}, \tag{3.5}$$

where H is the distribution of x, $m(H)$ and $C(H)$ are location and scatter equivariant functionals and $\|x\|_A := (x' A^{-1} x)^{1/2}$ for a symmetric positive definite matrix A. Suppose B is a $p \times p$ nonsingular matrix, b is a vector, and $\tilde{H}$ is the distribution of $Bx + b$, where $x \sim H$. Then we assume $m(\tilde{H}) = Bm(H) + b$ and $C(\tilde{H}) = BC(H)B'$.

REMARK 2. If the regression model lacks an intercept, then the one-step *GM*-estimator is regression equivariant without the centering in (3.5). If x has constant and binary components, then the weighting scheme should be modified. For instance, let $\{b_1, \dots, b_q\}$ be a linearly independent basis for the space orthogonal to the span of the constant and binary components of x, let $B = [b_1, \dots, b_q]$, set $z = B'x$, and replace x by z in (3.5).

In the analysis, we will make the following assumption about the function w.

ASSUMPTION A. The weights are of the form (3.4), where m and C are location and scatter equivariant, and where $w(u)$ satisfies (i) $0 < u < v < \infty$

implies $0 < w(v) \le w(u) \le k_0$ for some $k_0 < \infty$; (ii) $w(u)|u|^\kappa \le k_1$ for some $\kappa \ge 1$ and $k_1 < \infty$; (iii) $w(u)$ is continuous and (iv) $\lim_{u\to\infty} w(u)u^\kappa = k_2$ for some finite $k_2 > 0$.

EXAMPLE. A convenient class of weight functions satisfying A(i)–(iv) uses $w(u) = 1/(c + |u|^\kappa)$, where $c > 0$ and $\kappa \ge 1$.

Assumption A(ii) implies a downweighting of distant x's to limit their influence on the fit of the linear model to the central portion of the data. If the regression function is assumed to be only locally linear, then more severe downweighting is common; see, for example, Fan (1992). In addition to the influence-localizing effect of the weights, we require the residual score function to be bounded and relatively smooth. The following conditions are workable for Newton–Raphson.

ASSUMPTION B. (i) $\psi(u)$ is bounded, increasing, differentiable and odd, with a bounded, continuous derivative $\dot\psi(u) > 0$ for $-\infty < u < \infty$ and (ii) $\dot\psi$ has a bounded, continuous derivative $\ddot\psi$.

To analyze regression functionals, we use the maximum bias of (2.2) with the invariant norm $\|T-\beta\|_{C_0} := \{(T-\beta)'C_0^{-1}(T-\beta)\}^{1/2}$, where $C_0 = C_0(H_0)$ is an equivariant scatter functional. If the uncontaminated design distribution has a finite positive definite covariance matrix, then we may set $C_0 = \mathrm{Cov}_H(x)$ without loss of generality. Equivariance of T implies that $b_T(\varepsilon) = \sup_{F^*} \|T(F_{0,\varepsilon})\|$ if H_0 is standardized so that $C_0 = I_p$.

The bias of the one-step functional depends on the bias of the initial estimate T_0, the scale functional S, and the location and scatter functionals for x. We separate the bias of S into two parts, $b_S^-(\varepsilon) := S(F_0) - \inf_{F^*} S(F_\varepsilon)$ and $b_S^+(\varepsilon) := \sup_{F^*} S(F_\varepsilon) - S(F_0)$. Define $\varepsilon_S^+ := \inf\{\varepsilon\colon b_S^+(\varepsilon) = \infty\}$ and $\varepsilon_S^- := \inf\{\varepsilon\colon b_S^-(\varepsilon) = S(F_0)\}$. The breakdown point is the smaller of the two. See He (1989), Martin and Zamar (1989) and Rousseeuw and Croux (1994) for further discussion of breakdown of scale estimation. Here the scale estimate is applied to residuals from a preliminary fit. It will have the same breakdown point as the pure scale estimator if the preliminary regression estimator has a breakdown point at least as large as that of the preliminary scale estimator.

Let $H_\varepsilon = (1-\varepsilon)H_0 + \varepsilon H^*$ be the marginal distribution of x under the contamination model in which H^* is allowed to vary arbitrarily. Define the equivariant bias functions

$$b_m(\varepsilon) := \sup_{H^*} \|m(H_\varepsilon) - m(H_0)\|_{C(H_0)}$$

and

$$b_C(\varepsilon) := \sup_{H^*} \left|\lambda_{\max}^{1/2}\{C(H_0)C(H_\varepsilon)^{-1}\} - 1\right| + \left|\lambda_{\max}^{1/2}\{C(H_\varepsilon)C(H_0)^{-1}\} - 1\right|,$$

where $\lambda_{\max}(A)$ is the maximum eigenvalue of the matrix A. Also define the matrix norm $\|A\| := \sup_{\|u\|=1} \|Au\|$. If A is a symmetric matrix then $\lambda_{\max}(A) = \|A\|$.

In order for T_1 to be locally stable of order 1, we need $T_0(F)$ to be locally stable of order at least 1/2, and we need S, m and C to be globally stable. Global stability of T_1 follows from the global stability of T_0, S, m and C. Examples for T_0 and S include least median of squares, the S-estimators of Rousseeuw and Yohai (1984), the τ-estimates of Yohai and Zamar (1988) and the M-estimators with auxiliary scale studied by Yohai and Zamar (1997). These are known to have bias of order $\varepsilon^{1/2}$ as $\varepsilon \downarrow 0$. Examples for m and C include the multivariate S-estimates studied by Rouseeuw and Leroy (1987), Davies (1987) and Lopuhaä (1989), the P-estimators of Maronna, Stahel and Yohai (1992), and a locally and globally stable scatter functional developed by Davies (1992). Robust multivariate location and scatter is by no means a settled issue; see, for example, Rocke (1996) and Kent and Tyler (1996).

Under Assumption A we can relate properties of the regression weights to the bias of the location and scatter functionals. Let $m_\varepsilon = m(H_\varepsilon)$, $C_\varepsilon = C(H_\varepsilon)$, and $w_\varepsilon(x) = w(\|x - m_\varepsilon\|_{C_\varepsilon})$. The assumptions about w imply that if $a \le \kappa$ then

$$
\begin{aligned}
w_\varepsilon(x)\|x\|^a &\le \min\left(k_0, \frac{k_1}{\|x - m_\varepsilon\|^\kappa_{C_\varepsilon}}\right)(\|m_\varepsilon\| + \|x - m_\varepsilon\|)^a \\
&\le k_0\{d_\kappa + \Delta_\kappa(\varepsilon)\}^a,
\end{aligned} \tag{3.6}
$$

where $d_\kappa = \|m_0\| + (k_1/k_0)^{1/\kappa}\lambda^{1/2}_{\max}(C_0)$, and $\Delta_\kappa(\varepsilon) = \lambda^{1/2}_{\max}(C_0)\{b_m(\varepsilon) + (k_1/k_0)^{1/\kappa} b_C(\varepsilon)\}$. As long as the location and scatter functionals for x are globally stable, the weights satisfy

$$
\sup_{0\le\varepsilon\le\delta} \sup_x w_\varepsilon(x)\|x\|^a < \infty \quad \text{if } 0 \le \delta < \min(\varepsilon^*_m, \varepsilon^*_C). \tag{3.7}
$$

Because w is decreasing on $(0, -\infty)$, we can bracket the weight function as follows:

$$
0 < w^-_\varepsilon(x) \le w(\|x - m_\varepsilon\|_{C_\varepsilon}) \le w^+_\varepsilon(x) \le k_0, \tag{3.8}
$$

where $w^-_\varepsilon(x) = w[\{\|x - m_0\|_{C_0} + b_m(\varepsilon)\}\{1 + b_C(\varepsilon)\}]$ and $w^+_\varepsilon(x) = w[\{\|x - m_0\|_{C_0} - b_m(\varepsilon)\}_+ \{1 + b_C(\varepsilon)\}^{-1}]$. We need the the following additional properties.

PROPOSITION 1. *Assume the weight function satisfies condition A. If* $\delta < \min(\varepsilon^*_m, \varepsilon^*_C)$ *then*

$$
\sup_{0\le\varepsilon\le\delta} \sup_x \{w_\varepsilon(x)/w_0(x)\} < \infty \quad \textit{and} \quad \sup_{0\le\varepsilon\le\delta} \sup_x \{w_0(x)/w_\varepsilon(x)\} < \infty.
$$

If in addition $b_m(0+) = b_C(0+) = 0$, *then* $\lim_{\varepsilon\downarrow 0} \sup_x |w_\varepsilon(x) - w_0(x)|/w_0(x) = 0$ *and* $\lim_{\varepsilon\downarrow 0} \sup_x |w_\varepsilon(x) - w_0(x)|/w_\varepsilon(x) = 0$.

4. Generalized one-step functionals. A serious problem arises in the analysis of some of the one-step estimators. If the weights and α are such that $w(x)^{1-\alpha}\|x\|^2$ is unbounded, then the Hessian matrices in (3.3) may break down for arbitrarily small amounts of contamination. Of course, this need not cause the estimator to break down, since the estimator reverts to the initial estimator if the inverse Hessian is a zero matrix. However, it implies that a direct analysis of the one-step functional is not available. In order to properly define an extended version of the one-step functional when the Hessian breaks down, we replace the general distribution F in (3.3) by an empirical distribution sampled from F. The expectation with respect to the empirical distribution sampled from any proper distribution exists with probability 1. We therefore define the functional for weakly convergent sequences of empirical distributions. The limiting form provides an appropriate generalization of the one-step functional, which we then use in the sensitivity analysis of Section 4. If $w(x)^{1-\alpha}\|x\|^2$ is bounded, then the limiting form reduces to the previously defined one-step functional.

Suppose $(x', y)'$ are jointly distributed according to F, and let H be the marginal distribution function of x. Under the assumptions about w and ψ, we can define

$$d := \left[\dot{\psi}\left\{\frac{y - T_0(F)'x}{S(F)w(x;H)^{\alpha}}\right\}w(x;H)^{1-\alpha}\right]^{1/2} x. \tag{4.1}$$

In order to extend the definition of the one-step estimator, we construct a rotation of z that separates the components with finite second moments from those lacking finite second moments. Define

$$\mathscr{L}(F) = \{\nu \in R^p\text{: such that } E_F(\nu' d)^2 < \infty\}.$$

Then $\mathscr{L}$ is a subspace of R^p of dimension, say, $h = h(F) \in \{0, \dots, p\}$. If $h = 0$, then $T_1(F) = T_0(F)$. If $h \geq 1$, let $k = k(F) = p - h(F)$, let $U_1(F)$ be a $p \times k$ matrix whose columns form an orthonormal basis for $\mathscr{L}^{\perp}$ and let $U_2(F)$ be a $p \times h$ matrix whose columns form an orthonormal basis for $\mathscr{L}$. Then $U = [U_1 \quad U_2]$ is an orthogonal rotation matrix such that $U'U = UU' = I_p$, and $[s' \ t']' := [d'U_1 \ d'U_2]' = U'd$ is a rotation of d such that:

1. $E(\lambda' s)^2 = \infty$ for all $\lambda \neq 0$, $\lambda \in R^k$;
2. $E(t_i^2) < \infty$, $1 \leq i \leq h$.

We may therefore define the functional

$$R(F) = \begin{cases} 0_{pp}, & \text{if } h(F) = 0, \\ U_2(F)[E_F\{U_2(F)'dd'U_2(F)\}]^{-1}U_2(F)', & \text{if } h(F) > 0, \end{cases} \tag{4.2}$$

where 0_{ij} is the $i \times j$ matrix with all elements equal to zero, provided we require $E(t\,t')$ to be positive definite if $h > 0$. The extended one-step functional is given by

$$T_1(F) = T_0(F) + R(F)\Psi_0(F). \tag{4.3}$$

If $\dot{\Psi}_0(F)$ exists; that is, $h = p$, then the definition in (4.3) agrees with the one given in (3.3).

In the sequel we replace the one-step functional by the more general functional in (4.3). However, in order to justify that this is a proper interpretation of the one-step functional when $h(F) > 0$, we need the continuity property established in the following theorem.

THEOREM 1. *Let F_n be the empirical distribution for a random sample, $(x_1, y_1), \dots, (x_n, y_n)$, from F. Suppose Assumptions* A *and* B(i) *hold. Suppose z and U_2 are as described above. If $h > 0$, assume $E_F\{U_2(F)'dd'U_2(F)\}$ is positive definite. Then $\lim_{n\to\infty}\{\dot{\Psi}_0(F_n)\}^{-1} = -R(F)$ a.s.*

If $h(F) < p$ in (4.2), then we have a breakdown of the Hessian functional used in the updating algorithm. According to Theorem 2 of Section 5, this in itself does not cause a breakdown of the one-step estimator. However, if the Hessian breaks down, then the one-step fails to improve the local stability of the preliminary estimator. Thus, one would need to use a locally stable preliminary estimator. Perhaps more importantly, the covariance estimate in (3.4) breaks down to a subspace under arbitrarily small amounts of contamination. Thus, if $\kappa(1-\alpha) < 2$ and there are extreme design points, the sandwich estimate may lead to an over-optimistic assessment of precision. Simpson, Ruppert and Carroll (1992) referred to this phenomenon as variance breakdown.

5. Sensitivity analysis of *k*-step estimators. Armed with the extension given in Section 4, we state our main results concerning the local and global bias of the one-step estimator. Denote the gross-error sensitivity of the fully iterated *GM*-estimator by

$$\gamma^*_{GM} = S(F_0)\sup_{x,\,y}\left|\psi\{w_0(x)^{-\alpha}y/S(F_0)\}\right|\|w_0(x)A_0^{-1}x\|,$$

where $A_0 = E_0[\dot{\psi}\{w_0(x)^{-\alpha}y/S(F_0)\}w_0(x)^{1-\alpha}xx']$. We require the following additional conditions.

ASSUMPTION C. $E_0\{w_0(x)^{1-\alpha}\|x\|^2\} < \infty$ and $E_0\{w_0(x)^{1-\alpha}xx'\}$ is positive definite.

ASSUMPTION D. Here $b_{T_0}(0+) = b_S(0+) = b_m(0+) = b_C(0+) = 0$, and $F_0(y|x)$ is a symmetric distribution.

ASSUMPTION E. $\kappa(1-\alpha) \geq 2$ and $E_0\{w_0(x)^{1-2\alpha}\|x\|^3\} < \infty$.

THEOREM 2. (i) *Under Assumptions* A, B *and* C *there are nonnegative functions $A(\varepsilon)$ and $B(\varepsilon)$, bounded on $[0, \delta]$ for $\delta < \min(\varepsilon^*_{T_0}, \varepsilon^*_S, \varepsilon^*_m, \varepsilon^*_C)$, such that*

$$b_{T_1}(\varepsilon) \leq A(\varepsilon)b_{T_0}(\varepsilon) + \varepsilon B(\varepsilon)\gamma^*_{GM}.$$

(ii) *If, moreover,* D *is satisfied, then*

$$\limsup_{\varepsilon\downarrow 0} A(\varepsilon) \le 1 \quad \textit{and} \quad \limsup_{\varepsilon\downarrow 0} B(\varepsilon) \le 1,$$

and, if β *is one-dimensional, then* $\limsup_{\varepsilon\downarrow 0}\{A(\varepsilon)+B(\varepsilon)\} \le 1$.

(iii) *If, furthermore,* E *is satisfied, then* $A(\varepsilon) = O(\varepsilon) + o\{b_{T_0}(\varepsilon)\}$ *as* $\varepsilon \downarrow 0$.

As an immediate corollary we obtain a lower bound for the breakdown point of T_1.

COROLLARY 1. *Under the conditions of Theorem* 2(i), $\varepsilon^*_{T_1} \ge \min(\varepsilon^*_{T_0}, \varepsilon^*_S, \varepsilon^*_m, \varepsilon^*_C)$.

Under slightly stronger conditions we obtain an upper bound for the order r sensitivity.

COROLLARY 2. *Let* T_0 *be locally stable of order* r. *Then under the conditions of Theorem* 2(ii), T_1 *is also locally stable of order* r *and*

$$\gamma^{(r)}_{T_1} \le \begin{cases} \gamma^{(r)}_{T_0}, & \textit{if } 0 < r < 1, \\ \gamma^{(1)}_{T_0} + \gamma^*_{GM}, & \textit{if } r = 1. \end{cases}$$

If β *is one-dimensional, then* $\gamma^{(1)}_{T_1} \le \max(\gamma^{(1)}_{T_0}, \gamma^*_{GM})$.

Consequently, if any iteration is locally stable of order r, then subsequent iterations are locally stable of order at least r. Under a stronger condition on the weights, we find that iteration can improve the order of stability from 1/2 to 1.

COROLLARY 3. *Let* T_0 *be locally stable of order at least* 1/2. *Then under the conditions of Theorem* 2(iii), T_1 *is locally stable of order* 1 *and* $\gamma^{(1)}_{T_1} = \gamma^*_{GM}$.

6. Optimal local sensitivity. Suppose $F_0(y|x)$ is a symmetric distribution and $H_0(x)$ is spherical. Then the *GM*-estimator with the smallest sensitivity at F_0 has an influence function proportional to $\operatorname{sign}(y)x/\|x\|$; see Martin, Yohai and Zamar (1989). He and Simpson (1993) showed that this functional has optimal sensitivity over a larger class, namely, the locally linear functionals; see their Definition 2.1. We show that a one-step Hill–Ryan type estimator can have sensitivity arbitrarily close to the lower bound for locally linear functionals, which was given by He and Simpson (1993).

Define the weight function $w_c(u) = (c + |u|)^{-1}$, which satisfies Assumption A with $k_0 = c^{-1}$, $k_1 = k_2 = 1$ and $\kappa = 1$. Let $\tau_c(x) = c\tanh(x/c)$. This is an odd function that is strictly increasing from zero to c as its argument increases from zero to ∞. Its first two derivatives are $\dot\tau_c(x) = 1 - \tanh^2(x/c)$ and $\ddot\tau_c(x) = -(2/c)\tanh(x/c)\{1-\tanh^2(x/c)\}$, both of which are bounded and continuous. Thus τ_c satisfies Assumption B.

We consider a one-step Newton–Raphson functional as given in (3.2) and (3.3) with $\alpha = -1$, $\psi(x) = \tau_c(x)$, and weights as in (3.4) with $w(u) = w_c(u)$. Assume that $S(F_0) = 1$, $m(H_0) = 0$ and $C(H_0) = I_p$, where H_0 is the marginal distribution of x under F_0. As $k(1-\alpha) = 2$, Corollary 3 implies that the sensitivity of the one-step Hill–Ryan estimator is no larger than

$$\gamma_c^* = c \sup_x \{w_c(\|x\|)\|A_c^{-1}x\|\}, \tag{6.1}$$

where $A_c = E_0[\dot{\tau}_c\{w_c(\|x\|)y\}w_c^2(\|x\|)xx']$.

THEOREM 3. *Suppose that $F_0(x, y)$ has the conditional density $f_0(y|x) = f_0(y)$ almost surely, which is independent of x. Suppose f_0 is symmetric about zero and differentiable, and set $\phi(y) = (d/dy)\log\{f_0(y)\}$. Assume $E_0|\phi(y)| < \infty$, $E_0\|x\| < \infty$ and $P_0(\|x\| = 0) = 0$. Assume $H_0(x) = F_0(x, \infty)$ is spherical. Then*

$$\lim_{c\downarrow 0} \gamma_c^* = \frac{p}{E_0|\phi(y)|E_0\|x\|}. \tag{6.2}$$

The limit in (6.2) equals the lower bound for locally linear functionals in this setting; see He and Simpson (1993), Section 5.1. Hence, the sensitivity of the one-step Hill–Ryan estimator can be arbitrarily close to the optimum for locally linear functionals.

REMARK 3. This Hill–Ryan weighting method may be motivated by a heavy-tailed, heteroscedastic regression model in which the scale depends on the design, with larger scales being associated with more disparate regions of the design space. Such a model is one way to conceptualize uncertainty about the linear model in the extreme regions of the design space.

7. Numerical comparisons. To illustrate the effects of iteration we computed the maximum bias of one- and two-step *GM*-estimators for point mass contaminations. The initial regression estimator was Rousseeuw's (1984) least median of squares (LMS) estimate. For the scale estimate we used the median absolute deviation (MAD) of LMS residuals and for the covariance of x we used a P-estimator proposed by Maronna, Stahel and Yohai (1992).

We considered point mass contamination of a model specifying that $x_1, \ldots, x_p, y$ are independent $N(0,1)$ random variables, so $\beta = 0$. Because of the spherical symmetry, we need only consider point mass contaminations of the form $(x_0, 0, \ldots, 0, y_0)$.

Exact expressions for the LMS functional and scale functional were obtained as follows. Let F_0 be the distribution $x = (x_1, \ldots, x_p, y)$, a vector of $p+1$ independent N(0,1) random variables. We compute $T_0(x_0, y_0) = T_0(F_{x_0,y_0})$, where T_0 is the LMS estimate and $F_{x_0,y_0} = (1-\varepsilon)F_0 + \varepsilon\delta_{(x_0, 0\cdots 0, y_0)}$. It can be shown that the first component of $T_0(x_0, y_0)$ is independent of p, and the remaining components are all zero. It is therefore sufficient to consider $p = 1$.

Let $Q(\beta, u) = P_{x_0, y_0}\{(|y - \beta x|) \le u\}$, where P_{x_0, y_0} is the probability measure corresponding to F_{x_0, y_0} for $p = 1$. Then

$$Q(\beta, u) = \begin{cases} (1-\varepsilon)[2\Phi\{u/(1+\beta^2)^{1/2}\} - 1], & \text{if } u < |y_0 - \beta x_0|, \\ (1-\varepsilon)[2\Phi\{u/(1+\beta^2)^{1/2}\} - 1] + \varepsilon, & \text{if } u \ge |y_0 - \beta x_0|. \end{cases}$$

Define $d_1 = \Phi^{-1}\{(3/4 - \varepsilon)/(1-\varepsilon)\}$, $d_2 = \Phi^{-1}\{(3/4 - \varepsilon/2)/(1-\varepsilon)\}$ and $M(\beta, x_0, y_0)$, the median of $|y - \beta x|$ under F_{x_0, y_0}. Then

$$M(\beta, x_0, y_0) = \begin{cases} d_1(1+\beta^2)^{1/2}, & \text{if } |y_0 - \beta x_0| \le d_1(1+\beta^2)^{1/2}, \\ d_2(1+\beta^2)^{1/2}, & \text{if } |y_0 - \beta x_0| \ge d_2(1+\beta^2)^{1/2}, \\ |y_0 - \beta x_0|, & \text{if } d_1(1+\beta^2)^{1/2} \le |y_0 - \beta x_0| \le d_2(1+\beta^2)^{1/2}. \end{cases} \tag{7.1}$$

Let β_0 be the solution of the equation $d_1^2(1+\beta^2) = (y_0 - \beta x_0)^2$ which is closest to 0. Observe that for $|y_0| \ge d_1$ this equation has a solution. Then using (7.1), and the fact that $T_0(x_0, y_0) = T_0(F_{x_0, y_0}) = \operatorname{argmin}_\beta M(\beta, x_0, y_0)$, it is easy to show that

$$T_0(x_0, y_0) = \begin{cases} 0, & \text{if } |y_0| < d_1, \\ \beta_0, & \text{if } d_1 \le |y_0| \le d_2, \\ \beta_0, & \text{if } |y_0| > d_2 \text{ and } d_1(1+\beta_0^2)^{1/2} \le d_2, \\ 0, & \text{if } |y_0| > d_2 \text{ and } d_1(1+\beta_0^2)^{1/2} > d_2. \end{cases} \tag{7.2}$$

Let $S(x_0, y_0) = 1.4826\, M\{T_0(x_0, y_0), x_0, y_0\}$, the MAD scale of the LMS residuals. Then from (7.1) and (7.2) we get

$$S(x_0, y_0) = \begin{cases} 1.4826 d_1, & \text{if } |y_0| < d_1, \\ 1.4826 d_1(1+\beta_0^2)^{1/2}, & \text{if } d_1 \le |y_0| \le d_2, \\ 1.4826 d_1(1+\beta_0^2)^{1/2}, & \text{if } |y_0| > d_2 \text{ and } d_1(1+\beta_0^2)^{1/2} \le d_2, \\ 1.4826 d_2, & \text{if } |y_0| > d_2 \text{ and } d_1(1+\beta_0^2)^{1/2} > d_2. \end{cases}$$

In addition to our exact expressions for LMS and the scale functional, an exact formula for the covariance $C(x_0, y_0)$ is available from Maronna, Stahel and Yohai (1992). Given these expressions, we computed the one- and two-step functionals by insertion into the Newton–Raphson formula, as described in Section 3. The latter expression involves expectations with respect to y, x_1 and $z = \sum_{i=2}^p x_i^2$. The expectations were computed by Monte Carlo integration using 10,000 replications. In particular, we computed

$$T_k(x_0, y_0) = T_k\{(1-\varepsilon)F_0 + \varepsilon\delta_{(x_0, 0, \dots, 0, y_0)}\},$$

where T_k is the k-step Hill–Ryan estimator with $c = 0.2$. Then for selected ε, we computed

$$B_k^*(\varepsilon) = \sup_{x_0, y_0} \|T_k(x_0, y_0)\|, \qquad k = 1, 2. \tag{7.3}$$

TABLE 1
Maximum biases of one-step estimates

p	γ^*	$\varepsilon = 0.05$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$
2	2.49	0.20	0.52	0.91	1.36
3	2.96	0.23	0.58	0.97	1.44
4	3.15	0.26	0.62	1.01	2.21
5	3.48	0.29	0.64	1.05	3.19
10	4.89	0.34	0.73	1.92	5.12
15	6.54	0.39	1.06	2.94	7.93
20	7.66	0.42	1.27	3.35	8.44

We approximated the supremum in (7.3) by the maximum over a grid search on (x_0, y_0).

Table 1 gives the values for B_1^* for selected amounts of contamination, whereas Table 2 gives values for B_2^*. For comparison, Table 3 contains results for the minimax *GM*-estimator, using a *P*-type covariance estimator for x, and Table 4 summarizes results for LMS, the minimax *S*-estimator, and two types of *P*-estimators. The one- and two-step estimators are clearly compromises between the initial estimator (LMS) and the *GM*-estimator towards which they iterate. The one-step is closer to LMS, but improves substantially on small-ε, small-p performance of LMS. Although locally and for small p, the one-step does not do quite as well as the *GM*-estimator, it improves substantially on the bias of the *GM*-estimator for larger p and ε. The two-step estimator is more similar to the *GM*-estimator, with comparable small-ε bias. It eliminates the technical breakdown of the *GM*-estimator seen in Table 3. For $\varepsilon = 0.2$ and $p \geq 15$, however, its large bias indicates near breakdown.

In the example presented, the one-step–two-step strategy eliminated or reduced the impact of the worst features of the initial and *GM*-estimators: the local instability of LMS and the low breakdown point of the *GM*-estimator. We expect similar benefits with other choices of initial estimator and *GM*-estimator. Although the number of iterations had no effect on the sensitivity, it clearly had an impact on the bias for larger amounts of contamination. Fur-

TABLE 2
Maximum biases of two-step estimates

p	γ^*	$\varepsilon = 0.05$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$
2	2.49	0.13	0.31	0.66	1.15
3	2.96	0.14	0.38	0.79	1.54
4	3.15	0.17	0.45	0.93	2.13
5	3.48	0.20	0.52	1.21	3.29
10	4.89	0.29	0.86	2.50	7.50
15	6.54	0.41	1.33	4.71	16.65
20	7.66	0.50	1.82	6.52	20.65

TABLE 3
Maximum biases of the minimax GM-estimates with P-type covariance matrix estimates

		***GM*-estimates**			
p	γ^*	$\varepsilon = 0.05$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$
2	2.00	0.10	0.27	0.47	0.83
3	2.35	0.15	0.34	0.67	1.72
4	2.67	0.17	0.43	0.92	∞
5	2.94	0.18	0.49	1.29	∞
10	4.06	0.27	0.83	∞	∞
15	4.94	0.33	1.30	∞	∞
20	5.66	0.41	2.31	∞	∞

ther investigation may reveal strategies for selecting the number of iterations. Rousseeuw and Croux (1994) provided results on the effect of increasing the number of Newton steps in the location and scale problem. Jurečková and Malý (1995) studied the effect of iteration when the score function has jump points.

8. Recommendations. According to Theorem 2, an important condition for good stability is that $\kappa(1-\alpha) \geq 2$, where κ refers to the degree of downweighting and α determines how the weights interact with the scale of the residual score function; see (3.2) and Assumption A.

The one-step Hill–Ryan estimator, with $\alpha = -1$, has $\kappa(1-\alpha) \geq 2$ as long as $\kappa \geq 1$ so that for large $\|x\|$ the weights are of order $\|x\|^{-1}$ or smaller. It also satisfies the design conditions $E_0\{w_0(x)^2\|x\|^2\} < \infty$ and $E_0\{w_0(x)^3\|x\|^3\} < \infty$ regardless of the distribution of x, if $\kappa \geq 1$. Therefore, the one-step Hill–Ryan estimator is locally stable of order 1 provided the initial estimator is locally stable of order at least 1/2.

The one-step Mallows estimator, with $\alpha = 0$, requires more severe downweighting with $\kappa \geq 2$ and the design condition $E_0\{w_0(x)\|x\|^3\} < \infty$. With

TABLE 4
Maximum biases of S- and P-estimates

		***S*-estimates (all *p*)**			
	γ^*	$\varepsilon = 0.05$	$\varepsilon = 0.10$	$\varepsilon = 0.15$	$\varepsilon = 0.20$
Minimax	∞	0.49	0.77	1.05	1.37
LMS	∞	0.53	0.83	1.07	1.52
		***P*-estimates (all *p*)**			
MP	3.14	0.163	0.36	0.56	0.82
CMP	1.57	0.085	0.19	0.31	0.50

From Maronna and Yohai (1993).

weights of order $\|x\|^{-2}$, it is enough to have $E_0\|x\| < \infty$. With weights of order $\|x\|^{-1}$, which might be preferred due to efficiency considerations, we cannot guarantee local stability of order 1 unless the initial estimator is locally stable of order 1. One possibility is a two-step strategy starting with either a Hill–Ryan iteration or Mallows with squared weights.

The one-step Schweppe estimator, with $\alpha = 1$, cannot have $\kappa(1-\alpha) \geq 2$, so, if the starting value is not locally stable of order 1, we cannot guarantee local stability of the Schweppe estimator. Moreover, to ensure that it retains the stability of the starting value, we require $E_0\|x\|^2 < \infty$. If one prefers a Schweppe-type estimator, we recommend first iterating with a Hill–Ryan type or Mallows with squared weights. Coakley and Hettmansperger (1993) made a similar recommendation. Inferences based on the estimator may still be problematic because of the Hessian breakdown discussed in Section 4.

9. Proofs of main results. We present only the major steps in the proofs of Theorems 1 and 2 by stating a series of lemmas on which the proofs rely. We omit the proofs of all lemmas and propositions. These are in the technical report by Simpson and Yohai (1993), which we will be happy to supply on request.

The proof of Theorem 1 requires additional notation. Let v be a random vector of dimension k, and let z be a random vector of dimension h, both with finite second moments. Let H be the joint distribution of $(v', z')'$. Given $\lambda \in R^k$ and $\mu \in R^h$ we denote by $\beta(\lambda, \mu, H)$ the coefficient of the best linear predictor of $\mu' z$ based on $\lambda' v$; that is, $\beta(\lambda, \mu, H) = \operatorname{argmin}_{\beta \in R} E(\mu' z - \beta\lambda' v)^2$, and we let $\rho(\lambda, \mu, H)$ denote the corresponding "uncentered" correlation coefficient. Then we have

$$\beta(\lambda, \mu, H) = \frac{E_H\{(\lambda' v)(\mu' z)\}}{E_H(\lambda' v)^2} \quad \text{and} \tag{9.1}$$
$$\rho(\lambda, \mu, H) = \frac{E_H\{(\lambda' v)(\mu' z)\}}{\{E_H(\lambda' v)^2\}^{1/2}\{E_H(\mu' z)^2\}^{1/2}}.$$

We also define $\beta_i(H)$ to be the vector of coefficients of the best linear predictor of the ith component z_i of z based on v:

$$\beta_i(H) = \operatorname*{argmin}_{\beta \in R^k} E(z_i - \beta' v)^2.$$

Finally, let $S^k = \{\lambda \in R^k \colon \|\lambda\| = 1\}$.

LEMMA 1. *For each $n \geq 1$, let H_n be a distribution on R^{k+h} with second moments, and let Q_n and J_n be the marginal distributions of the first k and last h coordinates, respectively. Assume also that*

$$\lim_{n\to\infty} \inf_{\lambda \in S^k} E_{J_n}(\lambda' v)^2 = \infty \quad \textit{and} \quad \limsup_{n\to\infty} \sup_{\mu \in R^h} E_{Q_n}(\mu' z)^2 < \infty. \tag{9.2}$$

Then $\lim_{n\to\infty} \sup_{\lambda \in S^k,\, \mu \in S^h} |\beta(\lambda, \mu, H_n)| = 0$.

LEMMA 2. *Assume the conditions of Lemma* 1. *Then* $\lim_{n\to\infty} \beta_i(H_n) = 0$.

LEMMA 3. *Let* J *be a distribution on* R^k *such that* $E_J(\lambda' v)^2 = \infty$ *for all* $\lambda \in R^k$ *and let* J_n, $n \geq 1$ *be a sequence of distribution functions on* R^h *with finite second-order moments and such that* $J_n \to J$ *weakly. Then* $\lim_{n\to\infty} \inf_{\lambda \in S^k} E_{J_n}(\lambda' v)^2 = \infty$.

LEMMA 4. *Let* H *be a distribution on* R^{k+h} *and let* J *and* Q *be the marginal distributions of the first* k *and last* h *coordinates, respectively. Assume the following*:

(i) $E_J(\lambda' v)^2 = \infty$, *for all nonzero* $\lambda \in R^k$;
(ii) Q *has second-order moments and* $E_Q(zz')$ *is nonsingular*.

For each $n \geq 1$, *let* H_n *be a distribution on* R^{k+h} *with second moments, and let* J_n *and* Q_n *be the corresponding marginal distributions. Assume the following*:

(iii) $H_n \to H$ *weakly*;
(iv) $\lim_{n\to\infty} E_{Q_n}(zz') = E_Q(zz')$.

Then $\lim_{n\to\infty} \sup_{\lambda \in S^k, \mu \in S^h} |\rho(\lambda, \mu, H_n)| = 0$.

LEMMA 5. *Let* H, J *and* Q *satisfy conditions* (i) *and* (ii) *of Lemma* 4. *Let* $(v_1', z_1')', \ldots, (v_n', z_n')', \ldots$ *be a sequence of independent random vectors with distribution* H, *where* v_i *is of dimension* k *and* z_i *of dimension* h. *Let*

$$M_{n,vv} = \frac{1}{n}\sum_{i=1}^{n} v_i v_i', \qquad M_{n,zz} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i', \qquad M_{n,vz} = \frac{1}{n}\sum_{i=1}^{n} v_i z_i'$$

and

$$M_n = \begin{pmatrix} M_{n,vv} & M_{n,vz} \\ M_{n,vz}' & M_{n,zz} \end{pmatrix}.$$

Then

$$\lim_{n\to\infty} M_n^{-1} = \begin{bmatrix} 0_{kk} & 0_{kh} \\ 0_{hk} & \{E_Q(zz')\}^{-1} \end{bmatrix} \quad a.s.$$

PROOF OF THEOREM 1. The result follows from Lemma 5 applied to $d = (v', z')'$, where d is given in (4.1). □

Next we state several lemmas used to prove Theorem 2.

LEMMA 6. *Let* H_0 *be a distribution on* R^p, $p = k + h$, *with finite second-order moments such that* $E_{H_0}(xx')$ *is nonsingular. Let* H^* *be another distribution on* R^p, *and* J^* *and* Q^* *the corresponding marginal distributions of the first* k *and last* h *coordinates, respectively. Assume that* J^* *and* Q^* *satisfy* (i) $E_{J^*}(\lambda' v)^2 = \infty$ *for all nonzero* $\lambda \in R^k$; *and* (ii) $E_{Q^*}\|z\|^2 < \infty$. *Let*

$H = (1-\varepsilon)H_0 + \varepsilon H^*$. Then there is a sequence of distributions H_n^* on R^p with finite second-order moments such that $H_n^* \to H^*$ weakly, and such that

$$\lim_{n\to\infty}\{(1-\varepsilon)E_{H_0}(xx') + \varepsilon E_{H_n^*}(xx')\}^{-1} = \begin{bmatrix} 0_{kk} & 0_{kh} \\ 0_{hk} & \{E_Q(zz')\}^{-1} \end{bmatrix},$$

where Q is the marginal of H corresponding to the last h coordinates.

We establish some properties of the Hessian type functional,

$$A_0(F) = E_0\left[\dot\psi\left\{\frac{y - T_0(F)'x}{S(F)w(x;F^x)^\alpha}\right\}w(x;F^x)^{1-\alpha}xx'\right], \tag{9.3}$$

where E_0 denotes the expectation under F_0, $w(x;H) = w(\|x - m(H)\|_{C(H)})$ and F^x is the marginal distribution of x under F. In the following, let

$$\omega(\varepsilon) = \sup_{\delta\in[0,\,\varepsilon]}\sup_x \max\left\{\frac{w_\varepsilon(x)}{w_0(x)}, \frac{w_0(x)}{w_\varepsilon(x)}\right\}. \tag{9.4}$$

By Proposition 1(i), $\omega(\varepsilon)$ is finite for $\varepsilon < \min\{\varepsilon_m^*, \varepsilon_C^*\}$.

LEMMA 7. *If w satisfies Assumption* A *for some $\kappa \geq 1$ and ψ satisfies Assumption* B, *then*

$$\|A_0(F_\varepsilon)\| \leq \omega(\varepsilon)\sup_u |\dot\psi(u)| E_0\{w_0(x)^{1-\alpha}\|x\|^2\}.$$

If in addition $E_0\{w_0(x)^{1-\alpha}\|x\|^2\} < \infty$ and $b_{T_0}(0+) = b_S(0+) = b_m(0+) = b_C(0+) = 0$, then $\lim_{\varepsilon\downarrow 0}\|A_0(F_\varepsilon) - A_0(F_0)\| = 0$.

LEMMA 8. *Suppose w satisfies Assumption* A *for some $\kappa \geq 1$ and ψ satisfies Assumption* B. *Let*

$$M_\varepsilon(x, y) = \inf_{|v|\leq 1}\dot\psi\left[v\frac{|y| + b_{T_0}(\varepsilon)\|x\|}{\{S_0 - b_S^-(\varepsilon)\}\tilde w_\varepsilon(x)^\alpha}\right],$$

where $\tilde w_\varepsilon(x) = w_\varepsilon^+(x)$ if $\alpha < 0$, $= 1$ if $\alpha = 0$ and $= w_\varepsilon^-(x)$ if $\alpha > 0$. If $E_0\{w_0(x)^{1-\alpha}xx'\}$ is a finite, positive definite matrix and $0 \leq \varepsilon < \min(\varepsilon_{T_0}^*, \varepsilon_S^*, \varepsilon_m^*, \varepsilon_C^*)$, then

$$\lambda_{\min}\{A_0(F_\varepsilon)\} \geq a_0(\varepsilon) := \lambda_{\min}\left[E_0\{M_\varepsilon(x,y)w_\varepsilon(x)xx'\}\right] > 0. \tag{9.5}$$

If in addition $b_{T_0}(0+) = b_S(0+) = b_m(0+) = b_C(0+) = 0$, then

$$\lim_{\varepsilon\downarrow 0}\lambda_{\min}\{A_0(F_\varepsilon)\} = \lambda_{\min}\{A_0(F_0)\} > 0.$$

We next present error bounds for a key linearization. Define the functional

$$D_0(F) = S(F)E_0\left[\psi\left\{\frac{y - T_0(F)'x}{S(F)w^\alpha(x;F^x)}\right\}w(x;F^x)x\right] + A_0(F)T_0(F), \tag{9.6}$$

where A_0 is given in (9.3).

LEMMA 9. *Suppose w satisfies Assumption* A *for some $\kappa \geq 1$ and ψ is odd and satisfies Assumption* B. *Suppose the distribution of y under F_0 is symmetric about zero. Assume $b_{T_0}(0+) = b_S(0+) = b_m(0+) = b_C(0+) = 0$.*

(i) *If $E_0\{w_0(x)^{1-\alpha}\|x\|^2\} < \infty$, then*

$$\|D_0(F_\varepsilon)\| \leq b_{T_0}(\varepsilon)\omega(\varepsilon)\|\dot{\psi}\|_{\sup} E_0\{w_0(x)^{1-\alpha}\|x\|^2\} \tag{9.7}$$

and $\|D_0(F_\varepsilon)\| = o(b_{T_0}(\varepsilon))$ as $\varepsilon \downarrow 0$.

(ii) *If $E_0\{w_0(x)^{1-2\alpha}\|x\|^3\} < \infty$, then*

$$\|D_0(F_\varepsilon)\| \leq \frac{b^2_{T_0}(\varepsilon)\omega(\varepsilon)\|\ddot{\psi}\|_{\sup}}{S_0 - b^-_S(\varepsilon)} E_0\{w_0(x)^{1-2\alpha}\|x\|^3\}, \tag{9.8}$$

and $\|D_0(F_\varepsilon)\| = o\{b^2_{T_0}(\varepsilon)\}$ as $\varepsilon \downarrow 0$.

PROOF OF THEOREM 2. Let $T_\varepsilon = T_0(F_\varepsilon)$, $S_\varepsilon = S(F_\varepsilon)$, $m_\varepsilon = m(H_\varepsilon)$, $C_\varepsilon = C(H_\varepsilon)$, and $w_\varepsilon(x) = w(\|x - m_\varepsilon\|_{C_\varepsilon})$. Use (4.3), (9.3) and (9.6) to write

$$\begin{aligned} T_1(F_\varepsilon) &= T_\varepsilon + (1-\varepsilon)S_\varepsilon R(F_\varepsilon) E_0\left[\psi\left\{\frac{y - T'_\varepsilon x}{S_\varepsilon w^\alpha_\varepsilon(x)}\right\} w_\varepsilon(x) x\right] \\ &\quad + \varepsilon S_\varepsilon R(F_\varepsilon) E_{F^*}\left[\psi\left\{\frac{y - T'_\varepsilon x}{S_\varepsilon w^\alpha_\varepsilon(x)}\right\} w_\varepsilon(x) x\right] \\ &= \{I - (1-\varepsilon)R(F_\varepsilon)A_0(F_\varepsilon)\}T_\varepsilon + (1-\varepsilon)R(F_\varepsilon)D_0(F_\varepsilon) \\ &\quad + \varepsilon S_\varepsilon R(F_\varepsilon) E_{F^*}\left[\psi\left\{\frac{y - T'_\varepsilon x}{S_\varepsilon w^\alpha_\varepsilon(x)}\right\} w_\varepsilon(x) x\right]. \end{aligned} \tag{9.9}$$

By the definition of $R(F_\varepsilon)$ and Lemma 6, there is a distribution F^{**}, which depends on F_ε, such that $\|x\|$ has finite expectation and

$$\left\|R(F_\varepsilon) - (1-\varepsilon)^{-1}\{A_0(F_\varepsilon) + B_\varepsilon(F^{**})\}^{-1}\right\| \leq \varepsilon, \tag{9.10}$$

where

$$B_\varepsilon(F) = \frac{\varepsilon}{(1-\varepsilon)} E_F\left[\dot{\psi}\left\{\frac{y - T'_\varepsilon x}{S_\varepsilon w^\alpha_\varepsilon(x)}\right\} w_\varepsilon(x)^{1-\alpha} x x'\right]. \tag{9.11}$$

Using (9.11),

$$\begin{aligned} &\|I - (1-\varepsilon)R(F_\varepsilon)A_0(F_\varepsilon)\| \\ &\quad \leq \varepsilon(1-\varepsilon)\|A_0(F_\varepsilon)\| + \left\|\{A_0(F_\varepsilon) + B_\varepsilon(F^{**})\}^{-1} B_\varepsilon(F^{**})\right\| \\ &\quad \leq \varepsilon(1-\varepsilon)\|A_0(F_\varepsilon)\| + \min\left[1,\ \frac{\varepsilon\|\dot{\psi}\|_{\sup} \sup_x\{w_\varepsilon(x)^{1-\alpha}\|x\|^2\}}{(1-\varepsilon)\lambda_{\min}\{A_0(F_\varepsilon)\}}\right] \\ &\quad = A_1(\varepsilon), \text{ say.} \end{aligned} \tag{9.12}$$

Thus, the norm of the first term on the right in (9.9) is no larger than $A_1(\varepsilon)b_{T_0}(\varepsilon)$. If Assumption A holds with $\kappa(1-\alpha) \geq 2$, then $A_1(\varepsilon) = O(\varepsilon)$ as $\varepsilon \downarrow 0$. Otherwise we merely have $A_1(\varepsilon) = 1 + O(\varepsilon)$.

The representation in (9.10) also implies that

$$\|R(F_\varepsilon)\| \le \varepsilon + (1-\varepsilon)^{-1}\|A_0(F_\varepsilon)^{-1}\| = \varepsilon + \frac{1}{(1-\varepsilon)\lambda_{\min}\{A_0(F_\varepsilon)\}}. \tag{9.13}$$

Using Lemmas 7, 8 and 9(i), we have $(1-\varepsilon)\|R(F_\varepsilon)D_0(F_\varepsilon)\| \le A_2(\varepsilon)b_{T_0}(\varepsilon)$, where

$$A_2(\varepsilon) \le \left\{\varepsilon(1-\varepsilon) + \frac{1}{a_0(\varepsilon)}\right\}\omega(\varepsilon)\|\dot{\psi}\|_{\sup} E_0\{w_0(x)^{1-\alpha}\|x\|^2\}, \tag{9.14}$$

and $\lim_{\varepsilon\downarrow 0} A_2(\varepsilon) = 0$. If $E_0\{w_0(x)^{1-2\alpha}\|x\|^3\} < \infty$, then Lemma 10(ii) implies the accelerated convergence $A_2(\varepsilon) = o\{b_{T_0}(\varepsilon)\}$ as $\varepsilon \downarrow 0$.

It remains to bound the last term on the right side of (9.9). Using (9.10) again,

$$\begin{aligned} &\left\| R(F_\varepsilon) E_{F^*}\left[\psi\left\{\frac{y - T'_\varepsilon x}{S_\varepsilon w_\varepsilon^\alpha(x)}\right\} w_\varepsilon(x) x\right]\right\| \\ &\quad \le \|\psi\|_{\sup} \sup_x\{\|w_\varepsilon(x)A_0(F_0)^{-1}x\|\} \\ &\qquad \times \left\{\varepsilon\|A_0(F_0)\| + \frac{1}{1-\varepsilon}\left\|\{A_0(F_\varepsilon) + B_\varepsilon(F^{**})\}^{-1}A_0(F_0)\right\|\right\}. \end{aligned} \tag{9.15}$$

Moreover,

$$\left\|\{A_0(F_\varepsilon) + B_\varepsilon(F^{**})\}^{-1}A_0(F_0)\right\| \le 1 + \frac{\|A_0(F_\varepsilon) - A_0(F_0)\|}{a_0(\varepsilon)} \tag{9.16}$$

and

$$\begin{aligned} \sup_x \|w_\varepsilon(x)A_0(F_0)^{-1}x\| &\le \sup_x \|w_0(x)A_0(F_0)^{-1}x\| \\ &\quad + \frac{\sup_x |w_\varepsilon(x) - w_0(x)|\|x\|}{\lambda_{\min}\{A_0(F_0)\}}. \end{aligned} \tag{9.17}$$

Combining (9.15)–(9.17) shows that the last term in (9.9) is bounded by $\varepsilon B(\varepsilon)\gamma^*_{GM}$, where

$$\begin{aligned} B(\varepsilon) = \{1 + S_0^{-1}b_S^+(\varepsilon)\}&\left[1 + \frac{\sup_x \|\{w_\varepsilon(x) - w_0(x)\}A_0^{-1}x\|}{\sup_x \|w_0(x)A_0^{-1}x\|}\right] \\ &\times \left\{1 + \varepsilon\|A_0\| + \frac{\varepsilon + \|A_0(F_\varepsilon) - A_0\|}{(1-\varepsilon)a_0(\varepsilon)}\right\}. \end{aligned}$$

Proposition 1 and Lemma 9.7 imply that $B(\varepsilon)$ is finite if $0 \le \varepsilon < \min(\varepsilon^*_{T_0}, \varepsilon^*_S, \varepsilon^*_m, \varepsilon^*_C)$, and $\lim_{\varepsilon\downarrow 0} B(\varepsilon) = 0$. Setting $A(\varepsilon) = A_1(\varepsilon) + A_2(\varepsilon)$ completes the proof for general p.

In Theorem 2(ii), the slight improvement for $p = 1$ occurs because in this case $|\{A_0(F_\varepsilon) + B_\varepsilon(F^{**})\}^{-1}B_\varepsilon(F^{**})| = 1 - |\{A_0(F_\varepsilon) + B_\varepsilon(F^{**})\}^{-1}A_0(F_\varepsilon)|$. □

PROOF OF THEOREM 3. For each $\beta \in R^p$, let $F_\beta(x, y) = F_0(x, y - \beta' x)$ and let E_β denote expectation under F_β. Because of the symmetry,

$$E_\beta[\dot{\tau}_c\{w_c(\|x\|)(y - \beta' x)\}w_c(\|x\|)x] = 0.$$

Differentiating with respect to β and setting $\beta = 0$ yields the identity

$$\begin{aligned} A_c &= E_0\big[\dot{\tau}_c\big\{w_c(\|x\|)y\big\}w_c^2(\|x\|)xx'\big] \\ &= -E_0\big[\tau_c\big\{w_c(|x|)y\big\}\phi(y)w_c(\|x\|)xx'\big]. \end{aligned} \tag{9.18}$$

Next observe that for each nonzero x and y

$$\lim_{c\downarrow 0} w_c(\|x\|) = \|x\|^{-1}, \qquad \lim_{c\downarrow 0} c^{-1}\tau_c\big\{w_c(\|x\|)y\big\} = \operatorname{sign}(y), \tag{9.19}$$

and

$$\big|c^{-1}\tau_c\big\{w_c(\|x\|)y\big\}w_c(\|x\|) - \operatorname{sign}(y)\|x\|^{-1}\big| \le 2\|x\|^{-1}. \tag{9.20}$$

By assumption,

$$E_0\big\|\phi(y)\|x\|^{-1}xx'\big\| = E_0|\phi(y)|E_0\|x\| < \infty. \tag{9.21}$$

Using (9.18)–(9.21) we obtain

$$\lim_{c\downarrow 0} c^{-1}A_c = -E_0\{\operatorname{sign}(y)\phi(y)\}E_0(\|x\|^{-1}xx') = -E_0|\phi(y)|E_0(\|x\|^{-1}xx').$$

The spherical symmetry implies that

$$E_0(\|x\|^{-1}xx') = c_p I_p \tag{9.22}$$

for some scalar c_p. Taking the trace on both sides of (9.22) shows that $c_p = p^{-1}E_0\|x\|$.

To establish the convergence of γ_c^*, first observe that $\gamma_c^* \ge p\{E_0|\phi|E_0\|x\|\}^{-1}$ by Theorem 2.2 of He and Simpson (1993) because γ_c^* is the sensitivity of a locally linear functional. On the other hand,

$$\gamma_c^* = \sup_x \frac{\|cA_c^{-1}x\|}{c + \|x\|} = \sup_{\|x\|\ne 0} \frac{\|cA_c^{-1}x\|}{c + \|x\|} \le \sup_{\|x\|\ne 0} \frac{\|cA_c^{-1}x\|}{\|x\|} = \lambda_{\max}(cA_c^{-1}).$$

Moreover, by continuity, $\lim_{c\downarrow 0} \lambda_{\max}(cA_c^{-1}) = p\{E_0|\phi(y)|E_0\|x\|\}^{-1}$. □

REFERENCES

BICKEL, P. J. (1975). One–step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434.

BICKEL, P. J. (1981). Quelque aspects de la statistique robuste. *École d'Été de Probabilités de St. Flour* IX. *Lecture Notes in Math.* **876** 2–68. Springer, Berlin.

CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836.

COAKLEY, C. W. and HETTMANSPERGER, T. P. (1993). A bounded influence, high breakdown, efficient regression estimator. *J. Amer. Statist. Assoc.* **88** 872–880.

DAVIES, P. L. (1987). Asymptotic behaviour of *S*-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DAVIES, L. (1990). The asymptotics of S-estimators in the linear regression model. *Ann. Statist.* **18** 1651–1675.

DAVIES, P. L. (1992). An efficient Fréchet differentiable high breakdown multivariate location and dispersion estimator. *J. Multivariate Anal.* **40** 311–327.

DAVIES, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.* **21** 1843–1899.

DONOHO, D. L. and LIU, R. C. (1988). The "automatic" robustness of minimum distance functionals. *Ann. Statist.* **16** 552–586.

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87** 998–1004.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **62** 1179–1186.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

HE, X. (1989). Contributions to the theory of statistical breakdown. Ph.D. dissertation, Dept. Statistics, Univ. Illinois, Urbana-Champaign.

HE, X. and SIMPSON, D. G. (1992). Robust direction estimation. *Ann. Statist.* **20** 351–369.

HE, X. and SIMPSON, D. G. (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *Ann. Statist.* **21** 314–337.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

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

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

JUREČKOVÁ, J. and MALÝ, M. (1995). The asymptotics for studentized k-step M-estimators of location. *Sequential Anal.* **14** 229–245.

JUREČKOVÁ, J. and PORTNOY, S. (1987). Asymptotics for one-step M-estimators in regression with application to combining efficiency and high breakdown point. *Comm. Statist. Theory Methods* **16** 2187–2199.

KENT, J. T. and TYLER, D. E. (1996). Constrained M-estimation for multivariate location and scatter. *Ann. Statist.* **24** 1346–1370.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.* **77** 595–604.

LOPUHAÄ, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17** 1662–1663.

MALLOWS, C. L. (1975). On some topics in robustness. Technical memorandom, Bell Laboratories, Murray Hill, NJ.

MARONNA, R. A., BUSTOS, O. H. and YOHAI, V. J. (1979). Bias- and efficiency-robustness of general M-estimators for regression with random carriers. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.). Springer, New York.

MARONNA, R. A., STAHEL, W. A. and YOHAI, V. J. (1992). Bias-robust estimators of multivariate scatter based on projections. *J. Multivariate Anal.* **42** 141–161.

MARONNA, R. A. and YOHAI, V. J. (1993). Bias-robust estimates of regression based on projections. *Ann. Statist.* **21** 965–990.

MARTIN, R. D., YOHAI, V. J. and ZAMAR, R. H. (1989). Min-max bias robust regression. *Ann. Statist.* **17** 1608–1630.

MARTIN, R. D. and ZAMAR, R. H. (1989). Asymptotically min-max bias robust M-estimates of scale for positive random variables. *J. Amer. Statist. Assoc.* **84** 494–501.

ROCKE, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Ann. Statist.* **24** 1327–1345.

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

ROUSSEEUW, P. J. and CROUX, C. (1994). Bias of k-step M-estimators. *Statist. Probab. Lett.* **20** 411–420.

ROUSSEEUW, P. J. and LEROY, A. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

ROUSSEEUW, P. J. and YOHAI, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Hardle and R. D. Martin, eds.) 256–272. Springer, New York.

ROUSSEEUW, P. J. and VAN ZOMEREN, B. C. (1990). Unmasking multivariate outliers and leverage points (with discussion). *J. Amer. Statist. Assoc.* **85** 633–651.

RUPPERT, D. and SIMPSON, D. G. (1990). Comments on "Unmasking multivariate outliers and leverage points," by P. J. Rousseeuw and B. C. van Zomeren. *J. Amer. Statist. Assoc.* **85** 644–646.

SIMPSON, D. G. and CHANG, Y. C. (1997). Reweighting approximate *GM* estimators: asymptotics and residual-based graphics. *J. Statist. Plann. Inference* **57** 273–294.

SIMPSON, D. G., RUPPERT, D. and CARROLL, R. J. (1992). On one-step *GM*-estimates and stability of inferences in linear regression. *J. Amer. Statist. Assoc.* **87** 439–450.

SIMPSON, D. G. and YOHAI, V. J. (1993). Functional stability of one-step *GM* estimators in linear regression. Technical report 71, Dept. Statistics, Univ. Illinois, Urbana-Champaign.

YOHAI, V. and ZAMAR, R. (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83** 406–413.

YOHAI, V. and ZAMAR, R. (1997). Optimally bounding a generalized gross error sensitivity of unbounded influence M-estimates of regression. *J. Statist. Plann. Inference* **57** 233–244.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
101 ILLINI HALL
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820
E-MAIL: dgs@unic.edu

DEPARTAMENTO DE MATEMATICAS
CIUDAD UNIVERSITARIA PABELLON 1
1426 BUENOS AIRES
ARGENTINA
E-MAIL: vyohai@dm.uba.ar