BOUNDED INFLUENCE REGRESSION IN THE PRESENCE OF HETEROSKEDASTICITY OF UNKNOWN FORM

Miguel A. DELGADO
Department of Economics
Indiana University
Bloomington, IN 47405 USA

ABSTRACT. In a regression model with conditional heteroskedasticity of unknown form, we propose a general class of M-estimators scaled by nonparametric estimates of the conditional standard deviations of the dependent variable. We give regularity conditions under which these estimators are asymptotically equivalent to M-estimators scaled by the true conditional standard deviations. The practical performance of these estimators is investigated through a Monte Carlo experiment.

1. Introduction

Cross-sectional data are typically heteroskedastic, often containing aberrant observations or gross errors. Under these circumstances, the least squares estimation method produces inefficient estimates. It is difficult to determine the source of inefficiency. As Huber (1973) pointed out:

> 'In the regression case, uncontrollable inhomogeneity of variance among the disturbances and genuinely long tailed error distributions have almost indistinguishable effects both impairing the efficiency of the estimates'.

Hence, it seems worthwhile to correct for heteroskedasticity using robust regression analysis. Carroll and Ruppert (1982) considered M-estimators scaled differently for each observation. The scale estimators were obtained under prior information on the functional form of the heteroskedasticity. Simulations reported by Carroll and Ruppert (1982) are encouraging. These estimators are expected to be nonrobust when the parameterization of the heteroskedasticity is incorrect. Asymptotically efficient estimators in the presence of unknown heteroskedasticity have been obtained by Carroll (1982), Robinson (1987) and Delgado (1989). However, these estimators have an unbounded influence function to residuals and leverage.

In this paper we establish the asymptotic properties of M-estimators with bounded influence function in the presence of heteroskedasticity of unknown nature. The conditional scale estimates are consistent nonparametric estimates of the conditional standard deviation of the dependent variable. The rest of the paper is organized as follows. In the next section we discuss a general class of optimal M-estimators which scale by the true conditional standard deviation of the dependent variable. In section 3 we present our M-estimators.

In section 4 we give conditions under which these M-estimators are adaptive. In section 5 we report the results of a Monte Carlo experiment and in section 6 we sumarize our conclusions. Proofs are confined to the appendix.

2. Optimal Bounded Regression in the Presence of Heteroskedasticity

We consider the usual regression model. We have independent observations $\{(Y_i, X_i), 1 \le i \le n\}$ from a $\mathbb{R} \times^p \mathbb{R}$ random variable (Y, X) where,

$$Y_i = X_i' \theta^o + \varepsilon_i \, \sigma(X_i); \; i \ge 1 \qquad (1.2)$$

and θ^o is a px1 vector of unknown parameters, $\sigma(.)$ is an unknown function and ε_i are, conditionally on X_i, symmetric about zero with $Var(\varepsilon_i | X_i) = 1$. The class of weighted least squares estimators is defined as,

$$\hat{\theta}_n(a) = \left\{ \sum_i X_i X_i' \, a_i^{-2} \right\}^{-1} \sum_i X_i Y_i \, a_i^{-2} \qquad (2.2)$$

where $\{a_i, 1 \le i \le n\}$ are suitable weights. Under regularity conditions, $n^{1/2}[\hat{\theta}_n(a) - \theta^o_2]$'s limiting covariance matrix has lower Gauss-Markov bound $\Phi_o = \{E[XX'\sigma^{-2}(X)]\}^{-1}$, and it is achieved by the unfeasible generalized least squares estimator $\hat{\theta}_n(\sigma)$, where $\sigma_i = \sigma(X_i)$. When the errors are conditionally normally distributed, Φ_o is the Cramer-Rao bound, but the normal model is never exactly true. In the presence of departures from the normality hypothesis, $\hat{\theta}_n(\sigma)$ may be very inefficient. On the other hand outlying X_i observations can adversely affect estimators such as (2.2). Therefore, it seems reasonable to consider estimators which bound the influence of the data.

Maronna and Yohai (1981) proved the asymptotic properties of a very general class of M-estimators of location and scale, implicitly defined as the simultaneous solution to,

$$\sum_i \phi(X_i, (Y_i - X_i'\theta)/\sigma) \, X_i = 0 \quad \text{and} \quad \sum_i \chi(X_i, (Y_i - X_i'\theta)/\sigma) = 0,$$

where $E[\phi(X, (Y - X'\theta^o)/\sigma) \, X | X] = 0$ and $E[\chi(X, (Y - X'\theta^o)/\sigma) | X] = 0$. All known proposals for $\phi(.,.)$ (see Hampel et. al 1986) may be written in the form,

$$\phi(X, r) = w(X) \, \psi(r \, v(X)) \qquad (2.4)$$

for appropriate functions $\psi: \mathbb{R} \to \mathbb{R}$ and weight functions $w: \mathbb{R}^p \to \mathbb{R}^+$ and $v: \mathbb{R}^p \to \mathbb{R}^+$. The function $\psi(.)$ bounds the influence of residuals and w(.) and v(.), the influence of leverage. Relles (1968) and Huber (1973), uses w(X)=1, v(X)= 1. There are a large number of ψ-function proposals. A popular proposal is the Huber's ψ-function, i.e. $\psi(u) = u \min\{1, c/|u|\}$, where $c > 0$ is an appropriate chosen constant. In these cases, the corresponding M-estimators have bounded residual influence but the influence of leverage is unbounded. For a discussion on different choices of w(.) and v(.), see Hampel et. al (1986). Under conditional heteroskedasticity, estimators as (2.3) are not scale invariant. Carroll and Ruppert (1982) named $\breve{\theta}(\sigma)$, the optimal M-estimator (OME) under heteroskedasticity, where

$$Q_n\{\check{\theta}_n(a),\ a\}= 0, \tag{2.5}$$

and $Q_n\{\theta,\ a\}= \sum_i \phi(X_i,\ (Y_i- X_i'\theta)/a_i)\ X_i/\ a_i$.

The conditions for asymptotic normality of the OME are very similar to those given in Maronna and Yohai (1981). Let us introduce the following notation,

$\lambda(\Delta)= E\{X\ \sigma(X)^{-1}\ \phi(X,\ \varepsilon- \Delta' X\ \sigma(X)^{-1})\}$, $L(X)= \sup_U|\phi(X,\ U)|\ \|X\|$ and

$H(X)= \sup_U|\phi'(X,\ U)|$ where $\phi'(X,\ U)= \partial\phi(X,\ U)/\partial U$. We assume,

N1.- For each X, $\phi(X,\ .)$ is odd, uniformly continuous, nondecreasing, $\phi(X,\ U)> 0$ for $U> 0$ and the conditional distribution of ε is symmetric about zero.

N2.- $\Pr\{\sigma(X)\le \delta\}= 0$ for some $\delta> 0$.

N3.- $\lambda(\Delta)$ has a nonsingular derivative at 0, $-\ V(\sigma)$ say, (that is $|\lambda(\Delta)- \lambda(0)+ V(\sigma)\ \Delta|= o(\|\Delta\|)$).

N4.- $E\{H(X)\ \|X\|^2\ L(X)\}< \infty$.

N5.- $E\{H(X)\ \|X\|^2\}< \infty$.

N6.- $E\|X\|< \infty$.

N7.- $E\{L(X)^2\}< \infty$.

Condition N1 implies that $E\{\phi[X_1,\ \varepsilon_1 S(X_1)]|X_1\}= 0$ for any function $S(X_1)$ depending on X_1 $i\ge 1$. Condition N2 is required by Robinson (1987) and Delgado (1989) for the adaptation proof of the semiparametric weighted least squares estimator (SWLSE). Alternatively, N2 may be removed by multiplying $\|X\|$ by $\sigma(X)^{-1}$ in N4-N7. We have found convenient to set out the conditions in this way in order to make comparisons with conditions needed in the next section. Conditions N3-N7 are required by Maronna and Yohai (1981). N3 and N7 guarantee that the asymptotic covariance matrix of the OME is positive definite (p.d). Note that for bounded influence estimators, as those with ϕ-functions as (2.4), L(X) is bounded. However, for classical M-estimators (i.e. $w(X)=v(X)=1$), N4 implies that $E\|X\|^3< \infty$ which seems a quite strong requirement. A normality proof in this case is possible, assuming $E\|X\|^2< \infty$, by using the results in Yohai and Maronna (1979). When, in (2.4), $\|X\|\ w(X)$, $\psi(.)$ and $\psi'(.)$ are bounded, N4-N7 holds if $E\|X\|^2< \infty$.

Theorem 1.- If N1-N7 hold,

$$n^{1/2}[\check{\theta}_n(\sigma)- \theta^o] \xrightarrow{d} N(0,\ \Phi_1)$$

where, $\Phi_1= V(\sigma)^{-1}\ E[\phi(X,\ \varepsilon)^2\ XX'/\sigma^2(X)]\ V(\sigma)^{-1}$.

Proof.- See Appendix.

The purpose of this paper is to obtain estimators first order asymptotically equivalent to $\check{\theta}_n(\sigma)$. We propose to estimate $\sigma(X_1)$ by nonparametric regression.

3.- Bounded Influence Function Semiparametric Estimators

When the functional form of $\sigma(X_1)$ is known and a preliminary root-n-consistent estimator of θ^o, $\tilde{\theta}_n$ say, is available, consistent estimators of

$\sigma^2(X_i) = [\sigma(X_i)]^2$ are obtained by regressing $(Y_i - X_i' \tilde{\theta}_n)^2$ against the known design of $\sigma^2(X_i)$. Under regularity conditions, the corresponding weighted least squares estimators (WLSE) of θ^o are as first order efficient as $\hat{\theta}_n(\sigma)$.

Carroll and Ruppert (1982) suggested the use the of consistent parametric weights to construct robust estimators such as those defined in (2.5) but using $w(X)=v(X)=1$ and X fixed. They proved that these estimators are, under regularity conditions, asymptotically equivalent to $\breve{\theta}_n(\sigma)$. These estimators are nonrobust with respect to the assumed parameterization of $\sigma^2(X)$.

Rose (1978) proposed several nonparametric estimators of $\sigma^2(X_i)$. One of them is obtained by performing a nonparametric regression of $(Y_i - X_i \tilde{\theta}_n)^2$ against the regressors on which $\sigma^2(X_i)$ is known to depend. Carroll (1982) and Robinson (1987) proved, under different regularity conditions, that the corresponding SWLSE asymptotically achieves the Gauss-Markov bound. Carroll (1982) used kernel regression while Robinson (1987) used k nearest neighbors (k-nn) regression.

In this paper we follow Robinson's approach. If $\sigma(X)$ is known to depend on a subset dx1 ($d \le p$) vector $\mathring{X}_i = (\mathring{X}_{1i}, \ldots, \mathring{X}_{di})'$ of nondegenerate elements of X_i, and given a positive integer $k = k(n)$, the sequence of k-nn nonparametric weights $\{\omega_{ij}(k),\ i,j=1,\ldots,n\}$ is defined by

$$\omega_{ij}(k) = 1(i\neq j)\ r_{ij}^{-1} \quad \sum_{\tau=p_{ij}}^{p_{ij}+r_{ij}-1} c_\tau(k) \qquad 1\le i,j\le n \tag{3.1}$$

where for $1\le i\le n$, $c_i(k) > 0$, $1\le i\le k$; $c_i(k)=0$ $i>k$; $\sum_{i=1}^{k} c_i(k) = 1$, and

$$p_{ij} = 1 + \sum_t 1(\rho_{it} < \rho_{ij}); \qquad r_{ij} = 1 + \sum_t 1(\rho_{it} = \rho_{ij}) \quad i\neq j$$

where the sums are over $1\le t\le n$, $t\neq i,j$ and $1(.)$ is the indicator function.

$$\rho_{ij} = \Sigma_{m=1}^{d} (\mathring{X}_{im} - \mathring{X}_{jm})^2 / s_m^2, \qquad s_m^2 = (n-1)^{-1} \Sigma_i \{\mathring{X}_{im} - [n^{-1}\Sigma_j \mathring{X}_{jm}]\}^2.$$

The uniform weights $c_i(k) = k^{-1}$, $i=1,\ldots,k$, satisfy these conditions. Other weights satisfying these conditions can be found in Stone (1977). These k-nn weights do not use the own observation. This sample splitting is not required for the consistency of the k-nn weights but it is technically convenient in semiparametric estimation. This sample splitting technique has also been employed by Robinson (1987) and Delgado (1989). Given a preliminary root-n-consistent estimator of θ^o, $\tilde{\theta}_n$ say, a consistent estimate of σ_i^2 is,

$$\hat{\sigma}_i^2 = \sum_j (Y_j - X_j'\tilde{\theta}_n)^2\ \omega_{ij}(k). \tag{3.2}$$

We suggest estimating θ^o by the semiparametric weighted M-estimator (SWME), $\breve{\theta}_n(\hat{\sigma})$ where $\hat{\sigma}_i = [\hat{\sigma}_i^2]^{1/2}$. Unlike $\hat{\theta}_n(\sigma)$ and $\hat{\theta}_n(\hat{\sigma})$, $\breve{\theta}_n(\hat{\sigma})$ has bounded influence function.

One would expect that the higher order efficiency of $\breve{\theta}_n(\hat{\sigma})$ will improve by using an iterative procedure (i.e. computing new $\hat{\sigma}_i$ at each iteration). A full iterated SWLSE is obtained at once by using the following pure nonparametric estimator of $\sigma^2(X_i)$,

$$\breve{\sigma}_i^2 = \sum_j Y_j^2\ \omega_{ij}(k) - \{\sum_j Y_j\ \omega_{ij}(k)\}^2, \tag{3.3}$$

Delgado (1989) proved, under Robinson (1987) regularity conditions, that $n^{1/2}[\hat{\theta}_n(\breve{\sigma}) - \hat{\theta}_n(\sigma)] = o_p(1)$. We propose to use also $\breve{\theta}_n(\breve{\sigma})$ as a possible alternative to $\breve{\theta}_n(\hat{\sigma})$ where $\breve{\sigma}_1 = [\breve{\sigma}_1^2]^{1/2}$.

4.- Asymptotic Theory

Robinson (1987) noted that it is technically convenient to relate the moment conditions on (Y, X) and the rate of convergence of k. We assume that,

K1.- $\overline{\lim}\ \max_i c_i(k)\ k < \infty$ as $n \to \infty$.

K2.- $k^{-\upsilon/4} n \longrightarrow 0$ and $n^{-1}k \longrightarrow 0$ as $n \longrightarrow \infty$ for $\upsilon > 4$.

Besides N1-N3, we need stronger moment conditions in N4-N7. In particular for the υ in K2 we need,

R1.- (a) $E|H(X)\|X\|^2\ L(X)|^{\upsilon/(\upsilon-2)} < \infty$ and (b) $E|H(X)\|X\|\ L(X)|\varepsilon\ \sigma(X)||^{\upsilon/(\upsilon-2)} < \infty$.

R2.- (a) $E|H(X)\ \|X\|^2|^{\upsilon/(\upsilon-2)} < \infty$ and (b) $E|H(X)\ \|X\|\ |\varepsilon\ \sigma(X)||^{\upsilon/(\upsilon-2)} < \infty$.

R3.- $E\|X\|^{2\upsilon/(\upsilon-2)} < \infty$.

R4.- $E|L(X)|^{2\upsilon/(\upsilon-2)} < \infty$.

R5.- $E|\varepsilon\ \sigma(X)|^{\upsilon} < \infty$.

The difference between assumptions N and R are similar as in the weighted least squares estimator case. The asymptotic normality of $\hat{\theta}_n(\sigma)$ needs that $E\|X\|^2 < \infty$ and the SWLSE of Robinson (1987) needs $E\|X\|^{2\upsilon/(\upsilon-2)} < \infty$. This last condition implies R1-R4 when, in (2.4), $w(X)\ \|X\|$, $\psi(.)$ and $\psi'(.)$ are bounded. R3 and R5 are needed in Robinson (1987), Newey (1987) and Delgado (1989).

Theorem 2.- If N1-N3, K1, K2 and R1-R5 hold:

(a) $$n^{1/2}[\breve{\theta}_n(\hat{\sigma}) - \theta^o] \xrightarrow{d} N(0,\ \Phi_1),$$

(b) $$n^{1/2}[\breve{\theta}_n(\breve{\sigma}) - \theta^o] \xrightarrow{d} N(0,\ \Phi_1).$$

Proof.- See Appendix.

Maronna and Yohai (1981) recommended to construct interval estimators, estimating the asymptotic variance by its natural sample analog. In our case, Φ_1 is estimated by,

$$\hat{\Phi}_1 = \hat{V}_n(\hat{\sigma})^{-1}\ n^{-1}\textstyle\sum_i\ X_iX_i'\ \hat{\sigma}_i^{-2}\ \phi^2\left(X_i,\ [Y_i - X_i'\breve{\theta}_n(\hat{\sigma})]\ \hat{\sigma}_i^{-1}\right)\ \hat{V}_n(\hat{\sigma})^{-1} \tag{4.1}$$

where $\hat{V}_n(\hat{\sigma}) = n^{-1}\sum_i\ X_iX_i'\ \hat{\sigma}_i^{-2}\ \phi'\left(X_i,\ [Y_i - X_i'\breve{\theta}_n(\hat{\sigma})]\ \hat{\sigma}_i^{-1}\right)$.

We can substitute $\hat{\sigma}_1$ by $\breve{\sigma}_1$ in (4.1).

5.- Monte Carlo

The experiments follow the model

$$Y_i = \theta_1^o + \theta_2^o\ X_i + \varepsilon_i\sigma(X_i),\ i \geq 1 \tag{5.1}$$

with $\theta_1^0 = \theta_2^0 = 1.$ and X_i~iid Uniform(0, 2). The ε_i's were generated iid and independent of X_i as follows,

CONTAMINATED.- ε_i~ iid {.9 N(0, 1) + .1 N(0, 9)}/$\sqrt{1.8}$.

NORMAL.- ε_i~iid N(0, 1).

Note that Var(ε)=1 in all models. The residual conditional variances are constructed according to the models,

$$\text{Model 1.- } \sigma(X_i)= \exp(\gamma\ X_i)$$

$$\text{Model 2.- } \sigma(X_i)= |1.+ X|^{\delta}.$$

We have only used uniform weights; i.e. in (3.1) c_i= 1/k for all i≤ k. We report two choices of k, $\hat{\theta}_n(\hat{\sigma})^1$, $\hat{\theta}_n(\check{\sigma})^1$, $\check{\theta}_n(\hat{\sigma})^1$, $\check{\theta}_n(\check{\sigma})^1$ are computed with k= $[n^{1/2}]$, while $\hat{\theta}_n(\hat{\sigma})^2$, $\hat{\theta}_n(\check{\sigma})^2$, $\check{\theta}_n(\hat{\sigma})^2$, $\check{\theta}_n(\check{\sigma})^2$ are computed with k= $[n^{2/3}]$.

We compare the unfeasible estimator $\hat{\theta}_n(\sigma)$, the OLSE $\hat{\theta}_n(1)$, the SWLSE's $\hat{\theta}_n(\hat{\sigma})$ and $\hat{\theta}_n(\check{\sigma})$, the unfeasible OME $\check{\theta}_n(\sigma)$, the M-estimator with fixed scale $\check{\theta}_n(1)$ and the SWME's $\check{\theta}_n(\hat{\sigma})$ and $\check{\theta}_n(\check{\sigma})$. In order to save space we only report results for the slope coefficient. We have only considered the classical Huber's estimator, i.e. $\psi(U)$= U min{1, c/|U|} in (2.4), where we choose c= 1.345. The robust estimators were computed using reweighted least squares. This c produces, under normality, $\check{\theta}_n(\sigma)$ 95% as efficient as $\hat{\theta}_n(\sigma)$.We also report results for the least absolute deviation estimator (LADE) without scaling. This estimator is used as the starting point in the reweighted procedure for the Huber's estimators and for the estimation of the residuals in order to compute $\hat{\sigma}_i$.

The tables show the bias (BIAS), variance (VAR) and the relative efficiency (EFF) of the different estimators for sample sizes of n= 30, n= 100 and n= 500 with 10,000, 5,000 and 1,000 replications respectively. The efficiency is the ratio of the mean square error (MSE) of the estimator with respect to MSE of $\hat{\theta}_n(\sigma)$. We report results for the different disturbances and the heteroskedasticity models 1 and 2 with different parameter values for δ and γ which produce different degrees of heteroskedasticity. All the programs were written in FORTRAN-77 double precision and NAG-13 routines were used to generate the variates. The programs were run on the Indiana University VAXes.

The simulations strongly support the applicability of our theorem. Through the experiments, the SWME's are always more efficient than the SWLSE's. We observe important gains in efficiency of the semiparametric estimates as the sample size increases while the EFF of the others estimators is not significantly affected. It is obviously due to the fact that the nonparametric estimates of the residual variances become more accurate as the sample size increases. This also happens when working with parametric weights. Therefore, for the smallest sample size (n= 30) and when the heteroskedasticity is mild, the SWME's are, sometimes, more inefficient than the LADE. However, when n= 100 or 500, the SWME's EFF is typically greater than one under departures from normality. Samples sizes of 30 or 100 are small in a cross-section context where samples of several thousands of observations are common. In the normal case, the SWME's EFF is quite close to their asymptotic values when n=500 and always is greater than the SWLSE's EFF. It is observed, in general, that for similar degrees of heteroskedasticity, the semiparametric estimators based on $\hat{\sigma}_1$ behaves slightly better than those based on $\check{\sigma}_1$ when the conditional variances are small, while the latter perform better than the former when the conditional variances are larger. The choice of k does not seem to affect significantly the results. We proceed to discuss the tables in some detail.

Table 1 shows results for CONTAMINATED errors. Note that the contamination is quite moderate. Simulations using this distribution have been also reported by Huber (1973) and Carroll and Ruppert (1982). We have considered MODEL 1 and MODEL 2 with γ= 0, -1, -2 and δ= 1,3 5. So, in MODEL 1 we have small variances while in MODEL 2 the variances are larger. The severity of the heteroskedasticity is comparable in the two models when γ= -1, -2 and δ= 3, 5. When n=500 the semiparametric estimators efficiencies are fairly close to their asymptotic values. However, when n= 30, the semiparametric estimators are more inefficient than the LADE and OLSE, in some cases, especially when the heteroskedasticity is mild. When the heteroskedasticity is heavy, i.e. γ= -2 and δ= 5, the SWLSE's are more efficient than than the OLSE for n= 30, but they are more inefficient than the LADE in MODEL 1 (when γ= -2). As expected, the SWME's are always more efficient than the SWLSE's. However, the SWME's are generally more inefficient than the LADE for n=30 in MODEL 1. In MODEL 2, with n= 30, the SWME's appear to be more inefficient than the LADE only when δ= 0, 1 and k= $[n^{1/2}]$. When n=100, the SWLSE's are always more efficient than the OLSE but the SWLSE's are still more inefficient than the LADE in MODEL 1 and MODEL 2 for δ= 1,3. For n= 100, the SWME's are always more efficient than LADE, OLSE and SWLSE's. They, some times, appear to be more efficient than $\hat{\theta}_n$ (e.g. γ=0, -1, δ= 3), as the asymptotic theory predicts. This prediction is fully supported when n= 500. In this case, the SWME's EFF is always greater than 1 and the SWLSE's EFF is closer to 1, but they still behaves worse than the LADE when γ and δ are small. We have tried other distributions (e.g. Student and Laplace) with similar results.

In Table 2 we report results for NORMAL errors with MODEL 1 (γ= -2) and MODEL 2 (δ= 5). When n= 30, the LADE is more efficient than $\check{\theta}_n(\check{\sigma})$ and $\hat{\theta}_n(\check{\sigma})$ in MODEL 1. In the other cases, the SWME's are more efficient than LADE. As the sample size increases, the SWME's EFF approaches their asymptotic value (.95). Note that for n=500, the SWME's EFF are closer to their asymptotic value than the SWLSE's EFF. In fact the SWME's appear to be more efficient than the SWLSE's in all cases reported, though asymptotically it is 5% more inefficient. this fact is of practical importance. In Tables 3 and 4 we report results for LAPLACE and T_4 errors. The OME's EFF is smaller, in these cases, than in the CONTAMINATED case and, therefore, the SWME's when n= 500 are less spectacular. However, the differences in EFF between the SWME's and SWLSE's are larger than in the CONTAMINATED case. For n= 30, the SMWE's typically perform better than the LADE and always perform better than the SWLSE's.

6. Conclusions

We have seen that the introduction of robust methods in the estimation of semiparametric models is of practical relevance. The Monte Carlo reported shows that our method can be widely used, without great losses in efficiency when the data are close to normal but heteroskedastic.

APPENDIX

Proof of Theorem 1: In proving the theorem we do not need to assume that $\check{\theta}_n(\sigma)$ uniquely solves $Q_n(\theta, \sigma) = 0$ but rather that $\|Q_n(\check{\theta}_n(\sigma), \sigma)\|$ is majorized by twice the infimum of $\|Q_n(\theta, \sigma)\|$ over θ. Let define,

$$U_n(\Delta, a) = n^{-1/2} \sum_i \phi\left(X_i, (\varepsilon_i - n^{-1/2} \Delta' X_i \sigma_i^{-1})\sigma_i a_i^{-1}\right) X_i a_i^{-1}. \tag{a.1}$$

Since $Q_n(\check{\theta}_n(\sigma), \sigma) = U_n(n^{1/2}(\check{\theta}_n(\sigma) - \theta^o), \sigma) = 0$,

$$\|U_n(n^{1/2}(\check{\theta}_n(\sigma) - \theta^o), \sigma)\| \le 2\, \|U_n(\Delta^*, \sigma)\|, \tag{a.2}$$

where

$$\Delta^* = -V(\sigma)^{-1} U_n(0, \sigma) = O_p(1), \tag{a.3}$$

by N3 and since by Chevyshev's inequality $U_n(0, \sigma) = O_p\{[E\|U(0, \sigma)\|^2]^{1/2}\} = O_p(1)$ by N1 and N7. Then noting that, by the Lindeberg-Levy central limit theorem (clt), $\Delta^* \xrightarrow{d} N(0, \Phi_1)$, the theorem follows from (a.2) after establishing that,

$$\|U_n(\Delta^*, \sigma)\| = o_p(1), \tag{a.4}$$

$$n^{1/2}(\check{\theta}_n(\sigma) - \theta^o) = O_p(1). \tag{a.5}$$

We conclude (a.4) from (a.3) and

$$\sup_{\|\Delta\| \le M} \|U_n(\Delta, \sigma) - U_n(0, \sigma) + V(\sigma)\, \Delta\| = o_p(1). \tag{a.6}$$

(a.6) follows from N3 and,

$$\sup_{\|\Delta\| \le M} \|U_n(\Delta, \sigma) - U_n(0, \sigma) - E\{U_n(\Delta, \sigma)\}\| = o_p(1). \tag{a.7}$$

In order to prove (a.7), take for convenience M=1. For $0 < \delta < 1$ define,

$$S_n(\Delta, \sigma) = \sup_{\|\Delta^1 - \Delta\| \le \delta} \|U_n(\Delta^1, \sigma) - U_n(\Delta, \sigma)\|. \tag{a.8}$$

Since for each fixed δ we can cover the ball of radius 1 in $\mathbb{R}^p$ by a finite number of balls of radius δ, we conclude (a.7) from,

$$U_n(\Delta, \sigma) - U_n(0, \sigma) - E\left\{U_n(\Delta, \sigma) - U_n(0, \sigma)\right\} = o_p(1), \tag{a.9}$$

for each fixed Δ, and for all $0 < \delta < 1$, all n and all $\|\Delta\| \le 1$,

$$E\left\{S_n(\Delta, \sigma)\right\} \le \delta\, O(1), \tag{a.10}$$

$$S_n(\Delta, \sigma) - E\left\{S_n(\Delta, \sigma)\right\} = o_p(1), \tag{a.11}$$

(a.9) follows from,

$$\mathrm{Var}\left\{U_n(\Delta,\sigma)-U_n(0,\sigma)\right\}=n^{-1}\sum_i \mathrm{Var}\|X_i\sigma_i^{-1}\,\phi\left(X_i,(\varepsilon_i-n^{-1/2}\Delta' X_i\,\sigma_i^{-1})\right)\|$$

$$\le n^{-1}\sum_i E\|X_i\sigma_i^{-1}\,\phi\left(X_i,(\varepsilon_i-n^{-1/2}\Delta' X_i\,\sigma_i^{-1})\right)\|^2$$

$$\le n^{-1/2}\,E\left\{\sigma(X)^{-2}\,\|X\|^2\,H(X)\,L(X)\right\}=O(n^{-1/2}),$$

applying a mean value theorem (mvt) argument, N2 and N4. (a.10) follows from

$$E\left\{S_n(\Delta,\sigma)\right\}\le \delta\,E\left\{\sigma(X)^{-2}\,\|X\|^2\,H(X)\right\}\le \delta\,O(1),$$

by N2 and N5. (a.11) is proved using similar arguments as in (a.9), in particular,

$$\mathrm{Var}\left\{S_n(\Delta,\sigma)\right\}\le E\left\{S_n^2(\Delta,\sigma)\right\}\le \delta\,n^{-1/2}\,E\left\{\sigma(X)^{-2}\,\|X\|^2\,H(X)\,L(X)\right\}=O(n^{-1/2}).$$

(a.4) holds if for each $\eta>0$, $\tau>0$ and M, there exists a M_1 satisfying,

$$\Pr\left\{\inf_{\|\Delta\|\le M_1}\|U_n(\Delta,\sigma)\|>\eta\right\}>1-\tau, \qquad \text{(a.12)}$$

and (a.12) follows from(a.6) using same arguments as Jurečkovà's (1977) proof of her Lemma 5.2.

Proof of Theorem 2 (a): Let introduce the following notation,

$$\tilde{\sigma}_i^2=\sum_j (Y_j-X_j'\tilde{\theta}_n)^2\omega_{ij} \text{ and } \bar{\sigma}_i^2=\sum_j \sigma_j^2\,\omega_{ij}$$

where $\omega_{ij}=\omega_{ij}(k)$. We need the following Lemmas, proved in Robinson (1987),

Lemma 1.- Let f(.) be a Borel function such that $E|f(x)|^p<\infty$,

$$E\{\sum_i |f(X_i)-f(X_1)|^p\omega_{1i}\}=o(1).$$

Lemma 2.- $\Pr\{\min_i \sigma_i\le\delta\}=0$ all n and some $\delta>0$.

Lemma 3.- $\Pr\{\bar{\sigma}_1\le\delta\}=0$ all n and some $\delta>0$.

Lemma 4.- $\Pr\{\min_i \bar{\sigma}_i\le\delta\}=0$ all n and some $\delta>0$.

Lemma 5.- $\{\min_i \tilde{\sigma}_i^2\}^{-1}=O_p(1)$.

Lemma 6.- $\{\min_i \hat{\sigma}_i^2\}^{-1}=O_p(1)$.

Lemma 7.- $\sum_i|\hat{\sigma}_i^2-\tilde{\sigma}_i^2|^2=O_p(k^{-1})$.

Lemma 8.- $E\{|\tilde{\sigma}_1^2-\bar{\sigma}_1^2|^{\upsilon/2}\}=O(k^{-\upsilon/4})$.

Lemma 9.- $\max_i|\hat{\sigma}_i^2-\tilde{\sigma}_i^2|=O_p(k^{-1/2})$.

Lemma 10.- $\max_i |\tilde{\sigma}_i^2 - \bar{\sigma}_i^2| = O_p(n\ k^{-\upsilon/4})$.

We follow the same strategy of proof as in Theorem 1. We first prove that,

$$\sup_{\|\Delta\|\le M} \|U_n(\Delta, \hat{\sigma}) - U_n(0, \hat{\sigma}) + V(\sigma)\ \Delta\| = o_p(1). \qquad \text{(b.1)}$$

We conclude (b.1) from (a.6) and,

$$\sup_{\|\Delta\|\le M} \|U_n(\Delta, \hat{\sigma}) - U_n(0, \hat{\sigma}) - \{U_n(\Delta, \sigma) - U_n(0, \sigma)\}\| = o_p(1). \qquad \text{(b.2)}$$

(b.2) follows from,

$$\sup_{\|\Delta\|\le M} \|n^{-1/2} \textstyle\sum_i X_i\ (\hat{\sigma}_i^{-1} - \tilde{\sigma}_i^{-1})\ R_i(\Delta, \hat{\sigma})\| = o_p(1), \qquad \text{(b.3)}$$

$$\sup_{\|\Delta\|\le M} \|n^{-1/2} \textstyle\sum_i X_i\ (\tilde{\sigma}_i^{-1} - \bar{\sigma}_i^{-1})\ R_i(\Delta, \hat{\sigma})\| = o_p(1), \qquad \text{(b.4)}$$

$$\sup_{\|\Delta\|\le M} \|n^{-1/2} \textstyle\sum_i X_i\ (\bar{\sigma}_i^{-1} - \sigma_i^{-1})\ R_i(\Delta, \hat{\sigma})\| = o_p(1), \qquad \text{(b.5)}$$

$$\sup_{\|\Delta\|\le M} \|n^{-1/2} \textstyle\sum_i X_i\ \sigma_i^{-1}\ [R_i(\Delta, \hat{\sigma}) - R_i(\Delta, \tilde{\sigma})]\| = o_p(1), \qquad \text{(b.6)}$$

$$\sup_{\|\Delta\|\le M} \|n^{-1/2} \textstyle\sum_i X_i\ \sigma_i^{-1}\ [R_i(\Delta, \tilde{\sigma}) - R_i(\Delta, \bar{\sigma})]\| = o_p(1), \qquad \text{(b.7)}$$

$$\sup_{\|\Delta\|\le M} \|n^{-1/2} \textstyle\sum_i X_i\ \sigma_i^{-1}\ [R_i(\Delta, \sigma) - R_i(\Delta, \sigma)]\| = o_p(1), \qquad \text{(b.6)}$$

where $R_i(\Delta, a) = \phi\left(X_i, (\varepsilon_i - n^{-1/2}\ \Delta' X_i\ \sigma_i^{-1})\sigma_i a_i^{-1}\right) - \phi\left(X_i, \varepsilon_i \sigma_i a_i^{-1}\right)$.

Using a mvt argument, the left hand side of (b.3) is bounded by,

$$K\ n^{-1} \textstyle\sum_i \|X_i\|^2\ H(X_i)\ \hat{\sigma}_i^{-1}\ |\hat{\sigma}_i^{-1} - \tilde{\sigma}_i^{-1}| \le$$

$$K \max_i |\hat{\sigma}_i^2 - \tilde{\sigma}_i^2| \left\{\min_i \hat{\sigma}_i^2\ \min_i \tilde{\sigma}_i\ (\min_i \hat{\sigma}_i + \min_i \tilde{\sigma}_i)\right\}^{-1} n^{-1} \textstyle\sum_i \|X_i\|^2\ H(X_i) = O_p(k^{-1/2})$$

by Lemmas 5, 6 and 9 and R1 (where, henceforth, K is a generic constant). The left hand side of (b.4) is bounded by,

$$K\ n^{-1} \textstyle\sum_i \|X_i\|^2\ H(X_i)\ \hat{\sigma}_i^{-1}\ |\tilde{\sigma}_i^{-1} - \bar{\sigma}_i^{-1}| \le$$

$$K \max_i |\tilde{\sigma}_i^2 - \bar{\sigma}_i^2| \left\{\min_i \hat{\sigma}_i\ \min_i \tilde{\sigma}_i\ \min_i \bar{\sigma}_i\ (\min_i \tilde{\sigma}_i + \min_i \bar{\sigma}_i)\right\}^{-1} n^{-1} \textstyle\sum_i \|X_i\|^2\ H(X_i)$$

$$= O_p(n\ k^{-\upsilon/4}),$$

by Lemmas 4, 5, 6 and 10 and R1. The left hand side of (b.5) is bounded by,

$$\left\{\min_i \hat{\sigma}_i\ \min_i \sigma_i\ \min_i \bar{\sigma}_i\ (\min_i \sigma_i + \min_i \bar{\sigma}_i)\right\}^{-1} K\ n^{-1} \textstyle\sum_i \|X_i\|^2\ H(X_i) |\bar{\sigma}_i^2 - \sigma_i^2| = o_p(1),$$

by Lemmas 2,3 and 6 and Markov's inequality, noting that, by Holder's inequality,

$$E\left\{\|X_1\|^2\ H(X_1)|\bar\sigma_1^{2}-\sigma_1^2|\right\}\le\left\{E|\,\|X_1\|^2\ H(X_1)|^{\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon}\left\{E|\bar\sigma_1^{2}-\sigma_1^2|^{\upsilon/2}\right\}^{2/\upsilon}=o(1)$$

by Lemma 1 and R1.(b.6), (b.7) and (b.8) follow by using same arguments as in (b.3), (b.4) and (b.5), after bounding the former expressions using a mvt argument. Next we prove that,

$$\sup_{\|\Delta\|\le M}\|U_n(\Delta,\ \hat\sigma)-U_n(0,\ \sigma)+V(\sigma)\ \Delta\|=o_p(1). \qquad (b.9)$$

We conclude (b.9) from (b.1) and,

$$\|U_n(0,\ \hat\sigma)-U_n(0,\ \sigma)\|=o_p(1). \qquad (b.10)$$

(b.10) follows from,

$$\|n^{-1/2}\textstyle\sum_i X_i\ (\bar\sigma_i^{-1}-\sigma_i^{-1})\ \phi(X_i,\ \varepsilon_i)\|=o_p(1), \qquad (b.11)$$

$$\|n^{-1/2}\textstyle\sum_i X_i\ \bar\sigma_i^{-1}\ \{\phi(X_i,\ \varepsilon_i\sigma_i\bar\sigma_i^{-1})-\phi(X_i,\ \varepsilon_i)\}\|=o_p(1), \qquad (b.12)$$

$$\|n^{-1/2}\textstyle\sum_i X_i\ (\hat\sigma_i^{-1}-\tilde\sigma_i^{-1})\ \phi(X_i,\ \varepsilon_i\sigma_i\bar\sigma_i^{-1})\|=o_p(1), \qquad (b.13)$$

$$\|n^{-1/2}\textstyle\sum_i X_i\ (\tilde\sigma_i^{-1}-\bar\sigma_i^{-1})\ \phi(X_i,\ \varepsilon_i\sigma_i\bar\sigma_i^{-1})\|=o_p(1), \qquad (b.14)$$

$$\|n^{-1/2}\textstyle\sum_i X_i\ \hat\sigma_i^{-1}\ \{\phi(X_i,\ \varepsilon_i\sigma_i\hat\sigma_i^{-1})-\phi(X_i,\ \varepsilon_i\sigma_i\bar\sigma_i^{-1})\}\|=o_p(1). \qquad (b.15)$$

Using the fact that $\phi(X,.)$ is odd and ε is conditionally symmetric about zero, (b.11) follows from Chevyshev's inequality since,

$$E\|n^{-1/2}\textstyle\sum_i X_i\ (\bar\sigma_i^{-1}-\sigma_i^{-1})\ \phi(X_i,\ \varepsilon_i)\|^2=E\{\|X_1\|^2\ |\bar\sigma_1^{-1}-\sigma_1^{-1}|^2\ \phi(X_1,\ \varepsilon_1)^2\|\}$$
$$\le K\ E\{L(X_1)^2\ |\bar\sigma_1^2-\sigma_1^2|\},$$

by N2, Lemma 3 and 4; and by Holder's inequality,

$$E\{L(X_1)^2\ |\bar\sigma_1^2-\sigma_1^2|\}\le\{E|L(X_1)|^{2\upsilon/(\upsilon-2)}\}^{(\upsilon-2)/\upsilon}\ E\{|\bar\sigma_1^2-\sigma_1^2|^{\upsilon/2}\}^{2/\upsilon}=o_p(1),$$

by Lemma 1 and R4.We conclude (b.12) from Chevyshev's inequality, since, by triangle inequality, Lemma 3 and 4 and Holder's inequality,

$$E\|n^{-1/2}\textstyle\sum_i X_i\ \bar\sigma_i^{-1}\ \{\phi(X_i,\ \varepsilon_i\sigma_i\bar\sigma_i^{-1})-\phi(X_i,\ \varepsilon_i)\}\|^2\le$$

$$E|\,\|X_1\|^2\ \bar\sigma_1^{-2}\ \{\phi(X_1,\ \varepsilon_1\sigma_1\bar\sigma_1^{-1})-\phi(X_1,\ \varepsilon_1)\}^2|\ =$$

$$E|\,\|X_1\|^2\ \bar\sigma_1^{-2}(\bar\sigma_1^{-1}-\sigma_1^{-1})\ \varepsilon_1\sigma_1\{\phi(X_1,\ \varepsilon_1\sigma_1\bar\sigma_1^{-1})-\phi(X_1,\ \varepsilon_1)\}\times$$
$$\int_0^1\phi(X_1,\ \varepsilon_1\sigma_1(\tau\ \bar\sigma_1^{-1}+(1-\tau)\ \sigma_1^{-1}))\ d\tau\le$$

$$E|\,\|X_1\|\ |\varepsilon_1\sigma_1|\ |\bar\sigma_1-\sigma_1|\ H(X_1)\ L(X_1)|\le$$

$$\left\{E|\,\|X_1\|\ |\varepsilon_1\sigma_1|\ H(X_1)\ L(X_1)|^{\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon}\left\{E|\bar\sigma_1^{2}-\sigma_1^2|^{\upsilon/4}\right\}^{2/\upsilon}=o(1)$$

by Lemma 1. Now note that the left hand side of (b.13) is bounded by,

$$\left\{\min_i\hat{\sigma}_i \ \min_i\tilde{\sigma}_i \ (\min_i\tilde{\sigma}_i + \min_i\hat{\sigma}_i)\right\}^{-1} \left\{\sum_i |\hat{\sigma}_i^2 - \tilde{\sigma}_i^2|^2\right\}^{1/2} \left\{n^{-1}\sum_i L(X_i)^2\right\}^{1/2} = o_p(1)$$

by lemmas 5, 6 and 7. The left hand side of (b.14) is bounded by,

$$\{\min_i \tilde{\sigma}_i \ \min_i \bar{\sigma}_i\}^{-1} n^{-1/2} \sum_i L(X_i) \ |\tilde{\sigma}_i - \bar{\sigma}_i| = o_p(1),$$

by Lemmas 4 and 5 and Chevyshev's inequality, since,

$$E|n^{-1/2} \sum_i L(X_i) \ |\tilde{\sigma}_i - \bar{\sigma}_i||^2 \le C_1 + C_2$$

where, by Hölder's inequality,

$$C_1 = E\{L(X_1)^2 \ |\tilde{\sigma}_1 - \bar{\sigma}_1|^2\} \le \left\{E|L(X_1)|^{2(\upsilon-2)/\upsilon}\right\}^{\upsilon/(\upsilon-2)} \left\{E|\tilde{\sigma}_1^2 - \bar{\sigma}_1^2|^{\upsilon/2}\right\}^{2/\upsilon} = O(k^{-1/2})$$

and by Cauchy's and Hölder's inequalities,

$$C_2 = E\left\{ n^{-1}\sum_{i\neq j}\sum L(X_i) \ |\tilde{\sigma}_i - \bar{\sigma}_i| \ L(X_j) \ |\tilde{\sigma}_j - \bar{\sigma}_j|\right\} \le E\left[n^{-1} \sum_{i\neq j}\sum L(X_i)^2 \ |\tilde{\sigma}_j - \bar{\sigma}_j|^2\right]$$

$$\le \left\{E|L(X_1)|^{2\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon} \left\{n \ E|\tilde{\sigma}_1^2 - \bar{\sigma}_1^2|^{\upsilon/2}\right\}^{2/\upsilon} = O\left[(n \ k^{-\upsilon/4})^{2/\upsilon}\right]$$

by Lemma 8 and R4. Now note that,

$$\|n^{-1/2}\sum_i \phi(X_i, \ \varepsilon_i\sigma_i\hat{\sigma}_i^{-1}) - \phi(X_i, \ \varepsilon_i\sigma_i\bar{\sigma}_i^{-1})\| =$$

$$\|n^{-1/2}\sum_i \varepsilon_i\sigma_i \ (\hat{\sigma}_i^{-1} - \bar{\sigma}_i^{-1}) \int_0^1 \phi'(X_i, \ \varepsilon_i\sigma_i[\hat{\sigma}_i^{-1}\tau + \bar{\sigma}_i^{-1}(1-\tau)]) \ d\tau\|$$

$$\le K \ n^{-1/2}\sum_i\|X_i\| \ H(X_i) \ |\varepsilon_i\sigma_i| \ |\hat{\sigma}_i^{-1} - \tilde{\sigma}_i^{-1}| + K \ n^{-1/2}\sum_i\|X_i\| \ H(X_i) \ |\varepsilon_i\sigma_i| \ |\hat{\sigma}_i^{-1} - \tilde{\sigma}_i^{-1}|$$

$$= o_p(1),$$

using arguments in (b.13) and (b.14). Then by (b.9), $U(\Delta^*, \hat{\sigma}) = o_p(1)$ and

$$0 = \|U_n[n^{1/2}(\check{\theta}_n(\hat{\sigma}) - \theta^o), \ \hat{\sigma}]\| \le 2 \ \|U_n(\Delta^*, \ \sigma)\| = o_p(1), \qquad \text{(b.18)}$$

and given (b.9), $n^{1/2}(\check{\theta}_n(\hat{\sigma}) - \theta^o) = O_p(1)$, since for each $\eta > 0$, $\tau > 0$ and M, there exists a M_1 satisfying,

$$\Pr\left\{ \inf_{\|\Delta\| \le M_1} \|U_n(\Delta, \ \hat{\sigma})\| > \eta\right\} > 1 - \tau,$$

using same arguments as Jurečkovà's (1977) proof of her Lemma 5.2. By (b.18),

$$n^{1/2}(\check{\theta}_n(\hat{\sigma}) - \theta^o) = V(\sigma)^{-1} \ U_n(0, \ \sigma) + o_p(1) \ \longrightarrow_d \ N(0, \ \Phi_1),$$

by the clt.

Proof of Theorem 2. (b): Following arguments given in the proof to Theorem 2, it suffices to prove that,

$$\sup_{\|\Delta\| \le M} \|U_n(\Delta, \ \check{\sigma}) - U_n(0, \ \check{\sigma}) - \{U_n(\Delta, \ \sigma) - U_n(0, \ \sigma)\}\| = o_p(1), \qquad \text{(b.19)}$$

and

$$\|U_n(\Delta, \check{\sigma}) - U_n(0, \sigma)\| = o_p(1). \qquad \text{(b.20)}$$

(b.19) and (b.20) follow in a similar manner as (b.2) and (b.10) but using the following Lemmas proved in Delgado (1989),

Lemma 12.- $\Pr\{\min_i \mathring{\sigma}_i \leq \delta\} = 0$ all n and some $\delta > 0$.

Lemma 13.- $\{\min \check{\sigma}^2\}^{-1} = O(1)$.

Lemma 14.- $E\{|\check{\sigma}_1^2 - \mathring{\sigma}_1^2|^{\upsilon/2}\} = O(k^{-\upsilon/4})$.

Lemma 15.- $\max_i |\check{\sigma}_i^2 - \mathring{\sigma}_i^2| = O_p(k^{-1/2})$.

where $\mathring{\sigma}_i^2 = \sum_j E(Y_j^2|X_j)\,\omega_{ij} - \{\sum_j X_j' \theta^\circ \omega_{ij}\}^2$.

Acknowledgements: This work was started at the London School of Economics as part of my doctoral dissertation. I thank Peter Robinson for encouraging me to investigate this problem. Financial support from the Economic and Social Research Council (ESRC) reference numbers: B00232156 and R000231441 and from a Indiana University Faculty Grant is gratefully acknowledged.

References

Carroll, R.J. (1982), "Adapting for heteroscedasticity in linear models", Annals of Statistics 10, 1244-1233.

Carroll, R.J. and D. Ruppert (1982), "Robust estimation in heteroscedastic linear models", Annals of Statistics 10,429-441.

Delgado, M.A. (1989), "An alternative asymptotically efficient semiparametric generalised least squares estimator", Unpublished.

Hampel, F.R., E.M. Rouchetti, P.J. Rousseew and W.A. Stahel (1986), Robust Statistics, Wiley: New York.

Huber, P.J. (1973), "Robust regression: asymptotics, conjectures and Monte Carlo", Annals of Statistics 1, 799-821.

Jureckova, J. (1977), "Asymptotic relations of M-estimates and R-estimates in linear regression model", Annals of Statistics 5,464-472.

Maronna, R.A. and V.J. Yohai, "Asymptotic behaviour of general M-estimates for regression and scale with random carriers", Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 58, 7-20

Relles, D. (1968), Robust Regression by Modified Least Squares, Ph.D. Thesis, Yale University.

Robinson, P.M. (1987), "Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form", Econometrica 55, 875-891.

Rose, R.L. (1978), Nonparametric Estimation of Weights in Least-Squares Regression Analysis, Ph.D. Thesis, University of California at Davis.

Stone, C.J. (1977), "Consistent nonparametric regression", (with discussion) Annals of Statistics 5,595-645.

Yohai, V.J. and R.A. Maronna (1979), "Asymptotic behavior of M-estimators for the linear model", Annals of Statistics 7, 258-268.

TABLE 1

CONTAMINATED ERRORS

MODEL 1 ($\gamma = 0.$)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	-5E-5	.1028	1.0000	-.0045	.0317	1.0000	-6E-4	.0062	1.0000
$\hat{\theta}_n(1)$	-5E-5	.1028	1.0000	-.0045	.0317	1.0000	-6E-4	.0062	1.0000
LADE	-.0027	.1007	1.0204	.0022	.0282	1.1235	-.0016	.0059	1.0528
$\hat{\theta}_n(\hat{\sigma})^1$	-.0015	.3541	.2903	-.0043	.0569	.5564	6E-4	.0078	.7965
$\hat{\theta}_n(\hat{\sigma})^2$	.0031	.3073	.3348	-6E-4	.0488	.6483	-4E-4	.0069	.8961
$\hat{\theta}_n(\check{\sigma})^1$	.0026	.3703	.2776	-.0032	.0585	.5415	8E-4	.0079	.7908
$\hat{\theta}_n(\check{\sigma})^2$	.0132	.2881	.3566	-.0025	.0477	.6639	-3E-4	.0069	.9009
$\check{\theta}_n(\sigma)$	-9E-4	.0766	1.3416	-.0025	.0220	1.4392	-8E-4	.0043	1.4327
$\check{\theta}_n(1)$	-9E-4	.0766	1.3416	-.0025	.0220	1.4392	-8E-4	.0043	1.4327
$\check{\theta}_n(\hat{\sigma})^1$	2E-4	.1488	.6909	-.006	.030	1.0429	-7E-4	.0050	1.2353
$\check{\theta}_n(\hat{\sigma})^2$	3E-4	.1136	.9053	-.003	.027	1.1471	-.0012	.0049	1.2697
$\check{\theta}_n(\check{\sigma})^1$	6E-4	.1370	.7501	-.004	.029	1.0752	-7E-4	.0050	1.2413
$\check{\theta}_n(\check{\sigma})^2$	.0135	.1085	.9464	-.001	.027	1.177	-.0011	.0048	1.2757

MODEL 1 ($\gamma = -1.$)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	-2E-5	.0146	1.0000	4E-4	.0046	1.0000	-6E-4	88E-5	1.0000
$\hat{\theta}_n(1)$	.0026	.0339	.4310	-.0031	.0096	.4816	-7E-7	.0020	.4384
LADE	-5E-4	.0193	.7594	-5E-4	.0053	.8824	-9E-4	.0010	.8064
$\hat{\theta}_n(\hat{\sigma})^1$	-5E-4	.0529	.2767	.0014	.0086	.5391	-4E-4	.0011	.7808
$\hat{\theta}_n(\hat{\sigma})^2$	.0017	.0430	.3405	-2E-5	.0073	.6342	-7E-4	99E-4	.8942
$\hat{\theta}_n(\check{\sigma})^1$	.0044	.0578	.2533	.0014	.0091	.5129	-2E-4	.0011	.7785
$\hat{\theta}_n(\check{\sigma})^2$	.0094	.0430	.3402	.0010	.0069	.6669	-7E-4	97E-4	.9075
$\check{\theta}_n(\sigma)$	-.0010	.0109	1.3354	-3E-4	.0032	1.4432	-9E-4	62E-5	1.4167
$\check{\theta}_n(1)$	.0016	.0257	.5697	-.0026	.0072	.6453	-9E-4	.0014	.6098
$\check{\theta}_n(\hat{\sigma})^1$	-3E-4	.0217	.6738	-6E-4	.0046	1.0182	-.0013	71E-5	1.2382
$\check{\theta}_n(\hat{\sigma})^2$	5E-4	.0176	.8317	-3E-4	.0042	1.1051	-.0012	69E-5	1.2695
$\check{\theta}_n(\check{\sigma})^1$	.0044	.0210	.6959	3E-4	.0045	1.0351	-.0012	71E-5	1.2381
$\check{\theta}_n(\check{\sigma})^2$	.0087	.0187	.7822	4E-4	.0043	1.0675	-.0011	70E-5	1.2574

TABLE 1 (Cont.)

MODEL 2 (δ= 3.)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	-.0015	4.8231	1.0000	-.0361	1.3672	1.0000	.0197	.2781	1.0000
$\hat{\theta}_n(1)$	-.0653	23.390	.2061	-.0348	6.9974	.1975	-.0172	1.3980	.1992
LADE	-.0455	9.0602	.5322	.0054	2.4612	.5560	-.0051	.4824	.5772
$\hat{\theta}_n(\hat{\sigma})^1$	-.0397	19.105	.2524	-.0245	2.6323	.5197	.0322	.3846	.7222
$\hat{\theta}_n(\hat{\sigma})^2$	-.0320	16.698	.2888	-.0268	2.2223	.6156	.0226	.3381	.8226
$\hat{\theta}_n(\check{\sigma})^1$	-.0519	21.121	.2283	-.0249	2.7676	.4943	.0304	.3895	.7132
$\hat{\theta}_n(\check{\sigma})^2$	-.0290	16.021	.3010	-.0276	2.2027	.6211	.0220	.3375	.8238
$\check{\theta}_n(\sigma)$	.0139	3.6326	1.3276	-.0236	1.0115	1.3522	.0219	.1975	1.4067
$\check{\theta}_n(1)$	-.0487	8.7229	.5527	.0043	2.4801	.5517	-5E-4	.4954	.5622
$\check{\theta}_n(\hat{\sigma})^1$	-2E-4	7.7019	.6262	-.0132	1.4622	.9357	.0354	.2500	1.1084
$\check{\theta}_n(\hat{\sigma})^2$	-.0290	6.9172	.6971	-.0159	1.3926	.9825	.0250	.2406	1.1543
$\check{\theta}_n(\check{\sigma})^1$	.0117	7.3163	.6592	-.0082	1.4371	.9522	.0344	.2487	1.1144
$\check{\theta}_n(\check{\sigma})^2$	-.0063	6.5064	.7413	-.0163	1.3660	1.0016	.0245	.2401	1.1573

MODEL 2 (δ= 5.)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	.0194	39.714	1.0000	-.1011	9.8887	1.0000	.0592	1.8264	1.0000
$\hat{\theta}_n(1)$	-.5159	1461.4	.0271	-.2760	442.68	.0223	.1996	88.534	.0207
LADE	-.2619	258.41	.1536	.0603	56.269	.1759	.0139	9.4482	.1936
$\hat{\theta}_n(\hat{\sigma})^1$	-.1122	206.76	.1920	-.0536	21.595	.4583	.0848	2.6485	.6890
$\hat{\theta}_n(\hat{\sigma})^2$	-.1928	231.36	.1716	-.1033	19.826	.4990	.0598	2.3742	.7695
$\hat{\theta}_n(\check{\sigma})^1$	-.1644	216.73	.1832	-.0530	21.978	.4503	.0783	2.6674	.6844
$\hat{\theta}_n(\check{\sigma})^2$	-.2036	211.56	.1876	-.1026	19.534	.5065	.0569	2.3597	.7744
$\check{\theta}_n(\sigma)$	.0720	30.293	1.3107	-.0727	7.3305	1.3494	.0658	1.2893	1.4144
$\check{\theta}_n(1)$	-.2749	256.30	.1549	.0464	57.008	.1736	.0047	9.9333	.1842
$\check{\theta}_n(\hat{\sigma})^1$	.0353	82.535	.4812	-.0227	11.716	.8448	.1010	1.7162	1.0599
$\check{\theta}_n(\hat{\sigma})^2$	-.1151	99.400	.3995	-.0634	12.559	.7879	.0737	1.6991	1.0735
$\check{\theta}_n(\check{\sigma})^1$	.0584	75.235	.5276	-.0182	11.026	.8978	.0982	1.6752	1.0861
$\check{\theta}_n(\check{\sigma})^2$	-.0652	89.291	.4447	-.0620	12.217	.8100	.0719	1.6848	1.0827

TABLE 1 (Cont.)

MODEL 1 (γ= -2.)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	-3E-6	.0016	1.0000	6E-4	46E-5	1.0000	-1E-4	95E-6	1.0000
$\hat{\theta}_n(1)$	.0028	.0227	.0701	-.0028	.0063	.0725	-4E-4	.0013	.0697
LADE	3E-4	.0047	.3359	-5E-5	.0012	.3791	-5E-4	23E-5	.4103
$\hat{\theta}_n(\hat{\sigma})^1$	-3E-4	.0065	.2449	.0010	98E-5	.4681	-3E-5	12E-5	.7513
$\hat{\theta}_n(\hat{\sigma})^2$	6E-4	.0056	.2826	1E-4	78E-5	.5868	-1E-5	11E-5	.8546
$\hat{\theta}_n(\check{\sigma})^1$	.0041	.0172	.0931	9E-4	.0014	.3355	-3E-4	13E-5	.7229
$\hat{\theta}_n(\check{\sigma})^2$	.0107	.0192	.0829	6E-4	.0016	.2841	-4E-4	15E-5	.6048
$\check{\theta}_n(\sigma)$	-3E-4	.0012	1.3003	3E-4	32E-5	1.3991	-2E-4	68E-6	1.4010
$\check{\theta}_n(1)$	.0021	.0174	.0920	-.0022	.0047	.0974	-6E-4	97E-5	.0977
$\check{\theta}_n(\hat{\sigma})^1$	-3E-4	.0027	.599	4E-4	50E-5	.9189	-3E-4	79E-6	1.1982
$\check{\theta}_n(\hat{\sigma})^2$	5E-5	.0025	.6466	2E-4	46E-5	1.0006	-3E-4	77E-6	1.2366
$\check{\theta}_n(\check{\sigma})^1$	.0028	.0051	.3122	4E-4	74E-4	.6718	-5E-4	87E-6	1.0921
$\check{\theta}_n(\check{\sigma})^2$	.0084	.0068	.2326	5E-4	.0010	.4450	-6E-4	11E-5	.8253

MODEL 2 (δ= 1.)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	-4E-5	.3627	1.0000	-.0067	.1061	1.0000	.0015	.0222	1.0000
$\hat{\theta}_n(1)$	-.0057	.4716	.7691	-.0047	.1399	.7588	3E-4	.0283	.7836
LADE	-.0079	.3865	.9383	.0024	.1123	.9455	-.0022	.0227	.9779
$\hat{\theta}_n(\hat{\sigma})^1$	-.0053	1.2806	.2832	-.0056	.2016	.5269	.0042	.0286	.7750
$\hat{\theta}_n(\hat{\sigma})^2$	.0018	1.1202	.3237	-.0037	.1692	.6278	.0019	.0254	.8714
$\hat{\theta}_n(\check{\sigma})^1$	-.0061	1.3736	.2640	-.0048	.2065	.5143	.0042	.0289	.7661
$\hat{\theta}_n(\check{\sigma})^2$	.0089	1.0821	.3352	-.0025	.1670	.6358	.0019	.0254	.8738
$\check{\theta}_n(\sigma)$	5E-4	.2715	1.3357	-.0042	.0778	1.3637	-.0015	.0155	1.4262
$\check{\theta}_n(1)$	-.0070	.3113	1.1649	-.0010	.0914	1.1616	-.0013	.0189	1.1709
$\check{\theta}_n(\hat{\sigma})^1$	-4E-4	.5389	.6731	-.0040	.1101	.9640	.0031	.0186	1.1962
$\check{\theta}_n(\hat{\sigma})^2$	-.0036	.4175	.8688	-.0027	.0979	1.0842	-.0013	.0179	1.2353
$\check{\theta}_n(\check{\sigma})^1$	.0064	.5006	.7244	-.0037	.1072	.9902	.0030	.0185	1.1962
$\check{\theta}_n(\check{\sigma})^2$	.0087	.3956	.9166	-.0018	.0960	1.1063	-.0013	.0179	1.2353

TABLE 2

NORMAL ERRORS

MODEL 1 (γ= -2.)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	-2E-4	.0016	1.0000	-2E-4	46E-5	1.0000	-2E-4	92E-6	1.0000
$\hat{\theta}_n(1)$	.0028	.0448	.0707	-7E-4	.0069	.0675	-7E-4	.0013	.0681
LADE	1E-4	.0072	.2232	-7E-5	.0019	.2458	-6E-4	35E-5	.2581
$\hat{\theta}_n(\hat{\sigma})^1$	1E-4	.0049	.3277	-7E-5	64E-5	.7198	-1E-4	10E-5	.9057
$\hat{\theta}_n(\hat{\sigma})^2$	8E-4	.0039	.4121	6E-4	56E-5	.8225	-3E-4	97E-6	.9500
$\hat{\theta}_n(\check{\sigma})^1$	.0048	.0118	.1369	5E-4	.0010	.4638	-4E-4	11E-5	.8439
$\hat{\theta}_n(\check{\sigma})^2$	.0117	.0122	.1311	8E-4	.0013	.3543	-6E-4	14E-5	.6584
$\check{\theta}_n(\sigma)$	-3E-4	.0017	.9637	1E-4	48E-5	.9653	-3E-4	98E-6	.9443
$\check{\theta}_n(1)$	.0025	.0217	.0746	-6E-4	.0064	.0725	-8E-4	.0013	.0733
$\check{\theta}_n(\hat{\sigma})^1$	1E-4	.0031	.5201	-1E-4	58E-5	.8042	-3E-4	10E-5	.9087
$\check{\theta}_n(\hat{\sigma})^2$	9E-5	.0029	.5729	2E-5	54E-5	.8647	-3E-4	97E-6	.9487
$\check{\theta}_n(\check{\sigma})^1$	.0034	.0060	.2692	2E-4	85E-5	.5465	-5E-4	11E-5	.8598
$\hat{\theta}_n(\check{\sigma})^2$	.0084	.0079	.2039	7E-4	.0012	.3959	-6E-4	14E-5	.6583

MODEL 2 (δ= 5.)

	N= 30			N= 100			N=500		
	BIAS	VAR	EFF	BIAS	VAR	EFF	BIAS	VAR	EFF
$\hat{\theta}_n(\sigma)$	.0176	39.134	1.0000	-.0972	9.8403	1.0000	.0637	1.7540	1.0000
$\hat{\theta}_n(1)$	-.5878	1490.7	.0262	-.2371	441.68	.0223	.2241	87.482	.0201
LADE	-.3699	394.42	.0992	.0495	86.638	.1137	.0062	14.855	.1183
$\hat{\theta}_n(\hat{\sigma})^1$	.0112	160.79	.2434	-.0311	15.677	.6282	.0754	2.0997	.8350
$\hat{\theta}_n(\hat{\sigma})^2$	-.0714	162.41	.2409	-.1191	15.872	.6200	.0624	2.0867	.8409
$\hat{\theta}_n(\check{\sigma})^1$	-.0039	158.70	.2466	-.0149	15.011	.6561	.0688	2.0572	.8526
$\hat{\theta}_n(\check{\sigma})^2$	-.0192	142.78	.2741	-.1137	15.431	.6377	.0586	2.0709	.8475
$\check{\theta}_n(\sigma)$	.0546	40.576	.9644	-.0918	10.283	.9570	.0711	1.8492	.9481
$\check{\theta}_n(1)$	-.3630	391.76	.0999	-.0363	87.009	.1132	7E-4	15.251	.1153
$\check{\theta}_n(\hat{\sigma})^1$	.0569	99.315	.3940	-.0354	13.701	.7188	.0859	2.0444	.8568
$\check{\theta}_n(\hat{\sigma})^2$	-.0367	117.45	.3331	-.1047	14.854	.6626	.0668	2.0589	.8520
$\check{\theta}_n(\check{\sigma})^1$	.0663	91.335	.4284	-.0169	12.966	.7596	.0795	2.0009	.8758
$\hat{\theta}_n(\check{\sigma})^2$	.0094	104.85	.3732	-.0990	14.465	.6804	.0636	2.0433	.8587