The Annals of Statistics
2002, Vol. 30, No. 2, 583–616

A CLASS OF ROBUST AND FULLY EFFICIENT REGRESSION ESTIMATORS

BY DANIEL GERVINI[1] AND VÍCTOR J. YOHAI[2]

University of Zurich and University of Buenos Aires and CONICET

This paper introduces a new class of robust estimators for the linear regression model. They are weighted least squares estimators, with weights adaptively computed using the empirical distribution of the residuals of an initial robust estimator. It is shown that under certain general conditions the asymptotic breakdown points of the proposed estimators are not less than that of the initial estimator, and the finite sample breakdown point can be at most $1/n$ less. For the special case of the least median of squares as initial estimator, hard rejection weights and normal errors and carriers, the maximum bias function of the proposed estimators for point-mass contaminations is numerically computed, with the result that there is almost no worsening of bias. Moreover—and this is the original contribution of this paper—if the errors are normally distributed and under fairly general conditions on the design the proposed estimators have full asymptotic efficiency. A Monte Carlo study shows that they have better behavior than the initial estimators for finite sample sizes.

1. Introduction. In this paper we address the problem of robust and efficient estimation in the linear regression model. It is well known that the least squares estimator (LSE) of the regression parameter θ has the smallest variance among unbiased estimates when the errors are normally distributed. However, the LSE is extremely sensitive to atypical data. A single observation placed far enough from the bulk of the data can carry the LSE arbitrarily far from θ, no matter how big the sample size is. This lack of stability of the LSE is a serious problem in applications. Thus several estimators that possess some stability in variance and bias under deviations from the regression model have been proposed over the last 30 years. However, some loss in efficiency under the normal model has been the price of this stability.

The least median of squares estimator (LMSE), proposed by Hampel (1975) and further developed by Rousseeuw (1984), was the first equivariant regression estimator that attained the maximum asymptotic breakdown point $1/2$ (as defined

Received September 1999; revised October 2001.

[1]Supported in part by the University of Buenos Aires Graduate Scholarships G1464 and G1848.

[2]Supported in part by Grant TW92 from the University of Buenos Aires, PIP 4186/96 from CONICET and PICT 03-06277 from ANPCYT.

AMS 2000 subject classifications. Primary 62J05; secondary 62F35.

Key words and phrases. Adaptive estimation, efficient estimation, maximum breakdown point, weighted least squares.

in Section 3). But this estimator has an important drawback: its rate of convergence is $n^{-1/3}$ [see Davies (1990)] and hence its relative efficiency with respect to the LSE is 0. To obtain a more efficient estimator under normality, Rousseeuw and Leroy (1987) suggested computing a weighted LSE (WLSE), skipping those observations whose LMSE standardized absolute residuals are greater than some fixed cutoff value. However, He and Portnoy (1992) showed that, even though the weighting step does reduce variability, the rate of convergence remains the same and hence the asymptotic efficiency is still 0.

S-estimators, proposed by Rousseeuw and Yohai (1984), were the first high-breakdown regression estimators to achieve the usual $n^{1/2}$-consistency under appropriate regularity conditions. However, Hössjer (1992) showed that S-estimators cannot achieve simultaneously high breakdown point and high efficiency under the normal model. Regression estimators that can attain a nearly optimal efficiency and maximum breakdown point at the same time are MM-estimators [Yohai (1987)] and τ-estimators [Yohai and Zamar (1988)]. However, tuning up these estimators for high efficiency will be accompanied by an increase in bias as an unpleasant side-effect. And, in any case, they will never achieve *maximum* asymptotic efficiency and positive breakdown point simultaneously.

We introduce in this paper a new class of estimators that simultaneously attain the maximum breakdown point and *full* asymptotic efficiency under normal errors. They are WLS estimators computed from an initial robust estimator, but unlike Rousseeuw and Leroy's proposal, the cutoff values are adaptively calculated from the data. We call these new estimators REWLSEs (robust and efficient weighted least squares estimators).

A different approach to robust and efficient estimation in linear regression models is presented in Agostinelli and Markatou (1998). They also proposed a WLSE computed from an initial robust estimator, but their weighting scheme is based on a measure of disparity between the density of the errors under the model and the smoothed empirical density of the residuals. The method we propose is based on the empirical distribution instead, so it is theoretically more tractable.

This article is organized as follows. The REWLS estimator is defined in Section 2. Sections 3 and 4 analyze its robust and asymptotic behavior, respectively. A Monte Carlo study is reported in Section 5. Proofs of the main results are given in the Appendices, although the reader is referred to Gervini and Yohai (2000) for complete technical details.

2. The REWLS estimator. We are given a random sample $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, where $\mathbf{x}_i$ is a vector of p explanatory variables and y_i is the response variable. We assume they are linked by the linear relationship

$$y_i = \mathbf{x}_i'\theta + u_i, \tag{1}$$

where $\theta \in \mathbb{R}^p$ is the regression parameter we will primarily focus on, and the error terms $\{u_i\}$ are i.i.d. unobservable random variables with unknown distribution

$F_0(\cdot/\sigma)$ for some scale parameter $\sigma > 0$. Most of the results in this article will assume that F_0 is symmetric about 0. This assumption, which is usual in many papers on robust regression, simplifies the theoretical treatment of the consistency and asymptotic normality of the estimators. Moreover, one of the main applications of our theory is the construction of highly robust estimates with full asymptotic efficiency under normal errors, and for this application the symmetry assumption certainly holds.

Consider a pair of initial robust estimators of regression and scale, $\mathbf{T}_{0n}$ and S_n, respectively. If $S_n > 0$, the standardized residuals are defined as

$$r_i = \frac{y_i - \mathbf{x}_i'\mathbf{T}_{0n}}{S_n}.$$

A large value of $|r_i|$ would suggest that $(\mathbf{x}_i, y_i)$ is an outlier. Assuming a normal-error model, it seems reasonable to consider as outliers those points with $|r_i| \geq 2.5$, say. Following this idea, Rousseeuw and Leroy (1987) defined

$$w_i = \begin{cases} 1, & \text{if } |r_i| < t_0, \\ 0, & \text{if } |r_i| \geq t_0, \end{cases} \tag{2}$$

with $t_0 = 2.5$, and computed a WLS estimator $\mathbf{T}_{1n} = (X'WX)^{-1}X'W\mathbf{Y}$, where $W = \text{diag}(w_1, \ldots, w_n)$ and $\mathbf{Y} = (y_1, \ldots, y_n)'$. It is known that this weighting step improves the efficiency under normal errors and it maintains the breakdown point of the initial estimator. However, even if observations perfectly followed the assumed linear model, there would be a small probability that the standardized absolute residuals exceeded any given fixed cutoff value. Thus a WLSE computed with weights as in (2) cannot be asymptotically efficient. Of course, a very large cutoff value t_0 could be used in (2) so that for any sample size to appear in practice no observations would be downweighted, and the WLSE would still maintain the breakdown point of the initial estimator. But such a choice of t_0 would have an adverse effect on the maximum bias of the estimator (as defined in Section 3).

The estimator we propose uses adaptive cutoff values. These cutoff values are constructed in such a way that the resulting WLSE is asymptotically efficient under the normal-error model and is robust under some deviations from the linear model. In particular, it maintains the breakdown point of the initial estimator and it does not worsen the maximum bias function too much, as we shall see in Section 3.

The adaptive cutoff values are defined as follows. Let the empirical distribution function of the standardized absolute residuals be

$$F_n^+(t) = \frac{1}{n}\sum_{i=1}^{n} I(|r_i| \leq t).$$

To detect outliers, we could compare $F_n^+(t)$ with the distribution function of the absolute errors under the model, $F_0^+(t)$. If $F_n^+(t) < F_0^+(t)$, the sample proportion of absolute residuals that exceed t is greater than the theoretical proportion.

If this happens for a large t, it means that outliers are present in the sample. However, since the actual distribution of the errors is never known in practice, a hypothetical F must be used instead of F_0. Typically, $F = \Phi$ will be chosen, although we will only require that F have finite variance. As a measure of the proportion of outliers in the sample, we then define

$$d_n = \sup_{t \geq \eta} \{F^+(t) - F_n^+(t)\}^+, \tag{3}$$

where $\{\cdot\}^+$ denotes the positive part, F^+ denotes the distribution of $|X|$ when $X \sim F$ and η is some large quantile of F^+. A value $\eta = 2.5$ as chosen by Rousseeuw and Leroy seems reasonable. Note that if $|r|_{(1)} \leq \cdots \leq |r|_{(n)}$ are the order statistics of the standardized absolute residuals and $i_0 = \max\{i : |r|_{(i)} < \eta\}$, then

$$d_n = \max_{i > i_0} \left\{ F^+(|r|_{(i)}) - \frac{(i-1)}{n} \right\}^+.$$

Thus we eliminate those $\lfloor nd_n \rfloor$ observations with largest standardized absolute residuals (here $\lfloor a \rfloor$ is the largest integer less than or equal to a). The resulting cutoff value is

$$t_n = \min\{t : F_n^+(t) \geq 1 - d_n\}, \tag{4}$$

that is, $t_n = |r|_{(i_n)}$ with $i_n = n - \lfloor nd_n \rfloor$. Observe that $i_n > i_0$ and $t_n > \eta$. With this adaptive cutoff value, we define weights of the form $w_i = w(|r_i|/t_n)$ and the REWLSE is

$$\mathbf{T}_{1n} = \begin{cases} (X'WX)^{-1}X'W\mathbf{Y}, & \text{if } S_n > 0, \\ \mathbf{T}_{0n}, & \text{if } S_n = 0. \end{cases}$$

The most common weight function is the hard-rejection weight $w(u) = I(u < 1)$, as in (2). But, in general, we will only require:

W1. The weight function w: $[0, \infty) \to [0, 1]$ is nonincreasing, right continuous, continuous in a neighborhood of 0, $w(0) = 1$, $w(u) > 0$ for $0 < u < 1$ and $w(u) = 0$ for $u \geq 1$.

Property W1 ensures that $w_i = 0$ if $|r_i| \geq t_n$, so that observations with large residuals are completely eliminated in the weighting step. Since t_n remains bounded in the presence of outliers, as we show in Section 3, this implies that $\mathbf{T}_{1n}$ keeps the finite sample and asymptotic breakdown points of $\mathbf{T}_{0n}$. On the other hand, when F_0 is of unbounded support but of lighter tails than F, $t_n \to \infty$ under the model and then $w(|r_i|/t_n) \to 1$. The same happens if F_0 is of bounded support with lighter tails than F and $w(u)$ is the hard-rejection function. This will eventually make $\mathbf{T}_{1n}$ asymptotically equivalent to the LSE under the model. Precise conditions for this to happen are given in Section 4. But we will analyze first the robust theoretical properties of the REWLSE.

3. Robustness of the REWLSE. In this section we study the behavior of the REWLSE under certain deviations from the central model. First, we analyze asymptotic robust properties as given by the maximum bias function and the influence function. In Section 3.4 we turn our attention to finite-sample robust properties, specifically the finite-sample breakdown point.

3.1. *Maximum bias properties.* We say that the random vector $(\mathbf{X}, Y)$ follows the central model if

$$(\mathbf{X}, Y) \sim H_0 \qquad \text{with } H_0(\mathbf{x}, y) = G_0(\mathbf{x}) F_0\{(y - \mathbf{x}'\theta)/\sigma\}. \tag{5}$$

The kind of departures from (5) we will consider consist of distributions in the gross-error neighborhood

$$\mathcal{H}_\varepsilon = \{H = (1-\varepsilon)H_0 + \varepsilon H^* : H^* \text{ any distribution on } \mathbb{R}^{p+1}\}.$$

Although $\mathcal{H}_\varepsilon$ is not a neighborhood in the topological sense, this definition allows an intuitive interpretation: we can think of $H \in \mathcal{H}_\varepsilon$ as a distribution that produces a fraction ε of outliers. We then assume $0 \le \varepsilon < 0.5$, so the majority of the data will always follow the central model (5).

Most estimates of θ can be defined by functionals. Let $\mathbf{T}$ be an $\mathbb{R}^p$-valued functional defined on a subset of distributions in $\mathbb{R}^{p+1}$ which includes all the empirical distributions and the contamination neighborhoods $\mathcal{H}_\varepsilon$ for $0 \le \varepsilon < 0.5$. Given a sample $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ with empirical distribution H_n, the estimate of θ associated with $\mathbf{T}$ is $\mathbf{T}_n = \mathbf{T}(H_n)$. Similarly, scale estimates can be defined by nonnegative functionals S defined on a subset of distributions in $\mathbb{R}$. A regression estimating functional $\mathbf{T}(H)$ is Fisher consistent if $\mathbf{T}(H_0) = \theta$. Another desirable property of regression estimating functionals is regression, affine and scale equivariance. This equivariance property means that, given $(\mathbf{X}, Y)$ that satisfies (5), $\mathbf{b} \in \mathbb{R}^p$, a nonsingular $A \in \mathbb{R}^{p\times p}$ and $a \in \mathbb{R}$, if we define $Y^* = aY + \mathbf{X}'\mathbf{b}$ and $\mathbf{X}^* = A'\mathbf{X}$ and H_0^* denotes the distribution of $(\mathbf{X}^*, Y^*)$, then $\mathbf{T}(H_0^*) = A^{-1}\{a\mathbf{T}(H_0) + \mathbf{b}\}$. Note that $(\mathbf{X}^*, Y^*)$ satisfies model (5) with regression parameter $\theta^* = A^{-1}(a\theta + \mathbf{b})$, so this invariance requirement is natural. For a scale functional S it is natural to require scale equivariance; that is, if Y has distribution R and R^* is the distribution of $Y^* = aY$, then $S(R^*) = |a| S(R)$.

The distance between the regression estimator $\mathbf{T}(H)$ and the target parameter θ is given by the asymptotic bias

$$b(\mathbf{T}, H) = \{(\mathbf{T}(H) - \theta)' C(G_0)(\mathbf{T}(H) - \theta)\}^{1/2} / S(R_0), \tag{6}$$

where $C(G_0)$ is an affine equivariant scatter functional and $R_0 = F_0(\cdot/\sigma)$. This measure of bias is invariant under the transformations described in the preceding paragraph when $\mathbf{T}$ is Fisher consistent. An invariant measure of bias for the scale estimator is given by

$$b(S, R_H) = |\log(S(R_H)/S(R_0))|, \tag{7}$$

where R_H is the distribution of $Y - \mathbf{X}'\mathbf{T}(H)$ when $(\mathbf{X}, Y) \sim H$. Note that $b(S, R_H)$ accounts for both explosion and implosion of the scale estimator.

As a measure of the outlier resistance of $\mathbf{T}$, we consider the worst possible bias produced by a distribution $H \in \mathcal{H}_\varepsilon$. This is given by the maximum bias (maxbias) function

$$\mathcal{B}_{\mathbf{T}}(\varepsilon) = \sup\{b(\mathbf{T}, H) : H \in \mathcal{H}_\varepsilon\}. \tag{8}$$

The maxbias function for the scale estimator is defined analogously. We consider that an estimator is robust if $\mathcal{B}_{\mathbf{T}}(\varepsilon) < \infty$ for some $\varepsilon > 0$. Since any equivariant estimator will explode for ε large enough, the asymptotic breakdown point is defined as

$$\varepsilon^*_{\mathbf{T}} = \inf\{\varepsilon : \mathcal{B}_{\mathbf{T}}(\varepsilon) = \infty\}.$$

To specify the functional form of the REWLSE, let us consider a random vector $(\mathbf{X}, Y)$ with joint distribution H and initial estimators $\mathbf{T}_0(H)$ and $S_0(R_H)$ of regression and scale. If $S_0(R_H) = 0$, then we set $\mathbf{T}_1(H) = \mathbf{T}_0(H)$. If $S_0(R_H) > 0$, then the standardized residual is

$$r_H(\mathbf{X}, Y) = \frac{Y - \mathbf{X}'\mathbf{T}_0(H)}{S_0(R_H)}$$

and the distribution function of the standardized absolute residual is

$$F_H^+(t) = P_H(|r_H(\mathbf{X}, Y)| \le t). \tag{9}$$

The functional forms of (3) and (4) are then

$$\begin{aligned} d(H) &= \sup_{t \ge \eta}\{F^+(t) - F_H^+(t)\}^+, \\ t(H) &= \min\{t : F_H^+(t) \ge 1 - d(H)\}. \end{aligned} \tag{10}$$

The definition of $d(H)$ automatically implies that $t(H) > \eta$. The weights are of the form $w_H(\mathbf{X}, Y) = w(|r_H(\mathbf{X}, Y)|/t(H))$ with a function $w(u)$ that satisfies W1. Thus the REWLSE is defined as

$$\mathbf{T}_1(H) = \arg\min_{\mathbf{t}} E_H\{w_H(\mathbf{X}, Y)(Y - \mathbf{X}'\mathbf{t})^2\}. \tag{11}$$

Note that if

$$\gamma(H) = \arg\min_{\mathbf{t}} E_H\{w_H(\mathbf{X}, Y)(Y - \mathbf{X}'\mathbf{T}_0(H) - \mathbf{X}'\mathbf{t})^2\},$$

then

$$\mathbf{T}_1(H) = \mathbf{T}_0(H) + \gamma(H).$$

In Gervini and Yohai (2000) it is proved that $E_H\{w_H(\mathbf{X}, Y)(Y - \mathbf{X}'\mathbf{T}_0(H))^2\}$ is finite for any $H \in \mathcal{H}_\varepsilon$ with $\varepsilon < \min\{\varepsilon^*_{\mathbf{T}_0}, \varepsilon^*_{S_0}\}$ and hence $\gamma(H)$ is well defined. When $E_H\{w_H(\mathbf{X}, Y)\|\mathbf{X}\|^2\}$ is infinite it is more complicated to obtain an explicit expression for $\gamma(H)$, but in this paper we will not need it anyway. The interested reader can find it in Gervini and Yohai (2000). Assuming that the estimating functional $\mathbf{T}_0$ is Fisher consistent and F_0 is symmetric, the Fisher consistency of $\mathbf{T}_1$ follows from (11). The equivariance of $\mathbf{T}_1$ follows from the equivariance of $\mathbf{T}_0$ and S_0.

Theorem 3.1 shows that the asymptotic breakdown point of the REWLSE is not less than those of the initial regression and scale estimators, and therefore it attains the maximum 50% if the initial estimators are properly chosen. The proof of Theorem 3.1 can be found in Gervini and Yohai (2000). We make the following assumptions:

R1. $P_{G_0}(\mathbf{X}'\mathbf{v} = 0) < 1$ for every $\mathbf{v} \in \mathbb{R}^p$;
R2. $\Sigma = E_{G_0}(\mathbf{X}\mathbf{X}')$ is finite and positive definite;
R3. F_0 is strictly increasing.

THEOREM 3.1. *If the hypothetical distribution F has finite variance and* R1–R3 *and* W1 *hold, then the asymptotic breakdown point of the REWLSE is* $\varepsilon^*_{\mathbf{T}_1} \geq \min\{\varepsilon^*_{\mathbf{T}_0}, \varepsilon^*_{S_0}\}$.

For a more complete description of the robustness properties of the REWLSE, we would need to compute the maxbias function for all ε between 0 and $\varepsilon^*_{\mathbf{T}_1}$. For some estimators this is possible [see, for instance, Martin, Yohai and Zamar (1989)]. Unfortunately, we were not able to do that for the REWLSE. However, when the initial estimator is the LMS, it is possible to evaluate numerically the maxbias function for point-mass contaminations. This is discussed next.

3.2. *Point-maxbias function when the LMS is the initial estimator.* To illustrate the effect of the REWLS weighting scheme in the bias of the initial estimator, we consider the case of the LMS as initial estimator. This special situation is important because the maximum bias of the LMSE is close to the minimum maxbias attainable within the class of residual admissible estimators [see Yohai and Zamar (1993)]. Besides, the LMSE is perhaps the most popular robust estimator among practitioners, despite its shortcomings.

To obtain a numerically computable approximation of $\mathcal{B}_{\mathbf{T}}(\varepsilon)$, we restrict ourselves to a narrower neighborhood of H_0, where only point-mass contaminations are allowed, and consider the point-maxbias function

$$\mathcal{B}^*_{\mathbf{T}}(\varepsilon) = \sup\{b(\mathbf{T}, (1-\varepsilon)H_0 + \varepsilon\Delta_{\mathbf{z}}) : \mathbf{z} \in \mathbb{R}^{p+1}\}. \tag{12}$$

For the LMSE the maxbias and the point-maxbias functions coincide. This is proved in Martin, Yohai and Zamar (1989), where an explicit expression for $\mathcal{B}_{\mathrm{LMS}}(\varepsilon)$ is given. Here we take H_0 as the multivariate normal, $\eta = 2.5$ and the hard-rejection weight $w(u) = I(u < 1)$. See Gervini and Yohai (2000) for a detailed explanation of how (12) is computed. Table 1 displays some values

TABLE 1
Maximum biases for point-mass contaminations

Percentage contamination	**Estimator**		
	LMS	**REWLS**	τ
0.05	0.53	0.54	0.63
0.10	0.83	0.85	0.95
0.15	1.14	1.17	1.30
0.20	1.52	1.56	1.62

of (12) for LMSE and REWLSE. Maxbiases of the WLSE with cutoff value $t_0 = 2.5$ were also computed and they coincide with those of the REWLSE (up to two decimal places). We also include the maximum biases of a τ-estimator with 0.5 breakdown point and 0.95 relative efficiency [see Table 1 in Yohai and Zamar (1988)].

We see in Table 1 that the maximum biases of the REWLSE exceed those of the LMSE only slightly, and in all cases they are less than the maximum biases of the τ-estimator. In conclusion, we can say that the proposed weighting scheme does not affect the asymptotic bias of the initial estimator very seriously. In any case, the small losses in asymptotic bias are compensated by the gains in asymptotic efficiency. A similar behavior was observed for small samples, as the Monte Carlo study reported in Section 5 shows.

3.3. *Influence function.* Besides the maxbias function, a useful tool for evaluating stability of an estimator is the influence function. Given $(\mathbf{x}, y) \in \mathbb{R}^{p+1}$ and $\Delta_{(\mathbf{x},y)}$, the corresponding point-mass distribution, let $H_\varepsilon = (1-\varepsilon)H_0 + \varepsilon\Delta_{(\mathbf{x},y)}$. The influence function of $\mathbf{T}$ at $(\mathbf{x}, y)$ is defined as

$$\mathrm{IF}_{\mathbf{T}}(\mathbf{x}, y) = \lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\{\mathbf{T}(H_\varepsilon) - \theta\}.$$

The book by Hampel, Ronchetti, Rousseeuw and Stahel (1986) develops a theory of robust estimation based on the influence function, focusing on bounded-influence estimators. This approach, however, leaves out many estimators with good robust properties. We will say more on this after Theorem 3.2.

Theorem 3.2 gives the influence function for a general WLSE computed with cutoff values $t(H_\varepsilon)$ that converge to a certain t_0 when ε goes to 0. To apply this result to the REWLSE, we need to know the limiting behavior of $d(H)$ and $t(H)$. Let

$$d_0 = \sup_{t\geq\eta}\{F^+(t) - F_0^+(t)\}^+,$$
$$t_0 = \min\{t : F_0^+(t) \geq 1 - d_0\}. \tag{13}$$

Observe that $t_0 \geq \eta$.

LEMMA 3.1. *Let $d(H)$ and $t(H)$ be as in* (10). *If $\mathbf{T}_0$ and S_0 are Fisher consistent*, $\lim_{\varepsilon\downarrow 0}\mathcal{B}_{\mathbf{T}_0}(\varepsilon)=\lim_{\varepsilon\downarrow 0}\mathcal{B}_{S_0}(\varepsilon)=0$ *and $F_0(t)$ is continuous*, *then*:

(i) $\lim_{\varepsilon\downarrow 0}\sup_{H\in\mathcal{H}_\varepsilon}\|F_0^+-F_H^+\|_\infty=0$, *where* $\|\cdot\|_\infty$ *denotes the* sup norm *of functions*.

(ii) $\lim_{\varepsilon\downarrow 0}\sup_{H\in\mathcal{H}_\varepsilon}|d(H)-d_0|=0$.

(iii) *If* $t_0=\infty$, $\lim_{\varepsilon\downarrow 0}\inf_{H\in\mathcal{H}_\varepsilon}t(H)=\infty$. *If $t_0<\infty$ and F_0 is strictly increasing in its support*, *then* $\lim_{\varepsilon\downarrow 0}\sup_{H\in\mathcal{H}_\varepsilon}|t(H)-t_0|=0$.

For Theorem 3.2 we require additional smoothness properties:

W2. The function $h_1(s,t)=\int w(s|u-t|)\,dF_0(u/\sigma)$ is continuous at $(s,0)$ for every $s\geq 0$. Let

$$\tau_1=h_1\left(\frac{1}{\sigma t_0},0\right)=\int w\left(\frac{|u|}{t_0}\right)dF_0(u). \tag{14}$$

W3. The function $h_2(s,t)=\int w(s|u-t|)u\,dF_0(u/\sigma)$ is differentiable in the variable t for every $s\geq 0$ and $\partial h_2/\partial t$ is continuous and bounded in both variables. Let

$$\tau_2=\frac{\partial h_2}{\partial t}\left(\frac{1}{\sigma t_0},0\right). \tag{15}$$

Condition W2 holds if $w(u)$ is continuous or if $F_0(u)$ is absolutely continuous. Condition W3 holds if $w(u)$ is continuously differentiable, in which case

$$\tau_2=-\int w'\left(\frac{|u|}{t_0}\right)\frac{|u|}{t_0}\,dF_0(u),$$

or if $F_0(u)$ has a continuously differentiable density function $f_0(u)$, in which case

$$\begin{aligned}\tau_2&=\int w\left(\frac{|u|}{t_0}\right)\{f_0(u)+uf_0'(u)\}\,du\\&=\tau_1+\int w\left(\frac{|u|}{t_0}\right)uf_0'(u)\,du.\end{aligned}$$

THEOREM 3.2. *Let $\mathbf{T}_1(H)$ be a WLSE computed with arbitrary cutoff values $t(H)$ such that* $\lim_{\varepsilon\downarrow 0}t(H_\varepsilon)=t_0$. *If assumptions* R2 *and* W1–W3 *hold*, *the initial estimators satisfy* $\lim_{\varepsilon\downarrow 0}\mathbf{T}_0(H_\varepsilon)=\theta$ *and* $\lim_{\varepsilon\downarrow 0}S_0(R_{H_\varepsilon})=\sigma$ *and the error distribution F_0 is symmetric*, *then the influence function of $\mathbf{T}_1$ is*

$$\mathrm{IF}_{\mathbf{T}_1}(\mathbf{x},y)=\tau_1^{-1}\left\{w\left(\frac{|y-\mathbf{x}'\theta|}{\sigma t_0}\right)\Sigma^{-1}\mathbf{x}(y-\mathbf{x}'\theta)+\tau_2\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x},y)\right\}, \tag{16}$$

where $\mathrm{IF}_{\mathbf{T}_0}$ *is the influence function of the initial estimator and τ_1 is given by* (14) *and τ_2 by* (15).

When $t_0 = \infty$ we have $\tau_1 = 1$ and $\tau_2 = 0$, so the influence function of the REWLSE coincides with that of the LSE and then it is unbounded. In spite of that, the REWLSE is nonetheless robust. Yohai and Zamar (1993) proved that residual-admissible estimators (a broad class that includes, among others, the LMSE and S-estimators) share with the REWLSE this characteristic of positive (even maximum) asymptotic breakdown point but unbounded influence. Other classes of estimators, such as GM-estimators, have bounded influence but their breakdown points tend to 0 when p increases. Hence bounded influence is neither a necessary nor a sufficient condition for robustness.

3.4. *Finite-sample breakdown point.* Section 3.1 analyzes the robustness of the REWLSE from an asymptotic point of view. In particular, Theorem 3.1 gives a lower bound for the asymptotic breakdown point of the REWLSE under the gross-error model. An analogous but nonasymptotic measure of robustness is the finite-sample replacement breakdown point, defined by Donoho and Huber (1983) as follows. Given a random sample $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, let $\mathbf{z}_i = (\mathbf{x}_i, y_i)$ and $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$. For $m \le n$, let $\mathcal{Z}_m$ be the set of all corrupted samples $\mathbf{Z}^*$ obtained after replacing m data points of $\mathbf{Z}$ with arbitrary values. The finite-sample replacement breakdown point of a regression estimator $\mathbf{T}_n$ is defined as the smallest fraction of outliers that can carry the estimator beyond all bounds. Formally,

$$\varepsilon_n^*(\mathbf{T}_n, \mathbf{Z}) = \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{\mathbf{Z}^* \in \mathcal{Z}_m} \|\mathbf{T}_n(\mathbf{Z}^*)\| = \infty \right\}.$$

The following theorem gives a lower bound for $\varepsilon_n^*(\mathbf{T}_{1n}, \mathbf{Z})$ when $\mathbf{Z}$ is in general position and S_n is a scale M-estimator. We recall [see Rousseeuw and Yohai (1984)] that $\mathbf{Z}$ is said to be in general position if no hyperplane in $\mathbb{R}^p$ can contain more than p points of $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. Given $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$, an M-estimator of scale S_n is defined as

$$S_n(\mathbf{u}) = \inf \left\{ s > 0 : \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{u_i}{s}\right) \le b \right\}, \tag{17}$$

where ρ is even, nonnegative and nondecreasing for $u \ge 0$ and $\rho(0) = 0$. Usually, b is chosen equal to $E_\Phi(\rho)$ so as to make S_n consistent for σ when $u_1, \dots, u_n$ is a random sample of an $N(0, \sigma^2)$ distribution.

THEOREM 3.3. *Assume that* W1 *is satisfied*, F *has finite variance and the sample* $\mathbf{Z}$ *is in general position. Also assume that* S_n *is a scale* M*-estimator based on a* ρ*-function such that* $\rho(u) = a$ *if* $|u| \ge c$, *where* $c < \eta$. *Then*

$$\varepsilon_n^*(\mathbf{T}_{1n}, \mathbf{Z}) \ge \min\{\varepsilon_n^*(\mathbf{T}_{0n}, \mathbf{Z}), b/a, 1 - b/a - p/n\}.$$

Observe that if $b/a = (n-p)/2n$ this theorem implies that $\varepsilon_n^*(\mathbf{T}_{1n}, \mathbf{Z}) \geq \min\{\varepsilon_n^*(\mathbf{T}_{0n}, \mathbf{Z}), (n-p)/2n\}$. Since, for any regression equivariant estimate $\mathbf{T}_n$, $\varepsilon_n^*(\mathbf{T}_n, \mathbf{Z}) \leq (\lfloor (n-p)/2 \rfloor + 1)/n$ [see Rousseeuw and Leroy (1987)], we get $\varepsilon_n^*(\mathbf{T}_{1n}, \mathbf{Z}) \geq \varepsilon_n^*(\mathbf{T}_{0n}, \mathbf{Z}) - 1/n$ for any equivariant estimator $\mathbf{T}_{0n}$.

A popular scale estimator is the standardized MAD defined by

$$S_n(u_1, \ldots, u_n) = \text{median}(|u_1|, \ldots, |u_n|)/\Phi^{-1}(3/4),$$

which is the scale M-estimator corresponding to the choices

$$\rho(u) = I\big(|u| \geq \Phi^{-1}(3/4)\big),$$

$b = 1/2$, $a = 1$ and $c = \Phi^{-1}(3/4)$. Note that c is less than $\eta = 2.5$ as required by Theorem 3.3. In this case $b/a = 1/2$, which is close to the optimal value $(n-p)/2n$ for large n. Another common scale M-estimator corresponds to Tukey's biweight function

$$\rho_c(u) = \begin{cases} \dfrac{u^2}{2}\left(1 - \dfrac{u^2}{c^2} + \dfrac{u^4}{3c^4}\right), & \text{if } |u| \leq c, \\ \dfrac{c^2}{6}, & \text{if } |u| > c. \end{cases} \tag{18}$$

The tuning constant that simultaneously makes $b/a = 1/2$ and $b = E_\Phi(\rho_c)$ is $c = 1.547$, which is also less than $\eta = 2.5$.

4. Asymptotics of the REWLSE. This section studies the asymptotic behavior of the REWLSE under the central model. We show that, under fairly general assumptions on the error distribution F_0 and the design space, the REWLSE is asymptotically equivalent to the LSE and hence asymptotically efficient for the normal-error model. We assume throughout this section that the sample $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ follows the central linear model (1). The explanatory variables are assumed to be deterministic. However, all the results are still valid if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a random sample stochastically independent of the errors, because the assumptions on the design would hold with probability 1.

We consider first the regular case of $\mathbf{T}_{0n}$ being $n^{1/2}$-consistent and asymptotically linear. In that situation convergence in probability of the cutoff values is enough to ensure $n^{1/2}$-consistency and asymptotic normality of the REWLSE. The asymptotic behavior of the cutoff values (4) is established in Lemma 4.1, which makes use of the following assumptions:

A1. $\mathbf{T}_{0n} \to \theta$ and $S_n \to \sigma$ in probability;
A2. $\lim_{K\to\infty} \limsup n^{-1} \sum_{i=1}^n I(\|\mathbf{x}_i\| > K) = 0$.

These assumptions are very general. Condition A1 should be satisfied under general conditions by all estimators used in practice. For the special case of S-estimators, see Theorem 3 of Davies (1990). Condition A2 is always satisfied by random carriers with probability 1.

LEMMA 4.1. *If* A1 *and* A2 *are satisfied and* F_0 *is continuous*, *then*:

(i) $\|F_n^+ - F_0^+\|_\infty = o_P(1)$.
(ii) *If* d_n *is as in* (3) *and* d_0 *as in* (13), *then* $d_n \to d_0$ *in probability*.
(iii) *If*, *in addition*, F_0 *is strictly increasing in its support and* t_n *is as in* (4) *and* t_0 *as in* (13), *then also* $t_n \to t_0$ *in probability*.

Lemma 4.1 implies that if the tails of F_0 are lighter than the tails of the assumed distribution F, then no observations are downweighted in the limit, and the REWLSE becomes asymptotically equivalent to the LSE as shown in Theorem 4.1. Note, however, that if F_0 has heavier tails than the normal, then the LSE is not a good estimator. For this reason we recommend taking $F = \Phi$.

To obtain the asymptotic distribution of the REWLSE, we need to assume:

A3. The initial regression estimator admits the asymptotic linear expansion

$$\mathbf{T}_{0n} - \theta = \Gamma_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi\left(\frac{u_i}{\sigma}\right) \mathbf{x}_i + o_P(n^{-1/2}),$$

where ψ is a bounded odd function and $\Sigma_n^{1/2} \Gamma_n^{-1} \Sigma_n^{1/2}$ converges in probability to a symmetric positive-definite matrix Γ, where Σ_n is as in A4.

A4. Let $\Sigma_n = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'$. Then the sequence of smallest eigenvalues of Σ_n is bounded away from 0.

A5. $\lim_{K\to\infty} \limsup n^{-1} \sum_{i=1}^{n} \|\mathbf{x}_i\|^2 I(\|\mathbf{x}_i\| > K) = 0$.

W4. Let

$$h_3(s,t) = \int w(s|u-t|) u \psi(u/\sigma)\, dF_0(u/\sigma),$$

$$h_4(s_1,t_1,s_2,t_2) = \int w(s_1|u-t_1|) w(s_2|u-t_2|) u^2\, dF_0(u/\sigma).$$

These functions are bounded and continuous at $(s,0)$ and $(s_1,0,s_2,0)$, respectively, for every nonnegative s, s_1 and s_2.

Assumption A5 implies the condition called D1 by Davies (1990). Lemma B.1 in Appendix B shows that if $\mathbf{T}_{0n}$ is consistent and solves the estimating equation

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{x}_i' \mathbf{T}_{0n}}{S_n}\right) \mathbf{x}_i = \mathbf{0}, \tag{19}$$

then, under some regularity conditions on ψ, A3 is satisfied with

$$\Gamma = \frac{\sigma}{E_{F_0}(\psi')} \mathrm{I}. \tag{20}$$

The S- and τ-estimators are defined as global minimizers of complicated nonconvex functions. Except for very small p, finding the global minimum is computationally impractical or even infeasible. However, (19) is satisfied by any local minimum of an S-estimating function. Yohai and Zamar (1988) proved that local-minimum τ-estimators (including S-estimators) with high breakdown can be found by using an iterative weighted least squares algorithm starting with a consistent and high-breakdown regression estimator. Adrover, Bianco and Yohai (2001) showed that τ-estimators computed by a subsampling algorithm as proposed in Rousseeuw and Leroy (1987) are consistent and have the same breakdown point as the exact τ-estimator with probability as close to 1 as desired, provided enough subsamples are taken.

We state Theorems 4.1 and 4.3 below in terms of arbitrary cutoff values, not necessarily those given by (4). Our purpose in doing this is twofold: first to broaden the applicability of the results to include other types of WLSE's and second to pinpoint the hypotheses that are really necessary in each case. For instance, continuity of F_0 is not necessary for these theorems if w is smooth enough. For the REWLSE, though, we need the continuity of the error distribution to ensure convergence in probability of the cutoff values.

THEOREM 4.1. *Let* $\{\mathbf{T}_{1n}\}$ *be a sequence of WLSE's computed with arbitrary cutoff values* $\{t_n\}$ *that converge in probability to some* t_0 (*which may be infinite*). *Suppose that*

$$V(t_0) = \sigma^2 E_{F_0}\left\{ w\left(\frac{|U|}{t_0}\right) U \mathrm{I} + \tau_2 \psi(U) \frac{1}{\sigma} \Gamma \right\}^2 \tag{21}$$

is finite, *with* τ_2 *as in* (15). *If* F_0 *is symmetric and conditions* W1–W4, A1 *and* A3–A5 *are satisfied*, *then*

$$\sqrt{n}\, \Sigma_n^{1/2} (\mathbf{T}_{1n} - \theta) \to_D N\big(\mathbf{0}, V(t_0)/\tau_1^2\big), \tag{22}$$

with τ_1 *as in* (14). *Moreover*, *when* $t_0 = \infty$ *condition* A3 *is not necessary and* (22) *holds*, *provided* $\mathbf{T}_{0n} - \theta = O_P(n^{-1/2})$.

According to Theorem 4.1, when $t_0 = \infty$ the asymptotic variance of $\sqrt{n}\, \Sigma_n^{1/2} \times (\mathbf{T}_{1n} - \theta)$ comes down to $\sigma^2 E_{F_0}(U^2)\mathrm{I}$ and then the WLSE is asymptotically equivalent to the LSE. In such situations Theorem 4.1 is implicitly assuming that F_0 has finite variance. In the specific case of the REWLSE, when F_0 is strictly increasing in its support the asymptotic cutoff value is $t_0 = (F_0^+)^{-1}(1 - d_0)$, according to Lemma 4.1. If the support of F_0 is unbounded, $t_0 = \infty$ only if $d_0 = 0$. This in turn happens only if the tails of F_0 are lighter than the tails of the assumed distribution F, so $E_{F_0}(U^2)$ will automatically be finite if F is chosen with finite variance. If F_0 is of bounded support, the asymptotic cutoff $t_0 = (F_0^+)^{-1}(1)$ is finite and the REWLSE is asymptotically equivalent to the LSE only when hard-rejection weights are used.

Let us now turn to the case where the initial regression estimator is not $n^{1/2}$-consistent. This is the case of the LMSE, which is only $n^{1/3}$-consistent, although the companion scale estimator is still $n^{1/2}$-consistent [see Theorems 4 and 5 in Davies (1990)]. He and Portnoy (1992) proved that a WLSE with finite cutoff value converges at the same rate as the initial estimator. Theorem 4.2 shows that the REWLSE is able to produce the $n^{-1/2}$ rate of convergence when $t_0 = \infty$. The following lemma is necessary for the proof of Theorem 4.2.

LEMMA 4.2. *Suppose that the error distribution F_0 is symmetric and absolutely continuous, with a differentiable density function f_0 such that f_0' and $u^2 f_0'(u)$ are bounded. If* A5 *is satisfied and the initial estimators $\mathbf{T}_{0n}$ and S_n are n^τ-consistent with $\tau \geq 1/4$, then $\|F_n^+ - F_0^+\|_\infty = O_P(n^{-1/2})$.*

For Theorem 4.2 we need a third-moment condition on the design:

A6. $\lim_{K\to\infty} \limsup n^{-1} \sum_{i=1}^n \|\mathbf{x}_i\|^3 I(\|\mathbf{x}_i\| > K) = 0$.

THEOREM 4.2. *Let $\{\mathbf{T}_{1n}\}$ be a sequence of REWLSEs, computed with the cutoff values $\{t_n\}$ given by* (4). *Assume that* A6 *and all the conditions of Lemma* 4.2 *hold, $\mathbf{T}_{0n}$ is n^τ-consistent with $\tau > 1/4$ and the function $g(u) = u f_0(u)$ satisfies $g(u) \leq C(1 - F_0^+(u))^{1/2}$ for each $u \geq 0$ and some constant $C > 0$. If $w(u) = I(u < 1)$ and $t_0 = \infty$, with t_0 given by* (13), *then*

$$\sqrt{n}\, \Sigma_n^{1/2}(\mathbf{T}_{1n} - \theta) \to_D N(\mathbf{0}, \sigma^2 E_{F_0}(U^2)\mathrm{I}).$$

Although the preceding theorem is not as general as Theorem 4.1, it shows nonetheless that the weighting scheme we propose represents an important improvement over the classical WLSE with fixed cutoff value, since the REWLSE is now able to improve the rate of convergence of the initial estimator. The condition in Theorem 4.2 that $g(u) \leq C(1 - F_0^+(u))^{1/2}$ for each $u \geq 0$ is not too restrictive. Essentially, it says that $u^2 f_0^2(u)$ tends to 0 faster than the tail probabilities when $|u| \to \infty$, and this is satisfied by densities with exponential decrease. For instance, it holds for $F_0 = \Phi$. To see this, note that $2(1 - \Phi(u)) - u^2 f_0^2(u)$ is strictly decreasing on $[u_0, \infty)$ for some $u_0 > 0$ and tends to 0 when $u \to \infty$, so it cannot be negative in $[u_0, \infty)$. By adequately choosing $C > 1$ we can make $g(u) \leq C(1 - F_0^+(u))^{1/2}$ for every $u \geq 0$.

To finish this section, we consider the case of the initial regression estimator not being $n^{1/2}$-consistent and the asymptotic cutoff value being finite. This is essentially the same as doing a WLSE with a fixed finite cutoff value; hence we obtain basically the same result as the theorem on page 2166 of He and Portnoy (1992). The rate of convergence is not improved in this case.

THEOREM 4.3. *Let* $\{\mathbf{T}_{1n}\}$ *be a sequence of WLSE's computed with arbitrary cutoff values* $\{t_n\}$ *that converge in probability to some* $t_0 < \infty$. *Assume that* F_0 *is symmetric*, S_n *is consistent*, $\mathbf{T}_{0n}$ *is* n^{τ}*-consistent with* $\tau \geq 1/4$ *and conditions* A6 *and* W1–W3 *hold*. *In addition*, *for* $h_2(s,t)$ *the function given in* W3, *assume that* $\partial h_2/\partial t$ *is Lipschitz in the variable* t, *uniformly in* s. *Then*

$$\big(\tau_1 + o_P(1)\big)\Sigma_n(\mathbf{T}_{1n} - \theta) = \big(\tau_2 + o_P(1)\big)\Sigma_n(\mathbf{T}_{0n} - \theta) + O_P\big(n^{-1/2}\big),$$

with τ_1 *and* τ_2 *as in* (14) *and* (15), *respectively*.

Let $\nu = \tau_2/\tau_1$ as in He and Portnoy (1992). When F_0 has a differentiable density, $\nu = 1 + \tau_1^{-1} \int w(|u|/t_0) u f_0'(u)\,du$. If f_0 is unimodal the integral in the second term of ν is negative, so $\nu < 1$. Therefore, if $n^{\tau}(\mathbf{T}_{0n} - \theta)$ converges to a nondegenerate distribution $\mathcal{Z}$ with $\tau \in [1/4, 1/2)$, then $n^{\tau}(\mathbf{T}_{1n} - \theta) \to_D \nu\mathcal{Z}$ and the WLSE has less asymptotic variability than the initial estimator. So, in terms of asymptotic variability, there are still gains in doing the reweighting step.

5. Monte Carlo study and examples. In this section we report on a Monte Carlo study that was carried out to assess finite-sample efficiency and robustness of the REWLSE. As $\mathbf{T}_{0n}$ we used the LMSE and an S-estimator of regression. S-estimators of regression [Rousseeuw and Yohai (1984)] are defined as follows. Let S_n be an M-scale estimate as defined by (17). The corresponding regression S-estimator for model (1) is defined as

$$\mathbf{T}_n = \arg\min_{\mathbf{t}} S_n(y_1 - \mathbf{x}_1'\mathbf{t}, \ldots, y_n - \mathbf{x}_n'\mathbf{t}).$$

The scales used to standardize residuals were the standardized MAD when $\mathbf{T}_{0n}$ was the LMSE and the corresponding M-scale when $\mathbf{T}_{0n}$ was an S-estimator.

We compared the following estimators:

1. Least squares (LS).
2. Least median of squares (LMS).
3. An S-estimator based on Tukey's biweight ρ-function (18) with $c = 1.547$ (S). This estimator, as well as the LMSE, has finite-sample breakdown point equal to $(\lfloor n/2 \rfloor - p + 2)/n$ [see Rousseeuw and Leroy (1987)].
4. One-step weighted least squares with cutoff value $t_0 = 2.5$, starting from the LMSE (WLS-LMS).
5. Same as above, starting from the S-estimator described in estimator 3 (WLS-S).
6. REWLSE with hard-rejection weight $w(u) = I(u < 1)$ and $\eta = 2.5$, starting from the LMSE (REWLS-LMS).
7. Same as above, starting from the S-estimator described in estimator 3 (REWLS-S).
8. For the case of linear regression with t-distributed errors in Section 5.2 we also considered the corresponding maximum likelihood estimator.

We considered three different models:

- Regression with normal errors and no outliers.
- Regression with heavy-tailed errors (Student's t distribution with 3 degrees of freedom).
- Regression with normal errors, where some fraction of the sample is replaced by outliers.

The results for each of these models are reported in the sections that follow.

5.1. *Regression model with normal errors.* We considered regression models with intercept, normal carriers and normal errors. Specifically, let $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ be a random sample that follows the linear model (1) with $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip-1})'$ and such that $(x_{i1}, \ldots, x_{ip-1})'$ has an $N(\mu, \Sigma)$ distribution. Since all estimators are regression, affine and scale equivariant, without loss of generality we took $\mu = 0$, $\Sigma = \mathrm{I}$ and $\theta = \mathbf{0}$.

We considered sample sizes $n = 20, 50, 100, 200, 500, 1000$ and $p = 2, 5$. For each value of p and n we generated 1000 samples, and for each estimator $\mathbf{T}_n$ we computed the relative mean squared efficiency with respect to the LSE,

$$\mathrm{EFF} = \frac{\sum_{i=1}^{1000} \|\mathbf{T}_{ni}^{\mathrm{LS}}\|^2}{\sum_{i=1}^{1000} \|\mathbf{T}_{ni}\|^2},$$

where $\mathbf{T}_{ni}$ and $\mathbf{T}_{ni}^{\mathrm{LS}}$ are the ith generated values of $\mathbf{T}_n$ and LSE, respectively.

To compute the LMSE, we used a subsampling algorithm based on elemental sets [see Rousseeuw and Leroy (1987)] with 1000 subsamples. The SE was computed using an iterative reweighted LS algorithm, starting from the LMSE. The results are shown in Tables 2 and 3.

We observe that in order for the efficiencies of the REWLSEs to be close to 1, very large sample sizes are required. However, for $n \geq 100$ the REWLSEs are noticeably more efficient than the corresponding WLSE's. In Sections 5.2 and 5.3 we show that this improvement in efficiency is obtained without serious loss in robustness of the initial estimators.

TABLE 2
Efficiencies for normal errors and $p = 2$

Estimator	***n*** 20	50	100	200	500	1000
LMS	0.20	0.19	0.16	0.13	0.09	0.08
WLS-LMS	0.58	0.66	0.76	0.80	0.82	0.79
REWLS-LMS	0.61	0.68	0.79	0.86	0.91	0.93
S	0.27	0.28	0.29	0.26	0.27	0.26
WLS-S	0.61	0.73	0.86	0.83	0.89	0.87
REWLS-S	0.65	0.75	0.89	0.89	0.95	0.96

TABLE 3
Efficiencies for normal errors and $p = 5$

Estimator	*n* 20	50	100	200	500	1000
LMS	0.18	0.18	0.15	0.11	0.07	0.04
WLS-LMS	0.26	0.50	0.67	0.77	0.78	0.77
REWLS-LMS	0.26	0.51	0.70	0.83	0.88	0.92
S	0.18	0.23	0.25	0.25	0.27	0.26
WLS–S	0.23	0.50	0.71	0.80	0.86	0.87
REWLS-S	0.23	0.50	0.74	0.86	0.93	0.96

5.2. *Regression with heavy-tailed errors.* In the same situations as in Section 5.1, now the errors u_i's were generated according to a t distribution with 3 degrees of freedom. The efficiencies were calculated with respect to the maximum likelihood estimator and are shown in Tables 4 and 5.

TABLE 4
Efficiencies for Student errors with 3 *d.f. and $p = 2$*

Estimator	*n* 20	50	100	200	500	1000
LS	0.56	0.59	0.48	0.53	0.52	0.47
LMS	0.32	0.30	0.25	0.20	0.15	0.10
WLS-LMS	0.73	0.80	0.82	0.79	0.78	0.77
REWLS-LMS	0.72	0.80	0.83	0.81	0.81	0.80
S	0.45	0.48	0.49	0.53	0.55	0.54
WLS-S	0.77	0.86	0.89	0.88	0.88	0.90
REWLS-S	0.76	0.86	0.89	0.88	0.88	0.90

TABLE 5
Efficiencies for Student errors with 3 *d.f. and $p = 5$*

Estimator	*n* 20	50	100	200	500	1000
LS	0.61	0.60	0.54	0.51	0.51	0.51
LMS	0.21	0.23	0.17	0.10	0.05	0.03
WLS-LMS	0.38	0.70	0.75	0.74	0.65	0.56
REWLS-LMS	0.38	0.70	0.76	0.76	0.70	0.63
S	0.24	0.34	0.42	0.48	0.51	0.49
WLS-S	0.33	0.67	0.82	0.87	0.89	0.90
REWLS-S	0.33	0.67	0.83	0.87	0.89	0.89

We observe that the REWLSEs perform sometimes better than the WLSE's, and never worse. Both the WLS-SE and the REWLS-SE have an efficiency greater than 0.80 for $n \geq 50$ when $p = 2$ and for $n \geq 100$ when $p = 5$.

5.3. *Model with normal errors and some fraction of outlier contamination.* We considered the same normal-error and normal-carrier model as in Section 5.1, but now k observations in each sample were replaced by identical outliers of the form $(\mathbf{x}_0, y_0)$. Because of the sphericity of the normal distribution, without loss of generality we took $\mathbf{x}_0 = (1, x_0, 0, \dots, 0)'$. We chose $x_0 = 1, 10$, which correspond to low- and high-leverage outliers, respectively, and varied y_0 in the grid $\{0.1 j x_0 : j \text{ positive integer}\}$. Since the asymptotic behavior of these estimators under this contamination model is studied in Section 3, we only considered a small sample size $n = 50$. We took $p = 2$ and $k = 3, 5, 8, 10$. For each estimator $\mathbf{T}_n$ and each value of k, x_0 and y_0, we estimated the mean squared error based on 1000 Monte Carlo replications: $\text{MSE}(\mathbf{T}_n, k, x_0, y_0)$. Tables 6 and 7 report the values of $\max_{y_0} \text{MSE}(\mathbf{T}_n, k, x_0, y_0)$, which represents the worst performance of the estimator for that leverage and that number of outliers.

We observe that, in general, both the WLSE and the REWLSE behave similarly, and better than the LMSE and the SE. Estimators starting from SE outperformed those that used the LMSE as initial estimator.

5.4. *Examples.* We applied the REWLSE, starting with the LMS and the S-estimate, to several data sets included in Rousseeuw and Leroy (1987) that

TABLE 6
Maximum MSE with outliers with $x_0 = 1$

	Estimator					
k	**LMS**	**WLS-LMS**	**REWLS-LMS**	**S**	**WLS-S**	**REWLS-S**
3	0.34	0.09	0.09	0.24	0.09	0.08
5	0.48	0.17	0.16	0.40	0.16	0.15
8	0.90	0.42	0.42	0.93	0.42	0.41
10	1.45	0.79	0.79	1.59	0.80	0.80

TABLE 7
Maximum MSE with outliers with $x_0 = 10$

	Estimator					
k	**LMS**	**WLS-LMS**	**REWLS-LMS**	**S**	**WLS-S**	**REWLS-S**
3	0.34	0.19	0.18	0.24	0.15	0.15
5	0.56	0.38	0.38	0.35	0.26	0.25
8	1.14	0.92	0.92	0.55	0.45	0.45
10	1.89	1.62	1.62	0.69	0.57	0.57

contained influential outliers: (i) the international phone calls data (page 25), the Hertzprung–Russell diagram data (page 27), (iii) the salinity data (page 82) and (iv) the artificial data of Hawkins, Bradu and Kass (page 93). In all these cases the REWLSE with hard-rejection-weight function and $\eta = 2.5$ eliminates exactly the same observations as the nonadaptive weighted least squares estimate with fixed cutoff point 2.5, and it is not much influenced by outliers. These results are consistent with our Monte Carlo results of Section 5.3, which show that under outlier contamination both types of weighted least squares estimates behave similarly.

The program in MATLAB used to compute the REWLSE can be obtained from the authors upon request.

APPENDIX A

Proofs of robust results.

PROOF OF LEMMA 3.1. (i) If $\lim_{\varepsilon \downarrow 0} \sup_{H \in \mathcal{H}_\varepsilon} \|F_0^+ - F_H^+\|_\infty > 0$, then there would be a $\delta > 0$ and sequences $\varepsilon_n \downarrow 0$, $H_n \in \mathcal{H}_{\varepsilon_n}$ and $\{u_n\}$ such that $|F_0^+(u_n) - F_{H_n}^+(u_n)| \geq \delta$ for every n. Now

$$
\begin{aligned}
&|F_0^+(u_n) - F_{H_n}^+(u_n)| \\
&\quad \leq \left| F_0^+(u_n) - (1-\varepsilon_n) P_{H_0}\left(\frac{|Y - \mathbf{X}'\mathbf{T}_0(H_n)|}{S_0(R_{H_n})} \leq u_n \right) \right| + \varepsilon_n.
\end{aligned}
\tag{23}
$$

Since $\mathbf{T}_0(H_n) \to \theta$ and $S_0(R_{H_n}) \to \sigma$, continuity of F_0 and dominated convergence imply that

$$
P_{H_0}\left(\frac{|Y - \mathbf{X}'\mathbf{T}_0(H_n)|}{S_0(R_{H_n})} \leq t \right) \to F_0^+(t) \quad \text{uniformly in } t,
$$

so that

$$
\left| P_{H_0}\left(\frac{|Y - \mathbf{X}'\mathbf{T}_0(H_n)|}{S_0(R_{H_n})} \leq u_n \right) - F_0^+(u_n) \right| \to 0.
$$

Then (23) would imply that $|F_0^+(u_n) - F_{H_n}^+(u_n)| \to 0$, a contradiction.

(ii) Since $|d(H) - d_0| \leq \|F_0^+ - F_H^+\|_\infty$, the proof is immediate from (i).

(iii) First, let $t_0 < \infty$. If $\lim_{\varepsilon \downarrow 0} \sup_{H \in \mathcal{H}_\varepsilon} |t(H) - t_0| > 0$, then there would be a $\delta > 0$, sequences $\varepsilon_n \downarrow 0$ and $H_n \in \mathcal{H}_{\varepsilon_n}$ and a certain n_0 such that

$$
|t(H_n) - t_0| \geq \delta \qquad \text{for every } n \geq n_0. \tag{24}
$$

Since $F_{H_n}^+(t(H_n)) \geq 1 - d(H_n)$ by definition and $d(H_n) \to d_0$ by part (i), we have that $\liminf F_{H_n}^+(t(H_n)) \geq 1 - d_0$. Since $|F_0^+(t(H_n)) - F_{H_n}^+(t(H_n))| \to 0$ by part (i), we also have $\liminf F_0^+(t(H_n)) \geq 1 - d_0$ and then $\liminf t(H_n) \geq t_0$, which together with (24) imply that $\liminf t(H_n) \geq t_0 + \delta$. Then there exists a subsequence $\{t(H_{n_k})\}$ such that $t(H_{n_k}) > t_0 + \delta/2$ for every k. By the definition of $t(H)$ we would then have that $F_0^+(t_0+\delta/2) < 1-d(H_{n_k})$ for every k and taking the limit we obtain $F_0^+(t_0+\delta/2) \leq 1-d_0$. This cannot happen if $F_0(t)$ is strictly increasing in its support.

Now let $t_0 = \infty$. This implies that $d_0 = 0$ and F_0 is of unbounded support. If $\lim_{\varepsilon\downarrow 0}\inf_{H\in\mathcal{H}_\varepsilon} t(H) < \infty$, there would be a $K < \infty$, sequences $\varepsilon_n \downarrow 0$ and $H_n \in \mathcal{H}_{\varepsilon_n}$ and a certain n_0 such that $t(H_n) \leq K$ for every $n \geq n_0$. Then

$$1 - d(H_n) \leq F_{H_n}^+\big(t(H_n)\big) \leq F_{H_n}^+(K),$$

which in the limit implies that $1 = F_0^+(K)$. This contradicts the fact that F_0 is of unbounded support. $\square$

PROOF OF THEOREM 3.2. Given $\varepsilon > 0$ and a contaminating point $(\mathbf{x}_0, y_0)$, let $H_\varepsilon = (1-\varepsilon)H_0 + \varepsilon\Delta_{(\mathbf{x}_0,y_0)}$. Then

$$\text{(25)}\quad \mathbf{T}_1(H_\varepsilon) - \theta = \big[E_{H_\varepsilon}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}\mathbf{X}'\}\big]^{-1} E_{H_\varepsilon}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}(Y-\mathbf{X}'\theta)\}.$$

We have

$$E_{H_\varepsilon}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}\mathbf{X}'\} = (1-\varepsilon)E_{H_0}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}\mathbf{X}'\} + \varepsilon w_{H_\varepsilon}(\mathbf{x}_0,y_0)\mathbf{x}_0\mathbf{x}_0'$$

and

$$E_{H_0}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}\mathbf{X}'\} = E_{G_0}\left\{h_1\left(\frac{1}{S_0(R_{H_\varepsilon})t(H_\varepsilon)}, \mathbf{X}'\big(\mathbf{T}_0(H_\varepsilon)-\theta\big)\right)\mathbf{X}\mathbf{X}'\right\}.$$

So assumption W2 and dominated convergence imply that

$$\lim_{\varepsilon\downarrow 0} E_{H_\varepsilon}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}\mathbf{X}'\} = \tau_1\Sigma. \tag{26}$$

Similarly,

$$\begin{aligned}E_{H_\varepsilon}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}(Y-\mathbf{X}'\theta)\} &= (1-\varepsilon)E_{H_0}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}(Y-\mathbf{X}'\theta)\}\\ &\quad + \varepsilon w_{H_\varepsilon}(\mathbf{x}_0,y_0)\mathbf{x}_0(y_0-\mathbf{x}_0'\theta).\end{aligned}$$

Here

$$E_{H_0}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}(Y-\mathbf{X}'\theta)\} = E_{G_0}\left\{h_2\left(\frac{1}{t(H_\varepsilon)S_0(R_{H_\varepsilon})}, \mathbf{X}'\big(\mathbf{T}_0(H_\varepsilon)-\theta\big)\right)\mathbf{X}\right\}.$$

By W3 and the symmetry of F_0 we have

$$\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}h_2\left(\frac{1}{t(H_\varepsilon)S_0(R_{H_\varepsilon})}, \mathbf{x}'\big(\mathbf{T}_0(H_\varepsilon)-\theta\big)\right) = \tau_2\mathbf{x}'\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0,y_0).$$

Then by dominated convergence

$$\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}E_{H_0}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}(Y-\mathbf{X}'\theta)\} = \tau_2\,\Sigma\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0,y_0)$$

and hence

$$\begin{aligned}(27)\qquad &\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}E_{H_\varepsilon}\{w_{H_\varepsilon}(\mathbf{X},Y)\mathbf{X}(Y-\mathbf{X}'\theta)\}\\ &\quad=\tau_2\Sigma\mathrm{IF}_{\mathbf{T}_0}(\mathbf{x}_0,y_0)+w\biggl(\frac{|y_0-\mathbf{x}_0'\theta|}{\sigma t_0}\biggr)\mathbf{x}_0(y_0-\mathbf{x}_0'\theta).\end{aligned}$$

From (25), (26) and (27), we obtain expression (16). □

LEMMA A.1. *Let S_n be an M-scale defined by* (17) *and suppose that $\rho(u)=a$ for $|u|\ge c$, $\rho(0)=0$ and ρ is continuous at* 0. *Then*:

(i) *For any $\mathbf{u}=(u_1,\dots,u_n)$ we have*

$$\#\{i:|u_i|\ge cS_n(\mathbf{u})\}\le\frac{nb}{a}.$$

(ii) *Let $A(M,j)=\{\mathbf{u}\in\mathbb{R}^n:\#\{i:|u_i|\le M\}\ge j\}$. Then if $m<nb/a$, we have*

$$\sup_{\mathbf{u}\in A(M,n-m)}S_n(\mathbf{u})<\infty.$$

PROOF. Suppose that assertion (i) is not true. Then if $A=\{i:|u_i|\ge cS_n(\mathbf{u})\}$, we have $\#A>nb/a$ and then

$$\frac{1}{n}\sum_{i=1}^n\rho\biggl(\frac{u_i}{S_n(\mathbf{u})}\biggr)\ge\frac{1}{n}\sum_{i\in A}\rho\biggl(\frac{u_i}{S_n(\mathbf{u})}\biggr)\ge\frac{1}{n}\sum_{i\in A}\rho(c)>b,$$

which contradicts the definition of $S_n(\mathbf{u})$.

Now we prove (ii). Given $M\ge 0$ and $m<nb/a$, let $0<\varepsilon<nb/a-m$. Then there exists a $K>0$ such that

$$\rho\biggl(\frac{M}{K}\biggr)\le\frac{\varepsilon a}{n}.$$

We will prove that

$$\sup_{\mathbf{u}\in A(M,n-m)}S_n(\mathbf{u})\le K.\tag{28}$$

Take $\mathbf{u}\in A(M,n-m)$ and let $B=\{i:|u_i|\le M\}$. Then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\rho\biggl(\frac{u_i}{K}\biggr)&=\frac{1}{n}\sum_{i\in B}\rho\biggl(\frac{u_i}{K}\biggr)+\frac{1}{n}\sum_{i\notin B}\rho\biggl(\frac{u_i}{K}\biggr)\\ &<\frac{1}{n}\sum_{i\in B}\rho\biggl(\frac{M}{K}\biggr)+\frac{a}{n}\biggl(\frac{nb}{a}-\varepsilon\biggr)\\ &\le\frac{\varepsilon a}{n}+b-\frac{\varepsilon a}{n}\le b\end{aligned}$$

and therefore (28) must hold by the definition of $S_n(\mathbf{u})$. $\square$

PROOF OF THEOREM 3.3. Let $m_0 = \min\{\lceil n - nb/a - p\rceil, \lceil nb/a\rceil, n\varepsilon_n^* \times (\mathbf{T}_{0n}, \mathbf{Z})\}$, where $\lceil x\rceil$ denotes the smallest integer that is larger than or equal to x. If $n - nb/a \le p + 1$, $m_0 = 1$ and it is obvious that $\varepsilon_n^*(\mathbf{T}_{0n}, \mathbf{Z}) \ge m_0/n$. Then we will consider only the case $\lceil n - nb/a - p\rceil \ge 2$. Take $m < m_0$ and let

$$M_1 = \sup_{\mathbf{Z}^* \in \mathcal{Z}_m} \|\mathbf{T}_{0n}(\mathbf{Z}^*)\|, \qquad M_2 = \sup_{1\le i\le n} |y_i|, \qquad M_3 = \sup_{1\le i\le n} \|\mathbf{x}_i\|.$$

Then if $M = M_2 + M_1 M_3$, for any $\mathbf{Z}^* = \{\mathbf{z}_1^*, \dots, \mathbf{z}_n^*\} \in \mathcal{Z}_m$, $\mathbf{z}_i^* = (\mathbf{x}_i^*, y_i^*)$, we have

$$\sup_{1\le i\le n} |y_i - \mathbf{x}_i' \mathbf{T}_{0n}(\mathbf{Z}^*)| \le M,$$

and therefore $(y_1^* - \mathbf{x}_1^{*\prime}\mathbf{T}_{0n}(\mathbf{Z}^*), \dots, y_n^* - \mathbf{x}_n^{*\prime}\mathbf{T}_{0n}(\mathbf{Z}^*)) \in A(M, n-m)$, where $A(M, n-m)$ is defined in Lemma A.1(ii). Then

$$s_0 = \sup_{\mathbf{Z}^* \in \mathcal{Z}_m} S_n\bigl(y_1^* - \mathbf{x}_1^{*\prime}\mathbf{T}_{0n}(\mathbf{Z}^*), \dots, y_n^* - \mathbf{x}_n^{*\prime}\mathbf{T}_{0n}(\mathbf{Z}^*)\bigr) < \infty. \tag{29}$$

To simplify the notation, let $\mathbf{t}_0 = \mathbf{T}_{0n}(\mathbf{Z}^*)$, $s_0 = S_n(y_1^* - \mathbf{x}_1^{*\prime}\mathbf{t}_0, \dots, y_n^* - \mathbf{x}_n^{*\prime}\mathbf{t}_0)$ and $r_i^* = (y_i^* - \mathbf{x}_i^{*\prime}\mathbf{t}_0)/s_0$ and let $|r^*|_{(1)}, \dots, |r^*|_{(n)}$ be the ordered $|r_i^*|$'s. If $i_0 = \max\{i : |r^*|_{(i)} < \eta\}$, then

$$d_n = \max_{i > i_0} \left\{ F^+\bigl(|r^*|_{(i)}\bigr) - \frac{i-1}{n} \right\}^+ \tag{30}$$

and $t_n = |r^*|_{(i_n)}$ with $i_n = n - \lfloor nd_n \rfloor$. Remember that $i_n > i_0$ and $t_n > \eta$. Let $w_i = w(|r_i^*|/t_n)$ and $w_{(i)} = w(|r^*|_{(i)}/t_n)$. Since $\mathbf{T}_{1n}(\mathbf{Z}^*)$ minimizes $\sum_{i=1}^n w_i (y_i^* - \mathbf{x}_i^{*\prime}\mathbf{t})^2$,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n w_i \bigl(y_i^* - \mathbf{x}_i^{*\prime}\mathbf{T}_{1n}(\mathbf{Z}^*)\bigr)^2 &\le \frac{1}{n}\sum_{i=1}^n w_i (y_i^* - \mathbf{x}_i^{*\prime}\mathbf{t}_0)^2 \\ &= \frac{s_0^2}{n}\sum_{i=1}^n w_{(i)} |r^*|_{(i)}^2. \end{aligned} \tag{31}$$

We have $w_{(i)} = 0$ for $i \ge i_n$ and, according to (30),

$$|r^*|_{(i)}^2 \le \left\{ (F^+)^{-1}\left(\frac{i + \lfloor nd_n \rfloor}{n} \right) \right\}^2$$

for $i \geq i_0 + 1$. Then

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} w_{(i)} |r^*|_{(i)}^2 &\leq \eta^2 + \frac{1}{n}\sum_{i=i_0+1}^{n-\lfloor nd_n\rfloor -1} |r^*|_{(i)}^2 \\
&\leq \eta^2 + \frac{1}{n}\sum_{i=i_0+1}^{n-\lfloor nd_n\rfloor -1} \left\{(F^+)^{-1}\left(\frac{i+\lfloor nd_n\rfloor}{n}\right)\right\}^2 \\
&\leq \eta^2 + \frac{1}{n}\sum_{i=1}^{n}\left\{(F^+)^{-1}\left(\frac{i-1}{n}\right)\right\}^2 \qquad (32)\\
&\leq \eta^2 + \int_0^1 \{(F^+)^{-1}(u)\}^2\,du \\
&= \eta^2 + \int_{-\infty}^{\infty} u^2\,dF < \infty.
\end{aligned}
$$

Therefore by (29), (31) and (32) we have

$$
\sup_{\mathbf{Z}^* \in \mathcal{Z}_m} \frac{1}{n}\sum_{i=1}^{n} w_i \big(y_i^* - \mathbf{x}_i^{*\prime}\mathbf{T}_{1n}(\mathbf{Z}^*)\big)^2 < \infty. \qquad (33)
$$

By Lemma A.1(i) we have $\#\{|r_i^*| \leq c\} \geq n - na/b$. Since $m < n - nb/a - p$ we have $|r_i^*| \leq c$ for at least $p+1$ points of the original sample, say $\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{p+1}}$. For those points $w_i \geq w(c/\eta) > 0$. Then

$$
\frac{1}{n}\sum_{i=1}^{n} w_i\big(y_i^* - \mathbf{x}_i^{*\prime}\mathbf{T}_{1n}(\mathbf{Z}^*)\big)^2 \geq w\left(\frac{c}{\eta}\right)\frac{1}{n}\sum_{j=1}^{p+1}\big(y_{i_j} - \mathbf{x}_{i_j}^{\prime}\mathbf{T}_{1n}(\mathbf{Z}^*)\big)^2. \qquad (34)
$$

Let

$$
\delta(\mathbf{Z}) = \min_{\|\mathbf{v}\|=1} \min\{|\mathbf{x}_{i_j}^{\prime}\mathbf{v}| : 1 \leq i_1 < \cdots < i_{p+1} \leq n\}.
$$

By the general-position assumption, $\delta(\mathbf{Z}) > 0$. Then

$$
\begin{aligned}
\sum_{j=1}^{p+1}\big(y_{i_j} - \mathbf{x}_{i_j}^{\prime}\mathbf{T}_{1n}(\mathbf{Z}^*)\big)^2 &\geq \sum_{j=1}^{p+1}\left(\frac{1}{2}|\mathbf{x}_{i_j}^{\prime}\mathbf{T}_{1n}(\mathbf{Z}^*)|^2 - y_{i_j}^2\right) \qquad (35)\\
&\geq \frac{(p+1)}{2}\delta^2(\mathbf{Z})\|\mathbf{T}_{1n}(\mathbf{Z}^*)\|^2 - \sum_{i=1}^{n} y_i^2.
\end{aligned}
$$

From (33), (34) and (35) we deduce that

$$
\sup_{\mathbf{Z}^* \in \mathcal{Z}_m} \|\mathbf{T}_{1n}(\mathbf{Z}^*)\| < \infty
$$

and then $\varepsilon_n^*(\mathbf{T}_{1n}, \mathbf{Z}) > m/n$. $\square$

APPENDIX B

Proofs of asymptotic results. The proofs that follow will make use of stochastic process concepts that can be found in Pollard (1990). The notation then will be somewhat different from the rest of the article because we are going to follow Pollard's notation. We will assume that the i.i.d. errors $\{u_i(\omega)\}$ are defined on a common probability space $(\Omega, \mathcal{S}, \mathbb{P})$. $\mathbb{E}$ will denote expectation with respect to $\mathbb{P}$. To simplify the notation, we will also use $\beta_{0n} = \mathbf{T}_{0n} - \theta$ and $\nu_n = (S_n t_n)^{-1}$.

PROOF OF LEMMA 4.1. (i) Let us consider the family of functions

$$\{f_i(\omega, \mathbf{v}, s) = I(|u_i(\omega) - \mathbf{x}_i'\mathbf{v}| \le s) : (\mathbf{v}, s) \in T = \mathbb{R}^p \times \mathbb{R}^+\}.$$

For each $\omega \in \Omega$, $\{(f_1(\omega, \mathbf{v}, s), \dots, f_n(\omega, \mathbf{v}, s)) : (\mathbf{v}, s) \in T\}$ has pseudodimension at most $p + 2$ according to Lemma 5 in Gervini and Yohai (2000). Then by Corollary 4.10 of Pollard (1990) the process $\{f_i(\omega, \mathbf{v}, s) : (\mathbf{v}, s) \in T\}$ is Euclidean with envelope $\mathbf{F}$ given by $F_i = 1$. Let $S_n(\omega, \mathbf{v}, s) = \sum_{i=1}^n f_i(\omega, \mathbf{v}, s)$ and $M_n(\mathbf{v}, s) = \sum_{i=1}^n \mathbb{E} f_i(\cdot, \mathbf{v}, s)$, so that

$$F_n^+(t) = \frac{1}{n} S_n(\omega, \beta_{0n}, S_n t).$$

By the maximal inequality (7.10) of Pollard (1990), we have

$$\mathbb{E}\Big\{\sup_{(\mathbf{v},s)\in T} |S_n(\cdot, \mathbf{v}, s) - M_n(\mathbf{v}, s)|^2\Big\} \le Cn$$

for some constant C. This implies

$$\sup_{t\ge 0} \frac{1}{\sqrt{n}} |S_n(\omega, \beta_{0n}, S_n t) - M_n(\beta_{0n}, S_n t)| = O_P(1)$$

or, equivalently,

$$\sup_{t\ge 0}\left|F_n^+(t) - \frac{1}{n} M_n(\beta_{0n}, S_n t)\right| = O_P(n^{-1/2}). \tag{36}$$

So in order to prove $\|F_n^+ - F_0^+\|_\infty = o_P(1)$ we only have to show that

$$\sup_{t\ge 0}\left|\frac{1}{n} M_n(\beta_{0n}, S_n t) - F_0^+(t)\right| = o_P(1). \tag{37}$$

Note that

$$\frac{1}{n} M_n(\beta_{0n}, S_n t) - F_0^+(t) = \frac{1}{n}\sum_{i=1}^n \Big\{F_0\Big(\frac{S_n t + \mathbf{x}_i'\beta_{0n}}{\sigma}\Big) - F_0(t)\Big\}$$
$$- \frac{1}{n}\sum_{i=1}^n \Big\{F_0\Big(\frac{-S_n t + \mathbf{x}_i'\beta_{0n}}{\sigma}\Big) - F_0(-t)\Big\}.$$

Assumptions A1 and A2 and the continuity of F_0 imply convergence in probability to 0 for each t. That F_0 is a continuous distribution function implies that this convergence is uniform in t and hence (37) follows.

(ii) Since $|d_n - d_0| \le \|F_n^+ - F_0^+\|_\infty$ by definition, the result follows immediately from part (i)

(iii) Use the results in parts (i) and (ii) and proceed as in the proof of Lemma 3.1(ii). □

PROOF OF LEMMA 4.2. Consider again the sequence of processes $\{f_i(\omega, \mathbf{v}, s) : (\mathbf{v}, s) \in T\}$ as in the proof of Lemma 4.1. Since we have already established (36) it suffices to show that

$$\sup_{t\ge 0}\left|\frac{1}{n}M_n(\beta_{0n}, S_n t) - F_0^+(t)\right| = O_P(n^{-1/2}).$$

We will first show that

$$\sup_{t\ge 0}\left|\frac{1}{n}M_n(\beta_{0n}, S_n t) - F_0^+\left(\frac{S_n t}{\sigma}\right)\right| = O_P(n^{-1/2}) \tag{38}$$

and then that

$$\sup_{t\ge 0}\left|F_0^+\left(\frac{S_n t}{\sigma}\right) - F_0^+(t)\right| = O_P(n^{-1/2}). \tag{39}$$

Second-order Taylor expansions yield

$$\begin{aligned} &F_0\left(\frac{S_n t + \mathbf{x}_i'\beta_{0n}}{\sigma}\right) - F_0\left(\frac{S_n t}{\sigma}\right) \\ &\qquad = f_0\left(\frac{S_n t}{\sigma}\right)\frac{\mathbf{x}_i'\beta_{0n}}{\sigma} + f_0'(\xi_{in}(t))\frac{1}{2}\left(\frac{\mathbf{x}_i'\beta_{0n}}{\sigma}\right)^2 \end{aligned} \tag{40}$$

and

$$\begin{aligned} &F_0\left(\frac{-S_n t + \mathbf{x}_i'\beta_{0n}}{\sigma}\right) - F_0\left(\frac{-S_n t}{\sigma}\right) \\ &\qquad = f_0\left(\frac{-S_n t}{\sigma}\right)\frac{\mathbf{x}_i'\beta_{0n}}{\sigma} + f_0'(\zeta_{in}(t))\frac{1}{2}\left(\frac{\mathbf{x}_i'\beta_{0n}}{\sigma}\right)^2, \end{aligned} \tag{41}$$

with $|\xi_{in}(t)| \le |\mathbf{x}_i'\beta_{0n}|$ and $|\zeta_{in}(t)| \le |\mathbf{x}_i'\beta_{0n}|$. Now subtract (41) from (40) and use the symmetry of f_0 to obtain

$$M_n(\beta_{0n}, S_n t) - F_0^+\left(\frac{S_n t}{\sigma}\right) = \frac{1}{n}\sum_{i=1}^n\{f_0'(\xi_{in}(t)) - f_0'(\zeta_{in}(t))\}\frac{1}{2}\left(\frac{\mathbf{x}_i'\beta_{0n}}{\sigma}\right)^2.$$

Since f_0' is bounded and $\|\beta_{0n}\|^2 = O_P(n^{-2\tau})$ with $\tau \geq 1/4$, (38) follows. To prove (39), we again use Taylor expansions and the symmetry of f_0 to get

$$F_0^+\left(\frac{S_n t}{\sigma}\right) = F_0^+(t) + \frac{1}{2}\{f_0'(\xi_n(t)) - f_0'(\zeta_n(t))\}\left(\frac{S_n t}{\sigma} - t\right)^2$$

for $\xi_n(t)$ between $S_n t/\sigma$ and t and $\zeta_n(t)$ between $-S_n t/\sigma$ and $-t$. Then $|t/\xi_n(t)|$ and $|t/\zeta_n(t)|$ fall between 1 and σ/S_n, so that

$$\sup_{t\geq 0}\left|F_0^+\left(\frac{S_n t}{\sigma}\right) - F_0^+(t)\right| \leq \sup_u |u^2 f_0'(u)| \left(\frac{\sigma}{S_n} \vee 1\right)^2 \left(\frac{S_n}{\sigma} - 1\right)^2 = O_P(n^{-2\tau}),$$

which implies (39) because $\tau \geq 1/4$. □

LEMMA B.1. *Let* $\mathbf{T}_{0n}$ *be an estimator of* θ *that satisfies* (19). *Suppose that*:

1. ψ *is odd and differentiable*, *and* $\lim_{u\to\infty}\psi(u) = c < \infty$;
2. ψ' *is bounded and continuous*, $E_{F_0}(\psi') \neq 0$;
3. F_0 *is symmetric*;
4. A1, A4 *and* A5 *hold*.

Then $\mathbf{T}_{0n}$ *satisfies assumption* A3 *with* $\Gamma_n = \sigma^{-1}E_{F_0}(\psi')\Sigma_n$ *and* Γ *is given by* (20).

PROOF. See Gervini and Yohai (2000). □

LEMMA B.2. *If conditions* A1, A5, W1 *and* W2 *are satisfied and* $t_n \to t_0$ *in probability*, *then*

$$\frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i' - \tau_1 \Sigma_n = o_P(1).$$

PROOF. First, note that, given $K > 0$,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i' - \tau_1 \Sigma_n\right| \\ &\qquad \leq \left|\frac{1}{n}\sum_{i=1}^n (w_i - \tau_1)\mathbf{x}_i \mathbf{x}_i' I(\|\mathbf{x}_i\| \leq K)\right| + 2\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i\|^2 I(\|\mathbf{x}_i\| > K).\end{aligned} \tag{42}$$

By A5 we only have to show that the first term in (42) is $o_P(1)$ for each $K > 0$. Fix two coordinates j and k and consider the sequence of processes

$$\{f_i(\omega, \mathbf{v}, s) = w(s|u_i(\omega) - \mathbf{x}_i'\mathbf{v}|)x_{ij}x_{ik}I(\|\mathbf{x}_i\| \leq K) : (\mathbf{v}, s) \in T = \mathbb{R}^p \times \mathbb{R}^+\}.$$

The sequence of processes $\{f_i(\omega, \mathbf{v}, s) : (\mathbf{v}, s) \in T\}$ is Euclidean with envelope given by $F_i = K^2$ and then

$$\mathbb{E}\left\{\sup_{(\mathbf{v},s)\in T}\left|\sum_{i=1}^{n} f_i(\omega, \mathbf{v}, s) - M_n(\mathbf{v}, s)\right|^2\right\} \le CnK^4 \tag{43}$$

where

$$M_n(\mathbf{v}, s) = \sum_{i=1}^{n} h_1(s, \mathbf{x}_i'\mathbf{v}) x_{ij} x_{ik} I(\|\mathbf{x}_i\| \le K)$$

and h_1 is defined in W2. From (43) we obtain that

$$\frac{1}{n}\left|\sum_{i=1}^{n} w_i x_{ij} x_{ik} I(\|\mathbf{x}_i\| \le K) - M_n(\beta_{0n}, \nu_n)\right| = o_P(1).$$

By assumption W2 we also have that

$$\frac{1}{n}\left|M_n(\beta_{0n}, \nu_n) - \tau_1 \sum_{i=1}^{n} x_{ij} x_{ik} I(\|\mathbf{x}_i\| \le K)\right| = o_P(1)$$

and the proof is complete. $\square$

From now on we will work with the standardized explanatory variables

$$\mathbf{z}_{ni} = \frac{1}{\sqrt{n}} \Sigma_n^{-1/2} \mathbf{x}_i.$$

Note that $\sum_{i=1}^{n} \mathbf{z}_{ni}\mathbf{z}_{ni}' = \mathrm{I}$ and that A5 implies

$$\lim_{n\to\infty} \sum_{i=1}^{n} \|\mathbf{z}_{ni}\|^2 I(\|\mathbf{z}_{ni}\| > \varepsilon) = 0 \qquad \text{for any } \varepsilon > 0.$$

LEMMA B.3. *Consider the process*

$$\mathbf{W}_n(\omega, s, \mathbf{t}) = \sum_{i=1}^{n}\left\{w\big(s|u_i(\omega) - \mathbf{z}_{ni}'\mathbf{t}|\big)u_i(\omega)\mathrm{I} + \tau_2\psi\left(\frac{u_i(\omega)}{\sigma}\right)\Gamma\right\}\mathbf{z}_{ni},$$

with $\{\mathbf{z}_{ni}\}$ *as above*, Γ *a symmetric* $p \times p$ *matrix and* ψ *a bounded odd function. Let* $\mathbf{M}_n(s, \mathbf{t}) = \mathbb{E}\mathbf{W}_n(\cdot, s, \mathbf{t})$. *If* $\sqrt{n}\Sigma_n^{1/2}\beta_{0n} = O_P(1)$ *and conditions* A1, A4, A5 *and* W1–W4 *hold, then*

$$\mathbf{W}_n\big(\omega, \nu_n, \sqrt{n}\Sigma_n^{1/2}\beta_{0n}\big) - \mathbf{M}_n\big(\nu_n, \sqrt{n}\Sigma_n^{1/2}\beta_{0n}\big) \to_D N\big(\mathbf{0}, V(t_0)\big),$$

with covariance matrix $V(t_0)$ *given by* (21).

PROOF. If we take $\alpha \in \mathbb{R}^p$, then $\alpha'\mathbf{W}_n(\omega, s, \mathbf{t}) = \sum_{i=1}^n f_{ni}(\omega, s, \mathbf{t})$, where

$$f_{ni}(\omega, s, \mathbf{t}) = w\big(s|u_i(\omega) - \mathbf{z}_{ni}'\mathbf{t}|\big)u_i(\omega)\alpha'\mathbf{z}_{ni} + \tau_2\psi\left(\frac{u_i(\omega)}{\sigma}\right)\alpha'\Gamma\mathbf{z}_{ni}.$$

Let $\nu_0 = (\sigma t_0)^{-1}$ and take $S = [\nu_0/2, \nu_0 + 1]$, so that $\mathbb{P}(\nu_n \in S) \to 1$. Take also a compact K in $\mathbb{R}^p$ such that $\mathbb{P}(\sqrt{n}\Sigma_n^{1/2}\beta_{0n} \in K) \to 1$. Then the array of processes $\{f_{ni}(\omega, s, \mathbf{t}) : (s, \mathbf{t}) \in S \times K\}$ is Euclidean with envelope $\mathbf{F}_n$ given by

$$F_{ni} = \left(\frac{2}{\nu_0} + \sqrt{p}\sup_{\mathbf{t}\in K}\|\mathbf{t}\|\right)|\alpha'\mathbf{z}_{ni}| + |\tau_2|\,\|\psi\|_\infty|\alpha'\Gamma\mathbf{z}_{ni}|$$

if $\nu_0 > 0$, and

$$F_{ni}(\omega) = |u_i(\omega)|\,|\alpha'\mathbf{z}_{ni}| + |\tau_2|\,\|\psi\|_\infty|\alpha'\Gamma\mathbf{z}_{ni}|$$

if $\nu_0 = 0$. We have used that $\max_{1\le i\le n}\|\mathbf{z}_{ni}\|^2 \le \sum_{i=1}^n\|\mathbf{z}_{ni}\|^2 = p$. By assumption, the envelopes are square-integrable in both cases. Moreover, we have

$$\limsup\sum_{i=1}^n \mathbb{E}F_{ni}^2 < \infty$$

and the Lindeberg condition

$$\lim_{n\to\infty}\sum_{i=1}^n \mathbb{E}\{F_{ni}^2 I(F_{ni} > \varepsilon)\} = 0 \qquad \text{for each } \varepsilon > 0.$$

Let h_i for $i = 2, 3, 4$ be the functions given in conditions W2–W4. Also, define

$$h_5(s_1, t_1, s_2, t_2) = \int\big\{w(s_1|u - t_1|) - w(s_2|u - t_2|)\big\}^2 u^2\, dF_0(u/\sigma),$$

which may actually be written in terms of h_4 but we define it explicitly for convenience. The covariance functional of the array of processes defined above is given by

$$\begin{aligned}
&H_n\big((s_1, \mathbf{t}_1), (s_2, \mathbf{t}_2)\big)\\
&\quad= \sum_{i=1}^n\big[\mathbb{E}\{f_{in}(\cdot, s_1, \mathbf{t}_1)f_{in}(\cdot, s_2, \mathbf{t}_2)\} - \mathbb{E}f_{in}(\cdot, s_1, \mathbf{t}_1)\mathbb{E}f_{in}(\cdot, s_2, \mathbf{t}_2)\big]\\
&\quad= \sum_{i=1}^n h_4(s_1, \mathbf{z}_{ni}'\mathbf{t}_1, s_2, \mathbf{z}_{ni}'\mathbf{t}_2)(\alpha'\mathbf{z}_{ni})^2 + \tau_2^2 E_{F_0}(\psi^2)\sum_{i=1}^n(\alpha'\Gamma\mathbf{z}_{ni})^2\\
&\qquad+ \tau_2\sum_{i=1}^n h_3(s_1, \mathbf{z}_{ni}'\mathbf{t}_1)(\alpha'\mathbf{z}_{ni})(\alpha'\Gamma\mathbf{z}_{ni}) + \tau_2\sum_{i=1}^n h_3(s_2, \mathbf{z}_{ni}'\mathbf{t}_2)(\alpha'\mathbf{z}_{ni})(\alpha'\Gamma\mathbf{z}_{ni})\\
&\qquad- \left\{\sum_{i=1}^n h_2(s_1, \mathbf{z}_{ni}'\mathbf{t}_1)(\alpha'\mathbf{z}_{ni})\right\}\left\{\sum_{i=1}^n h_2(s_2, \mathbf{z}_{ni}'\mathbf{t}_2)(\alpha'\mathbf{z}_{ni})\right\}.
\end{aligned}$$

Note that $h_2(s,0)=0$ for any $s \geq 0$ by the symmetry of F_0, and recall that $\sum_{i=1}^n \mathbf{z}_{ni} \cdot \mathbf{z}'_{ni} = \mathrm{I}$. Then it follows from assumptions W1–W4 that $H_n((s_1,\mathbf{t}_1),(s_2,\mathbf{t}_2)) \to H((s_1,\mathbf{t}_1),(s_2,\mathbf{t}_2))$, where

$$H\big((s_1,\mathbf{t}_1),(s_2,\mathbf{t}_2)\big) = h_4(s_1,0,s_2,0)\|\alpha\|^2 + \tau_2^2 E_{F_0}(\psi^2)\|\alpha'\Gamma\|^2 + \tau_2 h_3(s_1,0)\alpha'\Gamma\alpha + \tau_2 h_3(s_2,0)\alpha'\Gamma\alpha.$$

Consider now the sequence of pseudometrics in $S \times K$ given by

$$\rho_n\big((s_1,\mathbf{t}_1),(s_2,\mathbf{t}_2)\big) = \left\{\sum_{i=1}^n \mathbb{E}|f_{ni}(\cdot,s_1,\mathbf{t}_1) - f_{ni}(\cdot,s_2,\mathbf{t}_2)|^2\right\}^{1/2} = \left\{\sum_{i=1}^n h_5(s_1,\mathbf{z}'_{ni}\mathbf{t}_1,s_2,\mathbf{z}'_{ni}\mathbf{t}_2)(\alpha'\mathbf{z}_{ni})^2\right\}^{1/2}.$$

Again it is not difficult to see that $\rho_n((s_1,\mathbf{t}_1),(s_2,\mathbf{t}_2)) \to \rho((s_1,\mathbf{t}_1),(s_2,\mathbf{t}_2))$, where

$$\rho\big((s_1,\mathbf{t}_1),(s_2,\mathbf{t}_2)\big) = \|\alpha\|^2 h_5(s_1,0,s_2,0) = \|\alpha\|^2 \int \big\{w(s_1|u|) - w(s_2|u|)\big\}^2 u^2\, dF_0(u/\sigma).$$

Note that ρ does not depend on $\mathbf{t}_1$ and $\mathbf{t}_2$. The continuity and boundedness of h_5 guarantee that the convergence of ρ_n to ρ is uniform and then, for any deterministic sequences $\{(s_{1n},\mathbf{t}_{1n})\}$ and $\{(s_{2n},\mathbf{t}_{2n})\}$ in $S \times K$,

$$\rho\big((s_{1n},\mathbf{t}_{1n}),(s_{2n},\mathbf{t}_{2n})\big) \to 0 \Rightarrow \rho_n\big((s_{1n},\mathbf{t}_{1n}),(s_{2n},\mathbf{t}_{2n})\big) \to 0.$$

Hence all conditions of the functional central limit theorem (10.6) of Pollard (1990) are satisfied. Let $\ell^\infty(S \times K)$ be the space of all bounded, real-valued functions on $S \times K$, equipped with the sup norm, and let $U_\rho(S \times K) \subset \ell^\infty(S \times K)$ be the subspace of uniformly ρ-continuous functions. Then $S \times K$ is totally bounded under the ρ pseudometric, the finite-dimensional distributions of $\alpha'(\mathbf{W}_n - \mathbf{M}_n)$ have Gaussian limits with zero mean and covariances given by H, which uniquely determine a Gaussian distribution P concentrated on $U_\rho(S \times K)$ and $\alpha'(\mathbf{W}_n - \mathbf{M}_n)$ converges in distribution to P. Let Z be a random element in $U_\rho(S \times K)$ with distribution P. To complete the proof, define a map g: $\ell^\infty(S \times K) \times (S \times K) \to \mathbb{R}$ as $g(x,(s,\mathbf{t})) = x(s,\mathbf{t})$. Since $(\alpha'(\mathbf{W}_n - \mathbf{M}_n),(\nu_n,\sqrt{n}\Sigma_n^{1/2}\beta_{0n}))$ converges in distribution to $(Z,(\nu_0,\mathbf{0}))$ in the product norm, with $Z \in U_\rho(S \times K)$, and the map g is continuous for any $x \in U_\rho(S \times K)$, we apply the continuous mapping theorem to obtain that

$$\alpha'\big\{\mathbf{W}_n\big(\omega,\nu_n,\sqrt{n}\Sigma_n^{1/2}\beta_{0n}\big) - \mathbf{M}_n\big(\nu_n,\sqrt{n}\Sigma_n^{1/2}\beta_{0n}\big)\big\} \to_D Z(\nu_0,\mathbf{0}).$$

Since $Z(\nu_0,\mathbf{0})$ has a zero-mean normal distribution with variance $H((\nu_0,\mathbf{0}),(\nu_0,\mathbf{0})) = \alpha' V(t_0)\alpha$ and $\alpha \in \mathbb{R}^p$ was arbitrary, the proof is complete. □

PROOF OF THEOREM 4.1. Let $\beta_{1n} = \mathbf{T}_{1n} - \theta$. Then

$$\sum_{i=1}^{n} w_i \mathbf{z}_{ni} u_i = \left(\sum_{i=1}^{n} w_i \mathbf{z}_{ni} \mathbf{z}_{ni}'\right) \sqrt{n}\, \Sigma_n^{1/2} \beta_{1n}.$$

We apply Lemma B.2 to the right-hand side of this equation to obtain that $\sum_{i=1}^{n} w_i \times \mathbf{z}_{ni} \mathbf{z}_{ni}' \to \tau_1 \mathrm{I}$ in probability. On the left-hand side we use Lemma B.3 and then we only have to prove that

$$\left\| \mathbf{M}_n(\nu_n, \sqrt{n}\, \Sigma_n^{1/2} \beta_{0n}) - \tau_2 \sum_{i=1}^{n} \psi\left(\frac{u_i}{\sigma}\right) \Gamma \mathbf{z}_{ni} \right\| = o_P(1)$$

or, equivalently,

$$\|\mathbf{M}_n(\nu_n, \sqrt{n}\, \Sigma_n^{1/2} \beta_{0n}) - \tau_2 \sqrt{n}\, \Sigma_n^{1/2} \beta_{0n}\| = o_P(1). \tag{44}$$

But $\mathbf{M}_n(s, \mathbf{t}) = \sum_{i=1}^{n} h_2(s, \mathbf{z}_{ni}' \mathbf{t}) \mathbf{z}_{ni}$ with h_2 as in W2. Recalling that $\sum_{i=1}^{n} \mathbf{z}_{ni} \cdot \mathbf{z}_{ni}' = \mathrm{I}$, we can rewrite (44) as

$$\left\| \sum_{i=1}^{n} \{h_2(\nu_n, \mathbf{z}_{ni}' \mathbf{t}_{0n}) - \tau_2 \mathbf{z}_{ni}' \mathbf{t}_{0n}\} \mathbf{z}_{ni} \right\| = o_P(1), \tag{45}$$

with $\mathbf{t}_{0n} = \sqrt{n}\, \Sigma_n^{1/2} \beta_{0n}$. Since $h_2(\nu_n, 0) = 0$ by symmetry of F_0 and $(\partial/\partial t) h_2(s, t)$ is bounded and continuous in s at $t = 0$, (45) follows immediately.

If $t_0 = \infty$, then $\tau_2 = 0$ and we only need to prove

$$\|\mathbf{M}_n(\nu_n, \sqrt{n}\, \Sigma_n^{1/2} \beta_{0n})\| = o_P(1).$$

This can be proved under the assumption $\mathbf{t}_{0n} = O_P(1)$, with no need of the linear expansion A3. □

PROOF OF THEOREM 4.2. Since the LSE is asymptotically normal under these hypotheses, it is enough to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - w_i) \mathbf{x}_i u_i = o_P(1).$$

Fix a coordinate j, take $\delta > 0$ and consider the sequence of processes

$$\{f_i(\omega, s, \mathbf{v}) = I(s|u_i(\omega) - \mathbf{x}_i' \mathbf{v}| \ge 1) x_{ij} u_i(\omega) : (s, \mathbf{v}) \in T(\delta)\},$$

with $T(\delta) = \{(s, \mathbf{v}) : 0 \le s < \delta, \|\mathbf{v}\| < \delta\}$. This sequence of processes is manageable for envelopes given by

$$F_i(\omega, \delta) = \left\{ I\left(|u_i(\omega)| \ge \frac{1}{2\delta}\right) + I\left(\|\mathbf{x}_i\| \ge \frac{1}{2\delta^2}\right) \right\} \|\mathbf{x}_i\| |u_i(\omega)|.$$

Then, if $S_n(\omega, s, \mathbf{v}) = \sum_{i=1}^n f_i(\omega, s, \mathbf{v})$ and $M_n(s, \mathbf{v}) = \sum_{i=1}^n \mathbb{E} f_i(\cdot, s, \mathbf{v})$, we have

$$\mathbb{E}\Big\{\sup_{T(\delta)} |S_n(\cdot, s, \mathbf{v}) - M_n(s, \mathbf{v})|^2\Big\} \le C \sum_{i=1}^n \mathbb{E} F_i^2(\cdot, \delta)$$

for some C independent of δ. Now take a deterministic sequence $\delta_n \downarrow 0$ such that $\mathbb{P}\{(\nu_n, \beta_{0n}) \in T(\delta_n)\} \to 1$. For (ν_n, β_{0n}) in $T(\delta_n)$ use Chebyshev's inequality and the fact that $n^{-1} \sum_{i=1}^n \mathbb{E} F_i^2(\cdot, \delta_n)$ tends to 0 to obtain

$$\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n (1 - w_i) x_{ij} u_i - M_n(\nu_n, \beta_{0n}) \right| = o_P(1).$$

To complete the proof, we must show that $M_n(\nu_n, \beta_{0n}) = o_P(n^{1/2})$. Note that

$$M_n(\nu_n, \beta_{0n}) = \sum_{i=1}^n \bigg[\int I\bigg(\frac{|u - \mathbf{x}_i' \beta_{0n}|}{S_n t_n} \ge 1 \bigg) g\Big(\frac{u}{\sigma}\Big)\, du \bigg] x_{ij}$$

and

$$\int I\bigg(\frac{|u - \mathbf{x}_i' \beta_{0n}|}{S_n t_n} \ge 1 \bigg) g\Big(\frac{u}{\sigma}\Big)\, du = -\int_{-S_n t_n}^{S_n t_n} g\bigg(\frac{u + \mathbf{x}_i' \beta_{0n}}{\sigma} \bigg)\, du.$$

Let

$$G(s, t) = \int_{-s}^{s} g\Big(\frac{u + t}{\sigma} \Big)\, du.$$

Then

$$\dot{G}(s, t) \triangleq \frac{\partial}{\partial t} G(s, t) = g\Big(\frac{s + t}{\sigma} \Big) - g\Big(\frac{-s + t}{\sigma} \Big),$$
$$\ddot{G}(s, t) \triangleq \frac{\partial^2}{\partial t^2} G(s, t) = \frac{1}{\sigma} g'\Big(\frac{s + t}{\sigma} \Big) - \frac{1}{\sigma} g'\Big(\frac{-s + t}{\sigma} \Big)$$

are continuous and $\ddot{G}(s, t)$ is bounded. Note that $G(s, 0) = 0$ and $\dot{G}(s, 0) = 2g(s/\sigma)$. Since

$$M_n(\nu_n, \beta_{0n}) = -\sum_{i=1}^n G(S_n t_n, \mathbf{x}_i' \beta_{0n}) x_{ij}$$

from a second-order Taylor expansion of $G(S_n t_n, \mathbf{x}_i' \beta_{0n})$, we get

$$|M_n(\nu_n, \beta_{0n})| \le 2g\Big(\frac{S_n t_n}{\sigma} \Big) \|\beta_{0n}\| \sum_{i=1}^n \|\mathbf{x}_i\|^2 + \frac{2}{\sigma} \|g'\|_\infty \|\beta_{0n}\| \sum_{i=1}^n \|\mathbf{x}_i\|^3.$$

Since $\|\beta_{0n}\| = o_P(n^{-1/4})$, in order to prove $M_n(\nu_n, \beta_{0n}) = o_P(n^{1/2})$ it suffices to show that $g(S_n t_n/\sigma) = O_P(n^{-1/4})$. By definition, $F_n^+(t_n) \geq 1 - d_n$. Then

$$\begin{aligned} g(S_n t_n/\sigma) &\leq C\big(1 - F_0^+(S_n t_n/\sigma)\big)^{1/2} \\ &\leq C\big(1 - F_0^+(t_n)\big)^{1/2} + C\big|F_0^+(t_n) - F_0^+(S_n t_n/\sigma)\big|^{1/2} \\ &\leq C\big(d_n + F_n^+(t_n) - F_0^+(t_n)\big)^{1/2} + C\big|F_0^+(t_n) - F_0^+(S_n t_n/\sigma)\big|^{1/2} \\ &\leq C\big(2\|F_n^+ - F_0^+\|_\infty\big)^{1/2} + \sup_u \big|u^2 f_0'(u)\big|^{1/2} (S_n^{-1} \vee \sigma^{-1})|S_n - \sigma|. \end{aligned}$$

From Lemma 4.2 and the assumption that $|S_n - \sigma| = O_P(n^{-\tau})$ with $\tau \geq 1/4$, it follows that $g(S_n t_n/\sigma) = O_P(n^{-1/4})$. □

PROOF OF THEOREM 4.3. Since

$$\frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i'(\beta_{1n} - \beta_{0n}) = \frac{1}{n}\sum_{i=1}^n w_i (u_i - \mathbf{x}_i'\beta_{0n})\mathbf{x}_i,$$

we only have to prove that

$$\frac{1}{n}\sum_{i=1}^n w_i (u_i - \mathbf{x}_i'\beta_{0n})\mathbf{x}_i = \big(\tau_2 - \tau_1 + o_P(1)\big)\Sigma_n \beta_{0n} + O_P\big(n^{-1/2}\big). \tag{46}$$

Let $S = [\nu_0/2, \nu_0 + 1]$ and take a compact K such that $\mathbb{P}(\beta_{0n} \in K) \to 1$. Fix a coordinate j and consider the family of processes

$$\big\{f_i(\omega, s, \mathbf{t}) = w\big(s|u_i(\omega) - \mathbf{x}_i'\mathbf{t}|\big)\big(u_i(\omega) - \mathbf{x}_i'\mathbf{t}\big)x_{ij} : (s, \mathbf{t}) \in T = S \times K\big\}.$$

Since we are assuming $\nu_0 > 0$, this family is Euclidean with envelopes

$$F_i = \frac{2}{\nu_0}|x_{ij}|.$$

So if $S_n(\omega, s, \mathbf{t}) = \sum_{i=1}^n f_i(\omega, s, \mathbf{t})$ and $M_n(s, \mathbf{t}) = \sum_{i=1}^n \mathbb{E} f_i(\cdot, s, \mathbf{t})$, we have a maximal inequality

$$\mathbb{E}\Big\{\sup_T |S_n(\cdot, s, \mathbf{t}) - M_n(s, \mathbf{t})|^2\Big\} \leq C\frac{2}{\nu_0}\sum_{i=1}^n \|\mathbf{x}_i\|^2.$$

By Chebyshev's inequality this proves that $|S_n(\cdot, \nu_n, \beta_{0n}) - M_n(\nu_n, \beta_{0n})| = O_P(n^{1/2})$, or, in other words,

$$\left|\frac{1}{n}\sum_{i=1}^n w_i(u_i - \mathbf{x}_i'\beta_{0n})x_{ij} - \frac{1}{n}\sum_{i=1}^n \{h_2(\nu_n, \mathbf{x}_i'\beta_{0n}) - h_1(\nu_n, \mathbf{x}_i'\beta_{0n})\mathbf{x}_i'\beta_{0n}\}x_{ij}\right| = O_P\big(n^{-1/2}\big)$$

with h_1 and h_2 given in W2 and W3, respectively. Therefore

$$\frac{1}{n}\sum_{i=1}^{n} w_i(u_i - \mathbf{x}_i'\beta_{0n})\mathbf{x}_i$$
$$= \frac{1}{n}\sum_{i=1}^{n} h_2(\nu_n, \mathbf{x}_i'\beta_{0n})\mathbf{x}_i - \big(\tau_1 + o_P(1)\big)\Sigma_n\beta_{0n} + O_P\big(n^{-1/2}\big).$$

Since $h(\nu_n, 0) = 0$ the mean value theorem yields

$$h_2(\nu_n, \mathbf{x}_i'\beta_{0n}) = \frac{\partial}{\partial t}h_2(\nu_n, \xi_{ni})\mathbf{x}_i'\beta_{0n},$$

with $|\xi_{ni}| \le |\mathbf{x}_i'\beta_{0n}|$. Thus

$$\frac{1}{n}\sum_{i=1}^{n} h_2(\nu_n, \mathbf{x}_i'\beta_{0n})\mathbf{x}_i = \big(\tau_2 + o_P(1)\big)\Sigma_n\beta_{0n} + \mathbf{R}_n,$$

with

$$\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial}{\partial t}h_2(\nu_n, \xi_{ni}) - \frac{\partial}{\partial t}h_2(\nu_n, 0)\right\}\mathbf{x}_i\mathbf{x}_i'\beta_{0n}.$$

Then

$$\|\mathbf{R}_n\| \le L\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i\|^3\|\beta_{0n}\|^2 = O_P(n^{-2\tau}).$$

Since $\tau \ge 1/4$, consolidating O_P and o_P terms, we get (46) and the proof is complete. □

Acknowledgments. We thank both referees for their helpful criticism and comments which contributed to the improvement of the paper.

REFERENCES

ADROVER, J., BIANCO, A. and YOHAI, V. J. (2001). Approximate τ-estimates for linear regression based on subsampling of elemental sets. In *Statistics in Genetics and in the Environmental Sciences* (L. T. Fernholz, S. Morgenthaler and W. Stahel, eds.) 141–151. Birkäuser, Basel.

AGOSTINELLI, C. and MARKATOU, M. (1998). A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. *Statist. Probab. Lett.* **37** 341–350.

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

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

GERVINI, D. and YOHAI, V. J. (2000). A class of robust and fully efficient regression estimators. Unpublished manuscript.

HAMPEL, F. R. (1975). Beyond location parameters: robust concepts and methods (with discussion). *Bull. ISI* **46** 375–391.

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. and PORTNOY, S. (1992). Reweighted LS estimators converge at the same rate as the initial estimator. *Ann. Statist.* **20** 2161–2167.

HÖSSJER, O. (1992). On the optimality of S-estimators. *Statist. Probab. Lett.* **14** 413–419.

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

POLLARD, D. (1990). *Empirical Processes*: *Theory and Applications*. IMS, Hayward, CA.

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

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. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

SIMPSON, D. G. and YOHAI, V. J. (1998). Functional stability of one-step GM-estimators in approximately linear regression. *Ann. Statist.* **26** 1147–1169.

YOHAI, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.* **15** 642–656.

YOHAI, V. J. and ZAMAR, R. H. (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. J. and ZAMAR, R. H. (1993). A minimax-bias property of the least α-quantile estimates. *Ann. Statist.* **21** 1824–1842.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF ZURICH
SUMATRASTRASSE 30
CH-8006 ZURICH
SWITZERLAND
E-MAIL: gervini@ifspm.unizh.ch

DEPARTAMENTO DE MATEMÁTICA
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
CIUDAD UNIVERSITARIA, PABELLÓN 1
1428 BUENOS AIRES
ARGENTINA
E-MAIL: vyohai@dm.uba.ar