High Breakdown Rank Regression

William H. Chang
Trilogy Consulting Corp.

Joseph W. McKean & Joshua D. Naranjo
Western Michigan University

Simon J. Sheather
University of New South Wales

June 17, 1998

Abstract

A weighted rank estimate is proposed which has 50% breakdown and is asymptotically normal at rate $\sqrt{n}$. Based on this theory, inferential procedures, including asymptotic confidence and tests, and diagnostic procedures, such as Studentized residuals, are developed. The influence function of the estimate is derived and it is shown to be continuous and bounded everywhere in $(\mathbf{x}, Y)$ space. Examples show that robustness against outlying high-leverage clusters may approach that of the LMS, while retaining more stability against inliers. The estimator uses weights that correct for both factor and response spaces. A Monte Carlo study shows that the estimate is more efficient than GR, a similar estimate where the weights only correct for factor space. When weights are constant, the estimate reduces to the regular Wilcoxon rank estimate.

KEY WORDS: Rank regression; Weights; GR-estimate; LMS; 50% breakdown; Inlier stability.

1 Introduction

Rank-based estimates of the regression coefficients of a linear model were proposed by Jureckova (1971) and Jaeckel (1972). These estimates based on the Wilcoxon scores generalize the Wilcoxon-Mann-Whitney tests and estimates for simple location problems. In particular, they have the same efficiency properties that the nonparametric procedures have in the simple location problems. For example, if the errors of the linear model have a normal distribution, these Wilcoxon regression estimates have asymptotic relative efficiency .955 relative to least squares estimates. Further, this relative efficiency is much higher for long tailed error distributions. See Hettmansperger, McKean and Sheather (1997) for a recent review article on these Wilcoxon estimates and inference procedures based on them.

The Wilcoxon regression estimates have bounded influence in the Y-space but unbounded influence in the $\mathbf{x}$-space. Hence the breakdown of the Wilcoxon estimate is $1/n$. The Wilcoxon estimates, though, can easily be generalized to classes of weighted estimates. Sievers (1983) and Naranjo and Hettmansperger (1994) extended the Wilcoxon estimates to a class of weighted estimates with weights based on the $\mathbf{x}$'s. We shall call these GR estimates. These estimates have bounded influence in both the Y and $\mathbf{x}$ spaces and while they have positive breakdown they do not achieve 50% breakdown. The breakdown point of GR estimates is less than 33% and decreases with the number of explanatory variables (Naranjo and Hettmansperger, 1994).

In this paper, we propose weighted Wilcoxon estimates where the weights are a function of both $\mathbf{x}$ and the residuals from an initial fit. As we show in Section 4, these estimates achieve 50% breakdown provided the initial estimates have 50% breakdown. Further, since the weights are a function of both $\mathbf{x}$ and the residuals, they will be more efficient than the GR estimates. This is demonstrated in a Monte Carlo study presented in Section 8. We call these estimates high breakdown R estimates (HBR). As with the Wilcoxon and GR estimates, the HBR estimates converge at rate $\sqrt{n}$; hence, as discussed in Section 5, they can be used to establish an inference on the regression coefficients, including asymptotic confidence intervals and tests. This theory is used in Section 5 to develop Studentized residuals similar to those proposed by McKean, Sheather and Hettmansperger (1990, 1993) for robust fits of linear models.

Our weighted Wilcoxon estimates are based on Wilcoxon regression scores, hence we do not need to assume symmetry of the error density. Other rank-based estimates of regression are based on signed-rank regression scores, which require symmetry in order to obtain asymptotic results. See, also, Hettmansperger and McKean (1983) for a discussion of the distinction between these scores. Tableman (1990) proposed a k-step bounded influence estimate based on signed-rank scores and Hössjer (1994) proposed a high breakdown estimate based on signed-rank scores.

In Section 6, we derive the influence function (IF) of the HBR estimate. The IF turns out to be a continuous and bounded function in both the $\mathbf{x}$ and Y spaces, and tapers to 0 as $(\mathbf{x}, Y)$ get large in any direction. The influence function facilitates comparison of the HBR estimate with other high breakdown estimates; see Croux, Rousseeuw and Hössjer (1994) for discussion of the influence functions for other high breakdown estimates. In contrast to many high breakdown estimates, including least median squares and generalized S estimators, the influence function of the HBR estimate is bounded everywhere. This helps explain its stability to inliers as discussed below.

As with the Wilcoxon estimates, the GR and HBR estimates are obtained by minimizing a convex function. Unlike the k-step GM estimates proposed by Simpson, Ruppert and Carroll (1992) and Coakley and Hettmansperger (1993), the GR and HBR estimates are fully iterated estimates. Hence, for them determination of the number of steps is void. For high breakdown, the expensive computing step is in computing the initial estimates used by the weights. Once the weights are obtained, simple Gauss-Newton type algorithms are all that are necessary to perform the minimization (see e.g. Kapenga, McKean and Vidmar, 1988).

High breakdown regression estimates often have problems detecting and fitting curvature; see Cook, Hawkins and Weisberg (1992), McKean, Sheather and Hettmansperger (1993), and Naranjo et al. (1994). In Section 3, we present two examples which illustrate this curvature problem. Recently, McKean, Naranjo and Sheather (1996a) proposed diagnostics which are helpful for investigating the discrepancies between a highly efficient and a high breakdown estimate. These diagnostics can also be based on the HBR estimates. Another problem of concern is the instability of some high breakdown estimates to slight changes to centrally located data; i.e., inliers. This was discussed for the least median of squares (LMS) estimate by Hettmansperger and Sheather (1992, 1993) and Sheather, McKean and Hettmansperger (1997). Due to their availability, the LMS estimates are convenient to use as initial estimates for the HBR weights. An alternative for initial estimates are least trimmed squares (LTS) estimates. In Section 7 we present the results of a stability study for the weighted Wilcoxon estimates. The HBR estimates appear to be much more stable than either the LMS or the LTS estimates which were used as initial estimates.

These weighted Wilcoxon estimates provide the user with a wide spectrum of estimates, from the highly efficient Wilcoxon estimates (weights equal to one) to the 50% breakdown HBR estimates. As discussed by McKean, Naranjo and Sheather (1996b), these allow exploratory procedures for determining optimal fitting methods.

2 High Breakdown, Ranked-Based Estimate

Consider the linear regression model $y_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i,\ i = 1, \ldots, n$ where $\boldsymbol{x}_i'$ is the ith row of a known $n \times p$ matrix of centered explanatory variables $\mathbf{X}$, and $\epsilon_1, \ldots, \epsilon_n$ are independent random variables with density f. Consider estimating $\boldsymbol{\beta}$ by minimizing the convex function

$$D(\boldsymbol{\beta}) = \sum\sum_{i<j} b_{ij}|e_i - e_j| \tag{2.1}$$

where $e_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ and b_{ij} is a weight function. The dispersion function (2.1) was originally proposed by Sievers(1983). Note that $D(\boldsymbol{\beta})$ is free of the intercept α which may be estimated later from the residuals.

When $b_{ij} \equiv 1$, (2.1) reduces to Jaeckel's (1972) rank dispersion function for Wilcoxon scores

$$D(\boldsymbol{\beta}) = \sum\sum_{i<j} |e_i - e_j| = 2\sum_{i=1}^{n} [R(e_i) - (n+1)/2]\, e_i\ , \tag{2.2}$$

where $R(e_i)$ is the rank of e_i among $e_1 \cdots e_n$. Let $\widehat{\boldsymbol{\beta}}_W$ denote the estimate obtained by minimizing (2.2). This is an R estimate based on a linear score function and is often referred to as the Wilcoxon estimate; hence, we have used the subscript W in denoting this estimate. Under regularity conditions (see Heiler and Willers, 1988), it can be shown that

$$\widehat{\boldsymbol{\beta}}_W \text{ is asymptotically distributed as } N(\boldsymbol{\beta}, \tau^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}), \tag{2.3}$$

where τ is the scale parameter defined by

$$\tau^{-1} = \sqrt{12}\int_{-\infty}^{\infty} f^2(x)\, dx\ . \tag{2.4}$$

The Wilcoxon estimate is a highly efficient estimate having asymptotic relative efficiency of .955 with respect to the least squares estimate when the errors have a normal distribution and generally having much higher efficiency than least squares for error distributions with longer tails than the normal. The estimate $\widehat{\boldsymbol{\beta}}_W$, however, has breakdown $1/n$ because of its sensitivity to outliers in the $\mathbf{X}$-space. One way of obtaining an estimate with positive breakdown is by a proper choice of the weights.

Now let $\psi(t) = 1, \ t, \quad \text{or} \quad -1$ according as $t \geq 1$, $-1 < t < 1$, or $t \leq -1$. Let $m_i = \psi[b/(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})]$ for some constant b, where $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{S}$ are robust estimates of location and covariance of $\boldsymbol{x}$; see the discussion in Section 3. Finally, let

$$b_{ij} = \psi\left[\left|\frac{c\, m_i m_j}{(e_i(\hat{\boldsymbol{\beta}}_0)/\widehat{\sigma})(e_j(\hat{\boldsymbol{\beta}}_0)/\widehat{\sigma})}\right|\right], \tag{2.5}$$

where $\hat{\boldsymbol{\beta}}_0$ is an initial estimate, $e_i(\hat{\boldsymbol{\beta}}_0)$ is the ith residual of the initial estimate, and c is a tuning constant. The estimator of error standard deviation $\widehat{\sigma}$ used to scale the residuals in (2.5) is given by

$$\text{MAD} = 1.483\ \text{med}_i|e_i(\hat{\boldsymbol{\beta}}_0) - \text{med}_j\{e_j(\hat{\boldsymbol{\beta}}_0)\}|.$$

The estimator which minimizes (2.1) with weights given by (2.5) will be called the HBR estimate (for high breakdown rank estimate). When the weights are allowed to depend on the **X**-matrix only, the resulting estimator is the generalized rank (GR) estimator discussed by Naranjo and Hettmansperger (1994). The GR estimator has limited breakdown. Depending on the weights chosen, the breakdown is approximately 29% for $p = 1$, decreasing to approximately 16% for $p = 4$ (McKean et al., 1996). In any case, the GR breakdown cannot exceed 33%. In the next section, we show that the HBR estimate can have breakdown as high as 50%, depending on whether the estimates used for the weights have 50% breakdown. First we note the equivariance properties of the estimate. For a definition of the following equivariance terms, see, for example, Rousseeuw and Leroy (1987).

Proposition 1 *The HBR estimate is regression equivariant, scale equivariant, and affine equivariant provided the initial estimate employed by the weights are also regression, scale, and affine equivariant.*

Proof: We need to show that (i) $\hat{\boldsymbol{\beta}}_{HBR}(\boldsymbol{Y} + \boldsymbol{X\delta}) = \hat{\boldsymbol{\beta}}_{HBR}(\boldsymbol{Y}) + \boldsymbol{\delta}$, for all $\boldsymbol{\delta}$, (ii) $\hat{\boldsymbol{\beta}}_{HBR}(a\boldsymbol{Y}) = a\hat{\boldsymbol{\beta}}_{HBR}(\boldsymbol{Y})$ for all $a > 0$, and (iii) $\hat{\boldsymbol{\beta}}_{HBR}(\boldsymbol{Y}, A\boldsymbol{X}) = A^{-1}\hat{\boldsymbol{\beta}}_{HBR}(\boldsymbol{Y}, \boldsymbol{X})$, for any nonsingular matrix A. It can be shown that the weights (2.5) do not change under any of the above transformations on $\boldsymbol{Y}$ and $\boldsymbol{X}$. The results follow immediately from the fact that we are minimizing $\sum b_{ij}|e_i - e_j|$.

3 Examples

High breakdown estimates and highly efficient estimates often give conflicting results. The high breakdown estimates are less sensitive to outliers and clusters of outliers in the **x**-

space; hence, for data sets where this is a problem the high breakdown estimate often gives a better fit than the highly efficient fit. On the other hand, Cook, Hawkins and Weisberg (1992), McKean et al. (1993, 1994), and Naranjo et al. (1994), showed that high breakdown estimates may be hampered in fitting and detecting curvature, and that this is not true of the highly efficient estimates. We chose two examples which illustrate these disagreements between the high breakdown HBR fit and the highly efficient Wilcoxon fit.

For the GR estimates in this section and Sections 7 and 8.3, we used the Mallows type weighting scheme $b_{ij} = w_i w_j$, where

$$w_i = \min\left\{1,\ \left[\frac{b}{(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})}\right]^{k/2}\right\}. \tag{3.1}$$

The terms $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{S}$ are the minimum volume ellipsoid (MVE) measures of location and scatter (Rousseeuw and van Zomeren, 1990), computed using the algorithm proposed by Simonoff and Hawkins (1993). The cutoff point b was set at the 95th percentile of $\chi^2(p)$, and the parameter k was set at 2. Note that the severity of downweighting increases with k, with $k = 0$ corresponding to the Wilcoxon R-estimator.

For the HBR estimates, the weights (2.5) were computed using LMS as the initial estimate $\hat{\boldsymbol{\beta}}_0$, computed using the algorithm written by Stromberg (1993). The Mallows tuning constant b was set at the 95th percentile of $\chi^2(p)$. The tuning constant c was set at $[\text{Med}\{a_i\} + 3\,\text{MAD}\{a_i\}]^2$ where $a_i = e_i(\hat{\boldsymbol{\beta}}_0)/(\widehat{\sigma}\ m_i)$ and $\widehat{\sigma} = \text{MAD}\{e_i(\hat{\boldsymbol{\beta}}_0)\}$.

The Wilcoxon R-estimate was computed using the algorithm discussed in Kapenga et al. (1988). Similar Gauss-Newton type algorithms were used to obtain the weighted R-estimates.

Example 1: Stars Data. The variables measured are log-temperature, x, and log-light intensity of 47 stars; see Rousseeuw and Leroy (1987). Figure 1, Panel 1, shows a scatter plot of the data, overlaid by the Wilcoxon, HBR, and LMS fits. The values of the estimates are displayed in Table 1. There is strong disagreement between the Wilcoxon and HBR fits. A close look at the plot reveals the source of disagreement as the 4 outlying points in the upper left corner of the plot. These are outliers in the $\mathbf{x}$-space and they draw the Wilcoxon fit towards them. On the other hand, the HBR estimate is much less sensitive to these outliers and it fits the bulk of the data, as the LMS fit does also. Subject matter reveals that these 4 outlying stars are giant stars much larger than the other stars. Hence, the HBR fits the main group of stars and ignores these giant stars, while the Wilcoxon fit is quite sensitive to the giant stars. The remaining panels of Figure

Table 1: Estimates of Coefficients for the Examples

	Stars Data		Quadratic Data		
Fit	Intercept	Slope	Intercept	Linear	Quadratic
Wilcoxon	7.20	-.477	-.665	5.95	-.652
HBR	-6.91	2.71	.422	4.64	-.375
LMS	-12.8	4.00	1.12	3.65	-.141

1 display the residual plots based on these fits. Note that the 4 giant stars are apparent as outliers in the HBR fit. The Wilcoxon residual plot shows a quadratic-like pattern. Due to subject matter, the HBR fit is correct here; however, if subject matter was not available then an argument could be made that the Wilcoxon linear fit detected a quadratic pattern in the data whereas the HBR fit did not. This leads to our second example.

Example 2: Quadratic Data. In order to demonstrate the problems that the high breakdown estimates have in fitting curvature, we simulated data from the following quadratic model:

$$Y_i = 5.5|x_i| - .6x_i^2 + e_i \ , \tag{3.2}$$

where the e_i's were simulated iid $N(0,1)$ variates and the x_i's were simulated contaminated normal variates with the contamination proportion set at .25 and the ratio of the variance of the contaminated part to the non-contaminated part set at 16, (although, we actually modeled $|x_i|$). The first panel of Figure 2 displays a scatter plot of the data overlaid with the Wilcoxon, HBR, and LMS fits. The estimated coefficients for these fits are in Table 1. As shown, the Wilcoxon fit is quite good in fitting the curvature of the data. Its estimates are close to the true values. On the other hand, the high breakdown fits are quite poor. The LMS fit missed the curvature in the data. This is true too for the HBR fit, although, the fit did correct itself somewhat from the poor LMS starting values. The remaining panels of Figure 2 contain the residual plots based on these three fits. Based on the Wilcoxon residual plot, no further models would be considered. The HBR residual plot shows as outliers the two points which were fitted poorly. It also has a mild linear trend in it, which is not helpful since a linear term was fit. This is true for the LMS residual plot, also; although, it gives an overall impression of the lack of a quadratic term in the model. In such cases in practice, a higher degree polynomial may be fitted, which in this case would be incorrect. Difficulties in reading residual plots from high breakdown fits, as encountered here, were discussed in detail in McKean et al. (1993).

Figure 1 For the Stars Data, clockwise from the upper left corner: scatter plot of data and fits; the Wilcoxon residual plot; the LMS residual plot; and the HBR residual plot.

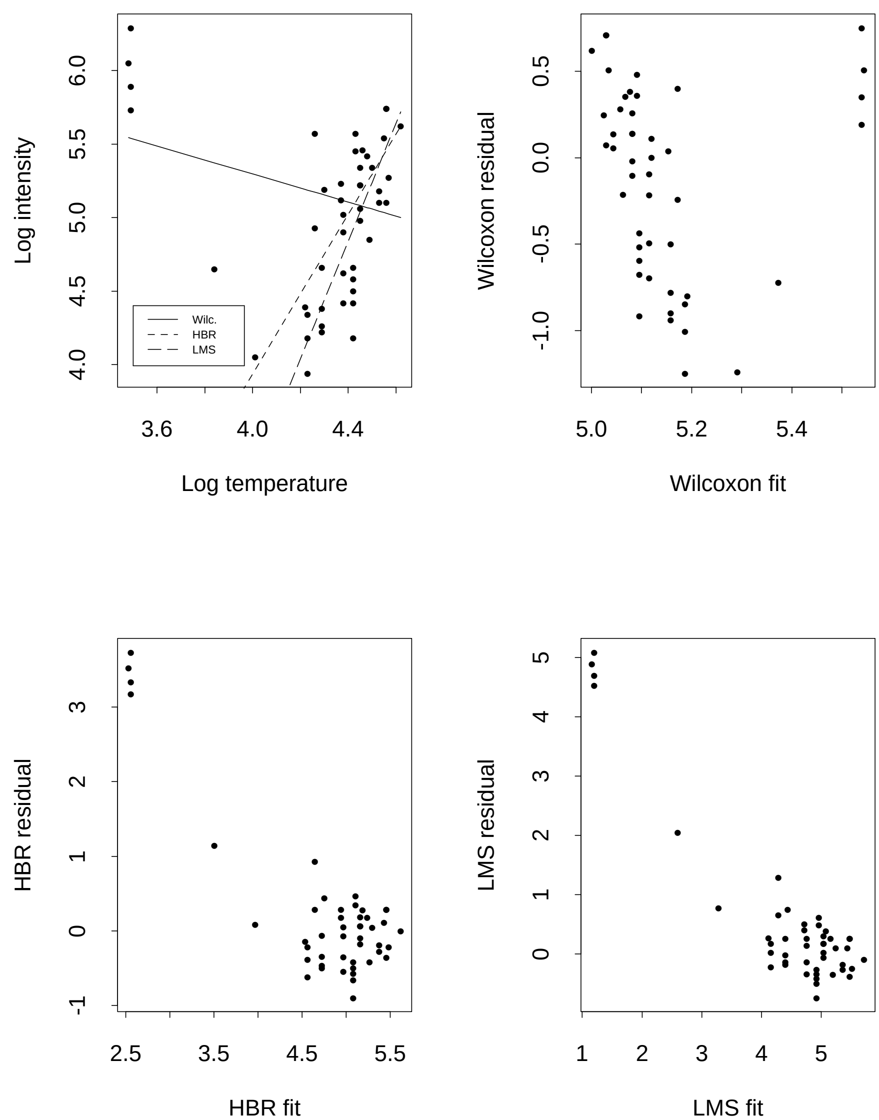
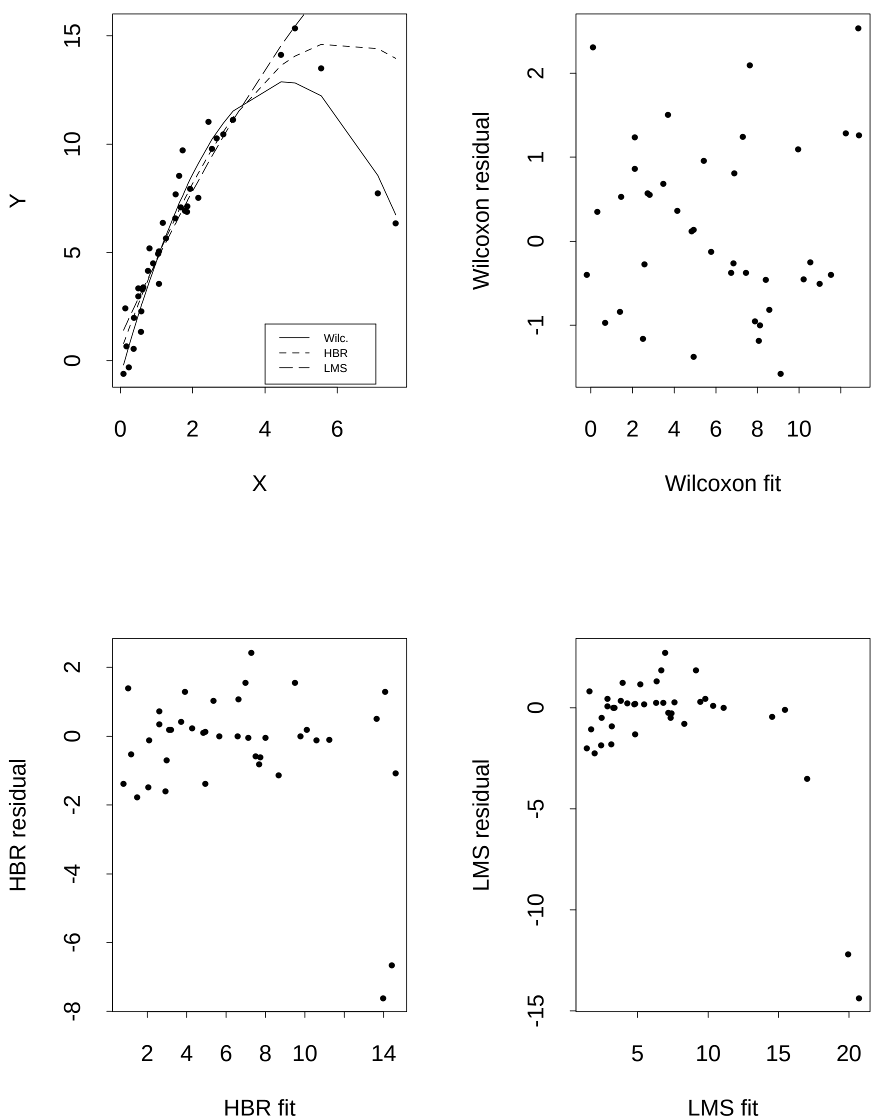


Figure 2 For the Quadratic Data, clockwise from the upper left corner: plot of data and fits; the Wilcoxon residual plot; the LMS residual plot; and the HBR residual plot.



McKean, Naranjo and Sheather (1996a) recently proposed diagnostics based on both a highly efficient robust estimator and a high breakdown estimator. These diagnostics can expose outliers and/or clusters of outliers which are highly influential to a fit of the model. They can also expose discrepancies due to curvature in the data. In a subsequent article, McKean, Naranjo and Sheather (1996b) proposed a model exploratory procedure based on these diagnostics. This procedure is a powerful exploratory tool for model criticism.

4 Breakdown

Let $\boldsymbol{Z} = \{z_i\} = \{(\boldsymbol{x}_i, y_i)\},\ i = 1, \dots, n$ denote the original sample of data points and $|\cdot|$ the Euclidean norm. Define the breakdown point (see page 98 of Hampel et al., 1986), of the estimator at sample $\boldsymbol{Z}$ as

$$\epsilon_n^*(\hat{\boldsymbol{\beta}},\ \boldsymbol{Z}) = \max\left\{\frac{m}{n}; \sup_{\boldsymbol{Z}'} |\hat{\boldsymbol{\beta}}(\boldsymbol{Z}') - \hat{\boldsymbol{\beta}}(\boldsymbol{Z})| < \infty\right\}$$

where the supremum is taken over all samples $\boldsymbol{Z}'$ that can result from replacing m observations in $\boldsymbol{Z}$ by arbitrary values.

We now state conditions under which the HBR estimate remains bounded.

Lemma 1 *Suppose there exist finite constants* $M_1 > 0$ *and* $M_2 > 0$ *such that the following conditions hold:*

(B1) $\inf_{|\boldsymbol{\beta}|=1} \sup_{ij}\{b_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\beta}\} = M_1.$

(B2) $\sup_{ij}\{b_{ij}|y_j - y_i|\} = M_2.$

Then

$$|\hat{\boldsymbol{\beta}}_{HBR}| < \frac{1}{M_1}\left[1 + 2\binom{n}{2}\right] M_2.$$

Proof:

$$\begin{aligned} D(\boldsymbol{\beta}) &\geq \sup_{ij}\{b_{ij}|y_j - y_i - (\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\beta}|\} \geq |\boldsymbol{\beta}|M_1 - M_2 \\ &\geq 2\binom{n}{2} M_2 \end{aligned}$$

whenever $|\boldsymbol{\beta}| \geq \frac{1}{M_1}\left[1 + 2\binom{n}{2}\right] M_2$. Since $D(0) = \sum\sum_{i<j} b_{ij}|y_j - y_i| \leq \binom{n}{2} M_2$ and $D(\boldsymbol{\beta})$ is a convex function of $\boldsymbol{\beta}$, it follows that $\hat{\boldsymbol{\beta}}_{HBR} = \text{argmin} D(\boldsymbol{\beta})$ satisfies

$$|\hat{\boldsymbol{\beta}}_{HBR}| < \frac{1}{M_1}\left[1 + 2\binom{n}{2}\right] M_2.$$

The lemma follows.

For our result, we need to further assume that the differences of the rows of $\boldsymbol{Z}$ are in **general position**; see Croux et al. (1994) for a discussion. This implies that $\boldsymbol{Z}$ is in general position, i.e. any subset of $(p+1)$ rows of Z determines a unique solution $\boldsymbol{\beta}$. In particular, this implies that neither all of the $\mathbf{x}_i$s are the same nor are all of the y_is are the same; hence, provided the weights have not broken down, this implies that both constants M_1 and M_2 of Lemma 1 are positive.

Theorem 4.1 *Let $\widehat{\boldsymbol{\mu}}$, $\boldsymbol{S}$, and $\hat{\boldsymbol{\beta}}_0$ be the initial estimates of location, covariance, and fit that are used to calculate the weights (2.5). Let $\epsilon_n^*(\widehat{\boldsymbol{\mu}}, \boldsymbol{Z})$, $\epsilon_n^*(\boldsymbol{S}, \boldsymbol{Z})$, and $\epsilon_n^*(\hat{\boldsymbol{\beta}}_0, \boldsymbol{Z})$ be the respective breakdown points at sample $\boldsymbol{Z}$. If the difference of the rows of $\boldsymbol{Z}$ are in genearl position, then the breakdown point of the HBR estimator is*

$$\epsilon_n^*(\hat{\boldsymbol{\beta}}_{HBR}, \boldsymbol{Z}) = min\{\epsilon_n^*(\hat{\boldsymbol{\beta}}_0, \boldsymbol{Z}),\ \epsilon_n^*(\widehat{\boldsymbol{\mu}}, \boldsymbol{Z}),\ \epsilon_n^*(\boldsymbol{S}, \boldsymbol{Z}),\ 1/2\}.$$

Proof: Corrupt m points in the data set $\boldsymbol{Z}$ and let $\boldsymbol{Z}'$ be the sample consisting of these corrupt points and the remaining $n-m$ points. Assume that the differences of the rows of $\boldsymbol{Z}'$ are in general position. Assume without loss of generality that $\widehat{\boldsymbol{\mu}}(\boldsymbol{Z}')$, $\boldsymbol{S}(\boldsymbol{Z}')$ and $\hat{\boldsymbol{\beta}}_0(\boldsymbol{Z}')$ have not broken down. Then the constants M_1 and M_2 of Lemma 1 are positive and finite . Hence, by Lemma 1, $\|\hat{\boldsymbol{\beta}}_{HBR}(\boldsymbol{Z}')\| < \infty$ and the theorem follows.

Based on this last result, the HBR-estimate has **50% breakdown** provided the initial estimates $\widehat{\boldsymbol{\mu}}$, $\boldsymbol{S}$ and $\hat{\boldsymbol{\beta}}_0$ all have 50% breakdown. Examples of 50% breakdown estimates for $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{S}$ are the MVE estimates of location and covariance discussed in Section 6; see Rousseeuw and Leroy (1987). For initial estimates of the regression coefficients, we again assume that the differences in the rows of $\boldsymbol{Z}$ are in general position, and that the LTS-estimates and the LMS-estimates have 50% breakdown; see Rousseeuw and Leroy (1987). Also, Hössjer (1994) proposed a rank-based 50% breakdown estimate. The HBR-estimates used for the examples discussed in Section 6 employ the MVE estimates of location and covariance and the LMS-estimate of the regression coefficients.

5 Asymptotic Normality

By the equivariance properties discuss in Section 2 for the estimate $\hat{\boldsymbol{\beta}}_{HBR}$, we can assume without loss of generality in this section that the true $\boldsymbol{\beta} = \mathbf{0}$. Let $\gamma_{ij} = B'_{ij}(0)/E(b_{ij})$, where $B_{ij}(t) = E[b_{ij}I(0 < y_i - y_j < t)]$ and $I(\cdot)$ is the indicator function. Consider the symmetric $n \times n$ matrix $\boldsymbol{A}_n = [a_{ij}]$ with off-diagonal element $a_{ij} = -\gamma_{ij}b_{ij},\ i < j$ and

ith diagonal element $a_{ii} = \sum_{k \neq i} \gamma_{ik} b_{ik}$. Let $\boldsymbol{C}_n = \boldsymbol{X}'\boldsymbol{A}_n\boldsymbol{X}$. Let $\boldsymbol{U}_i = (1/n)\sum_{j=1}^{n}(\boldsymbol{x}_j - \boldsymbol{x}_i)\ E(b_{ij}\text{sgn}(y_j - y_i)|y_i)$. Assume that

(N1) there exists a $p \times p$ matrix $\boldsymbol{C}$ such that $(1/n^2)\boldsymbol{C}_n \xrightarrow{p} \boldsymbol{C}$.

(N2) there exists a $p \times p$ matrix $\boldsymbol{\Sigma}$ such that $(1/n)\sum_{i=1}^{n} \text{Var}(\boldsymbol{U}_i) \to \boldsymbol{\Sigma}$.

The following theorem gives the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{HBR}$. Its proof is in the appendix.

Theorem 5.1 *Under assumptions (N1) and (N2) plus assumptions (A1)-(A6) in the appendix, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{HBR} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0},\ (1/4)\boldsymbol{C}^{-1}\boldsymbol{\Sigma}\boldsymbol{C}^{-1}).$$

It can be shown that $\boldsymbol{C}_n = \sum_{i<j} \gamma_{ij} b_{ij} (\boldsymbol{x}_j - \boldsymbol{x}_i)(\boldsymbol{x}_j - \boldsymbol{x}_i)'$. The following lemma shows another representation of the limiting matrix $\boldsymbol{C}$, and will be useful in deriving the influence function. Let $g_{ij}(\hat{\boldsymbol{\beta}}_0) \equiv b(\boldsymbol{x}_i, \boldsymbol{x}_j, y_i, y_j, \hat{\boldsymbol{\beta}}_0)$ denote the weights as a function of the initial estimator. Let $g_{ij}(\boldsymbol{0}) \equiv b(\boldsymbol{x}_i, \boldsymbol{x}_j, y_i, y_j)$ denote the weight function evaluated at the true value $\boldsymbol{\beta} = 0$.

Lemma 2

$$E\left[\frac{1}{2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{-\infty}^{\infty} b(\mathbf{x}_i, \mathbf{x}_j, y_i, y_j) f^2(y_j)\, dy_j (\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)'\right] \longrightarrow \boldsymbol{C}. \qquad (5.1)$$

Proof: By (4) of the Appendix,

$$E\left[\frac{1}{n^2}\boldsymbol{C}_n\right] = \frac{1}{2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} B'_{ij}(0)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)'$$

where $B_{ij}(t)$ is defined by (3) of the Appendix, and $B'_{ij}(t)$ is given in Lemma 3. Since $B'_{ij}(0)$ is uniformly bounded over all i and j, and the matrix $(1/n^2)\sum_i \sum_j (\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)'$ converges to a positive definite matrix, then the left hand side of (5.1) also converges. By Lemmas 3 and 5, we have

$$B'_{ij}(0) = \int b(\mathbf{x}_i, \mathbf{x}_j, y_j, y_j) f^2(y_j)\, dy_j + o(1) \qquad (5.2)$$

where the remainder term is uniformly small over all i and j. The result follows.

Note that Theorem 5.1 requires Assumption (A.5) of the Appendix; i.e, the initial estimate of regression coefficients are $\sqrt{n}$ consistent. Technically, then, when we want both asymptotic normality and high breakdown of the HBR estimate, we should use an estimate such as the LTS estimate which is $\sqrt{n}$ consistent. In practice, though, the LMS estimate is readily available and less expensive (time-wise) to compute than the LMS estimate. So for the examples of Section 3 and the Monte Carlo results of Section 8.3, we used LMS as the initial estimates for the weights. For the stability study we investigated the behavior of both LTS and LMS as the initial estimates for the weights, (the difference was slight).

5.1 Standard Errors and Studentized Residuals

5.1.1 Standard Errors

Using the asymptotic distribution of the HBR estimate as a guideline and upon substituting the estimated weights for the true weights we can estimate the asymptotic standard errors for these estimates. The asymptotic variance-covariance of $\widehat{\boldsymbol{\beta}}$ is a function of the two matrices $\mathbf{\Sigma}$ and $\mathbf{C}$. The matrix $\mathbf{\Sigma}$ is the variance-covariance matrix of the random vector $\mathbf{U}_i$. We can approximate $\mathbf{U}_i$ by the expression,

$$\widehat{\mathbf{U}}_i = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \mathbf{x}_i)\widehat{b}_{ij}(1 - 2F_n(\widehat{e}_i)) \ , \tag{5.3}$$

where $\widehat{b}_{ij}$ are the estimated weights, $\widehat{e}_i$ are the HBR residuals and F_n is the empirical distribution function of the residuals. The sample variance-covariance matrix of $\mathbf{U}_1, \ldots, \mathbf{U}_n$ is an estimate of $\mathbf{\Sigma}$. Call this estimate $\widehat{\mathbf{\Sigma}}$.

Note that the matrix $\mathbf{C}$ is the limit of $n^{-2}\boldsymbol{X}'\boldsymbol{A}_n\boldsymbol{X}$. Consider the results found in Lemma 2. Upon substituting the estimated weights for the weights, expression (5.2) simplifies to

$$B'_{ij}(0) \doteq \widehat{b}_{ij}\int f^2(t)\,dt = \widehat{b}_{ij}\frac{1}{\sqrt{12}\tau} \ , \tag{5.4}$$

where τ is defined in expression (2.4). A consistent estimate of τ based on residuals is presented in Koul, Sievers and McKean (1987). Call this estimate $\widehat{\tau}$. Substituting $\widehat{\tau}$ for τ in expression (5.4), we have as an estimate of the (i, j)th entry of the matrix $\boldsymbol{A}_n$,

$$\widehat{a}_{ij} = -\widehat{b}_{ij}/(\sqrt{12}\widehat{\tau}) \ . \tag{5.5}$$

Using these estimates of $\boldsymbol{A}_n$ and $\mathbf{\Sigma}$, estimates of the standard errors of $\widehat{\beta}_j$, $j = 1, \ldots, p$ are obtained.

5.1.2 Estimate of the Intercept

In practice an estimate of the intercept parameter α is needed. We will estimate α by the median of the residuals, i.e.,

$$\widehat{\alpha} = \text{med}\{y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{HBR}\} . \tag{5.6}$$

Without loss of generality, assume that the true median of e is 0. Because $\sqrt{n}\hat{\boldsymbol{\beta}}_{HBR}$ is bounded in probability and $\boldsymbol{X}$ is centered, it follows, using an argument very similar to the corresponding result for the R estimates (see McKean et al., 1990), that the joint asymptotic distribution of $\widehat{\alpha}$ and $\hat{\boldsymbol{\beta}}_{HBR}$ is given by

$$\sqrt{n}\left\{\begin{bmatrix} \widehat{\alpha} \\ \hat{\boldsymbol{\beta}}_{HBR} \end{bmatrix} \begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix}\right\} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \tau_S & \mathbf{0}' \\ \mathbf{0} & (1/4)\boldsymbol{C}^{-1}\boldsymbol{\Sigma}\boldsymbol{C}^{-1} \end{bmatrix}\right) , \tag{5.7}$$

where τ_S is defined by

$$\tau_S = 1/(2f(0)) ; \tag{5.8}$$

see Sheather and McKean (1987) for a discussion of estimation of τ_S.

5.1.3 Studentized Residuals

An important use of robust residual is in detection of outliers. This is easiest done when the residuals are correctly studentized by an estimate of their standard deviation. Let $\hat{\boldsymbol{\beta}}_{HBR}$ be the estimate of $\boldsymbol{\beta}$ and let $\widehat{\alpha}$, (5.6), be the estimate of α. Denote the residual for the ith case by

$$\widehat{e}_i^* = y_i - \widehat{\alpha} - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{HBR} , \tag{5.9}$$

and the vector of residuals by $\widehat{\mathbf{e}}^*$. Using Theorem 4.1 and (5.7), a first-order approximation of the standard deviation of the residuals, $\widehat{e}_i^*$, can be obtained following the development proposed by McKean, Sheather, and Hettmansperger (1990, 1993).

As briefly outlined in the Appendix, this development for the HBR residuals results in the first order approximation given by

$$\begin{aligned} \text{Var}(\widehat{\mathbf{e}}^*) &\doteq \sigma^2\mathbf{I} + \tau_S^2\mathbf{H_1} + \frac{1}{4}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^*\boldsymbol{X})^{-1}\boldsymbol{\Sigma}(\boldsymbol{X}'\boldsymbol{A}^*\boldsymbol{X})^{-1}\boldsymbol{X}' - 2\tau_S\kappa_1\mathbf{H_1} \\ &\quad -\sqrt{12}\tau\kappa_2\left\{\boldsymbol{A}^*\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^*\boldsymbol{X})^{-1}\boldsymbol{X} + \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^*\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{A}^*\right\} , \end{aligned} \tag{5.10}$$

where σ^2 is the variance of e_i, $\kappa_1 = E[|e_i|]$, $\kappa_2 = E[e_i(2F(e_i) - 1)]$, $\mathbf{H}_1 = n^{-1}\mathbf{1}\mathbf{1}'$, and $\boldsymbol{A}^*$ is defined above expression (A.1.6) of the Appendix.

We recommend estimating σ^2 by $MAD = 1.483\text{med}\{|e|_i^*\}$; κ_1 by

$$\widehat{\kappa}_1 = \frac{1}{n}\sum_{i=1}^{n} |\widehat{e}_i^*| \; ; \tag{5.11}$$

and κ_2 by

$$\widehat{\kappa}_2 = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{R(\widehat{e}_i^*)}{n+1} - \frac{1}{2}\right) \widehat{e}_i^* \; , \tag{5.12}$$

which is a consistent estimate of κ_2; see McKean et al. (1990). Replacing a_{ij}^* by $\widehat{a}_{ij}$, (5.5), yields an estimate of the matrix $\boldsymbol{A}^*$. Estimation of $\boldsymbol{\Sigma}$ was discussed in Section 5.1.1. Let $\widehat{\mathbf{V}}$ denote the estimate of $\text{Var}(\widehat{\mathbf{e}}^*)$.

Let $\widehat{\sigma}^2_{\widehat{e_i}}$ denote the the ith diagonal entry of $\widehat{\mathbf{V}}$. Define the Studentized residuals by

$$\widehat{e}_i^* = \frac{\widehat{e}_i}{\widehat{\sigma}_{\widehat{e_i}}} \; . \tag{5.13}$$

As in LS, these standard errors correct for both the underlying variance of the errors and location. For flagging outliers, appropriate benchmarks for these residuals are ± 2; see McKean et al. (1990, 1993) for discussion.

6 Influence Function

In order to derive the influence function, we start with the gradient equation $\mathbf{S}(\boldsymbol{\beta}) = \mathbf{0}$, written out as

$$\mathbf{0} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}\text{sgn}(e_j - e_i)(\mathbf{x}_j - \mathbf{x}_i).$$

Note by Lemma 5, that $b_{ij} = g_{ij}(\mathbf{0}) + O_p(1/\sqrt{n})$ so that the defining equation may be written as

$$\mathbf{0} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(\mathbf{0})\text{sgn}(e_j - e_i)(\mathbf{x}_j - \mathbf{x}_i). \tag{6.1}$$

ignoring a remainder term of magnitude $O_p(1/\sqrt{n})$.

Influence functions are derived at the model where both $\mathbf{x}$ and y are stochastic; hence, consider the linear model

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon \; , \tag{6.2}$$

where ϵ has density f, $\mathbf{x}$ is a $p \times 1$ random vector with density function m, and ϵ and $\mathbf{x}$ are independent. Let F and M denote the corresponding distribution functions of ϵ and $\mathbf{x}$. Let H and h denote the joint distribution function and density of y and $\mathbf{x}$. It then follows that

$$h(\mathbf{x}, y) = f(y - \mathbf{x}'\boldsymbol{\beta})m(\mathbf{x}) \; . \tag{6.3}$$

If we rewrite equation (6.1) using the Stieltjes integral notation of the empirical distribution of $(\mathbf{x}_i, y_i)$, for $i = 1, \ldots, n$, we see that the functional $\boldsymbol{\beta}(H)$ solves the equation

$$\mathbf{0} = \int\int b(\mathbf{x}_1, \mathbf{x}_2, y_1, y_2)\mathrm{sgn}\{y_2 - y_1 - (\mathbf{x}_2 - \mathbf{x}_1)'\boldsymbol{\beta}(H)\}(\mathbf{x}_2 - \mathbf{x}_1)dH(\mathbf{x}_1, y)dH(\mathbf{x}_2, y_2) \, .$$

Let $I(a < b) = 1$ or 0, depending on whether $a < b$ or $a > b$. Then using the fact that the sign function is odd and the symmetry of the weight function in its $\mathbf{x}$ and y arguments we can write the defining equation of the functional $\boldsymbol{\beta}(H)$ as

$$\mathbf{0} = \int\int \mathbf{x}_1 b(\mathbf{x}_1, \mathbf{x}_2, y_1, y_2)\left[I(y_2 - y_1 < (\mathbf{x}_2 - \mathbf{x}_1)'\boldsymbol{\beta}(H)) - \frac{1}{2}\right] dH(\mathbf{x}_1, y)dH(\mathbf{x}_2, y_2) \, .$$

Theorem 6.1 *The influence function for the estimate* $\widehat{\boldsymbol{\beta}}$ *is given by*

$$\boldsymbol{\Omega}(\mathbf{x}_0, y_0, \widehat{\boldsymbol{\beta}}) = \mathbf{C}^{-1}\frac{1}{2}\int\int(\mathbf{x}_0 - \mathbf{x}_1)b(\mathbf{x}_1, \mathbf{x}_0, y_1, y_0)sgn\{y_0 - y_1\}\, dF(y_1)dM(\mathbf{x}_1) \, , \quad (6.4)$$

where $\mathbf{C}$ *is the matrix in Lemma 2.*

Proof: Let $\delta_0(\mathbf{x}, y)$ denote the distribution function of the point mass at the point $(\mathbf{x}_0, y_0)$ and consider the contaminated distribution $H_t = (1-t)H + t\delta_0$ for $0 < t < 1$. Let $\boldsymbol{\beta}(H_t)$ denote the functional at H_t. Then $\boldsymbol{\beta}(H_t)$ satisfies

$$\mathbf{0} = \int\int \mathbf{x}_1 b(\mathbf{x}_1, \mathbf{x}_2, y_1, y_2)\left[I(y_2 - y_1 < (\mathbf{x}_2 - \mathbf{x}_1)'\boldsymbol{\beta}(H_t)) - \frac{1}{2}\right] dH_t(\mathbf{x}_1, y_1)dH_t(\mathbf{x}_2, y_2) \, . \quad (6.5)$$

We next implicitly differentiate (6.5) with respect to t to obtain the derivative of the functional. The value of this derivative at $t = 0$ is the influence function. Without loss of generality, we can assume that the true parameter $\boldsymbol{\beta} = \mathbf{0}$. Under this assumption $\mathbf{x}$ and y are independent. Substituting the value of H_t into (6.5) and expanding we obtain the four terms:

$$\begin{aligned}
\mathbf{0} \;=\; & (1-t)^2\int\int\int \mathbf{x}_1\left[\int_{-\infty}^{y_1+(\mathbf{x}_2-\mathbf{x}_1)'\boldsymbol{\beta}(H_t)} b(\mathbf{x}_1, \mathbf{x}_2, y_1, y_2)dF(y_2) - \frac{1}{2}\right] dM(\mathbf{x}_2)dM(\mathbf{x}_1)dF(y_1) \\
& +(1-t)t\int\int\int\int \mathbf{x}_1 b(\mathbf{x}_1, \mathbf{x}_2, y_1, y_2)\left[I(y_2 - y_1 < (\mathbf{x}_2 - \mathbf{x}_1)'\boldsymbol{\beta}(H_t)) - \frac{1}{2}\right] dM(\mathbf{x}_2)dF(y_2)d\delta_0(\mathbf{x}_1, y_1) \\
& +(1-t)t\int\int\int\int \mathbf{x}_1 b(\mathbf{x}_1, \mathbf{x}_2, y_1, y_2)\left[I(y_2 - y_1 < (\mathbf{x}_2 - \mathbf{x}_1)'\boldsymbol{\beta}(H_t)) - \frac{1}{2}\right] d\delta_0(\mathbf{x}_2, y_2)dM(\mathbf{x}_1)dF(y_1) \\
& +t^2\int\int\int\int \mathbf{x}_1 b(\mathbf{x}_1, \mathbf{x}_2, y_1, y_2)\left[I(y_2 - y_1 < (\mathbf{x}_2 - \mathbf{x}_1)'\boldsymbol{\beta}(H_t)) - \frac{1}{2}\right] d\delta_0(\mathbf{x}_2, y_2)d\delta_0(\mathbf{x}_1, y_1) \, .
\end{aligned}$$

Let $\dot{\boldsymbol{\beta}}$ denote the derivative of the functional evaluted at 0. Proceeding to implicitly differentiate this equation, evaluating the derivative at 0, and using the symmetry in the

$\mathbf{x}$ arguments and y arguments of the function b, we can simplify this expression to

$$\begin{aligned} \mathbf{0} &= -\left\{\frac{1}{2}\int\int\int(\mathbf{x}_2-\mathbf{x}_1)b(\mathbf{x}_1,\mathbf{x}_2,y_1,y_1)(\mathbf{x}_2-\mathbf{x}_1)'f^2(y_1)\,dy_1 dM(\mathbf{x}_1)dM(\mathbf{x}_2)\right\}\dot{\boldsymbol{\beta}} \\ &\quad +\int\int(\mathbf{x}_0-\mathbf{x}_1)b(\mathbf{x}_1,\mathbf{x}_0,y_1,y_0)\left[I(y_1<y_0)-\frac{1}{2}\right]dF(y_1)dM(\mathbf{x}_1)\ . \end{aligned}$$

Since $\boldsymbol{x}$ is stochastic, it is immediate that the matrix $\mathbf{C}$ in Lemma 2 is given by the expression in the braces. Using the relationship between the indicator function and the sign function we can rewrite this last expression as

$$\mathbf{0} = -\mathbf{C}\dot{\boldsymbol{\beta}} + \frac{1}{2}\int\int(\mathbf{x}_0-\mathbf{x}_1)b(\mathbf{x}_1,\mathbf{x}_0,y_1,y_0)\mathrm{sgn}\{y_0-y_1\}\,dF(y_1)dM(\mathbf{x}_1)\ .$$

Solving for $\dot{\boldsymbol{\beta}}$ leads to the desired result.

In order to show that the influence function correctly identifies the asymptotic distribution of the estimator, define W_i as

$$W_i = \int\int(\mathbf{x}_i-\mathbf{x}_1)b(\mathbf{x}_1,\mathbf{x}_i,y_1,y_i)\mathrm{sgn}(y_i-y_1)\,dF(y_1)dM(\mathbf{x}_1)\ . \tag{6.6}$$

Next write W_i in terms of a Stieltjes integral over the empirical distribution of $(\mathbf{x}_j, y_j)$ as

$$W_i^* = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_i-\mathbf{x}_j)b(\mathbf{x}_j,\mathbf{x}_i,y_j,y_i)\mathrm{sgn}(y_i-y_j)\ . \tag{6.7}$$

If we can show that $(1/\sqrt{n})\sum_{j=1}^{n} W_i^* \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Sigma})$, then we are done. From the proof of Theorem A.1.3 in the appendix, it will suffice to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i - W_i^*) \xrightarrow{p} 0\ , \tag{6.8}$$

where $U_i = (1/n)\sum_{j=1}^{n}(\mathbf{x}_i-\mathbf{x}_j)E[b_{ij}\mathrm{sgn}(y_i-y_j)|y_i]$. Writing the left hand side of 6.8 as

$$(1/n^{3/2})\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i-\boldsymbol{x}_j)\left\{E\left[b_{ij}\mathrm{sgn}(y_i-y_j)|y_i\right]-g_{ij}(\mathbf{0})\mathrm{sgn}(y_i-y_j)\right\},$$

where $g_{ij}(\mathbf{0}) \equiv b(\mathbf{x}_j,\mathbf{x}_i,y_j,y_i)$, the proof is analogous to the proof of Theorem A.1.3.

6.1 Discussion

The influence function, $\boldsymbol{\Omega}(\mathbf{x}_0, y_0, \widehat{\boldsymbol{\beta}})$, for the HBR estimate is a continuous function of $\mathbf{x}_0$ and y_0. With a proper choice of a weight function it is bounded in both the $\mathbf{x}$ and Y spaces. This is true for the weights given by (2.5); furthermore, for these weights $\boldsymbol{\Omega}(\mathbf{x}_0, y_0, \widehat{\boldsymbol{\beta}})$ goes to zero as $\mathbf{x}_0$ and y_0 get large in any direction.

The influence function $\mathbf{\Omega}(\mathbf{x}_0, y_0, \widehat{\boldsymbol{\beta}})$ is a generalization of the influence functions for the Wilcoxon and GR estimates. If the weights depend only on $\mathbf{x}$ then $\mathbf{\Omega}(\mathbf{x}_0, y_0, \widehat{\boldsymbol{\beta}})$ reduces to the influence function of the GR estimate discussed in Witt, McKean and Naranjo (1995). Further, if the weights are set at 1 then $\mathbf{\Omega}(\mathbf{x}_0, y_0, \widehat{\boldsymbol{\beta}})$ simplifies to the influence function of Wilcoxon estimate given by

$$\mathbf{\Omega}(\mathbf{x}_0, y_0, \widehat{\boldsymbol{\beta}}_{Wil}) = \mathbf{\Sigma}_W^{-1}\mathbf{x}_0 \left[F(y_0) - \frac{1}{2}\right] \; ; \qquad (6.9)$$

see Witt et al. (1995). Figure 3 shows the influence function of the HBR estimate for the special case where $(\mathbf{x}, Y)$ has a bivariate normal distribution with mean $\mathbf{0}$ and the identity matrix as the variance-covariance matrix. For this plot we used the weights given by (2.5) where $m_i = \psi(b/x_i^2)$ with the constants $b = c = 4$. For comparison purposes, the influence functions of the Wilcoxon and GR estimates are also shown in the figure. The Wilcoxon influence function is bounded in the Y space but is unbounded in the $\mathbf{x}$ space while the GR estimate is bounded in both spaces. Note because the weights of the GR estimate do not depend on Y, it does not taper to 0 as $y_0 \to \infty$ as the influence function of the HBR estimate does. For all three plots, we used the method of Monte Carlo, (10000 simulations for each of 1600 grid points), to perform the numerical integration. The plot of the Wilcoxon influence function is an easily verifiable check on the Monte Carlo because of its closed form, (6.9).

Many high breakdown estimates have unbounded influence functions. Such estimates can have instability problems, as discussed in Sheather et al. (1996) in the case of the LMS estimate which has unbounded influence in the $\mathbf{x}$ space at the quantiles of Y. The generalized S estimators discussed in Croux et al. (1994) also have unbounded influence functions in the $\mathbf{x}$ space at particular values of Y. In contrast, the influence function of the HBR estimate is bounded everywhere. This helps explain its more stable behavior than the LMS in the stability study discussed in Section 7.

7 Stability Study

To obtain its full 50% breakdown, the HBR estimates require initial estimates with 50% breakdown. We have discussed the LMS or the LTS as possible candidates. Hettmansperger and Sheather (1992, 1993) showed that slight changes to centrally located data can cause the LMS estimate to change by a large amount. This instability of the LMS was verified in a simulation study by Sheather, McKean and Hettmansperger (1997). Hence,

Figure 3a: Influence functions for the HB estimate.

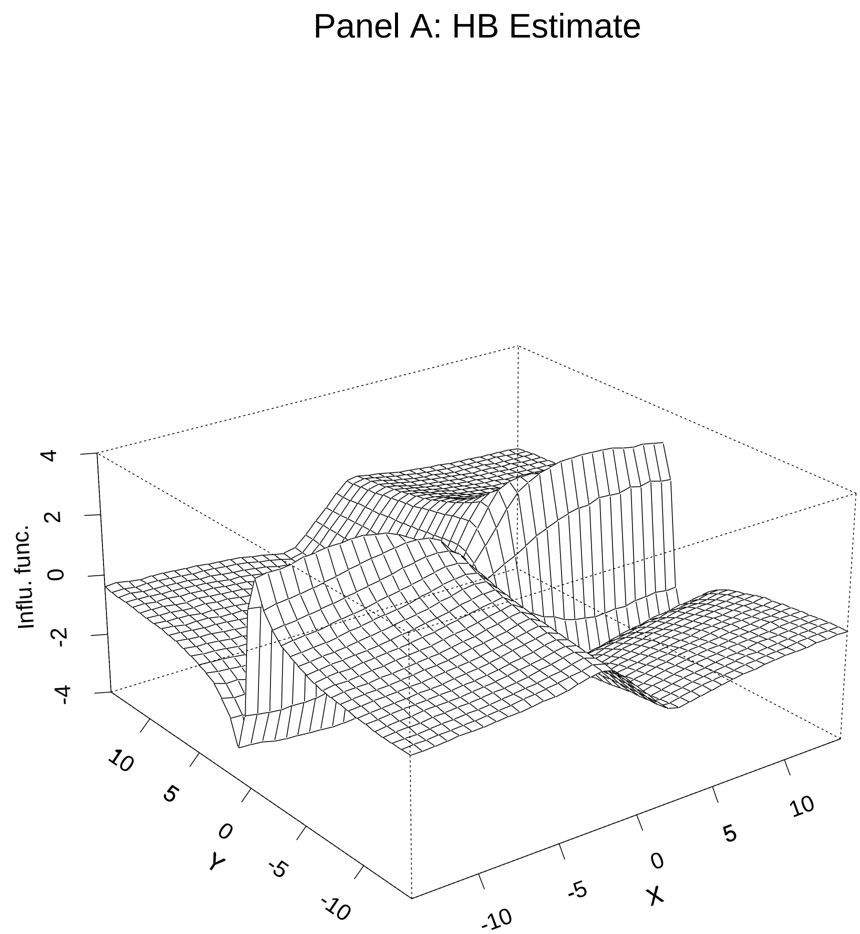


Figure 3b: Influence functions for the Wilcoxon estimate.

Panel B: Wilcoxon Estimate

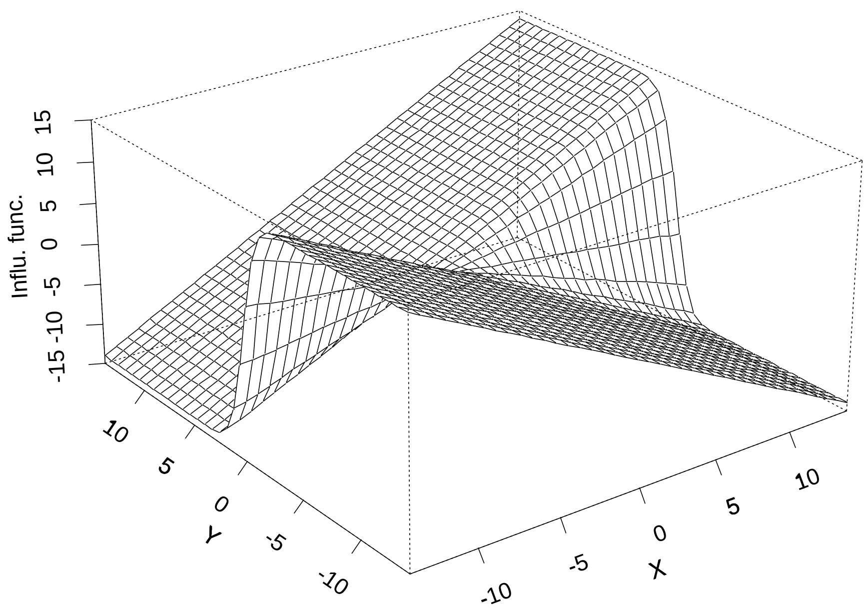
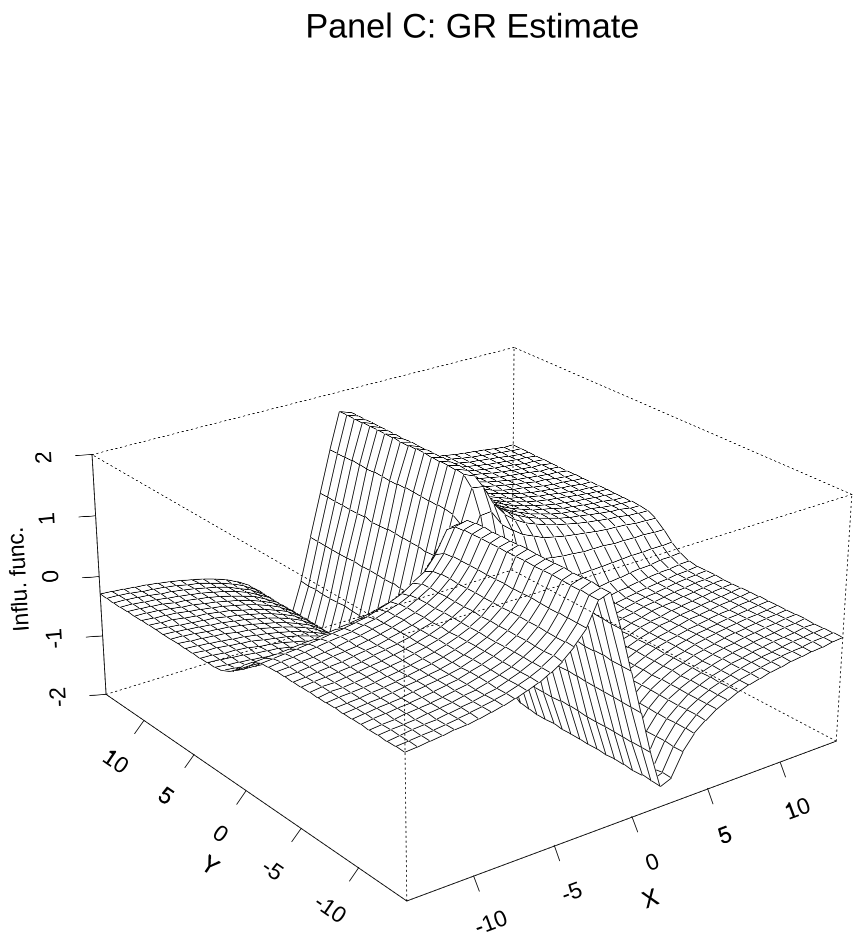


Figure 3c: Influence functions for the GR estimate.



the instability of the HBR estimates using LMS starts needs to be investigated. Since we have also discussed the LTS estimate as a possible starting value we also included it in our study.

We chose the simple linear model given by

$$y_i = \beta_0 + x_i\beta + e_i \ ,$$

where both the errors $e_1, \ldots, e_n$ and the predictors $x_1, \ldots, x_n$ are iid $N(0,1)$ and the predictors and errors are independent of one another. Since instability is a central data issue, the light tailed normals seem appropriate for the predictors. For the simulation we took $\beta_0 = \beta_1 = 0$ and set the sample size at $n = 30$.

As procedures, we selected LS, Wilcoxon, GR, HBR(with LMS starts), HBR(with LTS starts), LMS, and the LTS estimators. We used the exact algorithm developed in Stromberg (1993) to compute the LMS estimates. We used the feasible solution algorithm of Hawkins (1994) to compute the LTS estimates; although not exact, we did use the settings suggested by Hawkins so that the global optimum is attained with high probability. We measured instability by the effect that changes in x_i had on the estimate of β_1 for each of the procedures. We selected four changes: $x_i \pm .1$ and $x_i \pm .2$. Since the standard deviation of x_i is 1 these changes move x_i by one- and two-tenths of a standard deviation, respectively. For each estimate, we recorded the maximum of these 4 changes over all 30 data points, x_i. The LS standard deviation of $\widehat{\boldsymbol{\beta}}_1$ is a useful yard stick with which to measure these maximum values against. Since the sample variance of x_i converges to 1, a.s., this standard deviation is about $1/\sqrt{30} = .18$.

Our study consisted of 200 simulations of the above model. Both errors and predictors were simulated. Comparison boxplots of the 200 maximum changes for each estimator appear in Figure 4, while the five basic summary statistics, (extremes, quartiles, and medians), of these maximums for each estimator are displayed in Table 2. Note that the maximum for the LMS estimate is close to 10 standard deviations of the LS estimate of β_1. The median of these maximum values for the LMS estimate is .166 which is close to one standard deviation of the LS estimate of β_1. So in half the data sets, the LMS estimate of slope changed by at least 1 standard deviation of the LS estimate of β_1. The LTS estimate was a little more stable than the LMS. The maximum change for the LTS was 1.43 compared to 1.71 for the LMS. Based on the third quartile of the maximums of the LTS estimates, in 25% of the cases did the LTS estimate of slope changed by at least 1 standard deviation of the LS estimate of β_1.

Figure 4 Stability Study: Comparison boxplots of the maximum values for the 200 simulations: Upper Panel, all 7 Estimators; Lower Panel, all but LMS and LTS.

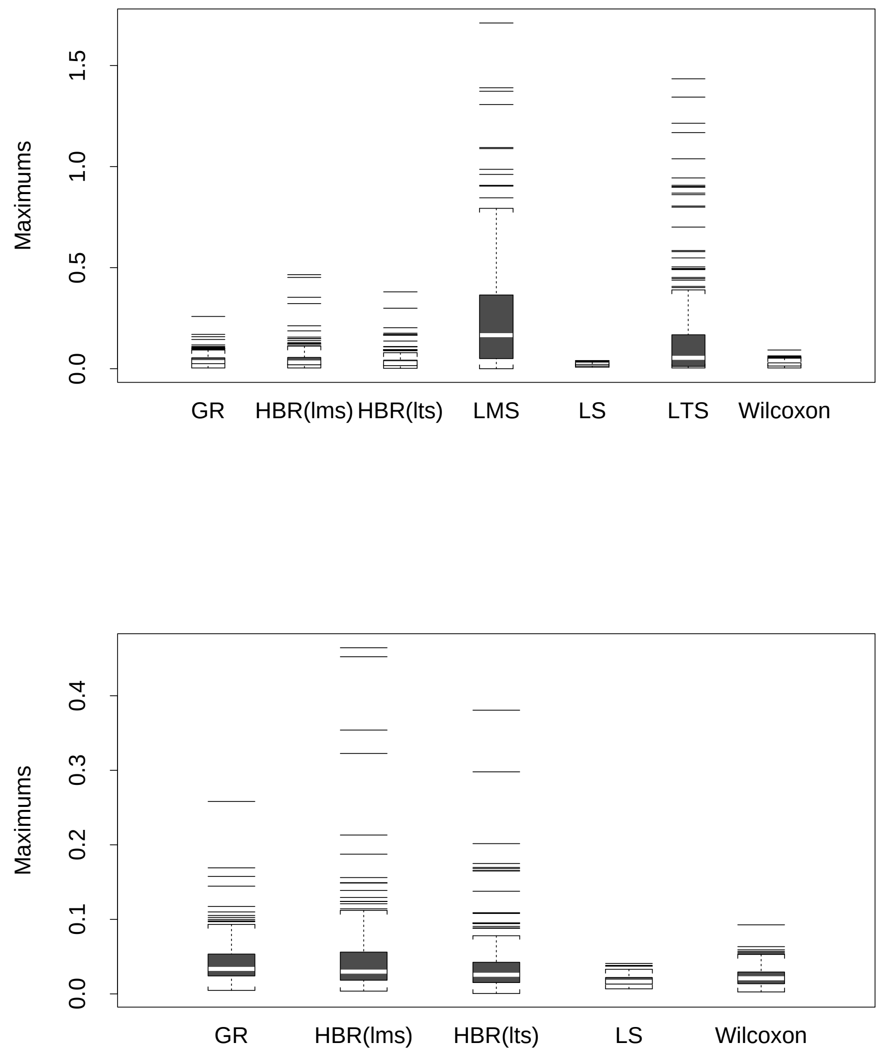


Table 2: Descriptive Statistics for the 200 Maximums for Each Procedure

Estimator	Min.	First Quartile	Median	Third Quartile	Max.
LS	.007	.013	.016	.022	.041
Wilcoxon	.003	.013	.021	.029	.093
GR	.004	.024	.033	.053	.258
HBR(LMS)	.003	.018	.030	.056	.465
HBR(LTS)	.001	.015	.026	.042	.380
LMS	.000	.050	.166	.366	1.71
LTS	.004	.012	.053	.170	1.43

It is clear that the LMS and LTS estimates are more unstable than the other estimators. To clarify the comparisons of the other procedures, comparison boxplots of the maximums for them appear in Figure 4. It seems that the HBR(LTS) estimate was slightly more stable than the HBR(LMS) but actually it is hard on the basis of these results to separate the stability behavior of the HBR(LMS), HBR(LTS), and GR estimates. For example, the HBR(LMS) estimate appears slightly more unstable than the GR estimate; although the median of its maximum values is slightly less than that of the GR. Note that the HBR estimates are more stable than their corresponding starting values. The Wilcoxon and LS estimates are more stable than the bounded influence estimates.

It appears on the basis of this study that the stability of the HBR estimator is not unduly influenced by the instability of its corresponding initial estimates, the LMS or the LTS. Its stability is much closer to that of the GR estimate.

8 Monte Carlo Study

One could presumably investigate the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_{HBR}$ to determine asymptotic relative efficiencies, (AREs). Note, however, that this expression is rather complicated. The matrix $\mathbf{C}$ is the derivative of an expectation involving the errors and the weight function while the matrix $\boldsymbol{\Sigma}$ involves a conditional expectation of the weight function and the errors. Since this seems intractable, in this section we instead discuss the results of a small Monte Carlo study. The main purposes of this study is to see if the HBR estimates recover the loss in efficiency which GR estimates lose to the Wilcoxon estimates. This is discussed in Section 8.1 (over different severity of contamination in factor space)

Table 8.1: Empirical AREs of the GR, HBR, and LMS Estimates Relative to the Wilcoxon Estimates over the Laplace Situations.

Estimator	Uniform	CN(.15, 16)	CN(.15, 64)	CN(.25, 100)
GR	.948	.525	.222	.181
HBR	1.06	.765	.523	.318
LMS	.413	.254	.131	.110

and Section 8.2. Also, we were interested in the efficiency comparisons between the HBR estimate and the LMS estimate, the initial regression estimate used in the weight function. We selected the LMS estimate because it is one of the most widely used high breakdown estimates and, further, in comparison to the LTS estimate it is relatively faster to compute (one-half the time). A second purpose is to compare the empirical ARE's of several competing estimators, which is discussed in Section 8.3.

For all but the two situations discussed in Section 8.2, the model simulated is:

$$Y_i = \beta_0 + \beta_1 x_i + e_i \ , \ \ i = 1, \ldots, 30 \ . \tag{8.1}$$

For all situations but those in Section 8.2, we set $\beta_0 = \beta_1 = 0$. As discussed below we selected various distributions for the random errors and the xs. We compare the estimators in terms of their empirical mean square errors (MSE) from the true slope parameter. We will call ratios of these MSEs, empirical asymptotic relative efficiencies (AREs). For each situation, 1000 simulations were run.

8.1 Comparisons over Severity of Factor Space Contamination

For this part of the study we simulated the basic model, (8.1) using the Laplace distribution (density $f(t) = 2^{-1}\exp\{-|t|/2\}$) to generate the random errors. We simulated four distributions for the xs: uniform$(0,1)$, $CN(.15,16)$, $CN(.15,64)$ and $CN(.25,100)$; where, $CN(\epsilon, \sigma^2)$ means a contaminated normal distribution with ϵ proportion of contamination and σ as the ratio of standard deviations between the contaminated and non contaminated parts. This sequence of distributions of x increase in terms of severity of contamination. In Table 8.1, we present the AREs (based on MSE), of the GR and HBR estimates relative to the Wilcoxon estimates for these situations.

Note that, though, the HBR estimates lose efficiency as the severity of contamination in factor space increases, the loss is much less than that of the GR estimates. Hence, the

Table 8.2: Empirical AREs of the GR, HBR, and LMS to Wilcoxon for the Situations of Section 8.2.

Estimator	Model (3.2)	Stars Type Data
GR	.201	5.10
HBR	.684	1.78
LMS	.701	1.44

HBR estimates have recovered some of the efficiency lost by the GR estimates. Note, also, that the LMS estimates are quite inefficient; hence, for these situations the HBR estimates recovered efficiency from inefficient starting values.

8.2 Models of Section 3

In Section 3, two examples were discussed which differentiated between highly efficient and high breakdown estimates. In this section, we present the results of a simulation study on models similar to these examples. Our first model, here, is Model (3.2) of Example 2. We ran 1000 simulations of this model. The first column of Table 8.2 displays the AREs relative to the Wilcoxon estimates for this situation. This is a linear model and the AREs are based on the MSEs from the true parameters. We only tabled the result for the quadratic term. The HBR estimate was three times as efficient as the GR estimate over this situation.

The second situation resembles the Stars Data, Example 1. As our base model we chose Model (8.1) with uniform $(0,1)$ xs and normal, $N(0,1)$, errors. Observations 5-30 followed this model with the true slope parameter $\beta_1 = 0$. For the first four cases, we set $x_1 = 2.9$, $x_2 = 2.8$, $x_3 = 2.7$ and $x_4 = 2.6$. At these four observations, we set $\beta_1 = 2$. We measured the MSE of an estimator in terms of its squared distance from 0 the value of the slope parameter at observations 5-30. The last column of Table 8.2 displays the empirical AREs of the GR, HBR and LMS estimates relative to the Wilcoxon estimate. While all three high breakdown estimates performed better than the Wilcoxon, clearly the GR estimate performed best. Note that the HBR estimate did gain efficiency over the poorer performing initial (LMS) estimates.

8.3 Comparison with Other Estimators

Besides, the GR, HBR and LMS high breakdown estimates, we included the GM estimate of Simpson, Ruppert and Carroll (1992) in our comparison study. This GM estimate is a k-step weighted M estimate which uses Hampel's redescending ψ function. We used LMS starting values, the same weights, (3.1), as employed by the GR and HBR estimates, and used 3 steps. The resulting estimate has 50% breakdown, is $\sqrt{n}$ consistent and should recover efficiency over its initial estimate; see Simpson et al. (1992).

Besides the situations discussed in Sections 8.1 and 8.2, we looked at normal and Cauchy distributed errors. For these, we selected uniform $(0,1)$ and $CN(.25,100)$ errors for factor space. We used the Model (8.1) with the true regression parameters at 0, except for the situations of Section 8.2. In all cases, 1000 simulations were run.

A summary of the results are displayed in Table 8.3. These are empirical AREs relative to LS, based on the ratios of empirical MSEs. Excluding, the last two situations, except for the Cauchy distribution, all the high breakdown estimates lose efficiency when the uniform distribution in factor space is replaced by the contaminated normal distribution. At the normal distribution, the loss is quite severe for all the high breakdown estimates. Over these situations, the best high breakdown estimates are the HBR and GM estimates; however, neither one dominates the other. For the last two situations, the GM estimate performs worse than the LMS estimate or the HBR estimate on the quadratic model, (3.2), while it is clearly dominated by the GR estimate on the Stars Type Data.

9 Conclusion

The HBR estimate is a high breakdown (50%) rank-based estimate which is asymptotically normal at rate $\sqrt{n}$. The asymptotic theory allows an inference based on the HBR estimates to be established for the regression coefficients of the model. Also, Studentized residuals can be formed as a useful diagnostic to detect outliers. The influence function of the HBR estimate is continuous and bounded in $(\mathbf{x}, Y)$ space and tapers to 0 in all directions. In contrast to many recent high breakdown estimates it is bounded everywhere in the plane.

For the weighting scheme used in the examples, its robustness against an outlying high-leverage cluster approaches that of the LMS, though not as severely (see Example 1, Section 3). As such, it shares the problems of the LMS in fitting curvature (see Example 2, Section 3). However, unlike the LMS, the HBR is more stable against inliers (see

Table 8.3: Empirical ARE of the Estimates Relative to the LS estimates. Ec means times 10^c.

Model		estimator				
Y	X	Wilcoxon	GR	HBR	GM	LMS
N	U	.931	.889	.777	.948	.253
N	CN(.25, 100)	.928	.146	.221	.273	.070
Lap.	U	1.34	1.27	1.42	1.32	.554
Lap.	CN(.15, 16)	1.14	.598	.872	.876	.289
Lap.	CN(.15, 64)	1.13	.251	.591	.499	.148
Lap.	CN(.25, 100)	1.18	.213	.375	.450	.130
Cau.	U	467	153	419	484	216
Cau.	CN(.25, 100)	9.1E3	4.6E3	11.E3	9.1E3	5.7E3
Model (3.2)		.996	.200	.681	.539	.698
Stars Type Data		.991	5.10	1.76	1.74	1.43

Section 7). This can be partially explained by the everywhere boundedness of the HBR influence function whereas the LMS influence function is unbounded at the quartiles of Y. In the study, the HBR is more stable against inliers than the LTS estimate also. In terms, of an initial estimate for weights, the HBR estimate was perhaps slightly more stable with the initial LTS estimate than the LMS estimate. In the Monte Carlo study, Section 8.3, in almost all situations the HBR estimate did recover efficiency lost by the GR estimate from the Wilcoxon estimate. Again in almost all situations, the HBR estimate was more efficient than the initial LMS estimate.

One other advantage of the HBR over LMS is the ability to control the degree of robustness by adjusting weight parameters. The weights allow varying robustness against outlying y-values and high leverage points (outlying $\boldsymbol{x}$). This allows for the development of dynamic graphics in exploring fits by examining the effects of increasing level of robustness on the fitted values and residuals. Also, as noted by one referee, we could reiterate the HBR estimate by using it to reestimate the weights. The additional (time-wise) would be minimal. This will be of interest in future studies.

9.0.1 Acknowledgements

The authors would like to kindly acknowledge the helpful comments of two referees and an associate editor on an earlier version of this paper.

A.1 Appendix

Notation:

1. $W_{ij}(\boldsymbol{\Delta}) = (1/2)[\mathrm{sgn}(z_j - z_i) - \mathrm{sgn}(y_j - y_i)]$, where $z_j = y_j - \boldsymbol{x}_j'\boldsymbol{\Delta}/\sqrt{n}$.
2. $t_{ij}(\boldsymbol{\Delta}) = (\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\Delta}/\sqrt{n}$.
3. $B_{ij}(t) = E[b_{ij}\boldsymbol{I}(0 < y_i - y_j < t)]$.
4. $\gamma_{ij} = B_{ij}'(0)/E(b_{ij})$.
5. $\boldsymbol{C}_n = \sum_{i<j} \gamma_{ij} b_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_i)(\boldsymbol{x}_j - \boldsymbol{x}_i)'$, a $p \times p$ matrix.
6. $\boldsymbol{R}(\boldsymbol{\Delta}) = n^{-3/2}\left[\sum_{i<j} b_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_i)W_{ij}(\boldsymbol{\Delta}) + \boldsymbol{C}_n\boldsymbol{\Delta}/\sqrt{n}\right]$, a $p \times 1$ vector.

Note that in Section 3 we used z_i for the vector $(\mathbf{x}_i, y_i)$ but in the Appendix, we will use z_i for the ith residual. Also, the definitions 2, 3, and 4 are the same as in Section 4. We have restated them for the benefit of the reader.

List of Assumptions:

(A1) b_{ij} are continuous with respect to y_i and y_j, for all i and j.

(A2) The error density $f(\cdot)$ is continuous and bounded.

(A3) There exists a positive definite matrix $\boldsymbol{\Gamma}$ such that $(1/n)\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{\Gamma}$.

(A4) $\sum_i (x_{ik} - \overline{x}_k)^2 / \max_i (x_{ik} - \overline{x}_k)^2 \to \infty$, as $n \to \infty$, for all k.

(A5) $\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Xi})$, where $\hat{\boldsymbol{\beta}}_0$ is the initial estimator, $\boldsymbol{\beta}_0$ is the true parameter value, and $\boldsymbol{\Xi}$ is a positive definite matrix.

(A6) The weight function $b_{ij} = g(\boldsymbol{x}_i, \boldsymbol{x}_j, y_i, y_j, \hat{\boldsymbol{\beta}}_0) \equiv g_{ij}(\hat{\boldsymbol{\beta}}_0)$ is continuous in the argument and ∇g_{ij} is bounded uniformly in i, j.

It can be shown that the weights (2.5) satisfy assumption (A6). Without loss of generality we will assume that the true $\boldsymbol{\beta}_0$ is $\boldsymbol{0}$.

Lemma 3 *Under assumptions (A1) and (A2),*

$$B_{ij}'(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} b(x_i, x_j, y_j + t, y_j, \hat{\boldsymbol{\beta}}_0)\ f(y_j + t)\ f(y_j) \prod_{k \neq i,j} f(y_k)\ dy_1 \cdots dy_n$$

is continuous in t.

Proof: Follows from (A1), (A2) and an application of Leibnitz's rule on differentiation of definite integrals.

Let $\boldsymbol{\Delta}$ be arbitrary but fixed. Denote $W_{ij}(\boldsymbol{\Delta})$ by W_{ij}, suppressing dependence on $\boldsymbol{\Delta}$.

Lemma 4 *Under assumptions (A1) and (A2), there exist constants* $|\xi_{ij}| < |t_{ij}|$ *such that* $E(b_{ij}W_{ij}) = -t_{ij}\ B'_{ij}(\xi_{ij})$.

Proof: Since $W_{ij} = 1, -1$, or 0 according as $t_{ij} < y_j - y_i < 0$, $0 < y_j - y_i < t_{ij}$, or otherwise, we have

$$E_{\boldsymbol{\beta}_0}(b_{ij}W_{ij}) = \int_{t_{ij}<y_j-y_i<0} b_{ij} f_{\boldsymbol{Y}}(\boldsymbol{y})\ d\boldsymbol{y} - \int_{0<y_j-y_i<t_{ij}} b_{ij} f_{\boldsymbol{Y}}(\boldsymbol{y})\ d\boldsymbol{y}.$$

When $t_{ij} > 0$, $E(b_{ij}W_{ij}) = -B_{ij}(t_{ij}) = B_{ij}(0) - B_{ij}(t_{ij}) = -t_{ij}\ B'_{ij}(\xi_{ij})$ by Lemma 3 and the Mean Value Theorem. The same result holds for $t_{ij} < 0$, which proves the lemma.

Lemma 5 *Under assumptions (A5) and (A6), we have*

$$b_{ij} = g_{ij}(\hat{\boldsymbol{\beta}}_0) = g_{ij}(\mathbf{0}) + [\nabla g_{ij}(\boldsymbol{\xi})]'\hat{\boldsymbol{\beta}}_0 = g_{ij}(\mathbf{0}) + O_p(1/\sqrt{n}),$$

uniformly over all i and j, where $\boldsymbol{\xi}$ is between $\hat{\boldsymbol{\beta}}_0$ and $\mathbf{0}$.

Proof: Follows from a multivariate Mean Value Theorem (see e.g. Apostol, 1974, p.355), and by (A5) and (A6).

Lemma 6 *Under assumptions (A1)-(A6),*

(i) $E(g_{ij}(\mathbf{0})g_{ik}(\mathbf{0})W_{ij}W_{ik}) \longrightarrow 0$, *as* $n \to \infty$

(ii) $E(g_{ij}(\mathbf{0})W_{ij}) \longrightarrow 0$, *as* $n \to \infty$

uniformly over i and j.

Proof: Without loss of generality, let $t_{ij} > 0$ and $t_{ik} > 0$, where the indices i, j and k are all different. Then

$$\begin{aligned} E(g_{ij}(\mathbf{0})g_{ik}(\mathbf{0})W_{ij}W_{ik}) &= E[g_{ij}g_{ik}\boldsymbol{I}(0 < y_j - y_i < t_{ij})\ \boldsymbol{I}(0 < y_k - y_i < t_{ik})] \\ &= \left| \int_{-\infty}^{\infty} \int_{y_i}^{y_i+t_{ik}} \int_{y_i}^{y_i+t_{ij}} g_{ij}g_{ik}\ f_i f_j f_k\ dy_j dy_k dy_i \right|. \end{aligned}$$

Assumptions (A3) and (A4) imply $(1/n)\max_i(x_{ik} - \overline{x}_k)^2 \to 0$ for all k, or equivalently $(1/\sqrt{n})\max_i|x_{ik} - \overline{x}_k| \to 0$ for all k, which implies that $t_{ij} \to 0$. Since the integrand is bounded, this proves (i).

Similarly, $E(g_{ij}(\mathbf{0})W_{ij}) = \int_{-\infty}^{\infty}\int_{y_i}^{y_i+t_{ij}} g_{ij} f_i f_j dy_j dy_i \to 0$, which proves (ii).

Lemma 7 *Under assumptions (A1)-(A6)*

(i) $Cov(b_{12}W_{12}, b_{34}W_{34}) = o(n^{-1})$.

(ii) $Cov(b_{12}W_{12}, b_{34}) = o(n^{-1})$.

(iii) $Cov(b_{12}W_{12}, b_{13}W_{13}) = o(1)$.

(iv) $Cov(b_{12}W_{12}, b_{13}) = o(1)$.

Proof: To prove (i), recall that $b_{12} = g_{12}(\mathbf{0}) + [\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0$. Thus

$$\begin{aligned} Cov(b_{12}W_{12}, b_{34}W_{34}) &= Cov(g_{12}(\mathbf{0})W_{12}, g_{34}(\mathbf{0})W_{34}) \\ &\quad +2\ Cov([\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\ W_{12}, g_{34}(\mathbf{0})W_{34}) \\ &\quad +Cov([\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\ W_{12}, [\triangledown g_{34}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\ W_{34}). \end{aligned}$$

Let I_1, I_2 and I_3 denote the three terms on the right hand side. The term I_1 is 0, by independence. Now,

$$\begin{aligned} I_2 &= 2E\left\{[\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\ W_{12}\ g_{34}(\mathbf{0})W_{34}\right\} - 2E\left\{[\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\ W_{12}\right\}\ E\left\{g_{34}(\mathbf{0})W_{34}\right\} \\ &= I_{21} - I_{22}. \end{aligned}$$

Write the first term above as

$$I_{21} = 2(1/n)E\left\{[\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\ g_{34}(\mathbf{0})(\sqrt{n}W_{12})(\sqrt{n}W_{34})\right\}.$$

The term $[\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0 = b_{12} - g_{12}(\mathbf{0})$ is bounded and of magnitude $o_p(1)$. If we can show that $\sqrt{n}W_{12}$ is integrable, then it follows using standard arguments that $I_{21} = o(1/n)$. Let F^* denote the cdf of $y_2 - y_1$ and f^* denote its pdf. Using the mean value theorem,

$$\begin{aligned} E[\sqrt{n}W_{12}(\boldsymbol{\Delta})] &= \sqrt{n}(1/2)E[\text{sgn}(y_2 - y_1 - (\boldsymbol{x}_2 - \boldsymbol{x}_1)')\boldsymbol{\Delta}/\sqrt{n}) - \text{sgn}(y_2 - y_1)] \\ &= \sqrt{n}(1/2)[2F^*(-(\boldsymbol{x}_2 - \boldsymbol{x}_1)')\boldsymbol{\Delta}/\sqrt{n}) - 2F^*(\mathbf{0})] \\ &= -\sqrt{n}f^*(\boldsymbol{\xi})(\boldsymbol{x}_2 - \boldsymbol{x}_1)')\boldsymbol{\Delta}/\sqrt{n} \le f^*(\boldsymbol{\xi})|(\boldsymbol{x}_2 - \boldsymbol{x}_1)'\boldsymbol{\Delta}| \end{aligned}$$

which is bounded. This proves that $I_{21} = o(1/n)$. Similarly,

$$I_{22} = 2(1/n)E\left\{[\triangledown g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\ (\sqrt{n}W_{12})\right\}\ E\left\{g_{34}(\mathbf{0})(\sqrt{n}W_{34})\right\} = o(1/n),$$

which proves $I_2 = 0$.

The term I_3 can be shown to be $o(n^{-1})$ similarly, which proves (i).

The proof of (ii) is analogous to (i).

To prove (iii), note that

$$
\begin{aligned}
Cov(b_{12}W_{12}, b_{13}W_{13}) &= Cov(g_{12}(\mathbf{0})W_{12}, g_{13}(\mathbf{0})W_{13}) \\
&+ \quad 2\,Cov([\nabla g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\, W_{12}, g_{13}(\mathbf{0})W_{13}) \\
&+ \quad Cov([\nabla g_{12}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\, W_{12}, [\nabla g_{13}(\boldsymbol{\xi})]' \cdot \hat{\boldsymbol{\beta}}_0\, W_{13}).
\end{aligned}
$$

The first term is $o(1)$ by Lemma 6. The second and third terms are clearly $o(1)$. This proves (iii). Result (iv) is analogously proved.

We are now ready to state and prove asymptotic linearity. Consider the negative gradient function

$$
S(\boldsymbol{\beta}) = -\nabla D(\boldsymbol{\beta}) = \sum\sum_{i<j} b_{ij}\mathrm{sgn}(z_j - z_i)(\boldsymbol{x}_j - \boldsymbol{x}_i). \tag{A.1.1}
$$

Theorem A.1.1 *Assume (A2)-(A6) hold. Then*

$$
\sup_{\|\sqrt{n}\boldsymbol{\beta}\| \le C} n^{-3/2}\,[\boldsymbol{S}(\boldsymbol{\beta}) - \boldsymbol{S}(\mathbf{0}) + 2\,\boldsymbol{C}_n\boldsymbol{\beta}] \xrightarrow{p} 0.
$$

Proof: Write $\boldsymbol{R}(\boldsymbol{\Delta}) = \left[\boldsymbol{S}(n^{-1/2}\boldsymbol{\Delta}) - \boldsymbol{S}(\mathbf{0}) + 2n^{-1/2}\,\boldsymbol{C}_n\boldsymbol{\Delta}\right]$. We will show that

$$
\sup_{\|\boldsymbol{\Delta}\| \le C} \boldsymbol{R}(\boldsymbol{\Delta}) = 2 \sup_{\|\boldsymbol{\Delta}\| \le C} \left\{ n^{-3/2}\sum\sum_{i<j} b_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_i)\, W_{ij}(\boldsymbol{\Delta}) + n^{-1/2}\boldsymbol{C}_n\boldsymbol{\Delta} \right\} \xrightarrow{p} 0.
$$

It will suffice to show that each component converges to 0. Consider the kth component

$$
\begin{aligned}
R_k(\boldsymbol{\Delta}) &= 2\left[n^{-3/2}\sum\sum_{i<j} b_{ij}(x_{jk} - x_{ik})\, W_{ij}(\boldsymbol{\Delta}) + \sum\sum_{i<j}\gamma_{ij}b_{ij}(x_{jk} - x_{ik})\, t_{ij}\right] \\
&= 2\, n^{-3/2}\sum\sum_{i<j}(x_{jk} - x_{ik})(b_{ij}W_{ij} + \gamma_{ij}t_{ij}b_{ij}).
\end{aligned}
$$

We will show that $E(R_k(\boldsymbol{\Delta})) \to 0$ and $Var(R_k(\boldsymbol{\Delta})) \to 0$. By Lemma 4 and the definition of γ_{ij},

$$
\begin{aligned}
E(R_k) &= 2\, n^{-3/2}\sum\sum_{i<j}(x_{jk} - x_{ik})[E(b_{ij}W_{ij}) + \gamma_{ij}t_{ij}E(b_{ij})] \\
&= 2\, n^{-3/2}\sum\sum_{i<j}(x_{jk} - x_{ik})t_{ij}[B'_{ij}(0) - B'_{ij}(\xi_{ij})] \\
&\le 2\, n^{-3/2}\left[\sum\sum_{i<j}(x_{jk} - x_{ik})^2\right]^{1/2}\left[\sum\sum_{i<j}t_{ij}^2\right]^{1/2}\sup_{i,j}|B'_{ij}(0) - B'_{ij}(\xi_{ij})| \\
&= 2\left[(1/n^2)\sum\sum_{i<j}(x_{jk} - x_{ik})^2\right]^{1/2}\left[(1/n)\sum\sum_{i<j}t_{ij}^2\right]^{1/2}\sup_{i,j}|B'_{ij}(0) - B'_{ij}(\xi_{ij})| \longrightarrow 0
\end{aligned}
$$

since $(1/n)\sum\sum_{i<j} t_{ij}^2 = (1/n)\boldsymbol{\Delta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\Delta} = O(1)$ and $\sup_{i,j} |B'_{ij}(0) - B'_{ij}(\xi_{ij})| \to 0$ by Lemma 3.

Next, we will show that $Var(R_k) \to 0$.

$$Var(R_k) = Var\left[2\, n^{-3/2} \sum\sum_{i<j} (x_{jk} - x_{ik})(b_{ij} W_{ij} + \gamma_{ij} t_{ij} b_{ij})\right]$$

$$= \quad Var\left[2\, n^{-3/2} \sum_{i=1}^{n}\sum_{j=1}^{n} (x_{jk} - \overline{x}_k)(b_{ij} W_{ij} + \gamma_{ij} t_{ij} b_{ij})\right]$$

$$= \quad 4\, n^{-3} \sum_{i=1}^{n}\sum_{j=1}^{n} (x_{jk} - \overline{x}_k)^2 Var(b_{ij} W_{ij} + \gamma_{ij} t_{ij} b_{ij})$$

$$+4\, n^{-3} \sum\sum\sum\sum_{(i,j)\neq(l,m)} (x_{jk} - \overline{x}_k)(x_{mk} - \overline{x}_k) Cov(b_{ij} W_{ij} + \gamma_{ij} t_{ij} b_{ij},\; b_{lm} W_{lm} + \gamma_{lm} t_{lm} b_{lm}).$$

The double sum term above goes to 0, since there there n^2 bounded terms in the double sum, multiplied by n^{-3}. There are two types of covariance terms in the quadruple sum, covariance terms with all four indices different, e.g. $((i,j),(l,m)) = ((1,2),(3,4))$, and covariance terms with one index of the first pair equal to one index of the second pair, e.g. $((i,j),(l,m)) = ((1,2),(1,3))$. Since there are of magnitude n^4 terms with all four indices different, we need to show that each covariance term is $o(n^{-1})$. This immediately follows from Lemma 7. Finally there are of magnitude n^3 covariance terms with one shared index, and we need to show each term is $o(1)$. Again, this immediately follows from Lemma 7. The theorem is proved.

Now, let

$$Q(\boldsymbol{\beta}) = D(\mathbf{0}) - \sum\sum_{i<j} b_{ij}\mathrm{sgn}(y_j - y_i)(\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{C}_n\boldsymbol{\beta}$$

$$D^*(\boldsymbol{\Delta}) = n^{-1} D(n^{-1/2}\boldsymbol{\Delta})$$

$$Q^*(\boldsymbol{\Delta}) = n^{-1} Q(n^{-1/2}\boldsymbol{\Delta}).$$

Note that minimizing $D^*(\boldsymbol{\Delta})$ and $Q^*(\boldsymbol{\Delta})$ is equivalent to minimizing $D(n^{-1/2}\boldsymbol{\Delta})$ and $Q(n^{-1/2}\boldsymbol{\Delta})$, respectively.

Theorem A.1.2 *Assume (A1)-(A6) hold. Then for a fixed constant C and for any $\epsilon > 0$,*

$$P\left(\sup_{\|\boldsymbol{\Delta}\|<C} |Q^*(\boldsymbol{\Delta}) - D^*(\boldsymbol{\Delta})| \geq \epsilon\right) \to 0.$$

Proof: Since $\frac{\partial Q^*}{\partial \boldsymbol{\Delta}} - \frac{\partial D^*}{\partial \boldsymbol{\Delta}} = 2n^{-3/2}\left[\sum\sum_{i<j} b_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_i)W_{ij} + \boldsymbol{C}_n(n^{-1/2}\boldsymbol{\Delta})\right] = R(\boldsymbol{\Delta})$, it follows from Theorem A.1.1 that for $\epsilon > 0$ and $C > 0$,

$$P\left(\sup_{\|\boldsymbol{\Delta}\|<C} \| \frac{\partial Q^*}{\partial \boldsymbol{\Delta}} - \frac{\partial D^*}{\partial \boldsymbol{\Delta}} \| \geq \epsilon/C\right) \to 0.$$

For $0 \leq t \leq 1$, let $\boldsymbol{\Delta}_t = t\,\boldsymbol{\Delta}$. Then

$$\begin{aligned}\left|\frac{d}{dt}[Q^*(\boldsymbol{\Delta}_t) - D^*(\boldsymbol{\Delta}_t)]\right| &= \left|\sum_{k=1}^{p} \Delta_k\left(\frac{\partial Q^*}{\partial \boldsymbol{\Delta}_{tk}} - \frac{\partial D^*}{\partial \boldsymbol{\Delta}_{tk}}\right)\right| \\ &\leq \|\boldsymbol{\Delta}\| \sup_{\|\boldsymbol{\Delta}\|<C} \| \frac{\partial Q^*}{\partial \boldsymbol{\Delta}} - \frac{\partial D^*}{\partial \boldsymbol{\Delta}} \| < \|\boldsymbol{\Delta}\| (\epsilon/C) < \epsilon\end{aligned}$$

with probability approaching 1. Now, let $h(t) = Q^*(\boldsymbol{\Delta}_t) - D^*(\boldsymbol{\Delta}_t)$. By the previous result, we have $|h'(t)| < \epsilon$ with high probability. Thus

$$|h(1)| = |h(1) - h(0)| = \left|\int_0^1 h'(t)\, dt\right| \leq \int_0^1 |h'(t)|\, dt < \epsilon,$$

with probability approaching one. This proves the theorem.

The next theorem states asymptotic normality of $\boldsymbol{S}(\mathbf{0})$.

Theorem A.1.3 *Under assumptions (A1)-(A6), we have*

$$n^{-3/2}\boldsymbol{S}(\mathbf{0}) \sim AN(\mathbf{0},\ \boldsymbol{\Sigma}).$$

Proof: Let $\boldsymbol{S}_P$ denote the projection of $\boldsymbol{S}^*(\mathbf{0}) = n^{-3/2}\boldsymbol{S}(\mathbf{0})$ onto the space of linear combinations of independent random variables. Then

$$\begin{aligned}\boldsymbol{S}_P = \sum_{k=1}^{n} E[S^*(\mathbf{0})|y_k] &= \sum_{k=1}^{n} E\left[n^{-3/2}\sum\sum_{i<j}(\boldsymbol{x}_j - \boldsymbol{x}_i)b_{ij}\text{sgn}(y_j - y_i) \mid y_k\right] \\ &= \sum_{k=1}^{n} n^{-3/2}\left[\sum_{i=1}^{k-1}(\boldsymbol{x}_k - \boldsymbol{x}_i)E[b_{ik}\text{sgn}(y_k - y_i)|y_k] + \sum_{j=k+1}^{n}(\boldsymbol{x}_j - \boldsymbol{x}_k)E[b_{kj}\text{sgn}(y_j - y_k)|y_k]\right] \\ &= n^{-3/2}\sum_{k=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_j - \boldsymbol{x}_k)E[b_{kj}\text{sgn}(y_j - y_k)|y_k] \\ &= (1/\sqrt{n})\sum_{k=1}^{n}\boldsymbol{U}_k,\end{aligned}$$

where $\boldsymbol{U}_k$ is defined in Section 5. By assumption (N2) and a multivariate extension of the Lindeberg-Feller theorem (Rao, 1973), it follows that $\boldsymbol{S}_P \sim AN(\mathbf{0},\ \boldsymbol{\Sigma})$. If we show that $E \| \boldsymbol{S}_P - \boldsymbol{S}^*(\mathbf{0}) \|^2 \to 0$, then it follows from the Projection theorem that $\boldsymbol{S}^*(\mathbf{0})$ has the same asymptotic distribution as $\boldsymbol{S}_P$, and the proof will be done. Equivalently, we may show that $E(S_{P,r} - S_r^*(\mathbf{0}))^2 \to 0$ for each component $r = 1, \ldots, p$. Since for each r we have $E(S_{P,r} - S_r^*(\mathbf{0})) = 0$, then

$$\begin{aligned}
E(S_{P,r} - S_r^*(\mathbf{0}))^2 &= Var(S_{P,r} - S_r^*(\mathbf{0})) \\
&= \ Var\left[n^{-3/2}\sum_{k=1}^{n}\sum_{j=1}^{n}(x_{jr} - x_{kr})\left\{E[b_{kj}\text{sgn}(y_j - y_k)|y_k] - b_{kj}\text{sgn}(y_j - y_k)\right\}\right] \\
&\equiv \ Var\left[n^{-3/2}\sum_{k=1}^{n}\sum_{j=1}^{n}T(y_j, y_k)\right] \\
&= \ n^{-3}\sum_{k=1}^{n}\sum_{j=1}^{n}Var(T(y_j, y_k)) + n^{-3}\sum_{k}\sum_{j}\sum_{l}\sum_{m}Cov[T(y_j, y_k),\ T(y_l, y_m)]
\end{aligned}$$

where the quadruple sum is taken over $(j, k) \neq (l, m)$. The double sum term goes to 0 since there are n^2 bounded terms divided by n^3. There are two types of covariance terms in the quadruple sum: terms with four different indices , and terms with three different indices (i.e., one shared index). Covariance terms with four different indices are zero (this can be shown by writing out the covariance in terms of expectations, and using symmetry to show that each covariance term is zero). Thus we only need to consider covariance terms with three different indices and show that the sum goes to 0. Letting k be the shared index (without loss of generality), and noting that $E\ T(y_j, y_k) = 0$ for all j, k, we have

$$\begin{aligned}
&n^{-3}\sum_{k}\sum_{j\neq k}\sum_{l\neq k,j} Cov[T(y_j, y_k),\ T(y_l, y_k)] \\
&= \ n^{-3}\sum_{k}\sum_{j\neq k}\sum_{l\neq k,j} E\left\{T(y_j, y_k)\cdot T(y_l, y_k)\right\} \\
&= \ n^{-3}\sum_{k}\sum_{j\neq k}\sum_{l\neq k,j} E\left\{[E(b_{kj}\text{sgn}(y_j - y_k)|y_k) - b_{kj}\text{sgn}(y_j - y_k)]\right. \\
&\quad \left.\cdot\ [E(b_{kl}\text{sgn}(y_l - y_k)|y_k) - b_{kl}\text{sgn}(y_l - y_k)]\right\} \\
&= \ n^{-3}\sum_{k}\sum_{j\neq k}\sum_{l\neq k,j} E\left\{[E(g_{kj}(\mathbf{0})\ \text{sgn}(y_j - y_k)|y_k) - g_{kj}(\mathbf{0})\ \text{sgn}(y_j - y_k)]\right. \\
&\quad \left.\cdot\ [E(g_{kl}(\mathbf{0})\ \text{sgn}(y_l - y_k)|y_k) - g_{kl}(\mathbf{0})\ \text{sgn}(y_l - y_k)]\right\} + o_p(1)
\end{aligned}$$

where the last equality follows from the relation $b_{kj} = g_{kj}(\mathbf{0}) + 0_p(1/\sqrt{n})$. Expanding the product, each term in the triple sum may be written as

$$\begin{aligned}
&E\left\{[E(g_{kj}(\mathbf{0})\ \text{sgn}(y_j - y_k)|y_k)]^2\right\} + E\left\{g_{kj}(\mathbf{0})\ \text{sgn}(y_j - y_k)g_{kl}(\mathbf{0})\ \text{sgn}(y_l - y_k)\right\} \\
&\quad -2\ E\left\{g_{kj}(\mathbf{0})\ \text{sgn}(y_j - y_k)\left[E(g_{kl}(\mathbf{0})\ \text{sgn}(y_l - y_k)|y_k)\right]\right\} \\
&= \ (1 + 1 - 2)\left\{[E(g_{kj}(\mathbf{0})\ \text{sgn}(y_j - y_k)|y_k)]^2\right\} = 0
\end{aligned}$$

where the first equality follows by taking conditional expectations with respect to k inside appropriate terms.

A similar method applies to terms where k is not the shared index. The theorem is proved.

Proof of Theorem 5.1:

Let $\tilde{\boldsymbol{\beta}}$ denote the value which minimizes $Q(\boldsymbol{\beta})$. Then $\tilde{\boldsymbol{\beta}}$ is the solution to

$$\mathbf{0} = \boldsymbol{S}(\mathbf{0}) - 2\boldsymbol{C}_n\boldsymbol{\beta}$$

so that $\sqrt{n}\tilde{\boldsymbol{\beta}} = (1/2)n^2\boldsymbol{C}_n^{-1}[n^{-3/2}\boldsymbol{S}(\mathbf{0})] \sim AN(\mathbf{0}, (1/4)\ \boldsymbol{C}^{-1}\boldsymbol{\Sigma}\boldsymbol{C}^{-1})$, by Theorem A.1.3 and Assumption (N1). It remains to show that $\sqrt{n}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) = o_p(1)$. This follows from Theorem A.1.2 and convexity of $D(\boldsymbol{\beta})$, using standard arguments as in Jaeckel (1972). The theorem is proved.

Studentized Residuals:

Assume without loss of generality that $\alpha = 0$, $\boldsymbol{\beta} = \mathbf{0}$, and $\operatorname{med} e_i = 0$. In this section we must further assume that the variance of e_i is finite, i.e., $\sigma^2 < \infty$. From the above proof of Theorem 5.1, asymptotically $\widehat{\boldsymbol{\beta}}$ can be expressed as

$$\sqrt{n}\widehat{\boldsymbol{\beta}} = \frac{1}{2}(n^{-2}\boldsymbol{X}'\boldsymbol{A}_n\boldsymbol{X})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{U}_k + o_p(1)\ , \tag{A.1.2}$$

where $\boldsymbol{U}_k$ is given in the first paragraph of Section 5. In this section, we will assume the weights b_{ij} are given. We use the approximation (5.3) in place of $\mathbf{U}_k$, with $a_{ij}^* = b_{ij}/(\sqrt{12}\tau)$, see (5.5). Hence, $\mathbf{U}_k$ is replaced by,

$$\widehat{\mathbf{U}}_k^* = -\frac{\sqrt{12}\tau}{n^2}\sum_{j=1}^{n}(\mathbf{x}_j - \mathbf{x}_k)a_{kj}^*(1 - 2F(e_k))\ . \tag{A.1.3}$$

The estimate of α given by (5.6) can be expressed asymptotically as

$$\widehat{\alpha} = \tau_S n^{-1/2}\sum_{i=1}^{n}\operatorname{sgn}(e_i) + o_p(1)\ ; \tag{A.1.4}$$

see McKean et al. (1990). Using (A.1.2) and (A.1.4) we have the following first order expression for the residuals $\widehat{e}_i^*$, (5.9),

$$\widehat{\mathbf{e}}^* \doteq \mathbf{e} - \tau_S\frac{1}{n}\sum_{i=1}^{n}\operatorname{sgn}(e_i)\mathbf{1} - \frac{1}{2\sqrt{n}}\boldsymbol{X}(n^{-2}\boldsymbol{X}'\boldsymbol{A}^*\boldsymbol{X})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{U}_k^*\ , \tag{A.1.5}$$

where $\boldsymbol{A}^* = [a_{ij}^*]$. Because $E[\mathbf{U}_k^*] = \mathbf{0}$ and $\operatorname{med} e_i = 0$, taking expectations of both sides of (A.1.5) leads to

$$E[\widehat{\mathbf{e}}^*] \doteq E[e_1]\mathbf{1}\ . \tag{A.1.6}$$

Write

$$\text{Var}(\widehat{\mathbf{e}}^*) \doteq E[(\widehat{\mathbf{e}}^* - E[e_1]\mathbf{1})(\widehat{\mathbf{e}}^* - E[e_1]\mathbf{1})'] \ . \tag{A.1.7}$$

The approximate variance-covariance matrix of $\widehat{\mathbf{e}}^*$ can then be obtained by subtituting the right side of expression (A.1.5) for $\widehat{\mathbf{e}}^*$ in expression (A.1.7) and then expanding and taking expectations term-by-term. This is a tedious derivation, but by making use of $E[\mathbf{U}_k^*] = \mathbf{0}$, $\text{med}\, e_i = 0$, $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$, and $\sum_{j=1}^n a_{ij}^* = \sum_{j=1}^n a_{ji}^* = 0$ we obtain expression (5.10) of Section 5.

BIBLIOGRAPHY

Apostal, T. M. (1974), *Mathematical Analysis, 2nd Edition*, Addison-Wesley: Reading, Massachusetts.

Coakley, C. and Hettmansperger, T. P. (1993), "A Bounded Influence, High Breakdown, Efficient Regression Estimator," *The Journal of the American Statistical Association*, 88, 872-880.

Cook, R. D., Hawkins, D. M. and Weisberg, S. (1992), "Comparison of Model Misspecification Diagnostics Using Residuals from Least Mean of Squares and Least Median of Squares Fits," *The Journal of the American Statistical Association*, 87, 419-424.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman and Hall, New York.

Croux, C. Rousseeuw, P. J. and Hössjer, O. (1994), "Generalized S-Estimators," *The Journal of the American Statistical Association*, 89, 1271-1281.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. J. (1986), *Robust Statistics, the Approach Based on Influence Functions*, New York: John Wiley and Sons.

Hawkins, D. M. (1994), "The Feasible Solution Algorithm for Least Trimmed Squares Regression," *Computational Statistics and Data Analysis*, 17, 185-196.

Heiler, S. and Willers, R. (1988), Asymptotic Normality of R-Estimates in the Linear Model, *Statistics*, **19**, 173-184.

Hettmansperger, T. P. and McKean, J. W. (1983), "A Geometric Interpretation of Inferences Based on Ranks in the Linear Model," *Journal of the American Statistical Association*, 78, 885-893.

Hettmansperger, T. P. McKean, J. W. and Sheather, S. J. (1997), "Rank-Based Analyses of Linear Models," In *Handbook of Statistics*, Eds: G. S. Maddala and C. R. Rao, 145-173, Amsterdam: Elsevier.

Hettmansperger, T. P. and Sheather, S. J. (1992), "A Cautionary Note on the Method of Least Median Squares", *The American Statistician*, 46, 79-83.

Hettmansperger, T. P. and Sheather, S. J. (1993), Reply to Comment by Appa and Land on "A cautionary note on the method of least median squares", *The American Statistician*, 47, 161-162.

Hössjer, O. (1994), "Rank-Based Estimates in the Linear Model with High Breakdown Point," *Journal of the American Statistical Association*, 89, 149-158.

Jaeckel, L. A. (1972), "Estimating Regression Coefficients by Minimizing the Dispersion of Residuals," *The Annals of Mathematical Statistics*, 43, 1449-1458.

Jureckova, J. (1971), "Nonparametric Estimate of Regression Coefficients, " *The Annals of Mathematical Statistics*, 42, 1328-1338.

Kapenga, J. A., McKean, J. W. and Vidmar, T. J. (1988), *RGLM: Users Manual*, Amer. Statist. Assoc. Short Course on Robust Statistical Procedures for the Analysis of Linear and Nonlinear Models, New Orleans.

Koul, H. L., Sievers, G. L. and McKean, J. W. (1987), "An Estimator of the Scale Parameter for the Rank Analysis of Linear Models under General Score Functions," *Scandinavian Journal of Statistics*, 14, 131-141.

McKean, J. W. Naranjo, J. D. and Sheather, S. J. (1996a), "Diagnostics to Detect Differences in Robust Fits of Linear Models," *Computational Statistics*, 11, 223-243.

McKean, J. W. Naranjo, J. D. and Sheather, S. J. (1996b), "An Efficient and High Breakdown Procedure for Model Criticism," *Communications in Statistics: Theory and Methods*, In press.

McKean, J. W. Sheather, S. J. and Hettmansperger, T. P. (1990), "Regression Diagnostics for Rank-Based Methods," *The Journal of the American Statistical Association*, 85, 1018-28.

McKean, J. W. Sheather, S. J. and Hettmansperger, T. P. (1993), "The Use and Interpretation of Residuals Based on Robust Estimation," *The Journal of the American Statistical Association*, 88, 1254-1263.

McKean, J. W. Sheather, S. J. and Hettmansperger, T. P. (1994), "Robust and High Breakdown Fits of Polynomial Models," *Technometrics*, 36, 409-415.

Naranjo, J. D. and Hettmansperger, T. P. (1994), "Bounded-Influence Rank Regression," *Journal of the Royal Statistical Society B*, 56, No. 1, 209-220.

Naranjo, J. D. McKean, J. W. Sheather, S. J. and Hettmansperger, T. P. (1994), "The Use and Interpretation of Rank-Based Residuals," *Nonparametric Statistics*, 3, 323-341.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, Wiley and Sons: New York.

Rousseeuw, P. J. and van Zomeren, B. C. (1990), "Unmasking Multivariate Outliers and Leverage Points," *The Journal of the American Statistical Association*, 85, 633-639.

Rousseeuw, P. J. and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, John Wiley and Sons, New York.

Sheather, S. J. and McKean, J. W. (1987), "A Comparison of Testing and Confidence Interval Methods for the Median," *Statistics and Probability Letters 6*, 31-36.

Sheather, S. J. McKean, J. W. and Hettmansperger, T. P. (1997), "Finite Sample Stability Properties of the Least Median of Squares Estimator," *Journal of Statistical Computation and Simulation*, 58, 371-383.

Sievers, G. L. (1983), "A Weighted Dispersion Function for Estimation in Linear Models," *Communications in Statistics, Theory and Methods*, 12(10), 1161-1179.

Simonoff, J. S. and Hawkins, D. M. (1993), "Algorithm AS 282: High Breakdown Regression and Multivariate Estimation," *Applied Statistics*, 42, 423-432.

Simpson, D. G. Ruppert, D. and Carroll, R. J. (1992), "On One-step GM-estimates and Stability of Inferences in Linear Regression," *Journal of the American Statistical Association*, 87, 439-450.

Stromberg, A. J. (1993), "Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression," *SIAM Journal of Scientific Computing*, 14, 1289-1299.

Tableman, M. (1990), "Bounded-Influence Rank Regression: A One-Step Estimator Based on Wilcoxon Scores," *Journal of the American Statistical Association*, 85, 508-513.

Witt, L., McKean, J. W. and Naranjo, J. (1995), "Influence Functions for Rank-Based Procedures in the General Linear Model," *Journal of Nonparametric Statistics*, 5, 339-358.