The Annals of Statistics
2002, Vol. 30, No. 2, 556–582

BOOTSTRAPPING ROBUST ESTIMATES OF REGRESSION

BY MATIAS SALIBIAN-BARRERA[1] AND RUBEN H. ZAMAR[2]

University of British Columbia

We introduce a new computer-intensive method to estimate the distribution of robust regression estimates. The basic idea behind our method is to bootstrap a reweighted representation of the estimates. To obtain a bootstrap method that is asymptotically correct, we include the auxiliary scale estimate in our reweighted representation of the estimates. Our method is computationally simple because for each bootstrap sample we only have to solve a linear system of equations. The weights we use are decreasing functions of the absolute value of the residuals and hence outlying observations receive small weights. This results in a bootstrap method that is resistant to the presence of outliers in the data. The breakdown points of the quantile estimates derived with this method are higher than those obtained with the bootstrap. We illustrate our method on two datasets and we report the results of a Monte Carlo experiment on confidence intervals for the parameters of the linear model.

1. Introduction. The standard error and sampling distribution of robust regression estimates can be estimated, in principle, using the bootstrap [Efron (1979)]. This method has been extensively studied for diverse models. In particular, the theory for the bootstrap distribution of robust estimates has been considered by Shorack (1982), Parr (1985), Yang (1985), Shao (1990, 1992), Liu and Singh (1992) and Singh (1998), among others.

The standard error of robust regression estimates can also be estimated using their asymptotic variances. However, the asymptotic distribution of robust regression estimates has been mainly studied under the central normal model which, of course, does not hold in most practical situations when robust methods would be recommended. When the distribution of the errors is symmetric, the estimates of the regression coefficients and of the scale of the errors are asymptotically independent. Because outliers need not be balanced on both sides of the regression line, many datasets with outliers do not satisfy this symmetry assumption. If one relaxes this condition, the calculation of the asymptotic distribution of robust location and regression estimates becomes involved [see Carroll (1978, 1979), Huber (1981), Rocke and Downs (1981), Carroll and Welsh (1988) and Salibian-Barrera (2000)].

We will focus on MM-estimates of regression [Yohai (1987)] calculated with an initial S-estimate [Rousseeuw and Yohai (1984)] but our method can, in principle,

Received October 2000; revised September 2001.

[1]Research partially funded by UBC and an NSERC operating grant.

[2]Research funded by an NSERC an operating grant.

AMS 2000 subject classifications. 62F35, 62F40, 62G09, 62G20, 62G35, 62J05.

Key words and phrases. Regression, breakdown point, confidence intervals.

be applied to other types of robust regression estimates (see Section 8). These estimates have desirable robustness properties and are available in the statistical software program S-plus. However, three problems arise when we want to use the bootstrap to estimate their asymptotic distribution:

- *Numerical instability*. The bootstrap distribution might be a very poor estimator of the distribution of the regression estimates because the proportion of outliers in the bootstrap samples can be higher than that in the original dataset.
- *Computational cost*. Due to the nonconvex optimization problems that have to be solved in order to calculate these robust regression estimates, for high-dimensional problems ($p > 10$ say) it may not be feasible to obtain a few thousand recalculated estimates.
- *Recalculating the residual scale estimate*. Recalculating the robust residual scale estimate with each bootstrap sample increases the already high computational demands of the method. If we do not recalculate the scale estimate for each bootstrap sample, the resulting distribution may not converge to the correct asymptotic distribution.

Intuitively, the reason for the *numerical instability* mentioned above is as follows. Outlying and nonoutlying observations have the same chance of belonging to any bootstrap sample. With a certain positive probability, the proportion of outliers in a bootstrap sample can be larger than the fraction of contamination the robust estimate tolerates. In other words, a certain proportion of the recalculated values of the robust estimate may be heavily influenced by the outliers in the data. Thus, the tails of the bootstrap distribution can be heavily influenced by the outliers, regardless of the robustness of the statistic being bootstrapped. Singh (1998) quantified this problem for the estimates of the quantiles of the asymptotic distribution of robust location estimates. He defined the breakdown point for bootstrap quantiles and showed that it is disappointingly low even for highly robust location estimates. He proposed drawing the bootstrap samples from the Winsorized observations and showed that the quantile estimates obtained with this method have the highest attainable breakdown point and that they converge to the quantiles of the asymptotic distribution of the estimate. Unfortunately, it is not clear how to extend Singh's proposal to the linear regression model.

The *recalculation of the residual scale estimate* and its impact on the overall *computational cost* of the method have not received much attention in the literature. Many proposals to recalculate robust estimates [e.g., Schucany and Wang (1991) and Hu and Kalbfleisch (2000)] ignore the fact that when the errors do not have a symmetric distribution the scale estimate is no longer asymptotically independent of the regression parameter estimates. Because of this, for the bootstrap procedure to be asymptotically correct we have to recalculate the scale estimate with each resample. When we bootstrap MM-regression estimates with initial S-regression estimates, for each bootstrap sample we have to solve a nonconvex minimization problem in p dimensions to determine the initial scale estimate. Moreover, the objective function of this optimization problem is defined implicitly. After the scale estimate is found, we have to find a local extreme of another nonconvex function (also in p dimensions) to determine the final MM-regression estimate. The number of bootstrap samples needed to obtain reliable distribution estimates naturally grows with the dimension of the statistic and hence makes the problem computationally even more expensive to solve. This large number of nonlinear optimization problems may render the method unfeasible for high-dimensional problems. As an example of the computational times that can be expected, the evaluation of 5000 bootstrap recalculations of an MM-regression estimate on a simulated dataset with 200 observations and 10 explanatory variables took 9120 CPU seconds ($\approx$2.5 hours) on a Sun Sparc Ultra Workstation. Note that due to the first problem discussed above, even if we wait for the results of this run, they might not be reliable.

The second problem mentioned above has received some attention in the literature. See, for example, Schucany and Wang (1991), Hu and Zidek (1995), Singh (1998) and Hu and Kalbfleisch (2000). Note, however, that simultaneous consideration of these problems seems to be missing in the literature. In particular, the need to recalculate the scale with each bootstrap sample has not yet been studied in the robustness literature.

Our basic idea is to bootstrap a reweighted representation of the estimate. This procedure, called "fast bootstrap," is computationally simple because for each bootstrap sample we only have to calculate a weighted average to recalculate the scale estimate and a weighted least squares estimate to obtain the bootstrapped regression estimate (this solves the problems associated with computational cost). The form of the weights for MM-regression estimates makes the procedure numerically stable and more robust to the presence of outliers (this solves the *numerical instability* problem mentioned above). We show that the breakdown point of the quantiles obtained with the fast bootstrap is higher than that of the quantiles obtained with the bootstrap and close to the maximum $1/2$ [see Singh (1998)]. The intuitive reason for the good breakdown point properties of the fast bootstrap is that outlying points will typically be associated with small weights and hence have small impact on the bootstrapped estimate.

To illustrate the gain in speed, consider the simulated dataset mentioned above. The same number of recalculations performed with the fast bootstrap required only 416 CPU seconds (approximately 7 minutes) (instead of the 2.5 hours needed by the bootstrap). In the context of data mining and other applications with extremely large datasets (both in the number of cases and in the number of covariates), full recalculation of robust estimates is rarely a feasible option.

The rest of the paper is organized as follows. Section 2 contains the definitions of the regression estimates considered in this work. Section 3 introduces the fast bootstrap. Section 4 illustrates its application with two examples. Sections 5 and 6 study its asymptotic and robustness properties, respectively. Section 7 presents some simulation results. Section 8 contains the concluding remarks. Finally, the Appendix gives the proofs of the main theorems discussed in this paper.

2. Definitions and notation. To fix ideas, we will explicitly consider the case of random explanatory variables and apply our method to regression MM-estimates [Yohai (1987)]. The case of fixed explanatory variables and other robust regression estimates is briefly discussed in Section 8.

Let $(y_1,\ \mathbf{z}_1')', \dots, (y_n,\ \mathbf{z}_n')'$ be independent random vectors with common distribution H, and set $\mathbf{x}_i = (1,\ \mathbf{z}_i')' \in \mathbb{R}^p$. We will consider the linear regression model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}_0 + \sigma_0 \varepsilon_i, \qquad i = 1, \dots, n. \tag{2.1}$$

Ideally, one would like to assume that y_i and $\mathbf{z}_i$ are independent, with $y_i \sim F_0$, $\mathbf{z}_i \sim G_0$, $(y_i, \mathbf{z}_i')' \sim H_0$ and F_0 being some specified symmetric distribution (typically the standard normal distribution). To allow for the occurrence of outliers and other departures from the classical model, we will assume that the actual distribution H of the data belongs to the contamination neighborhood

$$\mathcal{H}_\epsilon = \{H = (1-\epsilon)H_0 + \epsilon H^*\}, \tag{2.2}$$

where $0 \le \epsilon < 1/2$ and H^* is an arbitrary and unspecified distribution.

MM-estimates are based on two loss functions ρ_0 and ρ_1, which determine the breakdown point and the efficiency of the estimate, respectively. More precisely, the MM-estimate $\hat{\boldsymbol{\beta}}_n$ satisfies the equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho_1'\left(\frac{y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_n}{\hat{\sigma}_n}\right)\mathbf{x}_i = \mathbf{0}, \tag{2.3}$$

where $\hat{\sigma}_n$ is a scale S-estimate [Rousseeuw and Yohai (1984)]. That is, $\hat{\sigma}_n$ minimizes the M-scale $\hat{\sigma}_n(\boldsymbol{\beta})$ implicitly defined by the equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\hat{\sigma}_n(\boldsymbol{\beta})}\right) = b. \tag{2.4}$$

The asymptotic distribution of MM-estimates has been studied by Yohai (1987) under the assumption that $H = H_0$ (central parametric model). This assumption will not hold, however, in typical situations when one wishes to use highly robust MM-estimates. The fast bootstrap introduced in the next section yields a consistent estimate for the covariance of $\hat{\boldsymbol{\beta}}_n$ under rather general conditions, including the case $H \in \mathcal{H}_\varepsilon$.

3. The fast bootstrap. In what follows let $\hat{\boldsymbol{\beta}}_n$ be the MM-regression estimate that satisfies (2.3). The scale estimate $\hat{\sigma}_n$ is the S-scale obtained in (2.4). Let $\tilde{\boldsymbol{\beta}}_n$ be the associated S-regression estimate.

We are interested in making statistical inferences about the regression parameter $\boldsymbol{\beta}_0$. Based on the same "plug-in" principle behind the bootstrap [Efron (1979)], we propose using the following computer-intensive method to generate a large number of recalculated $\hat{\boldsymbol{\beta}}_n^*$'s. We will use the empirical distribution function of these recomputed statistics as our estimate of the sampling distribution of $\hat{\boldsymbol{\beta}}_n$. We could also use the empirical covariance matrix of the recalculated $\hat{\boldsymbol{\beta}}_n^*$'s to estimate the asymptotic covariance matrix of the sequence $\hat{\boldsymbol{\beta}}_n$. However, we prefer to use the first approach.

For each pair $(y_i, \mathbf{x}_i')'$ in the sample define the residuals associated with $\hat{\boldsymbol{\beta}}_n$ and $\tilde{\boldsymbol{\beta}}_n$: $r_i = y_i - \hat{\boldsymbol{\beta}}_n' \mathbf{x}_i$ and $\tilde{r}_i = y_i - \tilde{\boldsymbol{\beta}}_n' \mathbf{x}_i$. First note that $\hat{\boldsymbol{\beta}}_n$ and $\hat{\sigma}_n$ can be formally represented as the result of a weighted least squares fit. For $i = 1, \ldots, n$ define the weights ω_i and v_i as

$$\omega_i = \frac{\rho_1'(r_i/\hat{\sigma}_n)}{r_i}, \qquad v_i = \frac{\hat{\sigma}_n}{n\,b} \frac{\rho_0(\tilde{r}_i/\hat{\sigma}_n)}{\tilde{r}_i}. \tag{3.1}$$

Simple computations yield the following weighted average representation of (2.3) and (2.4):

$$\hat{\boldsymbol{\beta}}_n = \left[\sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \sum_{i=1}^n \omega_i \mathbf{x}_i y_i, \tag{3.2}$$

$$\hat{\sigma}_n = \sum_{i=1}^n v_i (y_i - \tilde{\boldsymbol{\beta}}_n' \mathbf{x}_i). \tag{3.3}$$

Let $\{(y_i^*, \mathbf{x}_i^{*'})', i = 1, \ldots, n\}$ be a bootstrap sample from the observations. Define the random variables $\hat{\boldsymbol{\beta}}_n^*$ and $\hat{\sigma}_n^*$ by

$$\hat{\boldsymbol{\beta}}_n^* = \left[\sum_{i=1}^n \omega_i^* \mathbf{x}_i^* \mathbf{x}_i^{*'}\right]^{-1} \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* y_i^*, \tag{3.4}$$

$$\hat{\sigma}_n^* = \sum_{i=1}^n v_i^* (y_i^* - \tilde{\boldsymbol{\beta}}_n' \mathbf{x}_i^*), \tag{3.5}$$

where $\omega_i^* = \rho_1'(r_i^*/\hat{\sigma}_n)/r_i^*$, $v_i^* = \hat{\sigma}_n \rho_0(\tilde{r}_i^*/\hat{\sigma}_n)/(nb\tilde{r}_i^*)$, $r_i^* = y_i^* - \hat{\boldsymbol{\beta}}_n' \mathbf{x}_i^*$ and $\tilde{r}_i^* = y_i^* - \tilde{\boldsymbol{\beta}}_n' \mathbf{x}_i^*$ for $1 \le i \le n$. Note that the estimates $\hat{\boldsymbol{\beta}}_n$, $\hat{\sigma}_n$ and $\tilde{\boldsymbol{\beta}}_n$ are not recalculated from each bootstrap sample.

The recalculated $\hat{\boldsymbol{\beta}}_n^*$ and $\hat{\sigma}_n^*$ obtained in (3.4) and (3.5) may not reflect the actual variability of the random vector $(\hat{\boldsymbol{\beta}}_n', \hat{\sigma}_n)'$ due to the fact that the estimates used in the weights ω_i and v_i are kept fixed. To fix this problem, we apply a linear correction to the recalculated $\hat{\boldsymbol{\beta}}_n^*$ and $\hat{\sigma}_n^*$ and combine them. Let

$$\mathbf{M}_n = \hat{\sigma}_n \left[\sum_{i=1}^n \rho_1''(r_i/\hat{\sigma}_n, \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i', \tag{3.6}$$

$$\mathbf{d}_n = a_n^{-1}\left[\sum_{i=1}^{n} \rho_1''(r_i/\hat{\sigma}_n, \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \sum_{i=1}^{n} \rho_1''(r_i/\hat{\sigma}_n, \mathbf{x}_i) r_i \mathbf{x}_i, \tag{3.7}$$

$$a_n = \frac{1}{n}\frac{1}{b}\sum_{i=1}^{n}[\rho_0'(\tilde{r}_i/\hat{\sigma}_n)\tilde{r}_i/\hat{\sigma}_n]. \tag{3.8}$$

The fast bootstrap recalculated $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_n^{R*} - \hat{\boldsymbol{\beta}}_n = \mathbf{M}_n(\hat{\boldsymbol{\beta}}_n^{*} - \hat{\boldsymbol{\beta}}_n) + \mathbf{d}_n(\hat{\sigma}_n^{*} - \hat{\sigma}_n).$$

Theorem 4.3 in Salibian-Barrera (2000) shows that, in general, the asymptotic behavior of the sequence $\hat{\boldsymbol{\beta}}_n$ depends on that of $\hat{\sigma}_n$. Hence, to obtain an estimate of the distribution of $\hat{\boldsymbol{\beta}}_n$, we must take into account the behavior of the scale estimate $\hat{\sigma}_n$.

REMARK 1. Note that to recalculate $\hat{\boldsymbol{\beta}}_n^{R*} - \hat{\boldsymbol{\beta}}_n$ we do not solve (2.4) and (2.3). For each bootstrap sample we only solve the linear system of equations (3.4) and calculate the weighted average (3.5). The correction factors $\mathbf{M}_n$, $\mathbf{d}_n$ and a_n arise from two linear systems and a weighted average, respectively, and are computed only once with the full sample.

REMARK 2. For MM-regression estimates $\hat{\boldsymbol{\beta}}_n$ with a redescending score function ρ_1' [i.e., $\rho_1'(r) \equiv 0$ for $|r| \geq c > 0$], the weights ω_i give the method stability in the presence of outliers. Outlying points will be associated with small weights in (3.2) and (3.3). Note that extreme outliers (those with an associated residual $|r_i| > c\hat{\sigma}_n$) will receive a zero weight, and hence will have no effect at all on the recalculated coefficients. Note that the weights v_i used in recalculating the scale are also decreasing in the absolute value of the residuals and hence outlying points are less influential in the recalculated $\hat{\sigma}_n^*$ as well.

4. Examples. We now illustrate the stability of the inference based on the fast bootstrap on a simple and a multiple linear regression analysis. In both cases we compare the inference obtained using the bootstrap and the fast bootstrap on the same robust regression estimates. These examples simultaneously illustrate the serious effect of the outliers on the inference derived from the bootstrap and the robustness of the fast bootstrap.

4.1. *Belgium international phone calls.* Consider the Belgium international calls dataset [see Rousseeuw and Yohai (1984)]. These data consist of the number of international phone calls (in tens of millions) originated in Belgium between 1950 and 1973. From 1964 to 1969 the observations were mistakenly recorded. Instead of the number of calls, their total duration in minutes was registered. The

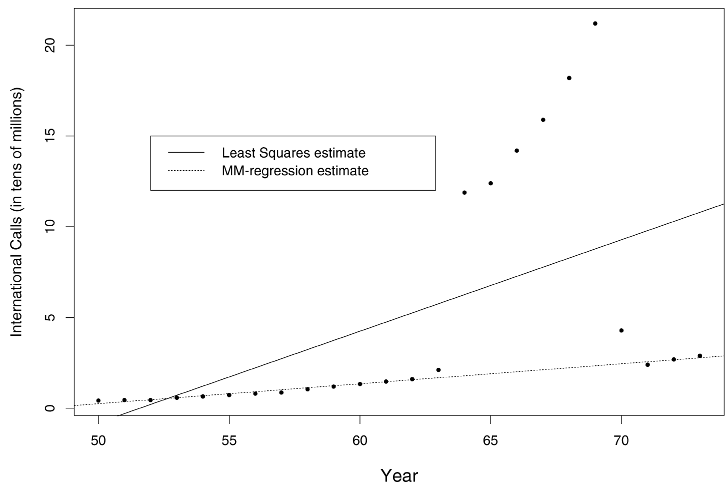


FIG. 1. *Least squares and robust regression fits to the Belgium international phone calls data.*

figure for 1970 is partly contaminated: some calls were recorded with their duration; others were registered according to the old convention. The linear regression model considered in the literature is

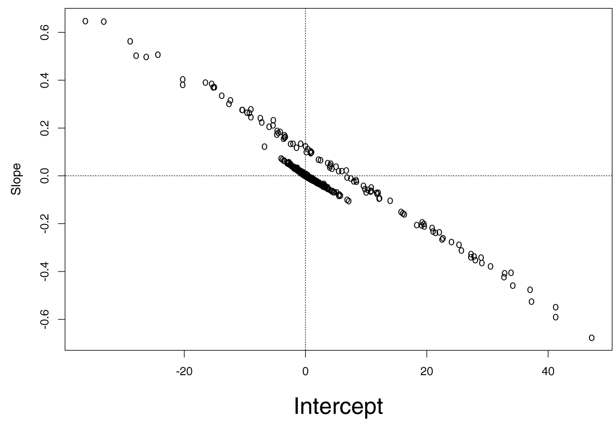
$$\#\text{Calls (in tens of millions)} = \alpha_0 + \beta_0 \text{Year} + \epsilon, \tag{4.1}$$

where α_0 and β_0 are the parameters of interest and the errors ϵ are assumed to be independent and identically distributed with mean 0 and unknown but constant variance σ^2. The MM-regression estimate with an S-scale gives $\hat{\alpha}_0 = -5.23$ and $\hat{\beta}_0 = 0.11$. Figure 1 displays the data with the robust and least squares fits. To obtain confidence intervals for the regression parameters $\boldsymbol{\beta}$, we use the bootstrap and fast bootstrap methods to estimate the distribution of the robust regression estimator. We performed 10,000 bootstrap recalculations. Scatterplots of $\hat{\boldsymbol{\beta}}_u^{R*} - \hat{\boldsymbol{\beta}}_n$ for the fast bootstrap and of $\hat{\boldsymbol{\beta}}_u^{*} - \hat{\boldsymbol{\beta}}_n$ for the bootstrap are presented in Figure 2. We clearly see that the fast bootstrap estimates are more stable. This is reflected in the length of the confidence intervals. Table 1 contains the 95% confidence intervals for the slope and intercept calculated with both the bootstrap and the fast bootstrap.

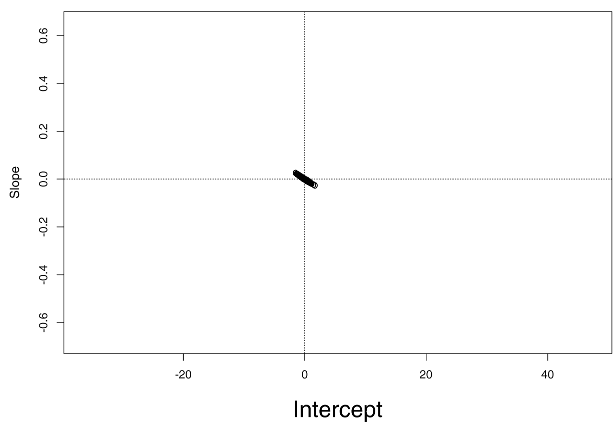
TABLE 1
95% confidence intervals for the regression coefficients of the Belgium international calls data

Coefficient	**Fast bootstrap**	**Classical bootstrap**
Intercept	(−10.32, −3.20)*	(−17.74, 0.35)
Year	(0.08, 0.20)*	(0.00, 0.28)

An asterisk indicates a coefficient significantly different at the 5% level.



(a) Classical bootstrap



(b) Fast bootstrap

FIG. 2. *Comparison of bootstrap and fast bootstrap distribution estimates for the Belgium international phone calls data*—10,000 *bootstrap samples.*

Using the fast bootstrap, we conclude that both regression coefficients are significantly different from 0 at the 5% level. On the other hand, the artificial variability introduced by the outliers in the bootstrap yields longer confidence intervals. As a consequence, using the bootstrap, we conclude that, at the 5% level, there is no significant linear relationship between the response and the predictor variable. The conclusion obtained with the fast bootstrap analysis is intuitively in agreement with the linear trend observed in the scatterplot of the data (see Figure 1).

4.2. *Verbal test score data.* These data contain observations drawn from 20 schools in the United States. They were first studied by Coleman et al. (1966) [see also Mosteller and Tukey (1977) and Rousseeuw and Leroy (1987) pages 79ff]. The data consist of verbal mean test scores for sixth-graders drawn from 20 schools in the Mid-Atlantic and New England states. The explanatory variables are: staff salaries per pupil (Staff Salary), percent age of white-collar fathers (White Collar), socioeconomic status composite deviation (Soc. Status), mean teacher's verbal test score (Teacher Score) and mean mother's educational level (Mother Ed.). We fit a multiple linear regression model to these data to find which variables have a significant effect on the mean verbal test score of the students. We used the classical least squares fit and a 50% breakdown point and 95% efficient MM-regression estimate with score functions in Tukey's family (5.1). Figure 3 contains the plot of the residuals obtained with the least squares and MM-regression estimates. From this plot it is clear that these data contain outliers and that the least squares fit is not appropriate.

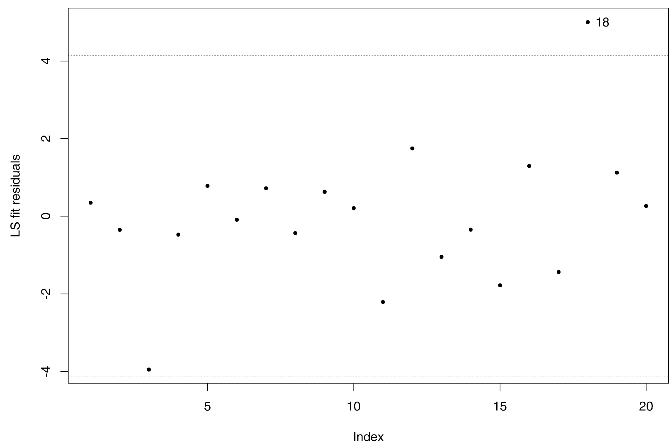
To determine which coefficients are significantly different from 0, we built 95% confidence intervals using both bootstrap and fast bootstrap methods to estimate the distribution of the robust MM-regression estimator. We used 5000 bootstrap samples to estimate the appropriate quantiles of the marginal distributions. The bootstrap calculations required 820 CPU seconds, whereas the fast bootstrap was done in 2.2 CPU seconds. The resulting confidence intervals are displayed in Table 2.

The only significant coefficients using the bootstrap (at the 5% level) are those of Soc. Status and Teacher Score. The confidence intervals constructed with the fast bootstrap indicate that all coefficients are significant at this level. What is most striking in this example is that the conclusions reached using the *bootstrap* are the *same* as those we would have obtained using a *nonrobust least squares analysis*. In other words, simply bootstrapping these highly robust estimates leads to the same qualitative conclusions yielded by the nonrobust least squares fit. The reason for the different behavior of the bootstrap and the fast bootstrap is again the serious effect of the outliers in the bootstrap samples. Also note that the lengths of

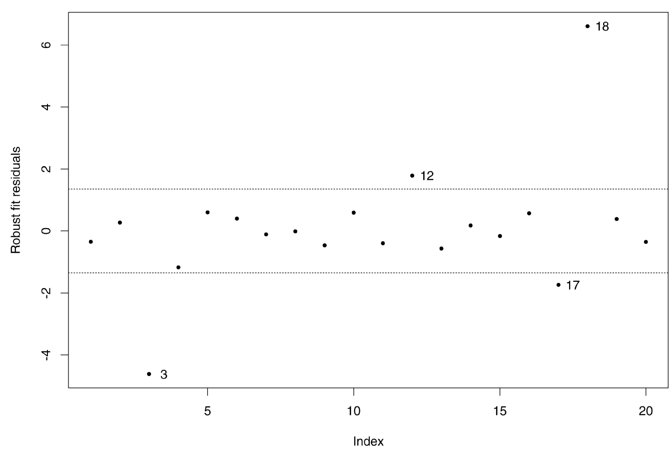
TABLE 2
95% confidence intervals for the regression coefficients of the verbal test score data

Coefficient	**Fast bootstrap**	**Classical bootstrap**
Intercept	(12.37, 31.66)*	(−5.60, 49.99)
Staff Salary	(−2.34, −0.23)*	(−3.75 1.26)
White Collar	(0.03, 0.11)*	(−0.03, 0.17)
Soc. Status	(0.57, 0.69)*	(0.41, 0.81)*
Teacher Score	(0.81, 1.56)*	(0.09, 2.27)*
Mother Ed.	(−4.35, −1.37)*	(−6.79, 0.97)

An asterisk indicates a coefficient significantly different from 0 at the 5% level.



(a) Least squares

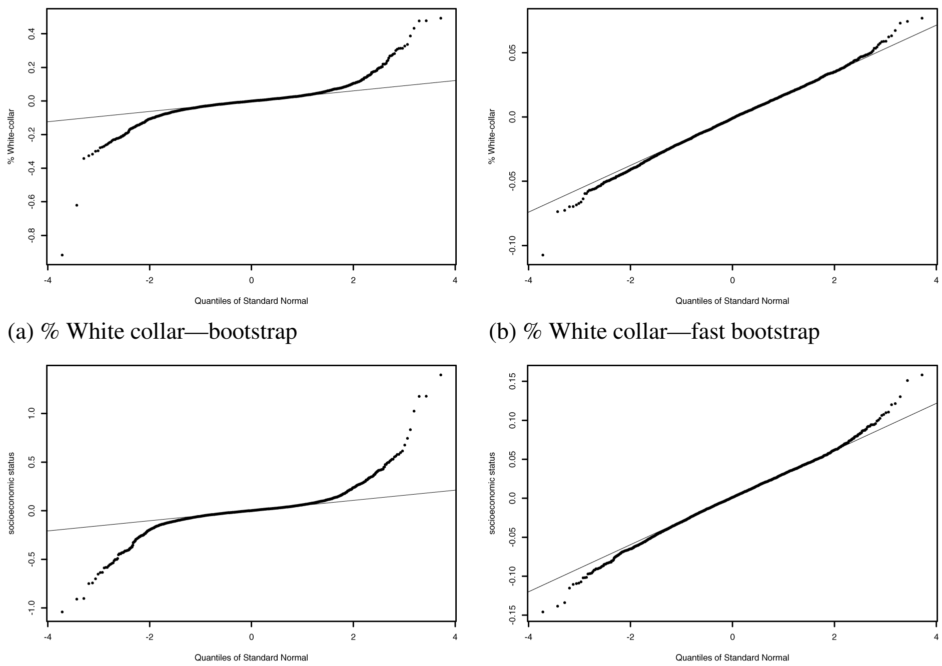


(b) Robust regression

FIG. 3. *Residual plots for the least squares and robust fits to the verbal test score data. The dotted lines indicate* $\pm 2\hat{\sigma}_n$.

the bootstrap confidence intervals are between 2.5 and 4 times longer than those obtained with the fast bootstrap.

To explore the shape of the estimates of the marginal distributions obtained with each method, we used QQ-plots of the marginal bootstrap distributions. Figure 4 contains these plots for two marginal distributions, the other marginal distributions being very similar. As expected, the marginal distributions of the bootstrap have heavier tails than those of the fast bootstrap, resulting in unduly long confidence intervals.



(a) % White collar—bootstrap (b) % White collar—fast bootstrap

(c) Socioeconomic status—bootstrap (d) Socioeconomic status—fast bootstrap

FIG. 4. *QQ-plots of the bootstrap and fast bootstrap marginal distributions for the verbal test score data.*

5. Asymptotic properties of the fast bootstrap. The following theorem shows that the asymptotic distribution of the fast bootstrap is the same as that of the MM-regression estimator.

In the rest of this paper we will assume the following regularity conditions for ρ_0 and ρ_1:

R1. $\rho_0(-u) = \rho_0(u)$ and $\rho_0(-u) = \rho_0(u)$ for all $u \in \mathbb{R}$;
R2. $\rho_0(0) = \rho_1(0) = 0$;
R3. ρ_0 and ρ_1 are continuously differentiable;
R4. $\sup_x \rho_0(x) = \sup_x \rho_1(x) = 1$;
R5. if $\rho_0(u) < 1$ and $0 \le v < u$ then $\rho_0(v) < \rho_0(u)$, and the same holds for ρ_1.

A widely used family of functions ρ that satisfy R1–R5 above was proposed by Beaton and Tukey (1974):

$$\rho(u) = \begin{cases} 3(u/d)^2 - 3(u/d)^4 + (u/d)^6, & \text{if } |u| \le d, \\ 1, & \text{if } |u| > d, \end{cases} \tag{5.1}$$

where $d > 0$ is a fixed constant.

THEOREM 1 (Convergence of the fast bootstrap distribution). *Let ρ_0 and ρ_1 be real functions satisfying* R1–R5. *Assume that they have continuous third* *derivatives. Let* $\hat{\boldsymbol{\beta}}_n$ *be the MM-regression estimator*, $\hat{\sigma}_n$ *the S-scale and* $\tilde{\boldsymbol{\beta}}_n$ *the associated S-regression estimator. Assume that they are consistent*, *that is*, $\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}$, $\hat{\sigma}_n \xrightarrow{P} \sigma$ *and* $\tilde{\boldsymbol{\beta}}_n \xrightarrow{P} \tilde{\boldsymbol{\beta}}$, *where* $\boldsymbol{\beta}$, σ *and* $\tilde{\boldsymbol{\beta}}$ *solve the following equations*:

$$E\big[\rho_1'\big((Y-\mathbf{X}'\boldsymbol{\beta})/\sigma\big)\big]=\mathbf{0},$$
$$E\big[\rho_0\big((Y-\mathbf{X}'\tilde{\boldsymbol{\beta}})/\sigma\big)\big]=b,$$
$$E\big[\rho_0'\big((Y-\mathbf{X}'\tilde{\boldsymbol{\beta}})/\sigma\big)\big]=\mathbf{0}.$$

If the following conditions hold:

1. *The following matrices exist and are finite*:

$$E\big[\rho_1'(r)/r\mathbf{X}\mathbf{X}'\big]^{-1},\ E\big[\rho_0'(r)/r\mathbf{X}\mathbf{X}'\big]^{-1},\ E\big[\rho_1'(r)\mathbf{X}\mathbf{X}'\big],\ E\big[\rho_1'(r)r\mathbf{X}\mathbf{X}'\big],$$
$$E\big[\rho_0''(r)\mathbf{X}\mathbf{X}'\big],\ E\big[\rho_1''(r)\mathbf{X}\mathbf{X}'\big]^{-1},\ E\big[\rho_0''(r)r\mathbf{X}\big],\ E\big[\rho_1''(r)r\mathbf{X}\big];$$

2. $E[\rho_0'(r)r] \neq 0$ *and finite*,
3. $\rho_1'(u)/u$, $\rho_1'(u)/u$, $(\rho_0'(u)-\rho_0''(u)u)/u^2$ *and* $(\rho_1'(u)-\rho_1''(u)u)/u^2$ *are continuous*;

then along almost all sample sequences $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{R*}-\hat{\boldsymbol{\beta}}_n)$ *converges weakly*, *as n goes to* ∞, *to the same limit distribution as* $\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})$.

REMARK 3. Assumption 3 above is satisfied for functions ρ in Tukey's family (5.1).

REMARK 4. Regarding the assumption of consistency of $\hat{\sigma}_n$, $\tilde{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}_n$, Salibian-Barrera (2000) found regularity conditions that suffice to prove consistency and asymptotic distribution of these estimates for any $F \in \mathcal{H}_\epsilon$ [see (2.2)].

6. Robustness properties of the fast bootstrap. We are interested in the robustness properties of the quantile estimates of our fast bootstrap. Let $t \in (0, 1)$ and let q_t be the tth upper quantile of a statistic $\hat{\boldsymbol{\theta}}_n$; that is, q_t satisfies $P[\hat{\boldsymbol{\theta}}_n > q_t] = t$.

Following Singh (1998), we define the upper breakdown point of a quantile estimate $\hat{q}_t$ as the minimum proportion of asymmetric contamination that can drive it over any finite bound.

There are two closely related scenarios in which the quantile estimates based on the fast bootstrap can break down. The first unfavorable situation is when the proportion of outliers in the original data is larger than the breakdown point of the estimate. In this case the estimate may already be unreliable, and so are the inferences we derive from it. The second case is related to the number of outliers appearing in the bootstrap samples. Let τ^* be the expected proportion of bootstrap samples that contain more outliers than the breakdown point of the estimate. In other words, we expect $\tau^* \times 100\%$ of the recalculated $\hat{\boldsymbol{\beta}}_n^*$'s to be unreliable. The estimate $\hat{q}_t$ may be severely affected by the outliers when $\tau^* > t$. The following theorem summarizes this discussion.

The breakdown point of robust regression estimates is related to the geometrical characteristics of the data. In the same way, these characteristics affect the breakdown point of the fast bootstrap quantile estimates. We need the following definition of general position [Rousseeuw and Leroy (1987)].

DEFINITION 1 (General position). We say that k points in $\mathbb{R}^p$ are in general position if no subset of size $p+1$ of them determines an affine subspace of dimension p. In other words, for every subset $\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p+1}}$, $1 \le i_j \le k$, $i_j \ne i_l$ if $j \ne l$, there are no vector $\mathbf{v}_0 \in \mathbb{R}^p \setminus \{\mathbf{0}\}$ and scalar $\alpha \in \mathbb{R}$ such that

$$\mathbf{x}_{i_j}' \mathbf{v}_0 = \alpha \quad \text{for } j = 1, \dots, p+1.$$

The main result of this section is the following theorem that establishes the breakdown point of the quantile estimates based on the fast bootstrap.

THEOREM 2 (Breakdown point of the fast bootstrap quantiles for the regression model). *Let* $(y_1, \mathbf{x}_1')', \dots, (y_n, \mathbf{x}_n')' \in \mathbb{R}^{p+1}$ *be a random sample following the linear model* (2.1). *Assume that the explanatory variables* $\mathbf{x}_1, \dots, \mathbf{x}_n$ *in* $\mathbb{R}^p$ *are in general position* (*see Definition* 1). *Let* $\hat{\boldsymbol{\beta}}_n$ *be an MM-regression estimate and let* ε^* *be its breakdown point. Then the breakdown point of the* t*th fast bootstrap quantile estimate of the regression parameters* β_j, $j = 1, \dots, p$, *is given by* $\min(\varepsilon^*, \varepsilon_R)$, *where* ε_R *satisfies*

$$\varepsilon_R = \inf\{\delta \in [0,1] : P[\text{Binomial}(n, 1-\delta) < p] \ge t\}. \tag{6.1}$$

It is easy to see that formula (6.1) is equivalent to

$$\varepsilon_R = \inf\{\delta \in [0,1] : P[\text{Binomial}(n, \delta) \ge n - p] \ge t\}. \tag{6.2}$$

Singh (1998) obtained the following formula for the upper breakdown point of the bootstrap estimate $\hat{q}_t$ of q_t:

$$\varepsilon_C = \inf\{\delta \in [0,1] : P[\text{Binomial}(n, \delta) \ge [\varepsilon^* n]] \ge t\}, \tag{6.3}$$

where $[x]$ denotes the smallest integer larger than or equal to x and ε^* is the breakdown point of the estimate being bootstrapped. Since $[\varepsilon^* n] \le [n/2] < n - 1$ for $n > 3$, we immediately see from (6.2) and (6.3) that $\varepsilon_C < \varepsilon_R$. Table 3 compares ε_C and ε_R for different sample sizes (n) and number of explanatory variables (p). We considered an MM-regression estimate with 50% breakdown point and 95% efficiency when the data are normally distributed. We compared the quantiles needed to construct 90%, 95% and 99% confidence intervals.

TABLE 3
Comparison of quantile upper breakdown points for MM-regression estimates with 50% breakdown point

		Fast bootstrap			Classical bootstrap		
p	n	$\hat{q}_{0.005}$	$\hat{q}_{0.025}$	$\hat{q}_{0.05}$	$\hat{q}_{0.005}$	$\hat{q}_{0.025}$	$\hat{q}_{0.05}$
1	10	0.500	0.500	0.500	0.191	0.262	0.304
	20	0.500	0.500	0.500	0.257	0.315	0.347
	30	0.500	0.500	0.500	0.293	0.343	0.370
2	10	0.456	0.500	0.500	0.128	0.187	0.222
	20	0.500	0.500	0.500	0.217	0.272	0.302
	30	0.500	0.500	0.500	0.265	0.313	0.339
5	10	0.191	0.262	0.304	0.011	0.025	0.036
	20	0.500	0.500	0.500	0.114	0.154	0.177
	30	0.500	0.500	0.500	0.185	0.226	0.249
10	20	0.257	0.315	0.347	0.005	0.012	0.018
	50	0.500	0.500	0.500	0.180	0.212	0.230
	100	0.500	0.500	0.500	0.294	0.322	0.336

Note that the only cases where the upper breakdown point for the fast bootstrap quantiles is significantly smaller than the breakdown point of the regression estimate (50%) are $n = 10,\ p = 5$ and $n = 20,\ p = 10$. These cases are not of interest from a practical point of view due to the extremely large dimension of the model relative to the number of observations available. Also note that our upper breakdown points are notably larger than those of the bootstrap quantiles estimate.

7. Simulation results. In this section we report the results of a Monte Carlo study on the finite sample properties of confidence intervals for the parameters $\boldsymbol{\beta}$ in the linear regression model (2.1). In what follows $\mathbf{h}_{(j)}$ denotes the jth coordinate of the vector $\mathbf{h}$.

We compare two methods for building confidence intervals for the coefficients $\boldsymbol{\beta}_{(j)},\ j = 1, \dots, p$. The first approach is to approximate the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ by its normal asymptotic distribution. We build an asymptotic $1 - \alpha$ confidence interval for $\boldsymbol{\beta}_{(j)}$ of the form $(\hat{\boldsymbol{\beta}}_{n(j)} - z_{\alpha/2}\hat{V}_n, \hat{\boldsymbol{\beta}}_{n(j)} + z_{\alpha/2}\hat{V}_n)$, where $\hat{V}_n^2$ is an estimate of the asymptotic variance of $\hat{\boldsymbol{\beta}}_{n(j)}$ and z_α is the quantile that leaves the area equal to α to its right under a standard normal curve. Let $\boldsymbol{\Sigma}(F, \boldsymbol{\beta}(F), \boldsymbol{\sigma}(F))$ be the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_n$. We use $\hat{V}_n^2$ equal to the corresponding diagonal element of the empirical version of the above asymptotic covariance matrix [namely, $\boldsymbol{\Sigma}(F_n, \hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n)$]. We refer to this method as "empirical asymptotic variance."

The second approach is based on directly estimating the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. In this case we focus on constructing confidence intervals for $\boldsymbol{\beta}_{(j)}$ of the form $(\mathcal{Z}^n_{1-\alpha/2}, \mathcal{Z}^n_{\alpha/2})$, where $\mathcal{Z}^n_\eta$ satisfies

$$P\big[\sqrt{n}(\hat{\boldsymbol\beta}_{n(j)} - \boldsymbol\beta_{(j)}) > \mathcal{Z}^n_\eta\big] = \eta \qquad \text{for } \eta \in (0, 1).$$

We can use bootstrap methods to obtain estimates of $\mathcal{Z}^n_\eta$. The basic idea is to use the empirical distribution of the recalculated $\hat{\boldsymbol\beta}^*_{n(j)}$'s to obtain estimated quantiles $\hat{\mathcal{Z}}^n_\eta$ [see, e.g., Davison and Hinkley (1997), page 18]. Note that with this method we do not use the symmetry assumption that underlies the asymptotic normal approximation in the previous approach. We generated many recalculated $\hat{\boldsymbol\beta}^{R*}_n - \hat{\boldsymbol\beta}_n$ with the fast bootstrap and used the empirical distribution of each projection to obtain estimates of the distribution of $\hat{\boldsymbol\beta}_{n(j)} - \boldsymbol\beta_{(j)}$ for each coordinate $j = 1, \dots, p$. With this empirical distribution we obtained estimates $\hat{\mathcal{Z}}^n_\eta$ of the quantiles needed to build the confidence intervals. In this context, the bootstrap demands so much computer time that it becomes almost unfeasible; hence, we did not include it in our study.

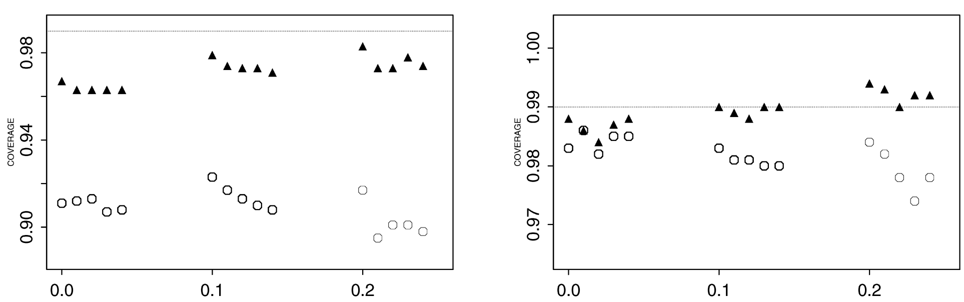
We considered sample sizes $n = 30$, 50 and 100 with $p = 2$ and $p = 5$ explanatory variables. These independent variables included an intercept: $x_1 \equiv 1$ and $x_i \sim N(0, 1)$ for $i = 2, \dots, p$. Finally, the errors followed the gross-error contamination model with distributions $F_\epsilon = (1 - \epsilon)\Phi(x) + \epsilon V(x)$, where $V(x) = 0.5\Phi((x - x_0)/0.1) + 0.5\Phi((x + x_0)/0.1)$ and $\Phi(x)$ denotes the standard normal cumulative distribution function. We used $\epsilon = 0.00$, 0.10 and 0.20. The contamination point x_0 was set at 3, 4 and 10. Here we report the results obtained for $x_0 = 4$, the others being very similar.

We generated 5000 datasets from the above distributions and built 99% confidence intervals for the parameters of the model. We used MM-regression estimates obtained with $\psi = \rho'_{4.685}$ in Tukey's family (5.1). The S-scale was obtained with $\rho_{1.54764}$ also in Tukey's family. This choice yields estimates with simultaneous 50% breakdown point and 95% efficiency when the data are normally distributed.

Table 4 tabulates the results of the simulation for $p = 5$. It is easier to see the difference between these methods by looking at Figure 5. These pictures show at a glance that the levels obtained with the fast bootstrap are better than the ones yielded by the empirical asymptotic variance (AV) estimate. Both methods are very close only for the case of $n = 100$ and $\epsilon = 0.00$. In all the other cases the fast bootstrap yields notably better coverage levels. The behavior for the case $n = 100$ and $\epsilon = 0.00$ is naturally expected because both methods are asymptotically equivalent (and correct), and hence will behave similarly for large sample sizes. Note that for $n = 100$ and $\epsilon = 0.20$, however, the empirical asymptotic variance method compares unfavorably with the fast bootstrap. The reason for this seems to be that the empirical asymptotic variance formula is numerically unstable (especially for contaminated datasets). The fast bootstrap being more stable shows a better performance.

TABLE 4
Coverage and length of 99% confidence intervals for the linear regression model, $p = 5$

n	ϵ	Parameter	Fast bootstrap	Empirical AV
30	0.00	β_0	0.967 (1.340)	0.911 (0.901)
		β_1	0.963 (1.408)	0.912 (0.924)
		β_2	0.963 (1.410)	0.913 (0.923)
		β_3	0.963 (1.417)	0.907 (0.923)
		β_4	0.963 (1.395)	0.908 (0.924)
		σ	0.993 (2.165)	0.992 (2.167)
30	0.20	β_0	0.983 (2.523)	0.917 (1.393)
		β_1	0.973 (2.641)	0.895 (1.429)
		β_2	0.973 (2.610)	0.901 (1.424)
		β_3	0.978 (2.688)	0.901 (1.427)
		β_4	0.974 (2.707)	0.898 (1.427)
		σ	0.995 (3.455)	0.995 (3.456)
100	0.00	β_0	0.988 (0.555)	0.983 (0.526)
		β_1	0.986 (0.563)	0.986 (0.530)
		β_2	0.984 (0.562)	0.982 (0.530)
		β_3	0.987 (0.564)	0.985 (0.530)
		β_4	0.988 (0.562)	0.985 (0.531)
		σ	0.995 (0.659)	0.995 (0.660)
100	0.20	β_0	0.994 (1.050)	0.984(0.903)
		β_1	0.993 (1.090)	0.982 (0.911)
		β_2	0.990 (1.095)	0.978 (0.910)
		β_3	0.992 (1.090)	0.974 (0.910)
		β_4	0.992 (1.093)	0.978 (0.910)
		σ	0.994 (1.024)	0.994 (1.025)



(a) $n = 30$ (b) $n = 100$

FIG. 5. *Average coverage of 99% confidence intervals for the linear regression model with $p = 5$. Solid triangles are levels of the confidence intervals for the intercept and the coefficients of $\mathbf{x}_1, \dots, \mathbf{x}_4$ calculated with the fast bootstrap; circles represent the corresponding levels for the confidence intervals obtained with the empirical asymptotic variance estimate. Across the horizontal axis, the three groups correspond to $\epsilon = 0.0$, 0.1 and 0.2, respectively. The horizontal line indicates the nominal level.*

8. Some concluding remarks. We have presented the "fast bootstrap" which provides an alternative to the bootstrap for estimating the distribution of robust regression estimates. Unlike the bootstrap, the fast bootstrap is fast to compute and resistant to outliers in the data. In particular, we have built confidence intervals for the coefficients of the linear model (2.1) based on MM-regression estimates and have used the fast bootstrap to estimate their endpoints. We showed that these confidence intervals are computationally feasible and have good coverage properties across different sample sizes and amounts of contamination.

We have studied the case of random explanatory variables. In principle, we could also apply our method to the case of fixed design as follows. Let $e_j = y_j - \hat{\boldsymbol{\beta}}_n' \mathbf{x}_j$, $1 \le j \le n$, be the residuals of the MM-estimate. The bootstrapped y_i^*'s are

$$y_i^* = \hat{\boldsymbol{\beta}}_n' \mathbf{x}_i + e_i^*,$$

where e_i^*, $1 \le i \le n$, is a random sample from the residuals. Now $\hat{\boldsymbol{\beta}}_n^*$ and $\hat{\sigma}_n^*$ are defined by

$$\hat{\boldsymbol{\beta}}_n^* = \left[\sum_{i=1}^n \omega_i^* \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \omega_i^* \mathbf{x}_i y_i^*, \tag{8.1}$$

$$\hat{\sigma}_n^* = \sum_{i=1}^n v_i^* \big(y_i^* - \tilde{\boldsymbol{\beta}}_n' \mathbf{x}_i \big), \tag{8.2}$$

where $\omega_i^* = \rho_1' \left(r_i^* / \hat{\sigma}_n \right) / r_i^*, v_i^* = \hat{\sigma}_n \rho_0(\tilde{r}_i^* / \hat{\sigma}_n)/(n \;\, b \;\, \tilde{r}_i^*), r_i^* = y_i^* - \hat{\boldsymbol{\beta}}_n' \mathbf{x}_i$ and $\tilde{r}_i^* = y_i^* - \tilde{\boldsymbol{\beta}}_n' \mathbf{x}_i$ for $1 \le i \le n$. The correction factors $\mathbf{A}_n$, $\mathbf{v}_n$ and a_n are defined as before, and so is $\hat{\boldsymbol{\beta}}_n^{R*} - \hat{\boldsymbol{\beta}}_n$. The main difference lies in the proposed resampling procedure. Following Freedman (1981), we intend it to best resemble the underlying process generating the data. Under suitable regularity conditions (which include that the robust estimate itself be asymptotically normally distributed), the consistency of the fast bootstrap for fixed designs can be obtained using a similar proof to the one provided for Theorem 1.

The fast bootstrap could also be applied to other types of robust regression estimates that can be represented as the solution of a smooth fixed-point equation $\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n) = \hat{\boldsymbol{\theta}}_n$. To obtain asymptotically correct distribution estimates with this method, it is paramount that $\hat{\boldsymbol{\theta}}_n$ above includes the scale estimate. For example, in the case of GM estimates given as the solution of

$$\sum_{i=1}^n \eta \left(\frac{y_i - \hat{\boldsymbol{\beta}}_n' \mathbf{x}_i}{\hat{\sigma}_n}, \mathbf{x}_i \right) \mathbf{x}_i = \mathbf{0}$$

for a certain function η: $\mathbb{R}^p \times \mathbb{R} \to \mathbb{R}_+$ [see Hampel, Ronchetti, Rousseeuw and Stahel (1986), Chapter 6], we have to consider the equations that define $\hat{\sigma}_n$. If $\hat{\sigma}_n$ is an S-estimate, we have

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{y_i-\tilde{\boldsymbol{\beta}}_n'\mathbf{x}_i}{\hat{\sigma}_n}\right)=b,$$

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0'\left(\frac{y_i-\tilde{\boldsymbol{\beta}}_n'\mathbf{x}_i}{\hat{\sigma}_n}\right)\mathbf{x}_i=\mathbf{0}.$$

Then $\hat{\boldsymbol{\theta}}_n=(\hat{\boldsymbol{\beta}}_n',\hat{\sigma}_n,\tilde{\boldsymbol{\beta}}_n')'\in\mathbb{R}^{2p+1}$ and the function $\mathbf{g}_n$: $\mathbb{R}^{2p+1}\to\mathbb{R}^{2p+1}$ includes the corresponding three equations above. The details regarding theoretical and numerical performance of the fast bootstrap for these estimates deserve further study.

APPENDIX

Proofs. The following lemma is needed for the proof of Theorem 2.

LEMMA 1. *Let $(y_1,\mathbf{x}_1')',\dots,(y_n,\mathbf{x}_n')'$ be $n\ge p$ points in $\mathbb{R}^p$ such that if*

$$\mathbf{X}_n=\begin{bmatrix}\mathbf{x}_1'\\ \vdots\\ \mathbf{x}_n'\end{bmatrix}, \tag{A.1}$$

then $\mathbf{X}_n'\mathbf{X}_n$ has full rank. For a given $(y_{n+1},\mathbf{x}_{n+1}')'$ let $\hat{\boldsymbol{\beta}}_{n+1}$ be the least squares regression coefficients determined by the $n+1$ points. There exists a finite constant K such that $\|\hat{\boldsymbol{\beta}}_{n+1}\|\le K$ for any $(y_{n+1},\mathbf{x}_{n+1}')'$ with $|y_{n+1}|\le c$. (The constant K only depends on the first n points and on the constant c.)

PROOF. We will show that the regression parameters $\hat{\boldsymbol{\beta}}_{n+1}$ obtained when adding a new point $(y_{n+1},\mathbf{x}_{n+1})$ are bounded for any $\mathbf{x}_{n+1}$ if y_{n+1} is bounded. Let $\mathbf{X}_n\in\mathbb{R}^{n\times p}$ be the design matrix in (A.1). Note that $\mathbf{X}_n$ has rank p by hypothesis. As a consequence, both $(\mathbf{X}_n'\mathbf{X}_n)$ and its inverse are positive definite. Let $\mathbf{C}$ be a nonsingular matrix in $\mathbb{R}^{p\times p}$ and let $\mathbf{h}\in\mathbb{R}^p$. Use the following formula [see, e.g., Seber (1984), page 519]

$$(\mathbf{C}+\mathbf{h}\mathbf{h}')^{-1}=\mathbf{C}^{-1}-\mathbf{C}^{-1}\mathbf{h}\mathbf{h}'\mathbf{C}^{-1}(1+\mathbf{h}'\mathbf{C}^{-1}\mathbf{h})^{-1}$$

to obtain

$$\hat{\boldsymbol{\beta}}_{n+1}=\left[\mathbf{I}-\frac{\mathbf{V}\mathbf{x}_{n+1}\mathbf{x}_{n+1}'}{1+\mathbf{x}_{n+1}'\mathbf{V}\mathbf{x}_{n+1}}\right]\hat{\boldsymbol{\beta}}_n+\left[\mathbf{V}-\frac{\mathbf{V}\mathbf{x}_{n+1}\mathbf{x}_{n+1}'\mathbf{V}}{1+\mathbf{x}_{n+1}'\mathbf{V}\mathbf{x}_{n+1}}\right]\mathbf{x}_{n+1}y_{n+1},$$

where $\mathbf{V}=(\mathbf{X}_n'\mathbf{X}_n)^{-1}$ is positive definite and $(y_{n+1},\mathbf{x}_{n+1}')'$ is the new point to be added to the regression. To simplify the notation, let

$$\mathbf{u}=\mathbf{x}_{n+1},\qquad \mathbf{A}=\mathbf{I}-\frac{\mathbf{V}\mathbf{u}\mathbf{u}'}{1+\mathbf{u}'\mathbf{V}\mathbf{u}},\qquad \mathbf{B}=\mathbf{V}-\frac{\mathbf{V}\mathbf{u}\mathbf{u}'\mathbf{V}}{1+\mathbf{u}'\mathbf{V}\mathbf{u}}.$$

The last equation can then be written as

$$\boldsymbol{\beta}_{n+1} = \mathbf{A}\boldsymbol{\beta}_n + \mathbf{Bu}y_{n+1}.$$

First, we will show that every entry in $\mathbf{A}$ is bounded for $\|\mathbf{u}\| \to \infty$. The (i, j) element is given by

$$\begin{aligned}\mathbf{A}_{(i,j)} &= \delta_{i,j} - \frac{u_j \sum_k v_{ik} u_k}{1 + \sum_k \sum_l v_{kl} u_k u_l} \\ &= \delta_{i,j} - \frac{v_{ij} u_j^2 + u_j (\sum_{k \neq j} v_{i\,k} u_k)}{1 + \sum_k \sum_l v_{kl} u_k u_l}.\end{aligned}$$

It is easy to see (e.g., by dividing both the numerator and the denominator by $\|\mathbf{u}\|^2$) that the denominator has the same order as the numerator, so that the fraction will remain bounded as $\|\mathbf{u}\| \to \infty$. Note that the denominator is bounded away from 0, so that the whole expression is bounded for any $\mathbf{u}$. We now show that the rth element of $\mathbf{Bu}$ goes to 0 as $\|\mathbf{u}\| \to \infty$. Note that

$$\mathbf{Bu} = \frac{\mathbf{Vu}}{1 + \mathbf{u}'\mathbf{Vu}}.$$

The rth element is then

$$\frac{\sum_i v_{ri} u_i}{1 + \sum_{ij} v_{ij} u_i u_j}.$$

Divide both the numerator and the denominator by $\|\mathbf{u}\|^2$ and use that

$$\frac{|\mathbf{u}_j|}{\|\mathbf{u}\|} \le 1 \qquad \text{for } 1 \le j \le p \tag{A.2}$$

to conclude that the denominator is bounded when $\|\mathbf{u}\| \to \infty$ and that the numerator goes to 0 when $\|\mathbf{u}\| \to \infty$. The latter can be seen noting that (A.2) implies

$$\frac{|\mathbf{u}_j|}{\|\mathbf{u}\|^2} \to 0 \qquad \text{for } 1 \le j \le p.$$

□

PROOF OF THEOREM 2. Let $(y_1, \mathbf{x}_1')', \dots, (y_n, \mathbf{x}_n')' \in \mathbb{R}^{p+1}$ be n observations following model (2.1). We assume that $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ are in general position (see Definition 1). This assumption guarantees that any subset of size p of them will determine a bounded least squares estimate.

We assume that there is a certain proportion of observations that do not necessarily follow the linear regression model (2.1). We will show that any bootstrap sample that contains at least p points that are not outliers yields a bounded $\hat{\boldsymbol{\beta}}_n^{R*}$. It follows that the only samples that can produce unbounded fast bootstrap coefficients are those that contain at most $p - 1$ points that are not outliers. The fast bootstrap $\hat{\boldsymbol{\beta}}_n^{R*}$ is given by

$$\hat{\boldsymbol{\beta}}_n^{R*} = \mathbf{M}_n(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) + \mathbf{d}_n(\hat{\sigma}_n^* - \hat{\sigma}_n).$$

Note that the matrix $\mathbf{M}_n$ and the vector $\mathbf{d}_n$ are not recalculated with each bootstrap sample, and as long as the robust regression estimate $\hat{\boldsymbol{\beta}}_n$ does not break down, they remain bounded. It is also easy to see that $\hat{\sigma}_n^*$ also remains bounded for any bootstrap sample. Hence, the problem becomes determining under which circumstances $\hat{\boldsymbol{\beta}}_n^*$ can be driven beyond any finite bound. Recall that

$$\hat{\boldsymbol{\beta}}_n^* = \left[\sum_{i=1}^n \omega_i^* \mathbf{x}_i^* \mathbf{x}_i^{*\prime}\right]^{-1} \sum_{i=1}^n \omega_i^* \mathbf{x}_i^* \, y_i^*,$$

where the weights $\omega_i^* = \rho_1'(r_i^*/\hat{\sigma}_n)/r_i^*$ are bounded. The above expression can be rewritten as

$$\hat{\boldsymbol{\beta}}_n^* = \left[\sum_{i=1}^n \tilde{\mathbf{x}}_i^* \tilde{\mathbf{x}}_i^{*\prime}\right]^{-1} \sum_{i=1}^n \tilde{\mathbf{x}}_i^* \tilde{y}_i^*,$$

where $\tilde{\mathbf{x}}_i^* = \sqrt{\omega_i^*}\mathbf{x}_i^*$ and $\tilde{y}_i^* = \sqrt{\omega_i^*} y_i^*$. We consider the case of having at least p data points that are not outliers. It is enough to have a bound on the effect of one outlier and that that bound does not depend on the outlier. In what follows we show how to obtain such a bound. To simplify the notation, we use the same symbols $\mathbf{x}_i$ and y_i for the weighted points $\tilde{\mathbf{x}}_i$ and $\tilde{y}_i$.

Let $(y_1, \mathbf{x}_1')', \ldots, (y_n, \mathbf{x}_n')'$ be a bootstrap sample of $n \geq p$ good data points and let $(y_{n+1}, \mathbf{x}_{n+1}')'$ be an arbitrary outlier included in this sample. Let $\hat{\boldsymbol{\beta}}_n$ be the MM-estimate based on the full data. Without loss of generality, assume that $\hat{\boldsymbol{\beta}}_n = \mathbf{0} \in \mathbb{R}^p$. The data can always be transformed to satisfy this assumption. In particular, if

$$\tilde{y}_i = y_i - \hat{\boldsymbol{\beta}}_n' \mathbf{x}_i, \qquad i = 1, \ldots, n,$$

then the points $(\tilde{y}_1, \mathbf{x}_1')', \ldots, (\tilde{y}_n, \mathbf{x}_n')'$ have a zero regression estimate.

We now show that the outlier $(y_{n+1}, \mathbf{x}_{n+1}')'$ will only have an effect on $\hat{\boldsymbol{\beta}}_{n+1}^*$ for a bounded range of y_{n+1}. Let $c > 0$ be the constant of the function ψ_c used for the MM-estimate in (2.3) and let $\sigma_n^+ = \sup \hat{\sigma}_n$ be the largest possible value of $\hat{\sigma}_n$ for a sample of size n. Any point $(y_{n+1}, \mathbf{x}_{n+1}')'$ satisfying $|y_{n+1}| > c\sigma_n^+$ has zero weight in the fast bootstrap recalculations. Hence, it is not possible to upset $\hat{\boldsymbol{\beta}}_{n+1}^*$ with this type of contamination. In what follows we consider the case $|y_{n+1}| \leq \sigma_n^+ c$. Lemma 1 gives a bound for the effect of $(y_{n+1}, \mathbf{x}_{n+1}')'$ on $\hat{\boldsymbol{\beta}}_{n+1}^*$. This bound only depends on the first n pairs.

Given a bootstrap sample of size n, assume that the first k observations are "good" and the remaining $n - k$ are arbitrary outliers. Applying Lemma 1 $n - k$ times, we see that the new $\hat{\boldsymbol{\beta}}_{n+1}^*$ can only be modified by a finite amount. This amount depends on the k first observations of this bootstrap sample, but it does not depend on the values of the $n - k$ outliers. Considering all the possible bootstrap samples that contain at least p points that are not outliers, we find a bound that only depends on the original dataset. To drive the tth fast bootstrap quantile estimate above any bound, we need to have at least $t\%$ of the bootstrap samples containing less than p "good" points. The proportion ϵ of outliers in the original sample should then satisfy $P[\text{Binomial}\,(n, 1-\epsilon) < p] \geq t$. □

LEMMA 2 [Serfling (1980), page 253]. *Let $X_1, \dots, X_n$ be a sequence of independent identically distributed random variables and let $g(x,t)$: $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuous in t uniformly on $x \in \mathbb{R}$. Let θ_n be a sequence of random variables such that $\theta_n \xrightarrow[n\to\infty]{P} \theta$, a constant. Then*

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i, \theta_n) \xrightarrow[n\to\infty]{P} E\big[g(X,\theta)\big]. \tag{A.3}$$

If $\theta_n \xrightarrow[n\to\infty]{a.s.} \theta$, then (A.3) *holds a.s. as well.*

PROOF OF THEOREM 1. First, note that the estimates $\hat{\boldsymbol\beta}_n$, $\hat\sigma_n$ and $\tilde{\boldsymbol\beta}_n$ satisfy the following equations:

$$\frac{1}{n}\sum_{i=1}^{n} \rho_1'\biggl(\frac{r_i(\hat{\boldsymbol\beta}_n)}{\hat\sigma_n}\biggr)\mathbf{x}_i = \mathbf{0},$$

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\biggl(\frac{r_i(\tilde{\boldsymbol\beta}_n)}{\hat\sigma_n}\biggr) = b,$$

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0'\biggl(\frac{r_i(\tilde{\boldsymbol\beta}_n)}{\hat\sigma_n}\biggr)\mathbf{x}_i = \mathbf{0}.$$

Simple calculations yield the following reweighted version of the estimates:

$$\begin{aligned} \hat{\boldsymbol\beta}_n &= \mathbf{A}_n(\hat{\boldsymbol\beta}_n, \hat\sigma_n)^{-1}\, \mathbf{v}_n(\hat{\boldsymbol\beta}_n, \hat\sigma_n), \\ \hat\sigma_n &= \hat\sigma_n u_n(\tilde{\boldsymbol\beta}_n, \hat\sigma_n), \\ \tilde{\boldsymbol\beta}_n &= \mathbf{B}_n(\tilde{\boldsymbol\beta}_n, \hat\sigma_n)^{-1}\mathbf{w}_n(\tilde{\boldsymbol\beta}_n, \hat\sigma_n), \end{aligned} \tag{A.4}$$

where

$$\mathbf{A}_n(\boldsymbol\beta_1, \sigma) = \frac{1}{n}\sum_{i=1}^{n} \frac{\rho_1'(r_i/\sigma)}{r_i}\mathbf{x}_i\mathbf{x}_i',$$

$$\mathbf{v}_n(\boldsymbol\beta_1, \sigma) = \frac{1}{n}\sum_{i=1}^{n} \frac{\rho_1'(r_i/\sigma)}{r_i} y_i\mathbf{x}_i,$$

$$u_n(\boldsymbol{\beta}_2, \sigma) = \sum_{i=1}^{n} \frac{\rho_0(\tilde{r}_i/\sigma)}{nb\tilde{r}_i} \tilde{r}_i,$$

$$\mathbf{B}_n(\boldsymbol{\beta}_2, \sigma) = \frac{1}{n} \sum_{i=1}^{n} \frac{\rho_0'(\tilde{r}_i/\sigma)}{\tilde{r}_i} \mathbf{x}_i \mathbf{x}_i',$$

$$\mathbf{w}_n(\boldsymbol{\beta}_2, \sigma) = \frac{1}{n} \sum_{i=1}^{n} \frac{\rho_0'(\tilde{r}_i/\sigma)}{\tilde{r}_i} y_i \mathbf{x}_i.$$

Equations (A.4) can be expressed as the fixed point of a conveniently chosen function. Consider $\mathbf{f}$: $\mathbb{R}^{2p+1} \to \mathbb{R}^{2p+1}$ defined for $\boldsymbol{\beta}_1 \in \mathbb{R}^p$, $\sigma \in \mathbb{R}$ and $\boldsymbol{\beta}_2 \in \mathbb{R}^p$ by

$$\mathbf{f}(\boldsymbol{\beta}_1, \sigma, \boldsymbol{\beta}_2) = \begin{pmatrix} \mathbf{A}_n(\boldsymbol{\beta}_1, \sigma)^{-1} \mathbf{v}_n(\boldsymbol{\beta}_1, \sigma) \\ \sigma u_n(\boldsymbol{\beta}_2, \sigma) \\ \mathbf{B}_n(\boldsymbol{\beta}_2, \sigma)^{-1} \mathbf{w}_n(\boldsymbol{\beta}_2, \sigma) \end{pmatrix}.$$

To simplify the notation, we do not explicitly indicate the dependence of $\mathbf{f}$ on n. We have

$$\mathbf{f}(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n, \tilde{\boldsymbol{\beta}}_n) = (\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n, \tilde{\boldsymbol{\beta}}_n)'.$$

Using the differentiability of ρ_0 and ρ_1, we can calculate a Taylor expansion of $\mathbf{f}$ about the limiting values of the estimates $(\boldsymbol{\beta}, \sigma, \tilde{\boldsymbol{\beta}})$:

$$\text{(A.5)} \qquad \begin{pmatrix} \hat{\boldsymbol{\beta}}_n \\ \hat{\sigma}_n \\ \tilde{\boldsymbol{\beta}}_n \end{pmatrix} = \mathbf{f}(\boldsymbol{\beta}, \sigma, \tilde{\boldsymbol{\beta}}) + \nabla \mathbf{f}(\boldsymbol{\beta}, \sigma, \tilde{\boldsymbol{\beta}}) \begin{pmatrix} \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \\ \hat{\sigma}_n - \sigma \\ \tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}} \end{pmatrix} + R_n,$$

where R_n is the remainder term and $\nabla \mathbf{f}(\cdot) \in \mathbb{R}^{(2p+1)\times(2p+1)}$ is the matrix of partial derivatives,

$$\begin{array}{c|ccc} & p & 1 & p \\ \hline p & \partial[\mathbf{A}_n^{-1}\mathbf{v}_n]/\partial\boldsymbol{\beta} & \partial[\mathbf{A}_n^{-1}\mathbf{v}_n]/\partial\sigma & \partial[\mathbf{A}_n^{-1}\mathbf{v}_n]/\partial\tilde{\boldsymbol{\beta}} \\ 1 & \partial[\sigma u_n]t/\partial\boldsymbol{\beta} & \partial[\sigma u_n]/\partial\sigma & \partial[\sigma u_n]/\partial\tilde{\boldsymbol{\beta}} \\ p & \partial[\mathbf{B}_n^{-1}\mathbf{w}_n]/\partial\boldsymbol{\beta} & \partial[\mathbf{B}_n^{-1}\ \mathbf{w}_n]/\partial\sigma & \partial[\mathbf{B}_n^{-1}\mathbf{w}_n]/\partial\tilde{\boldsymbol{\beta}} \end{array}$$

Tedious but straightforward calculations show that each entry in R_n is a linear combination of quadratic forms $\mathbf{x}_n' H_n \mathbf{x}_n$, where $\mathbf{x}_n = \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$ or $\mathbf{x}_n = \hat{\sigma}_n - \sigma$ or $\mathbf{x}_n = \tilde{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}$. Note that $\|\mathbf{x}_n\| = O_P(1/\sqrt{n})$. The matrix H_n is a combination of products of matrices of the form $\frac{1}{n}\sum(\rho_s^{(k)}(r_i)/r_i^l)\mathbf{x}_i\mathbf{x}_i'$, with $k = 0, 1, 2, 3$, $s = 0, 1$, and $l = 0, 1, 2$. The continuity of the derivatives of ρ_0 and ρ_1 together with assumptions 1 and 3 and Lemma 2 shows that $\|H_n\| = O_P(1)$. We have $|\mathbf{x}_n' H_n \mathbf{x}_n| = o_P(1/\sqrt{n})$. Hence, $\|R_n\| = o_p(1/\sqrt{n})$ in (A.5).

To simplify the notation, let $\boldsymbol{\tau}_n = (\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n, \tilde{\boldsymbol{\beta}}_n)'$ and $\boldsymbol{\tau} = (\boldsymbol{\beta}, \sigma, \tilde{\boldsymbol{\beta}})'$. Equation (A.5) becomes

$$\sqrt{n}(\boldsymbol{\tau}_n - \boldsymbol{\tau}) = \big[\mathbf{I} - \nabla \mathbf{f}(\boldsymbol{\tau})\big]^{-1} \sqrt{n}\big[\mathbf{f}(\boldsymbol{\tau}) - \boldsymbol{\tau}\big] + o_P(1)\,. \tag{A.6}$$

We will now show that the correction factors $\mathbf{M}_n$ and $\mathbf{d}_n$ in (3.6) and (3.7) are the corresponding first p rows of the estimate $[\mathbf{I} - \nabla \mathbf{f}(\boldsymbol{\tau}_n)]^{-1}$ of the matrix $[\mathbf{I} - \nabla \mathbf{f}(\boldsymbol{\tau})]^{-1}$ in (A.6). It is easy to see that $\mathbf{I} - \nabla \mathbf{f}(\boldsymbol{\tau}_n)$ has the form

$$\big[\mathbf{I} - \nabla \mathbf{f}(\boldsymbol{\tau}_n)\big] = \left[\begin{array}{ccc} & & 0\cdots 0 \\ \mathcal{A} & v & \vdots \quad \vdots \\ & & 0\cdots 0 \\ \hline 0\cdots 0 & a & 0\cdots 0 \\ \hline 0\cdots 0 & & \\ \vdots \quad \vdots & w & \mathcal{B} \\ 0\cdots 0 & & \end{array}\right],$$

where

$$\mathcal{A} = \mathbf{I} - \frac{\partial}{\partial \boldsymbol{\beta}}[\mathbf{A}_n^{-1}\mathbf{v}_n], \qquad v = -\frac{\partial}{\partial \sigma}[\mathbf{A}_n^{-1}\mathbf{v}_n], \qquad a = 1 - \frac{\partial}{\partial \sigma}[\sigma u_n],$$

$$w = -\frac{\partial}{\partial \sigma}[\mathbf{B}_n^{-1}\,\mathbf{w}_n], \qquad \mathcal{B} = \mathbf{I} - \frac{\partial}{\partial \tilde{\boldsymbol{\beta}}}[\mathbf{B}_n^{-1}\mathbf{w}_n].$$

That

$$\frac{\partial}{\partial \tilde{\boldsymbol{\beta}}}[u_n] = (0, \ldots, 0)$$

follows from the fact that $\hat{\sigma}_n$ attains the minimum of the S-scale.

Now note that the estimate of the correction factor in (A.6) has the following form:

$$\big[\mathbf{I} - \nabla \mathbf{f}(\boldsymbol{\tau}_n)\big]^{-1} = \left[\begin{array}{c|c|c} & & 0\cdots 0 \\ \mathcal{A}^{-1} & -\mathcal{A}^{-1}v/a & \vdots \quad \vdots \\ & & 0\cdots 0 \\ \hline 0\cdots 0 & 1/a & 0\cdots 0 \\ \hline 0\cdots 0 & & \\ \vdots \quad \vdots & -\mathcal{B}^{-1}w/a & \mathcal{B}^{-1} \\ 0\cdots 0 & & \end{array}\right]. \tag{A.7}$$

Note that in (A.6) we are only interested in the first $p+1$ coordinates of $\boldsymbol{\tau}_n$ (the remaining p correspond to the S-regression estimate). From $[\mathbf{I} - \nabla \mathbf{f}(\boldsymbol{\tau}_n)]^{-1}$

in (A.7) we see that the last p coordinates of $\mathbf{f}$ are not involved in determining the first $p+1$ coordinates of $\boldsymbol{\tau}_n - \boldsymbol{\tau}$. Hence, when we apply this method in practice we do not need to bootstrap $\tilde{\boldsymbol{\beta}}_n$.

It also follows from (A.7) that we only need to calculate $\mathcal{A}$, v and a. We need to find the derivatives of

$$\big[\mathbf{A}_n(\boldsymbol{\beta},\sigma)^{-1}\,\mathbf{v}_n(\boldsymbol{\beta},\sigma)\big].$$

One way to calculate them is to differentiate the vector $\boldsymbol{\alpha}_n$ defined implicitly by

$$\mathbf{A}_n(\boldsymbol{\beta},\sigma)\boldsymbol{\alpha}_n(\boldsymbol{\beta},\sigma) = \mathbf{v}_n(\boldsymbol{\beta},\sigma).$$

Drop the arguments $(\boldsymbol{\beta},\sigma)$ and the subscripts to simplify the notation. Differentiating both sides of the equation

$$\frac{\partial}{\partial\boldsymbol{\beta}}[\mathbf{A}\boldsymbol{\alpha}] = \frac{\partial}{\partial\boldsymbol{\beta}}\mathbf{v},$$

we obtain

$$\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\alpha} = \mathbf{A}^{-1}\left[\frac{\mathbf{A} - \sum_{i=1}^{n}\rho_1''(r_i/\hat{\sigma}_n)}{\hat{\sigma}_n\mathbf{x}_i\mathbf{x}_i'}\right].$$

It follows that

$$\mathcal{A} = \mathbf{I} - \frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\alpha}|_{\hat{\boldsymbol{\beta}},\hat{\sigma}_n} = \mathbf{A}^{-1}\frac{1}{\hat{\sigma}_n}\sum_{i=1}^{n}\rho_1''(r_i/\hat{\sigma}_n)\mathbf{x}_i\mathbf{x}_i'.$$

Then we have

$$\mathcal{A}^{-1} = \hat{\sigma}_n\left(\sum_{i=1}^{n}\rho_1''\left(\frac{r_i}{\hat{\sigma}_n}\right)\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\mathbf{A} \tag{A.8}$$

and

$$\frac{-\mathcal{A}^{-1}v}{a} = \frac{bn\hat{\sigma}_n\big[\sum_{i=1}^{n}\rho_1''(r_i/\hat{\sigma}_n)\mathbf{x}_i\mathbf{x}_i'\big]^{-1}\sum_{i=1}^{n}\rho_1''(r_i/\hat{\sigma}_n)r_i/\hat{\sigma}_n\mathbf{x}_i}{\sum_{i=1}^{n}\rho_0'(r_i/\hat{\sigma}_n)r_i/\hat{\sigma}_n}. \tag{A.9}$$

It is easy to see that $\mathbf{M}_n$ in (3.6) is equal to (A.8) and that $\mathbf{d}_n$ in (3.7) is $-\mathcal{A}^{-1}\,v/a$ in (A.9).

We will now show that the bootstrap distribution of $\sqrt{n}[\mathbf{f}^*(\boldsymbol{\tau}_n) - \boldsymbol{\tau}_n]$ converges to the same limiting distribution as that of the sequence $\sqrt{n}[\mathbf{f}^*(\boldsymbol{\tau}) - \boldsymbol{\tau}]$.

First, note that

$$[\mathbf{f}^*(\boldsymbol{\tau}_n) - \boldsymbol{\tau}_n] = \begin{pmatrix} \hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n \\ \hat{\sigma}_n^* - \hat{\sigma}_n \\ \tilde{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}_n \end{pmatrix} = \begin{pmatrix} \mathbf{A}_n^{*-1}\,\mathbf{v}_n^* - \hat{\boldsymbol{\beta}}_n \\ \hat{\sigma}_n\, u_n^* - \hat{\sigma}_n \\ \mathbf{B}_n^{*-1}\,\mathbf{w}_n^* - \tilde{\boldsymbol{\beta}}_n \end{pmatrix},$$

where * denotes the bootstrap version of these quantities. It is easy to see that

$$\mathbf{v}_n^*(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n) = \sum_{i=1}^n \rho_1'(r_i^*/\hat{\sigma}_n)\mathbf{x}_i^* + \mathbf{A}_n^*(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n)\hat{\boldsymbol{\beta}}_n$$

and

$$\mathbf{w}_n^*(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n) = \sum_{i=1}^n \rho_0'(\tilde{r}_i^*/\hat{\sigma}_n)\mathbf{x}_i^* + \mathbf{B}_n^*(\hat{\boldsymbol{\beta}}_n, \hat{\sigma}_n)\tilde{\boldsymbol{\beta}}_n.$$

Then

$$\text{(A.10)} \qquad \begin{pmatrix} \hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n \\ \hat{\sigma}_n^* - \hat{\sigma}_n \\ \tilde{\boldsymbol{\beta}}_n^* - \tilde{\boldsymbol{\beta}}_n \end{pmatrix} = \begin{pmatrix} \mathbf{A}_n^{*-1} \sum_{i=1}^n \rho_1'(r_i^*/\hat{\sigma}_n)\mathbf{x}_i^* \\ \hat{\sigma}_n u_n^* - \hat{\sigma}_n \\ \mathbf{B}_n^{*-1} \sum_{i=1}^n \rho_0'(\tilde{r}_i^*/\hat{\sigma}_n)\mathbf{x}_i^* \end{pmatrix}.$$

This last expression can be expressed as a function of means. Consider the function $\mathbf{g}$: $\mathbb{R}^{p\times p} \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^{p\times p} \times \mathbb{R}^p \to \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p$,

$$\mathbf{g}(\bar{\mathbf{A}}, \bar{\mathbf{v}}, \bar{u}, \bar{\mathbf{B}}, \overline{\mathbf{w}}) = (\bar{\mathbf{A}}^{-1}\bar{\mathbf{v}}, \bar{u}, \bar{\mathbf{B}}^{-1}\overline{\mathbf{w}}).$$

Then (A.10) can be written as $\mathbf{g}(\mathbf{A}_n^*, \overline{\mathbf{z}}^*, u_n^*, \mathbf{B}_n^*, \overline{\mathbf{w}}^*)$, where $\mathbf{A}_n^*$, u_n^* and $\mathbf{B}_n^*$ are as before, $\mathbf{z}_i = \rho_1'(r_i^*/\hat{\sigma}_n)\mathbf{x}_i^*$ and $\mathbf{w}_i = \rho_0'(\tilde{r}_i^*/\hat{\sigma}_n)\mathbf{x}_i^*$ for $1 \le i \le n$. This function is differentiable (this can be seen by thinking of it as a composition of differentiable functions). We have that the statistic we are bootstrapping is of the form

$$\mathbf{g}(\bar{\mathbf{y}}_n(\boldsymbol{\tau}_n)) - \mathbf{g}(\boldsymbol{\mu}(\boldsymbol{\tau}_n)),$$

where $\mathbf{y}_i$ for $1 \le i \le n$ is a vector of the bootstrapped dimension and $\boldsymbol{\tau}_n$ is a consistent estimate of the vector of parameters $\boldsymbol{\tau}$. We have to show that the asymptotic distribution of

$$\text{(A.11)} \qquad \sqrt{n}(\bar{\mathbf{y}}_n(\boldsymbol{\tau}_n) - \boldsymbol{\mu}(\boldsymbol{\tau}_n))$$

is the same as that of

$$\text{(A.12)} \qquad \sqrt{n}(\bar{\mathbf{y}}_n(\boldsymbol{\tau}) - \boldsymbol{\mu}(\boldsymbol{\tau})).$$

The proof of this last statement is based on bounding the distance d_2 [see Bickel and Freedman (1981)] between the distribution functions of (A.11) and (A.12) using the fact that $\boldsymbol{\tau}_n \to \boldsymbol{\tau}$ almost surely. Lemma 8.1 of Bickel and Freedman (1981) and the regularity conditions of $\mathbf{g}$ show that the bootstrap distribution of

$\mathbf{g}(\bar{\mathbf{y}}_n(\boldsymbol{\tau}_n)) - \mathbf{g}(\boldsymbol{\mu}(\boldsymbol{\tau}_n))$ converges to the same limit as that of the sequence $\mathbf{g}(\bar{\mathbf{y}}_n(\boldsymbol{\tau})) - \mathbf{g}(\boldsymbol{\mu}(\boldsymbol{\tau}))$. $\square$

REFERENCES

BEATON, A. E. and TUKEY, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16** 147–185.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

CARROLL, R. J. (1978). On almost sure expansions for M-estimates. *Ann. Statist.* **6** 314–318.

CARROLL, R. J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *J. Amer. Statist. Assoc.* **74** 674–679.

CARROLL, R. J. and WELSH, A. H. (1988). A note on asymmetry and robustness in linear regression. *Amer. Statist.* **42** 285–287.

COLEMAN, J. (1966). *Equality of Educational Opportunity*. Office of Education, U.S. Department of Health, Washington, DC.

DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge Univ. Press.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

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.

HU, F. and KALBFLEISCH, J. D. (2000). The estimating function bootstrap. *Canad. J. Statist.* **28** 449–499.

HU, F. and ZIDEK, J. V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika* **82** 263–275.

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

LIU, R. Y. and SINGH, K. (1992). Efficiency and robustness in resampling. *Ann. Statist.* **20** 370–384.

MOSTELLER, F. and TUKEY, J. W. (1977). *Data Analysis and Regression*. Addison–Wesley, Reading, MA.

PARR, W. C. (1985). The bootstrap: Some large sample theory and connections with robustness. *Statist. Probab. Lett.* **3** 97–100.

ROCKE, D. M. and DOWNS, G. W. (1981). Estimating the variances of robust estimators of location: Influence curve, jackknife and bootstrap. *Comm. Statist. Simulation Comput.* **10** 221–248.

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

ROUSSEEUW, P. J. and YOHAI, V. J. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis*. *Lecture Notes in Statist.* **26** 256–272. Springer, Berlin.

SALIBIAN-BARRERA, M. (2000). Contributions to the theory of robust inference. Ph.D. thesis, Dept. Statist., Univ. British Columbia, Vancouver.

SCHUCANY, W. R. and WANG, S. (1991). One-step bootstrapping for smooth iterative procedures. *J. Roy. Statist. Soc. Ser. B* **53** 587–596.

SEBER, G. A. F. (1984). *Multivariate Observations*. Wiley, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SHAO, J. (1990). Bootstrap estimation of the asymptotic variances of statistical functionals. *Ann. Inst. Statist. Math.* **42** 737–752.

SHAO, J. (1992). Bootstrap variance estimators with truncation. *Statist. Probab. Lett.* **15** 95–101.

SHORACK, G. R. (1982). Bootstrapping robust regression. *Comm. Statist. Theory Methods* **11** 961–972.

SINGH, K. (1998). Breakdown theory for bootstrap quantiles. *Ann. Statist.* **26** 1719–1732.

YANG, S.-S. (1985). On bootstrapping a class of differentiable statistical functionals with applications to L- and M-estimates. *Statist. Neerlandica* **39** 375–385.

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

DEPARTMENT OF STATISTICS
333-6356 AGRICULTURAL ROAD
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA
CANADA V6T 1Z2
E-MAIL: ruben@stat.ubc.ca