Fast and stable bootstrap methods for robust estimates

Matias Salibian-Barrera
School of Mathematics and Statistics − Carleton University

Abstract

The standard error and sampling distribution of robust estimates can, in principle, be estimated using the bootstrap. However, two problems arise when we want to use bootstrap with robust estimates on moderately large data sets: the bootstrap estimates may be unrealiable because the proportion of outliers in many bootstrap samples could be higher than that in the original data set, and the high computational demand of robust regression estimates may render the method unfeasible for moderately high-dimensional problems. Recently, Salibian-Barrera and Zamar (2002) have proposed a new bootstrap method called *robust bootstrap* to estimate the asymptotic distribution and asymptotic variance of MM-estimates. This method overcomes the problems mentioned above, namely: it is fast and stable (it can resist large proportion of outliers on the bootstrap samples). Unfortunately, its convergence seems to be only of order $O_p(1/\sqrt{n})$. Another way to estimate the asymptotic variance of robust estimates on large datasets is to bootstrap a one-step Newton-Raphson iteration of their estimating equations. This method will typically be fast (and hence feasible on moderately large datasets). In this paper we compare the performance of this method with that of the robust bootstrap for the simple location-scale model.

1 Introduction

Many robust point estimates have been proposed in the last decades. Unfortunately much less attention has been paid to the problem of inference based on robust estimates.

The standard error of robust estimates can be estimated using their asymptotic variances. However, the asymptotic distribution of robust estimates has been mainly studied under the central normal model which, of course, does not hold in most practical situations when robust methods are recommended. The calculation of the asymptotic distribution of robust estimates for asymmetric errors is involved (see Carroll, 1978, 1979; Huber, 1981; Rocke and Downs, 1981; Carroll and Welsh, 1988; and Salibian-Barrera, 2000).

The sampling distribution of robust estimates and their standard errors can also be estimated, in principle, using the bootstrap (Efron, 1979). 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. Two problems arise when we bootstrap robust estimates: *numerical instability* – the bootstrap distribution might be a very poor estimator of the distribution of the robust estimates because the proportion of outliers in the bootstrap samples can be higher than that in the original data; and *computational cost* – due to the non-convex optimization problems that have to be solved in order to calculate robust regression estimates, it may not be feasible to obtain a few thousand re-calculated estimates for high dimensional problems.

The first problem appears because outlying and non-outlying observations have the same chance of being in the bootstrap samples. In particular, a certain proportion of these bootstrap samples can have enough outliers to severely affect the re-calculated estimate, regardless of its robustness properties. This can affect heavily the tails of the bootstrap distribution, which are of much interest when building confidence intervals, for example. 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 to draw 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. Singh's proposal, however, does not address the problem of computational demand for problems in higher dimensions nor does it generalise easily to multivariate or linear regression models.

The second problem mentioned above has received some attention in the literature. See for example Schucany and Wang (1991), Hu and Zidek (1995), Hu and Kalbfleisch (2000). Unfortunately, due to the regularity conditions they require these proposals cannot be used with estimates with good robustness properties.

Recently, Salibian-Barrera and Zamar (2002) have proposed a new bootstrap method called "robust bootstrap" to estimate the asymptotic distribution and asymptotic variance of robust regression MM-estimates. This method also applies to location M-estimates calculated with S-estimates of scale (which we also call MM-estimates, see Section 2). MM-estimates have good robustness properties, are highly efficient when the errors are normal, and are available in S-PLUS (library `robust` in S-PLUS 6).

The robust bootstrap simultaneously overcomes both problems mentioned above, namely: it is fast and it can resist large proportion of outliers on the bootstrap samples. The basic idea is to bootstrap a re-weighted representation of the estimate. This robust bootstrap is computationally simple because for each bootstrap sample we only have to calculate a weighted average (weighted least squares in the linear regression case) to obtain the bootstrapped estimate (this solves the problems associated to *computational cost*). The form of the weights makes the procedure numerically stable and robust to the presence of outliers in the bootstrap samples (this solves the *numerical instability* mentioned above). Because the robust bootstrap works with a linearization of the robust estimate, we need to apply a correction factor (that is consistently estimable from the data). Unfortunately, this correction seems to affect the convergence rate of the robust bootstrap which appears to be only of order $O_p(1/\sqrt{n})$.

Another way to estimate the asymptotic variance of robust estimates on large datasets is to bootstrap a one-step Newton-Raphson iteration of their estimating equations. This method will typically be fast (and hence feasible on moderately large datasets). Since no correction will be needed in this case, this method may inherit the higher-order of convergence of the bootstrap.

Let $\hat{\theta}_n$ be the estimate of interest and assume that is defined by an estimating equation of the form $g_n(\hat{\theta}_n) = \sum_{i=1}^n f_i(x_i, \hat{\theta}_n) = 0$. Consider the Newton-Raphson iterations to solve it:

$$\hat{\theta}_n^{(i+1)} = \hat{\theta}_n^{(i)} - g_n(\hat{\theta}_n^{(i)}) / g_n'(\hat{\theta}_n^{(i)}) \,,$$

where $g_n'(\hat{\theta}_n) = \sum_{i=1}^n f_i'(x_i, \hat{\theta}_n)$ and $f_i'(x, \theta) = \partial f_i(x, \theta)/\partial\theta$ and $\hat{\theta}_n^{(0)}$ is a suitable initial estimate. Let $x_1^*, \ldots, x_n^*$ be a bootstrap sample drawn from the data. The Newton-Raphson bootstrap re-calculates $\hat{\theta}_n$

$$\hat{\theta}_n^* = \hat{\theta}_n - g_n^*(\hat{\theta}_n) / \, {g_n'}^* (\hat{\theta}_n) \,, \tag{1}$$

where $g_n^*(\hat{\theta}_n) = \sum_{i=1}^n f_i(x_i^*, \hat{\theta}_n)$ and similarly ${g_n'}^* (\hat{\theta}_n)$. Typically, the estimating equations of robust estimates are based on functions $f_i(x, \hat{\theta}_n)$ such that $f_i'(x, \hat{\theta}_n)$ vanishes for large values of $|x - \hat{\theta}_n|$. This may produce numerical problems since ${g_n'}^*$ could become very small for bootstrap samples with a large proportion of outliers. A simple modification to solve this problem is not to bootstrap g_n' in the denominator above, that is to use:

$$\hat{\theta}_n^* = \hat{\theta}_n - g_n^*(\hat{\theta}_n) / g_n'(\hat{\theta}_n) \,. \tag{2}$$

In Section 2 we introduce the robust estimates we consider in this paper. Approaches (1) and (2) are discussed in Section 3. Simulation results are presented in Section 4, and Section 5 contains our conclusions.

2 Definitions

To fix ideas we focus on robust estimates for the simple location-scale model. Let $x_1, \ldots, x_n$ be n observations on the real line satisfying

$$x_i = \mu + \sigma \, \epsilon_i \qquad i = 1, \ldots n,$$

where ϵ_i are independent and identically distributed observations with variance equal to 1. The interest is in estimating μ and the scale σ is a nuisance parameter. We consider M-location estimates (Huber, 1964) $\hat{\mu}_n$ defined as the solution of an estimating equation of the form

$$\sum_{i=1}^n \psi \left((x_i - \hat{\mu}_n) / \, \hat{\sigma}_n \right) = 0 \,; \tag{3}$$

where $\hat{\sigma}_n$ is a robust estimate of the residuals scale, and $\psi : \mathbb{R} \to \mathbb{R}$ is a bounded, non-decreasing and odd real function. In what follows, and for technical reasons we will consider *score functions* ψ in (3) that are continuously differentiable. An example of such a function is given by

$$\psi_c(u) = \text{sign}(u) \begin{cases} |u/c| & \text{if} \quad |u| \leq 0.8\, c \\ p_4(|u/c|) & \text{if} \quad 0.8\, c < |u| \leq c \\ p_4(1) & \text{if} \quad |u| > c \end{cases}, \tag{4}$$

where $c > 0$ is a user-chosed tuning constant, and $p_4(u) = 38.4 - 175\, u + 300\, u^2 - 225\, u^3 + 62.5\, u^4$ (see Fraiman *et al.* (2001), and also Bednarski and Zontek (1996), for other choices of smooth functions ψ).

The scale estimate $\hat{\sigma}_n$ (3) is an S-estimate of scale (Rousseeuw and Yohai, 1984) defined as follows. Let $\rho : \mathbb{R} \to \mathbb{R}_+$ be a bounded, continuous and even function satisfying $\rho(0) = 0$ and let $b \in (0, 1)$. The S-scale $\hat{\sigma}_n$ is defined by

$$\hat{\sigma}_n = \inf_{t \in \mathbb{R}} s_n(t) \;, \tag{5}$$

where, for each $t \in \mathbb{R}$, $s_n(t)$ is the solution of

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left((x_i - t) / s_n(t) \right) = b. \tag{6}$$

Naturally associated with this family are the *S-location* estimates $\tilde{\mu}_n$ given by

$$\tilde{\mu}_n = \arg \inf_{t \in \mathbb{R}} s_n(t) \,.$$

Beaton and Tukey (1974) proposed a family of functions ρ_d given by

$$\rho_d(u) = \begin{cases} 3\,(u/d)^2 - 3\,(u/d)^4 + (u/d)^6 & \text{if } |u| \leq d\,, \\ 1 & \text{if } |u| > d\,, \end{cases} \tag{7}$$

where the tuning constant d is positive. Following Yohai (1987) we will call these M-location estimates obtained with an S-scale *MM-location* estimates. In particular, the estimates $\hat{\mu}_n$, $\hat{\sigma}_n$ and $\tilde{\mu}_n$ solve the following system of equations:

$$\begin{aligned} \sum_{i=1}^{n} \psi\left((x_i - \hat{\mu}_n) / \hat{\sigma}_n \right) &= 0\,, \\ \frac{1}{n} \sum_{i=1}^{n} \left[\rho\left((x_i - \tilde{\mu}_n) / \hat{\sigma}_n \right) - b \right] &= 0\,, \\ \frac{1}{n} \sum_{i=1}^{n} \rho'\left((x_i - \tilde{\mu}_n) / \hat{\sigma}_n \right) &= 0\,. \end{aligned}$$

The robust location estimates $\hat{\mu}_n$ defined by (3) with $\hat{\sigma}_n$ as in (5) have simultaneously high breakdown and high efficiency at the central model. For example, the choice $d = 1.548$ for ρ_d in (7), $b = 0.5$ in (6), and $c = 1.525$ for ψ_c in (4) yields a location estimate $\hat{\mu}_n$ with 50% breakdown point and 95% efficiency when the errors have a normal distribution.

3 Newton-Raphson Bootstrap

A simple alternative to obtain a bootstrap method that does not involve fully re-calculating the estimate is to bootstrap the Newton-Rapshon iterations of its estimating equations.

In general the asymptotic distribution of $\hat{\mu}_n$ depends on that of the scale estimate $\hat{\sigma}_n$ (see Carroll, 1978, 1979; and Salibian-Barrera, 2000, for example). In turn, the distribution of smooth scale estimates (like the S-scales described in Section 2) depend on an auxiliary location estimate $\tilde{\mu}_n$. Hence, to obtain a bootstrap method that is consistent to the asymptotic distribution of $\hat{\mu}_n$ for arbitrary error distributions we need to re-calculate $\hat{\mu}_n$ jointly with $\hat{\sigma}_n$ and $\tilde{\mu}_n$.

Let $\hat{\boldsymbol{\theta}}_n \in \mathbb{R}^p$ be the robust estimate and assume that it satisifies an equation of the form

$$\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}\,, \tag{8}$$

for some function $\mathbf{g}_n : \mathbb{R}^p \to \mathbb{R}^p$, that in general depends on the sample $x_1, \ldots, x_n$. Typically, $\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)$ is of the form $\mathbf{g}_n(\hat{\boldsymbol{\theta}}_n) = \sum_{i=1}^n \mathbf{f}_i(x_i, \hat{\boldsymbol{\theta}}_n)$. For location MM-location estimates, we have $p = 3$, $\hat{\boldsymbol{\theta}}_n = (\hat{\mu}_n, \hat{\sigma}_n, \tilde{\mu}_n)'$ and

$$\mathbf{f}_i(x_i, \hat{\boldsymbol{\theta}}_n) = \mathbf{f}(x_i, \hat{\boldsymbol{\theta}}_n) = \begin{pmatrix} \psi((x_i - \hat{\mu}_n)/\hat{\sigma}_n) \\ \\ \rho((x_i - \tilde{\mu}_n)/\hat{\sigma}_n) - b \\ \\ \rho'((x_i - \tilde{\mu}_n)/\hat{\sigma}_n) \end{pmatrix}.$$

Then, the Newton-Raphson iterations to solve equation (8) are

$$\hat{\boldsymbol{\theta}}_n^{(i)} = \hat{\boldsymbol{\theta}}_n^{(i-1)} - \nabla \mathbf{g}_n(\hat{\boldsymbol{\theta}}_n^{(i-1)})^{-1}\, \mathbf{g}_n(\hat{\boldsymbol{\theta}}_n^{(i-1)})\,, \tag{9}$$

where $[\nabla \mathbf{g}_n(\boldsymbol{\theta})]_{j\,k} = \partial \sum_{i=1}^n \mathbf{f}_{i\,j}(x_i, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}_k$, $j, k = 1, \ldots, p$ is the matrix of first partial derivatives and $\mathbf{f}_{i\,j}$ denotes the j-th coordinate of $\mathbf{f}_i$. Let $x_1^*, \ldots, x_n^*$ be a bootstrap sample from the data $x_1, \ldots, x_n$. The Newton-Raphson bootstrap re-calculates $\hat{\boldsymbol{\theta}}_n$ based on equation (9):

$$\hat{\boldsymbol{\theta}}_n^* = \hat{\boldsymbol{\theta}}_n - \nabla \mathbf{g}_n^*(\hat{\boldsymbol{\theta}}_n)^{-1}\, \mathbf{g}_n^*(\hat{\boldsymbol{\theta}}_n)\,,$$

where $\mathbf{g}_n^*(\hat{\boldsymbol{\theta}}_n) = \sum_{i=1}^n \mathbf{f}_i(x_i^*, \hat{\boldsymbol{\theta}}_n)$. Note that we do not re-calculate $\hat{\boldsymbol{\theta}}_n$. For the particular case of location MM-estimates, it is easy to see that we get

$$\hat{\mu}_n^* = \hat{\mu}_n + \hat{\sigma}_n \left[\frac{\overline{\psi(u^*)}}{\overline{\psi'(u^*)}} + \frac{\overline{\psi(u^*)\, u^*}}{\overline{\psi'(u^*)}} \times \frac{\overline{\rho(u^*) - b}}{\overline{\rho'(u^*)\, u^*}} \right], \tag{10}$$

where $\overline{\psi(u^*)} = \sum \psi((x_i^* - \hat{\mu}_n)/\hat{\sigma}_n)/n$ and similarly the others.

In general, the inverse of the matrix of first derivatives $[\nabla \mathbf{g}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}$ is numerically unstable. For example, in the case of MM-estimates for location, we have $\psi'(x) = 0$ and $\rho'(x) = 0$ for large $|x|$. As these functions appear in the denominator of (10) this can cause numerical instability. An alternative is to use the following scheme:

$$\hat{\boldsymbol{\theta}}_n^* = \hat{\boldsymbol{\theta}}_n - \nabla \mathbf{g}_n\left(\hat{\boldsymbol{\theta}}_n\right)^{-1} \mathbf{g}_n^*\left(\hat{\boldsymbol{\theta}}_n\right),$$

in other words, we now do not bootstrap $\nabla \mathbf{g}_n^{-1}$. For location MM-estimates we get

$$\hat{\mu}_n^* = \hat{\mu}_n + \hat{\sigma}_n \left[\frac{\overline{\psi\left(u^*\right)}}{\overline{\psi'\left(u\right)}} + \frac{\overline{\psi\left(u\right) u}}{\overline{\psi'\left(u\right)}} \times \frac{\overline{\rho\left(u^*\right) - b}}{\overline{\rho'\left(u\right) u}} \right] . \tag{11}$$

This alternative is numerically more stable, but it may suffer from the same lower convergence rate as the robust bootstrap of Salibian-Barrera and Zamar (2002).

The next Section describes a small simulation experiment that compares the Newton-Raphson bootstrap, the modified Newton-Raphson bootstrap and the robust bootstrap for location MM-estimates.

4 Simulation results

We considered robust M-location estimates calculated with an S-scale (see Section 2). The tuning constants where chosen to obtain a breakdown point of 1/2 and an efficiency of 95% for normal errors.

Let $x_1, \ldots, x_n$ satisfy the simple location-scale model

$$x_i = \mu + e_i \quad i = 1, \ldots, n \, ,$$

where e_i are iid random errors and μ is the parameter of interest. We generated samples of sizes $n = 30$, 50 and 100. These samples where contaminated with 10%, 20% and 30% of outliers tightly clustered around 4. The distribution function F_ϵ for the contaminated random variables x_i was

$$F_\epsilon(x) = (1 - \epsilon) \, \Phi(x) + \epsilon \, \Phi((x - 4)/.1) \, ,$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $\epsilon = 0.10$, 0.20 and 0.30.

For each combination of n and ϵ we generated 5,000 samples. For each sample we obtained 2,000 bootstrap re-calculated $\hat{\mu}_n^*$ with the Newton-Raphson method of equation (10) [NR], the modified Newton Raphson as in (11) [NR2] and the robust bootstrap [RB]. We decided not to include the classical bootstrap in this study because it is already known that when the data contain outliers its performance is not satisfactory regardless of the robustness properties of the estimate being bootstrapped (Singh, 1998; Salibian-Barrera, 2000).

The sample variance (`var`) and the MAD^2 (`mad2`) (square of the Median Absolute Deviation from the median) of these 2,000 bootstrapped $\hat{\mu}_n^*$ were used as estimates of the variance of $\hat{\mu}_n$. Additionally, 95% confidence intervals for μ were constructed from the empirical quantiles of these 2,000 bootstrapped $\hat{\mu}_n^*$. Tables 1 to 6 contain the results on variance estimation. Tables 7 and 8 display the results for the confidence intervals.

To measure the performance of each variance estimate we use the mean (and SE) of the following three loss functions. To have a measure that penalises equaly over- and under-estimation of V we first considered

$$\text{loss}_{L1}\left(x, V\right) = \begin{cases} -\log\left(x/V\right) & \text{if } x < V \\ \log\left(x/V\right)^2 & \text{if } x \geq V \end{cases} \, , \tag{12}$$

where $x > 0$ and V is the actual asymptotic variance of $\hat{\mu}_n$. We also used

$$\text{loss}_{L2}(x, V) = \log(x/V)^2 \ , \tag{13}$$

and

$$\text{loss}_{Q}(x, V) = (x/V - 1)^2 \ . \tag{14}$$

A graphical comparison of the behaviour of these loss functions can be found in Figure 1. Note that loss_{L1} and loss_{L2} penalise over-estimation of V less than loss_Q since $\text{loss}_{Li}(x, V)/\,\text{loss}_Q(x, V) \to 0$ for $i = 1, 2$ when $x \to \infty$. On the other hand, both loss_{L1} and loss_{L2} penalise under-estimation more severely than loss_Q. The function loss_{L2} is the most severe loss function for under-estimation of the variance since $\text{loss}_{L2}(x, V)/\,\text{loss}_{L1}(x, V) \to \infty$ when $x \to 0$.

In Tables 1 and 2 we show the mean (and SE) of the 5,000 loss_{L1}-losses obtained with the `var` and `mad2` variance estimates respectively. Tables 3 and 4 display the same results for loss_{L2} and Tables 5 and 6 with loss_Q.

The results are very similar in all cases, but much more extreme for the quadratic loss (14) due to the large variance estimates that are obtained with the numerically unstable methods NR and NR2.

With the log-based loss functions loss_{L1} and loss_{L2} and for $\epsilon = 0.0$ and 0.10 the robust bootstrap [RB], the Newton-Raphson [NR] and the modified Newton-Raphson [NR2] are mostly comparable (altough RB seems to do better overall). For larger ϵ (0.20 and 0.30) RB is notably better than NR and NR2, the worst being NR in almost all cases. Note that the estimates obtained with RB are more stable than those of the other two methods (as measured by the SE of these 5,000 estimates). For $\epsilon = 0.20$ and 0.30 the SE of the NR and NR2 are most of the times between 1.5 and 3 times those of the RB (in one case the NR has an SE more than 7 times higher than that of the RB).

From Tables 5 and 6 we see that the NR and NR2 produce estimates that are much more variable than those obtained with RB. The robust bootstrap is consistently the best option, followed by NR2, both in terms of mean error and variability. The worst performing method is NR.

Finally Table 7 contains the mean (and SE) of the lengths of the 5,000 confidence intervals for μ. The levels were about the same in all cases (see Table 8), and the main difference lies in their lengths. Here the RB is clearly the best method, both in mean lenght and SE of the lengths. The modified Newton-Raphson (NR2) is sometimes marginally better ($\epsilon = 0.00$ and 0.10), than the RB but in these cases the mean of the robust bootstrap lengths (RB) is generally within one SE of the mean length of NR2.

5 Conclusions

We have studied two bootstrap methods to estimate the variance and sampling distribution of robust estimates. These methods are based on re-sampling the Newton-Raphson iterations used to solve the estimating equations that define the robust estimates. The first method consists of completely bootstrapping the Newton-Raphson iterations. Motivated by numerical stability considerations and by the

n	ϵ	RB	NR	NR2
30	0.00	0.192 (0.003)	0.190 (0.003)	0.194 (0.003)
	0.10	0.330 (0.006)	0.357 (0.008)	0.337 (0.007)
	0.20	0.494 (0.006)	0.672 (0.012)	0.589 (0.010)
	0.30	0.638 (0.007)	1.066 (0.053)	0.801 (0.015)
50	0.00	0.131 (0.002)	0.128 (0.002)	0.133 (0.002)
	0.10	0.220 (0.004)	0.222 (0.004)	0.221 (0.004)
	0.20	0.414 (0.006)	0.518 (0.010)	0.481 (0.008)
	0.30	0.504 (0.006)	0.616 (0.017)	0.598 (0.009)
100	0.00	0.085 (0.001)	0.083 (0.001)	0.085 (0.001)
	0.10	0.133 (0.002)	0.132 (0.002)	0.134 (0.002)
	0.20	0.272 (0.004)	0.304 (0.005)	0.295 (0.005)
	0.30	0.356 (0.005)	0.436 (0.005)	0.432 (0.005)

Table 1: Mean (SE) of 5,000 "var" bootstrap estimates of the variance of $\hat{\mu}_n$. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap. Log loss "loss_{L1}".

fact that for robust MM-estimates the derivatives of the estimating equations vanish for outlying observations, the second alternative does not resample the matrix of first partial derivatives. We compared the performance of these methods with that of the robust bootstrap (Salibian-Barrera and Zamar, 2002). We focused on the resulting variance estimates and in the length of the percentile confidence intervals for the location parameter. We found that generally the modified (more stable) Newton-Raphson method performs better than the standard one, but that in most cases the robust bootstrap gave the best results (in particular for larger proportions of contamination).

The study of these Newton-Raphson methods was motivated by the low rate of convergence of the robust bootstrap. However we found that numerical stability issues shadow in practice the difference in asymptotic speed. In other words, the robust bootstrap, a slightly "asymptotically slower" method, performs better in practice because it is significantly more stable.

Comparison of these three methods for other estimates and models are currently under way, as well as the study of their asymptotic distribution and quantile breakdown point.

References

[1] Bednarski, T. and Zontek, S. (1996). Robust estimation of parameters in mixed unbalanced models. *The Annals of Statistics*, **24**, 1493-1510.

[2] Carroll, R.J. (1978). On almost sure expansions for M-estimates. *The Annals of Statistics*, **6**, 314-318.

n	ϵ	RB	NR	NR2
30	0.00	0.194 (0.003)	0.190 (0.003)	0.194 (0.003)
	0.10	0.332 (0.006)	0.337 (0.007)	0.338 (0.007)
	0.20	0.503 (0.006)	0.603 (0.010)	0.595 (0.010)
	0.30	0.641 (0.007)	0.846 (0.016)	0.798 (0.015)
50	0.00	0.133 (0.002)	0.130 (0.002)	0.133 (0.002)
	0.10	0.223 (0.004)	0.220 (0.004)	0.223 (0.004)
	0.20	0.419 (0.006)	0.487 (0.009)	0.485 (0.008)
	0.30	0.506 (0.006)	0.648 (0.010)	0.596 (0.009)
100	0.00	0.089 (0.001)	0.086 (0.001)	0.088 (0.001)
	0.10	0.136 (0.002)	0.134 (0.002)	0.136 (0.002)
	0.20	0.276 (0.004)	0.297 (0.005)	0.297 (0.005)
	0.30	0.357 (0.005)	0.452 (0.006)	0.433 (0.005)

Table 2: Mean (SE) of 5,000 “mad” bootstrap estimates of the variance of $\hat{\mu}_n$. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap. Log loss “loss_{L1}”.

[3] Carroll, R.J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *Journal of the American Statistical Association*, **74**, 674-679.

[4] Carroll, R. J. and Welsh, A. H. (1988) A Note on Asymmetry and Robustness in Linear Regression. *The American Statistician*, **42**, 285-287.

[5] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* **7**, 1-26.

[6] Fraiman, R., Yohai, V. J. and Zamar, R. H. (2001). Optimal robust M-estimates of location. *The Annals of Statistics*. To appear.

[7] Hu, F. and Kalbfleisch, J.D. (2000). The estimating function bootstrap. *The Canadian Journal of Statistics*, **28**, ???-???.

[8] Hu, F. and Zidek, J.V. (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika*, **82**, 263-275.

[9] Huber, P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, **35**, 73-101.

[10] Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

[11] Parr, W.C. (1985). The bootstrap: some large sample theory and connections with robustness. *Statistics and Probability Letters*, **3**, 97-100.

[12] Rocke, D.M. and Downs, G.W. (1981). Estimating the variances of robust estimators of location: influence curve, jackknife and bootstrap. *Communications in Statistics, Part B – Simulation and Computation*, **10**, 221-248.

[13] Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series*. (J. Franke, W. Hardle and

n	ϵ	RB	NR	NR2
30	0.00	0.115 (0.002)	0.122 (0.003)	0.117 (0.002)
	0.10	0.252 (0.006)	0.285 (0.008)	0.258 (0.007)
	0.20	0.433 (0.006)	0.614 (0.013)	0.529 (0.010)
	0.30	0.680 (0.013)	1.067 (0.054)	0.850 (0.018)
50	0.00	0.065 (0.001)	0.067 (0.001)	0.065 (0.001)
	0.10	0.142 (0.003)	0.149 (0.004)	0.142 (0.003)
	0.20	0.343 (0.006)	0.450 (0.010)	0.410 (0.009)
	0.30	0.466 (0.008)	0.573 (0.018)	0.573 (0.010)
100	0.00	0.030 (0.001)	0.031 (0.001)	0.030 (0.001)
	0.10	0.070 (0.002)	0.071 (0.002)	0.069 (0.002)
	0.20	0.202 (0.004)	0.237 (0.005)	0.224 (0.005)
	0.30	0.286 (0.006)	0.379 (0.006)	0.377 (0.006)

Table 3: Mean (SE) of 5,000 "var" bootstrap estimates of the variance of $\hat{\mu}_n$. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap. Quadratic log loss function "loss_{L2}".

D. Martin, eds.). *Lecture Notes in Statist.*, **26** 256-272. Berlin: Springer-Verlag.

[14] Salibian-Barrera, M. (2000). Contributions to the Theory of Robust Inference. Unpublished Ph.D. Thesis. University of British Columbia, Department of Statistics, Vancouver, BC. Available on-line at `ftp://ftp.stat.ubc.ca/pub/matias/Thesis`.

[15] Salibian-Barrera, M. and Zamar, R.H. (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, **30**, No. 2, April 2002.

[16] Schucany, W.R. and Wang, S. (1991). One-step bootstrapping for smooth iterative procedures. *Journal of the Royal Statistical Society, Series B*, **53**, 587-596.

[17] Shao, J. (1990). Bootstrap estimation of the Asymptotic variances of statistical functionals. *Annals of the Institute of Statistical Mathematics*, **42**, 737-752.

[18] Shao, J. (1992). Bootstrap variance estimators with truncation. *Statistics and Probability Letters*, **15**, 95-101.

[19] Shorack, G.R. (1982). Bootstrapping robust regression. *Communications in Statistics, Part A – Theory and Methods*, **11**, 961-972.

[20] Singh, K. (1998). Breakdown theory for bootstrap quantiles. *The Annals of Statistics*, **26**, 1719-1732.

[21] Yang, S-S (1985). On bootstrapping a class of differentiable statistical functionals with applications to L- and M-estimates. *Statistica Neerlandica*, **39**, 375-385.

n	ϵ	RB	NR	NR2
30	0.00	0.118 (0.003)	0.118 (0.002)	0.119 (0.003)
	0.10	0.253 (0.006)	0.261 (0.007)	0.260 (0.007)
	0.20	0.443 (0.007)	0.539 (0.011)	0.535 (0.011)
	0.30	0.689 (0.013)	0.969 (0.019)	0.844 (0.018)
50	0.00	0.067 (0.001)	0.067 (0.001)	0.067 (0.001)
	0.10	0.144 (0.003)	0.145 (0.003)	0.144 (0.003)
	0.20	0.348 (0.006)	0.416 (0.009)	0.414 (0.009)
	0.30	0.469 (0.008)	0.690 (0.013)	0.571 (0.010)
100	0.00	0.032 (0.001)	0.032 (0.001)	0.032 (0.001)
	0.10	0.071 (0.002)	0.071 (0.002)	0.071 (0.002)
	0.20	0.205 (0.004)	0.226 (0.005)	0.226 (0.005)
	0.30	0.288 (0.006)	0.419 (0.007)	0.378 (0.006)

Table 4: Mean (SE) of 5,000 "mad" bootstrap estimates of the variance of $\hat{\mu}_n$. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap. Quadratic log loss function "loss_{L2}".

[22] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642-656.

n	ϵ	RB	NR	NR2
30	0.00	0.130 (0.004)	0.159 (0.006)	0.132 (0.004)
	0.10	0.624 (0.038)	1.107 (0.128)	0.779 (0.079)
	0.20	0.857 (0.026)	2.934 (0.184)	1.756 (0.094)
	0.30	0.252 (0.003)	6225 (2048)	8.846 (2.154)
50	0.00	0.070 (0.002)	0.077 (0.002)	0.070 (0.002)
	0.10	0.243 (0.012)	0.285 (0.017)	0.244 (0.014)
	0.20	0.728 (0.023)	1.741 (0.099)	1.319 (0.069)
	0.30	0.197 (0.003)	107.9 (62.61)	2.192 (0.764)
100	0.00	0.031 (0.001)	0.032 (0.001)	0.031 (0.001)
	0.10	0.093 (0.003)	0.098 (0.004)	0.092 (0.003)
	0.20	0.391 (0.013)	0.583 (0.026)	0.516 (0.023)
	0.30	0.138 (0.002)	0.411 (0.011)	0.359 (0.009)

Table 5: Mean (SE) of 5,000 "var" bootstrap estimates of the variance of $\hat{\mu}_n$. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap. Square loss function "loss_Q".

n	ϵ	RB	NR	NR2
30	0.00	0.134 (0.004)	0.142 (0.004)	0.137 (0.004)
	0.10	0.623 (0.039)	0.855 (0.092)	0.809 (0.084)
	0.20	0.879 (0.027)	1.958 (0.114)	1.843 (0.101)
	0.30	0.254 (0.003)	10.79 (2.868)	9.588 (2.432)
50	0.00	0.071 (0.002)	0.075 (0.002)	0.072 (0.002)
	0.10	0.242 (0.012)	0.259 (0.015)	0.250 (0.014)
	0.20	0.740 (0.023)	1.406 (0.076)	1.347 (0.069)
	0.30	0.199 (0.003)	2.756 (1.007)	2.165 (0.759)
100	0.00	0.033 (0.001)	0.033 (0.001)	0.033 (0.001)
	0.10	0.094 (0.003)	0.096 (0.004)	0.094 (0.003)
	0.20	0.394 (0.013)	0.530 (0.024)	0.525 (0.024)
	0.30	0.139 (0.002)	0.377 (0.009)	0.365 (0.009)

Table 6: Mean (SE) of 5,000 "mad" bootstrap estimates of the variance of $\hat{\mu}_n$. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap. Square loss function "loss_Q".

n	ϵ	RB	NR	NR2
30	0.00	0.739 (0.002)	0.756 (0.002)	0.737 (0.002)
	0.10	0.973 (0.004)	1.005 (0.004)	0.967 (0.004)
	0.20	1.433 (0.007)	1.570 (0.009)	1.471 (0.008)
	0.30	1.793 (0.007)	2.613 (0.069)	2.055 (0.018)
50	0.00	0.573 (0.001)	0.579 (0.001)	0.572 (0.001)
	0.10	0.736 (0.002)	0.746 (0.002)	0.732 (0.002)
	0.20	1.127 (0.005)	1.190 (0.006)	1.149 (0.006)
	0.30	1.479 (0.005)	1.743 (0.016)	1.636 (0.009)
100	0.00	0.403 (0.0005)	0.404 (0.0005)	0.402 (0.0005)
	0.10	0.515 (0.001)	0.518 (0.001)	0.513 (0.001)
	0.20	0.786 (0.003)	0.804 (0.003)	0.791 (0.003)
	0.30	1.124 (0.003)	1.263 (0.005)	1.237 (0.005)

Table 7: Mean length (SE) of 5,000 95% percentile bootstrap confidence intervals for μ. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap.

n	ϵ	RB	NR	NR2
30	0.00	0.936	0.938	0.936
	0.10	0.941	0.944	0.940
	0.20	0.893	0.896	0.900
	0.30	0.816	0.716	0.827
50	0.00	0.940	0.944	0.940
	0.10	0.944	0.950	0.946
	0.20	0.920	0.928	0.927
	0.30	0.801	0.726	0.808
100	0.00	0.944	0.945	0.943
	0.10	0.948	0.951	0.947
	0.20	0.943	0.952	0.949
	0.30	0.802	0.770	0.807

Table 8: Empirical coverage of 5,000 95% percentile bootstrap confidence intervals for μ. Based on 2,000 bootstrap samples. RB denotes the robust bootstrap, NR the Newton-Raphson bootstrap and NR2 the modified Newton-Raphson bootstrap.

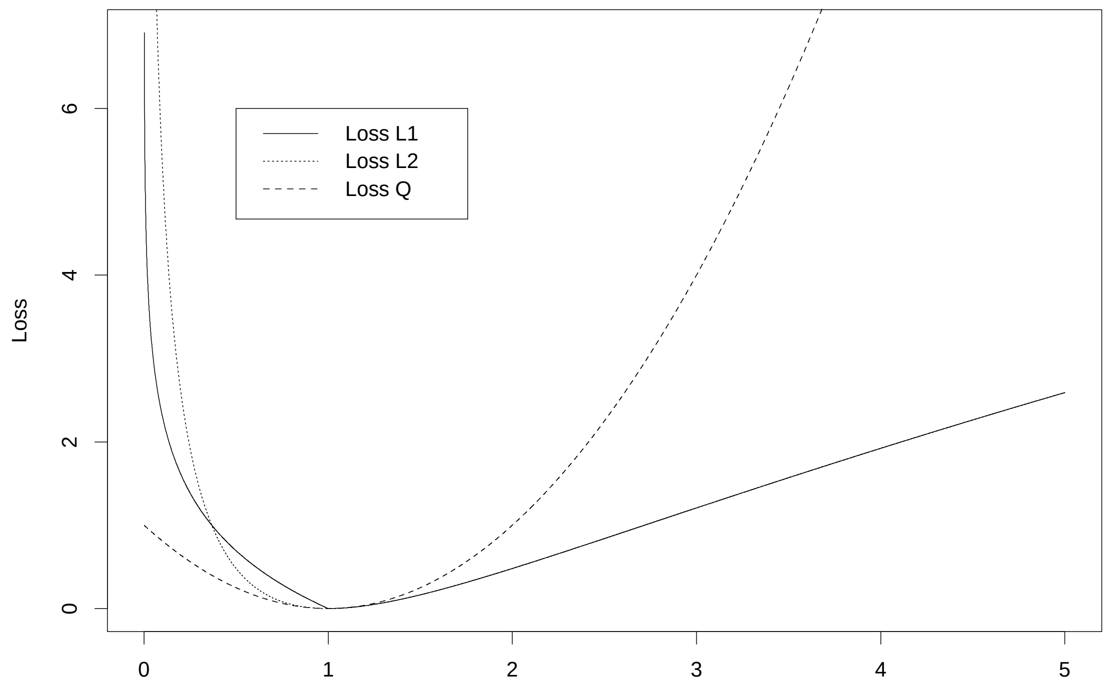

Figure 1: Different loss functions. See equations (12) [for Loss L1], (13) [for Loss L2] and (14) [for Loss Q].