Statistica Sinica **15**(2005), 981-1001

MULTIVARIATE REGRESSION S-ESTIMATORS FOR ROBUST ESTIMATION AND INFERENCE

Stefan Van Aelst and Gert Willems

Ghent University and University of Antwerp

Abstract: In this paper we consider S-estimators for multivariate regression. We study the robustness of the estimators in terms of their breakdown point and influence function. Our results extend results on S-estimators in the context of univariate regression and multivariate location and scatter. Furthermore we develop a fast and robust bootstrap method for the multivariate S-estimators to obtain inference for the regression parameters. Extensive simulation studies are performed to investigate finite-sample properties. The use of the S-estimators and the fast, robust bootstrap method is illustrated on some data.

Key words and phrases: Breakdown point, influence function, robust bootstrap.

1. Introduction

Consider the multivariate regression model given by $\mathbf{y} = \mathcal{B}^t\mathbf{x} + \boldsymbol{\epsilon}$ where $\mathbf{x}$ is the p-variate predictor and $\mathbf{y}$ the q-variate response. It is assumed that the q-variate error term $\boldsymbol{\epsilon}$ has an elliptically contoured density with zero center and scatter parameter $\Sigma \in \mathrm{PDS}(q)$. The unknown parameters $\mathcal{B}$ and Σ are to be estimated from an observed data set $\mathcal{Z}_n = \{\mathbf{z}_i := (\mathbf{x}_i^t, \mathbf{y}_i^t)^t, i = 1, \ldots, n\} \subset I\!R^{p+q}$. We assume that the errors $\boldsymbol{\epsilon}_i$ are i.i.d. and independent of the predictors. Note that this model generalizes the univariate regression model ($q = 1$) as well as the multivariate location/scatter model ($p = 1$; $x_i = 1$ for $i = 1, \ldots, n$).

The classical estimator for this model is the least squares estimator (see e.g., Johnson and Wichern (1988, p.301)). It is well known however that this estimator is extremely sensitive to outliers in the data. Therefore, several robust alternatives have been investigated in the literature, although most of the research has been limited to the univariate regression case. An overview of strategies for robust multivariate regression is given by Maronna and Yohai (1997) in the context of simultaneous equations models. An M-type method was proposed by Koenker and Portnoy (1990), but their estimator lacks affine equivariance. Methods based on the robust estimation of the location and scatter of the joint distribution of the $(\mathbf{x}, \mathbf{y})$ variables have been introduced by Rousseeuw et al. (2004) and Ollila et al. (2002, 2003). Agulló et al. (2002) investigated a multivariate generalization of the least trimmed squares estimator (MLTS).

In this paper we investigate S-estimators for multivariate regression. S-estimators for univariate multiple regression were introduced by Rousseeuw and Yohai (1984), and S-estimators for multivariate location and scatter have been studied by Davies (1987), Rousseeuw and Leroy (1987) and Lopuhaä (1989). Bilodeau and Duchesne (2000) introduced S-estimators in the context of Seemingly Unrelated Regression, which encompasses the multivariate regression model as a special case.

We study the robustness of these S-estimators and also develop a method for robust inference concerning the regression parameters $\mathcal{B}$. The standard errors of the S-estimates can be approximated by using their asymptotic variances (see Section 3). However, the asymptotic results only hold for some specified underlying model distribution such as the central normal model. They are not likely to yield accurate approximations in situations with outliers where robust estimators are actually recommended. The sampling distribution of S-estimators can also be estimated by the bootstrap method (Efron (1979)), but performing classical bootstrap on S-estimators can be extremely time-consuming and suffers from a lack of robustness. Recently Salibian-Barrera and Zamar (2002) introduced a fast and robust bootstrap method for MM-estimators of univariate regression. We adapt their method to S-estimators of multivariate regression.

The rest of the paper is organized as follows. In Section 2 we define the multivariate regression S-estimators, investigate their breakdown points and derive their influence functions. The asymptotic variance and corresponding efficiency results are given in Section 3. In Section 4 we describe the fast and robust bootstrap method for S-estimators and investigate its performance through simulations. Some examples are presented in Section 5, some remarks in Section 6. Proofs are omitted and can be found in Van Aelst and Willems (2004).

2. Definitions and Robustness Properties

The S-estimators for multivariate regression are a natural extension of the corresponding estimators for univariate regression and multivariate location and scatter.

Definition 1. Let $\mathcal{Z}_n = \{(\mathbf{x}_i^t, \mathbf{y}_i^t)^t; i = 1, \ldots, n\} \subset \mathbb{R}^{p+q}$ with $n \geq p + q$. The S-estimators of multivariate regression $(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n)$ minimize the determinant $\det(C)$ subject to

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left([(\mathbf{y}_i - B^t\mathbf{x}_i)^t C^{-1}(\mathbf{y}_i - B^t\mathbf{x}_i)]^{\frac{1}{2}}\right) = b$$

among all $(B, C) \in \mathbb{R}^{p\times q} \times \mathrm{PDS}(q)$.

The constant b can be chosen such that $b = E_F[\rho(\|\mathbf{r}\|)]$, which assures consistency at the model with error distribution F. In order to obtain positive

breakdown robust estimates, the function ρ is assumed to satisfy the following properties:

1. ρ is symmetric, twice continuously differentiable and $\rho(0) = 0$;
2. ρ is strictly increasing on $[0, c]$ and constant on $[c, \infty)$ for some $c < \infty$.

A popular choice is Tukey's biweight ρ-function:

$$\rho(t) = \begin{cases} \frac{t^2}{2} - \frac{t^4}{2c^2} + \frac{t^6}{6c^4}, & |t| \le c, \\ \frac{c^2}{6}, & |t| \ge c. \end{cases} \tag{2.1}$$

Following Lopuhaä (1989), multivariate regression S-estimators satisfy a first order condition given by the following equations:

$$\frac{1}{n}\sum_{i=1}^{n} u(d_i)\mathbf{x}_i(\mathbf{y}_i - B^t\mathbf{x}_i)^t = \mathbf{0}, \tag{2.2}$$

$$\frac{1}{n}\sum_{i=1}^{n} \left\{ q\, u(d_i)(\mathbf{y}_i - B^t\mathbf{x}_i)(\mathbf{y}_i - B^t\mathbf{x}_i)^t - v(d_i)C \right\} = \mathbf{0}, \tag{2.3}$$

where $d_i^2 = (\mathbf{y}_i - B^t\mathbf{x}_i)^t C^{-1}(\mathbf{y}_i - B^t\mathbf{x}_i)$, $u(t) = \rho'(t)/t$ and $v(t) = \rho'(t)t - \rho(t) + b$.

To investigate the global robustness of the S-estimators, we compute their breakdown point. For a given data set $\mathcal{Z}_n$ the finite-sample breakdown point (Donoho and Huber (1983)) of a regression estimator T_n is defined as the smallest fraction of observations of $\mathcal{Z}_n$ that need to be replaced to carry T_n beyond all bounds. Formally,

$$\epsilon_n^*(T_n, \mathcal{Z}_n) = \min\{\frac{m}{n} : \sup_{\mathcal{Z}_n'} \|T_n(\mathcal{Z}_n) - T_n(\mathcal{Z}_n')\| = \infty\},$$

where the supremum is over all possible collections $\mathcal{Z}_n'$ that differ from $\mathcal{Z}_n$ in at most m points. The breakdown point of a covariance estimator is the smallest fraction of outliers that can make the first eigenvalue arbitrarily large or the last eigenvalue arbitrarily small.

The breakdown point of multivariate regression S-estimators is given below. This result extends the results for multivariate location and scatter given by Lopuhaä and Rousseeuw (1991) and the results for univariate regression given by Rousseeuw and Yohai (1984).

Theorem 1. *Let $\mathcal{Z}_n \subset \mathbb{R}^{p+q}$. Denote by $k(\mathcal{Z}_n)$ the maximal number of observations lying on the same hyperplane of $\mathbb{R}^{p+q}$, and take $r := b/\rho(\infty)$. If $k(\mathcal{Z}_n) < \lceil n - nr \rceil$,*

$$\epsilon_n^*(\widehat{\mathcal{B}}_n, \mathcal{Z}_n) = \epsilon_n^*(\widehat{\Sigma}_n, \mathcal{Z}_n) = \frac{1}{n}\min(\lceil nr \rceil, \lceil n - nr \rceil - k(\mathcal{Z}_n)).$$

The maximal breakdown point is achieved when $r = (n - k(\mathcal{Z}_n))/(2n)$, in which case $\epsilon_n^* = \lceil (n-k(\mathcal{Z}_n))/2) \rceil /n$. If we assume that $\lim_{n\to\infty} k(\mathcal{Z}_n)/n = 0$, we can define the asymptotic breakdown point as $\epsilon^* = \lim_{n\to\infty} \epsilon_n^*$. The condition $\lim_{n\to\infty} k(\mathcal{Z}_n)/n = 0$ is satisfied when the data are *in general position*, which means that $k(\mathcal{Z}_n) = p + q - 1$ (Rousseeuw and Leroy (1987, p.117)). For the S-estimators we then have that $\epsilon^* = r = b/\rho(\infty)$, if $r \leq 0.5$. Hence, in order to achieve a specified breakdown point ϵ^* as well as consistency at the normal model, the constant c in Tukey's biweight (2.1) should be chosen as the solution to $E_\Phi[\rho(\|\mathbf{r}\|)]/(c^2/6) = \epsilon^*$. Here, Φ indicates the multivariate standard normal distribution. It is easy to see that such a solution for c can always be found. Furthermore, note that the value of c only depends on the dimension q of the responses. In the remainder of this paper we refer to the 25% *and* 50% *breakdown S-estimators* without explicitly mentioning that we use the Tukey biweight and have chosen c and b in order to obtain consistency at the normal model.

In order to obtain the influence function we first introduce the functional form of the multivariate regression S-estimators. Let $\mathcal{H}$ denote the class of all distributions on $\mathbb{R}^{p+q}$.

Definition 2. The S-functional $\mathbf{S} : \mathcal{H} \to (\mathbb{R}^{p\times q} \times \mathrm{PDS}(q))$ is the solution $\mathbf{S}(H) = (\mathcal{B}_S(H), \Sigma_S(H))$ to the problem of minimizing $\det(C)$, subject to

$$\int \rho\left([(\mathbf{y} - B^t\mathbf{x})^t C^{-1} (\mathbf{y} - B^t\mathbf{x})]^{1/2} \right) dH(\mathbf{z}) = b,$$

among all $(B, C) \in \mathbb{R}^{p\times q} \times \mathrm{PDS}(q)$.

The influence function of a functional T at a distribution H measures the effect on T of an infinitesimal contamination at a single point (Hampel et al. (1986)). It can be viewed as a measure of *local* robustness, in contrast with the global robustness measured by the breakdown point. If we denote the point mass at $\mathbf{z} = (\mathbf{x}^t, \mathbf{y}^t)^t$ by $\Delta_\mathbf{z}$ and consider the contaminated distribution $H_{\epsilon,\mathbf{z}} = (1-\epsilon)H + \epsilon\Delta_\mathbf{z}$, then the influence function is given by

$$IF(\mathbf{z}; T, H) = \lim_{\epsilon\to 0} \frac{T(H_{\epsilon,\mathbf{z}}) - T(H)}{\epsilon} = \frac{\partial}{\partial\epsilon} T(H_{\epsilon,\mathbf{z}})|_{\epsilon=0}.$$

Now consider a model with unimodal elliptically symmetric error distribution F_Σ. That is, errors have a density function of the form $f_\Sigma(u) = \det(\Sigma)^{-1/2} g(u^t \Sigma^{-1} u)$, where $\Sigma \in \mathrm{PDS}(q)$ and the function g has a strictly negative derivative. Note that, provided it exists, the covariance matrix of F_Σ equals Σ except for a multiplicative constant. It can easily be seen that the S-estimators are equivariant under regression transformations and affine transformations of the regressors or responses. Therefore it suffices to compute the influence function at a distribution

H_0 of $\mathbf{z} = (\mathbf{x}^t, \mathbf{y}^t)^t$ for which $\mathcal{B} = 0$, and at the error distribution $F_0 = F_{I_q}$. Since $\mathcal{B} = 0$ the error distribution is also the distribution of $\mathbf{y}$.

Theorem 2. *For a model H_0 as described above, the influence functions of the S-estimators for multivariate regression are given by*

$$\mathrm{IF}(\mathbf{z}; \mathcal{B}_S, H_0) = \mathrm{E}_{H_0}[\mathbf{x}\mathbf{x}^t]^{-1}\mathbf{x}\mathrm{IF}(\mathbf{y}; M_q, F_0)^t, \tag{2.4}$$

$$\mathrm{IF}(\mathbf{z}; \Sigma_S, H_0) = \mathrm{IF}(\mathbf{y}; S_q, F_0), \tag{2.5}$$

with (M_q, S_q) the q-dimensional S-estimators for location and scatter.

The influence functions of (M_q, S_q) are given in Lopuhaä (1989). In particular, for the location part we have that

$$\mathrm{IF}(\mathbf{y}; M_q, F_0) = \frac{1}{\beta}\rho'(\|\mathbf{y}\|)\frac{\mathbf{y}}{\|\mathbf{y}\|},$$

where $\beta = E_{F_0}[(1 - 1/q)u(\|\mathbf{y}_0\|) + (1/q)\rho''(\|\mathbf{y}_0\|)]$ and $u(t) = \rho'(t)/t$ as before. For the covariance matrix, Lopuhaä (1989) obtained

$$\mathrm{IF}(\mathbf{y}; S_q, F_0) = \frac{1}{\gamma_1}\rho'(\|\mathbf{y}\|)\|\mathbf{y}\|q\Big(\frac{\mathbf{y}\mathbf{y}^t}{\|\mathbf{y}\|^2} - \frac{1}{q}I_q\Big) + \frac{2}{\gamma_3}(\rho(\|\mathbf{y}\|) - b)I_q,$$

where $\gamma_1 = E_{F_0}[\rho''(\|\mathbf{y}_0\|)\|\mathbf{y}_0\|^2 + (q+1)\rho'(\|\mathbf{y}_0\|)\|\mathbf{y}_0\|]/(q+2)$ and $\gamma_3 = E_{F_0}[\rho'(\|\mathbf{y}_0\|)\|\mathbf{y}_0\|]$. Note that the influence function of $\mathcal{B}_S$ is bounded in $\mathbf{y}$ but unbounded in $\mathbf{x}$, so good leverage points can have a high effect on the S-estimator. The latter can also be seen from Figure 2.1 where we plotted the influence function at the normal model in the case of univariate simple regression with intercept ($q = 1$, $p = 2$).

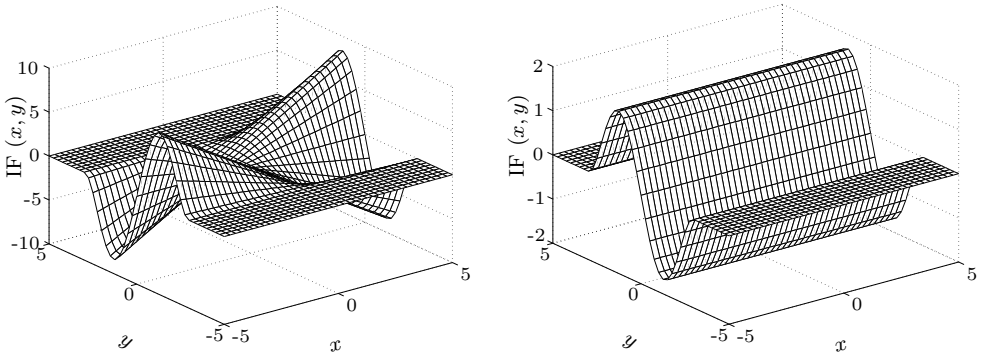

Figure 2.1. Influence function at the normal model of the 25% breakdown biweight S-estimator for simple regression: slope (left) and intercept (right).

3. Efficiency

As noted by Bilodeau and Duchesne (2000), the multivariate regression S-estimators satisfy first-order conditions of M-estimators as defined in Huber (1981). Hence we conclude that the estimators are asymptotically normal with convergence rate $n^{1/2}$ (by Theorem 3.1, Chap. 6 of Huber (1981)).

The asymptotic variance-covariance matrix of $\mathcal{B}_S$ at the model distribution H_0 can be computed by means of the influence function, as

$$ASV(\mathcal{B}_S, H_0) = \mathrm{E}_{H_0}[IF(\mathbf{z}; \mathcal{B}_S, H_0) \otimes IF(\mathbf{z}; \mathcal{B}_S, H_0)^t]$$

(see Hampel et al. (1986, pp.85 and 226). Here $\otimes$ denotes the Kronecker product. Denoting $\Sigma_{\mathbf{x}} := \mathrm{E}_{H_0}[\mathbf{x}\mathbf{x}^t]$, it follows from (2.4) that

$$ASV(\mathcal{B}_S, H_0) = K_{pq}(ASV(M_q, F_0) \otimes \Sigma_{\mathbf{x}}^{-1}), \tag{3.1}$$

where K_{pq} is the commutation matrix, the permutation matrix satisfying K_{pq} $\mathrm{vec}(A^t) = \mathrm{vec}(A)$ where A is a $p \times q$ matrix and vec is the operator which stacks the columns of a matrix on top of each other.

From (3.1) we find that the asymptotic variance of $(\mathcal{B}_S)_{jk}$ is

$$ASV((\mathcal{B}_S)_{jk}, H_0) = (\Sigma_{\mathbf{x}}^{-1})_{jj} ASV((M_q)_k, F_0),$$

while the asymptotic covariances, for $j \neq j'$, are given by

$$ASC((\mathcal{B}_S)_{jk}, (\mathcal{B}_S)_{j'k}, H_0) = (\Sigma_{\mathbf{x}}^{-1})_{jj'} ASV((M_q)_k, F_0),$$

and all other asymptotic covariances (for $k \neq k'$) equal 0. The asymptotic variances of M_q can also be found in Lopuhaä (1989). We now compute the asymptotic relative efficiency of S-estimators with respect to the least squares (LS) estimator. Due to affine equivariance, we can assume without loss of generality that $\Sigma_{\mathbf{x}} = I_p$. In this case all asymptotic covariances are zero, and $ASV((\mathcal{B}_S)_{jk}, H_0) = ASV((M_q)_k, F_0)$. We then obtain

$$\begin{aligned} ARE((\mathcal{B}_S)_{jk}, H_0) &= \frac{ASV((\mathcal{B}_{LS})_{jk}, H_0)}{ASV((\mathcal{B}_S)_{jk}, H_0)} = \frac{ASV((\overline{X})_k, F_0)}{ASV((M_q)_k, F_0)} \\ &= ARE((M_q)_k, F_0) \end{aligned}$$

for all $j = 1, \ldots, p$ and $k = 1, \ldots, q$. Hence the asymptotic relative efficiency of a multivariate regression S-estimator does not depend on the dimension p of the carriers nor on their distribution, but only on the dimension q and distribution of the errors. Table 3.1 lists some relative efficiencies for the normal distribution, as well as for multivariate Student distributions T_ν with degrees of freedom $\nu = 3$

and $\nu = 8$. Note that for the functional $\overline{X}$, corresponding to the empirical mean, $ASV((\overline{X})_k, \Phi) = 1$ and $ASV((\overline{X})_k, T_\nu) = \nu/(\nu - 2)$. We see that the asymptotic efficiency of S-estimators at the normal model is quite high and increases with the dimension q. Moreover, at the heavy tailed Student distributions the S-estimators are generally even more efficient than the classical least squares estimator. It is also interesting to note that, while at the normal model the 25% breakdown S-estimator is always more efficient than its 50% counterpart, this is not true anymore at the Student distributions. When the number of response variables increases, the 50% breakdown estimator gains efficiency with respect to the 25% estimator, and eventually becomes more efficient.

Table 3.1. Asymptotic relative efficiencies for $(\mathcal{B}_S)_{jk}$ w.r.t. the LS estimator at normal and Student distributions.

		$q=1$	$q=2$	$q=3$	$q=5$	$q=10$	$q=30$	$q=50$
$\epsilon^* = 25\%$	Φ	0.759	0.912	0.951	0.976	0.990	0.997	0.998
	T_8	0.894	1.059	1.108	1.141	1.162	1.173	1.174
	T_3	1.738	2.035	2.137	2.222	2.289	2.336	2.346
$\epsilon^* = 50\%$	Φ	0.287	0.580	0.722	0.846	0.933	0.981	0.989
	T_8	0.390	0.739	0.897	1.038	1.153	1.228	1.250
	T_3	0.904	1.601	1.903	2.177	2.410	2.583	2.620

We conclude this section with some finite-sample relative efficiencies, obtained through simulation. For several dimensions and sample sizes, we generated $m = 1,000$ random samples with both the errors and the predictors drawn from the standard multivariate normal distribution. An intercept term was included as well. The entries in the matrix $\mathcal{B}$ were all set equal to 1, but this particular choice does not matter due to affine equivariance. For each sample we computed the S-estimates. The Monte Carlo variance of $\widehat{\mathcal{B}}_n$ is measured here as $n \operatorname*{ave}_{j,k}(\widehat{\mathrm{Var}}((\widehat{\mathcal{B}}_n)_{jk}))$ for $j = 1, \ldots, p$ and $k = 1, \ldots, q$, where $\widehat{\mathrm{Var}}((\widehat{\mathcal{B}}_n)_{jk})$ is the empirical variance over the m estimates. The finite-sample relative efficiency is then estimated by the inverse of this variance estimate for the normal distribution, and by $\nu/(\nu - 2)$ over the variance estimate for the T_ν distribution. Table 3.2. lists finite-sample relative efficiencies for the 25% breakdown S-estimator for the normal and T_3 model. The results for $q = 2$ and $q = 5$ are, respectively, obtained from simulations using the multivariate regression model with $p = 2$ and $p = 5$. We see that the finite-sample relative efficiencies are generally slightly lower than the asymptotic relative efficiencies. Results for the 50% breakdown estimator were found to be similar.

Table 3.2. Finite-sample relative efficiencies for the 25% breakdown S-estimator $\widehat{\mathcal{B}}_n$ w.r.t. the LS estimator at the normal and T_3 distribution.

		$n = 30$	$n = 50$	$n = 100$	$n = 200$	$n = \infty$
Φ	$q = 2$	0.840	0.875	0.938	0.908	0.912
	$q = 5$	0.778	0.865	0.917	0.956	0.976
T_3	$q = 2$	1.704	1.889	2.005	2.003	2.035
	$q = 5$	1.392	1.706	1.887	1.951	2.222

4. Robust Inference

4.1. Fast and robust bootstrap

We now consider inference about the regression parameter $\mathcal{B}$. Inference can be based on the asymptotic variance, but this is not expected to give accurate results when the actual errors are not symmetric, e.g., when outliers are present. An alternative approach is given by the nonparametric bootstrap (Efron (1979)) which is, unlike the asymptotic approach, not particularly based on stringent distributional assumptions. The use of the bootstrap method is increasing enormously nowadays, due to increasing computer power. The basic idea is to generate a large number of samples from the original data set, and to recalculate the estimates for each resample. Then the distribution of, e.g., $\sqrt{n}(\widehat{\mathcal{B}}_n - \mathcal{B})$ can be approximated by the sample distribution of $\sqrt{n}(\widehat{\mathcal{B}}^*_n - \widehat{\mathcal{B}}_n)$, where $\widehat{\mathcal{B}}^*_n$ is the value of the recalculated estimator. However, there are two important drawbacks of the classical bootstrap method applied to S-estimators. First, although Ruppert (1992) provided a reasonably fast algorithm to compute S-estimators, they still are computer intensive. When the classical bootstrap is used to obtain percentile confidence intervals, for example, many resamples are to be generated and the S-algorithm has to be applied on each of those bootstrap samples. The minimum number of resamples needed for sufficiently accurate confidence intervals is often taken to be about 1,000, and this number should be adjusted exponentially when interested in the joint distribution of several parameters. Hence, the method may not be feasible due to the computational cost, especially for high dimensional data.

The second problem that arises concerns the robustness of the method. Even if the estimator is resistant to the proportion of outlying observations in the original data set, when taking a bootstrap sample this proportion can become high enough to break down the estimator for that particular resample. Consequently, inference based on the resulting bootstrap distribution can break down even if the S-estimate in the original sample does not. Singh (1998) and Stromberg (1997) quantified this problem in the context of robust location estimation.

Salibian-Barrera and Zamar (2002), in the context of MM-estimators for univariate regression, proposed a procedure to compute bootstrap values of $\widehat{\mathcal{B}}^*_n$

without explicitly calculating the actual MM-estimate each time. For each resample they compute an approximation of $\widehat{\mathcal{B}}_n^*$ based on a fixed-point representation of the estimator. Using this approximation rather than the actual MM-algorithm in each bootstrap sample results in a considerable gain in computation time. Furthermore, since outliers are downweighted no matter how many outliers are present in a resample, the method is more robust than the classical bootstrap. Their method can, in principle, be used for any estimator that can be written as a smooth fixed-point equation, such as S-estimators.

Suppose an estimator $\widehat{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$ can be represented by $\mathbf{g}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\theta}}$, where the function $\mathbf{g}$ involves the sample $\mathcal{Z}_n$. Then, using the smoothness of $\mathbf{g}$, we can calculate a Taylor expansion about the limiting value of the estimate $\widehat{\boldsymbol{\theta}}$,

$$\widehat{\boldsymbol{\theta}} = \mathbf{g}(\boldsymbol{\theta}) + \nabla\mathbf{g}(\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + R, \tag{4.1}$$

where R is the remainder term and $\nabla\mathbf{g}(\cdot)$ is the matrix of partial derivatives. Supposing that the remainder term is small, (4.1) can be rewritten as $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \approx [\mathbf{I} - \nabla\mathbf{g}(\boldsymbol{\theta})]^{-1}\sqrt{n}(\mathbf{g}(\boldsymbol{\theta}) - \boldsymbol{\theta})$. Taking bootstrap equivalents on both sides and estimating the matrix $[\mathbf{I} - \nabla\mathbf{g}(\boldsymbol{\theta})]^{-1}$ by $[\mathbf{I} - \nabla\mathbf{g}(\widehat{\boldsymbol{\theta}})]^{-1}$ yields

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}) \approx [\mathbf{I} - \nabla\mathbf{g}(\widehat{\boldsymbol{\theta}})]^{-1}\sqrt{n}(\mathbf{g}^*(\widehat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\theta}}), \tag{4.2}$$

where the function $\mathbf{g}^*$ is the function $\mathbf{g}$ computed with a bootstrap sample instead of the original sample $\mathcal{Z}_n$. For each bootstrap sample, we can then calculate the right-hand side of (4.2) instead of the left-hand side. Hence, we approximate the actual estimator in each sample by computing the function $\mathbf{g}^*$ in $\widehat{\boldsymbol{\theta}}$ and applying a linear correction.

We now apply this procedure to our multivariate regression S-estimators. We rewrite the estimating equations (2.2) and (2.3) in the following way:

$$\widehat{\mathcal{B}}_n = \mathbf{A}_n(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n)^{-1}\mathbf{B}_n(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n), \tag{4.3}$$

$$\widehat{\Sigma}_n = \mathbf{V}_n(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n) + w_n(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n)\widehat{\Sigma}_n, \tag{4.4}$$

where

$$\mathbf{A}_n(B, C) = \sum_{i=1}^{n} u(d_i)\mathbf{x}_i\mathbf{x}_i^t \qquad (p \times p), \tag{4.5}$$

$$\mathbf{B}_n(B, C) = \sum_{i=1}^{n} u(d_i)\mathbf{x}_i\mathbf{y}_i^t \qquad (p \times q), \tag{4.6}$$

$$\mathbf{V}_n(B, C) = \frac{1}{nb}\sum_{i=1}^{n} q\, u(d_i)(\mathbf{y}_i - B^t\mathbf{x}_i)(\mathbf{y}_i - B^t\mathbf{x}_i)^t \quad (q \times q), \tag{4.7}$$

$$w_n(B, C) = \frac{1}{nb}\sum_{i=1}^{n} w(d_i) \qquad (1 \times 1), \tag{4.8}$$

and $w(t) = \rho(t) - \rho'(t)t$. Note that there are different ways to transform (2.2) and (2.3) into fixed-point equations. We prefer (4.3) and (4.4) because other formulations turned out to be numerically unstable in certain situations.

Now let

$$\boldsymbol{\theta} := \begin{pmatrix} \text{vec}(\mathcal{B}) \\ \text{vec}(\Sigma) \end{pmatrix} \quad \text{and} \quad \mathbf{g}\begin{pmatrix} \text{vec}(B) \\ \text{vec}(C) \end{pmatrix} := \begin{pmatrix} \text{vec}(\mathbf{A}_n^{-1}\mathbf{B}_n) \\ \text{vec}(\mathbf{V}_n + w_n C) \end{pmatrix}.$$

Expressions for the partial derivatives of this function, contained in the matrix $\nabla\mathbf{g}(\cdot)$, are given in Van Aelst and Willems (2004).

For a bootstrap sample $\{(\mathbf{x}_i^*, \mathbf{y}_i^*), i = 1, \ldots, n\}$ we then have that

$$\mathbf{g}^*(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} \text{vec}(\mathbf{A}_n^*(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n)^{-1}\mathbf{B}_n^*(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n)) \\ \text{vec}(\mathbf{V}_n^*(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n) + w_n^*(\widehat{\mathcal{B}}_n, \widehat{\Sigma}_n)\widehat{\Sigma}_n) \end{pmatrix}, \tag{4.9}$$

where $\mathbf{A}_n^*, \mathbf{B}_n^*, \mathbf{V}_n^*$ and w_n^* are the bootstrap versions of quantities (4.5) to (4.8), that is, with $(\mathbf{x}_i, \mathbf{y}_i)$ replaced by $(\mathbf{x}_i^*, \mathbf{y}_i^*)$. Thus, in order to get the values of $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})$ for each bootstrap sample, we calculate (4.9), apply the linear correction given by the matrix of partial derivatives, and use approximation (4.2).

To generate bootstrap samples in a regression setup, one can either use case resampling or error resampling (see e.g., Davison and Hinkley (1997)). The former assumes random explanatory variables, the latter assumes a fixed design. In this paper we use the case resampling method, which means that we generate resamples by drawing with replacement from the observations $\{(\mathbf{x}_i, \mathbf{y}_i), i = 1, \ldots, n\}$, but the method can also be applied in the case of error resampling. In particular, Salibian-Barrera (2003a, 2003b) recently derived the validity of the robust bootstrap for fixed design, in case of univariate MM-estimators.

Let us now focus on confidence intervals resulting from this fast bootstrap procedure. One way to characterize the robustness of bootstrap confidence intervals is to define the breakdown point of a bootstrap quantile estimate for a statistic T_n. For $t \in [0, 1]$, let Q_t^* denote the tth quantile of the bootstrap sample distribution of T_n^*:

$$Q_t^* = \min\{x : \frac{1}{R} \times \#\{T_n^{*j} \geq x; j = 1, \ldots, R\} \leq t\},$$

where R is the number of bootstrap samples drawn. Singh (1998) defined the upper breakdown point of a statistic as the minimum proportion of asymmetric contamination that can carry the statistic over any bound. Let us now define the *expected* upper breakdown point of the bootstrap quantile Q_t^* as the minimum proportion of asymmetric contamination that is expected to be able to carry Q_t^* over any bound, where the expectation is taken over the distribution of drawing

R samples with replacement. It is easy to see that the expected upper breakdown point for the classical bootstrap quantile is given by the following formula:

$$\epsilon_n^E(C) = \inf\{\delta \in [0,1] : P(\mathrm{Bin}(n,\delta) \geq \lceil \epsilon_n^* n \rceil) \geq t\}$$

(Singh (1998)). Here ϵ_n^* is the breakdown point for the bootstrapped estimator. In other words, the bootstrap quantile estimate Q_t^* can be severely affected by outliers when $\tau^* > t$, where τ^* denotes the expected proportion of bootstrap samples containing more than $\epsilon_n^* n$ outliers. This result can be applied to the multivariate regression S-estimators.

As for the fast bootstrap, it can be shown that the recalculation in the bootstrap sample will not break down as long as that bootstrap sample contains at least p non-outlying observations in general position. Denote by $B_D(n,\delta)$ the number of distinct non-outlying observations in a resample of size n, drawn with replacement from a sample of size n with a proportion δ of outliers.

Theorem 3. *Let $\mathcal{Z}_n \subset \mathbb{R}^{p+q}$ and assume $k(\mathcal{Z}_n) = p + q - 1$. Let ϵ_n^* be the breakdown point of an S-estimate $\widehat{\mathcal{B}}_n$. Then the expected upper breakdown point of the tth fast bootstrap quantile of any regression parameter $\mathcal{B}_{jk}$, $j = 1,\ldots,p$; $k = 1,\ldots,q$, is given by $min(\epsilon_n^*, \epsilon_n^E(R))$, where*

$$\epsilon_n^E(R) = inf\{\delta \in [0,1] : P(B_D(n,\delta) < p) \geq t\}.$$

Table 4.1. Expected upper breakdown values for classical bootstrap and fast and robust bootstrap on maximal breakdown S-estimators.

		$p=2, q=1$				$p=8, q=2$			
	n	10	30	50	100	20	30	50	100
$Q^*_{0.05}$	$\epsilon_n^E(C)$	0.15	0.31	0.36	0.40	0.10	0.22	0.30	0.37
	$\epsilon_n^E(R)$	0.50	0.50	0.50	0.50	0.30	0.50	0.50	0.50
$Q^*_{0.005}$	$\epsilon_n^E(C)$	0.08	0.24	0.30	0.36	0.00	0.16	0.25	0.33
	$\epsilon_n^E(R)$	0.40	0.50	0.50	0.50	0.20	0.40	0.50	0.50

Table 4.1 lists some values for $\epsilon_n^E(C)$ and $\epsilon_n^E(R)$ for different dimensions and sample sizes, for the S-estimator with maximal breakdown point. Two different quantiles are considered, $Q^*_{0.05}$ and $Q^*_{0.005}$, which can, respectively, be used to construct 90% and 99% percentile confidence intervals. We see that both bootstrap methods can pose robustness problems when the ratio $n/(p+q)$ is very small. Otherwise the fast and robust bootstrap clearly yields a gain in breakdown point over the classical bootstrap. In particular the expected upper breakdown point for the fast bootstrap generally equals the breakdown value of the S-estimator itself. Analogous results can be found for the expected *lower* breakdown value.

Concerning the convergence of the fast bootstrap method, the univariate regression case is covered by the result of Salibian-Barrera and Zamar (2002), since S-estimators are a special case of MM-estimators. An extension of the proof to the multivariate setting is fairly straightforward, although some tedious calculations are needed to obtain the regularity conditions.

4.2. Simulation results

To investigate the performance of the fast bootstrap we carried out an extensive simulation study. We are primarily interested in the performance of confidence intervals for the regression coefficients based on the fast bootstrap. We would like those intervals to have a coverage that is close to the nominal value, while being relatively short.

Simulations were performed for sample sizes $n = 30, 50, 100$ and 200. We considered actual multivariate regression models ($p = 2$; $q = 2$ and $p = 5$; $q = 5$), univariate regression models ($p = 2$; $q = 1$ and $p = 5$; $q = 1$), and multivariate location and scatter models ($p = 1$; $q = 2$ and $p = 1$; $q = 5$). An intercept term was included in each regression model by setting the first entry in $\mathbf{x}_i$ equal to 1. The remaining predictor variables were generated from the $(p-1)$-variate Gaussian distributions $N(\mathbf{0}, I_{p-1})$. The true value of the parameter $\mathcal{B}$ was set to $\mathbb{1}_{p,q}$, the $p \times q$ matrix having 1 for each entry. As before, this choice does not affect the performance results due to the equivariance properties of the S-estimators. We considered the following situations in the simulation:

- normal errors, generated from $N(\mathbf{0}, I_q)$;
- long-tailed errors, generated from the multivariate Student distribution with 3 d.f. (T_3) and 1 d.f. (T_1, Cauchy distribution);
- vertical outliers, proportion $1-\delta$ of the errors generated from $N(\mathbf{0}, I_q)$ and proportion δ generated from $N(5\sqrt{\chi^2_{q;0.99}}\,\mathbb{1}_{q,1}, (1.5)^2 I_q)$, for $\delta = 0.15$ and $\delta = 0.30$;
- bad leverage points, proportion $1-\delta$ of the errors generated from $N(\mathbf{0}, I_q)$ and proportion δ generated from $N(10\,\mathbb{1}_{q,1}, 10\,I_q)$, with corresponding predictors substituted by predictors generated from $N(-10\,\mathbb{1}_{p-1,1}, 10\,I_{p-1})$, for $\delta = 0.15$ and $\delta = 0.30$.

The latter situation obviously was not applied to the location/scatter models. For each of these situations, and for each of the sample sizes and dimensions given above, we constructed 1,000 datasets and computed S-estimates with Tukey's biweight function. Both the 25% and the 50% breakdown estimators were considered. Next we applied the fast bootstrap to generate $R = 1,000$ recalculated values $(\widehat{\mathcal{B}}^*_n, \widehat{\Sigma}^*_n)$.

Bootstrap confidence intervals for the components $\mathcal{B}_{jk}$ were constructed using the bias corrected and accelerated (BCA) method (see e.g., Davison and Hinkley (1997, p.202). The bootstrap intervals are compared with confidence intervals based on the asymptotic normality of the S-estimator. The latter are of the form $[(\widehat{\mathcal{B}}_n)_{jk} - \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{V}_{jk}/n}, (\widehat{\mathcal{B}}_n)_{jk} + \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{V}_{jk}/n}]$ for a $100(1-\alpha)\%$ confidence interval, where $\widehat{V}_{jk}$ denotes the empirical version of the asymptotic variance (EASV) of the (j,k)th component of $\widehat{\mathcal{B}}_n$. Note that a comparison with the classical bootstrap method is not made due to the computational cost. Some of the results are presented in Figures 4.1−4.3, respectively, for multivariate regression ($p = 5$, $q = 5$), univariate regression ($p = 5$, $q = 1$), and

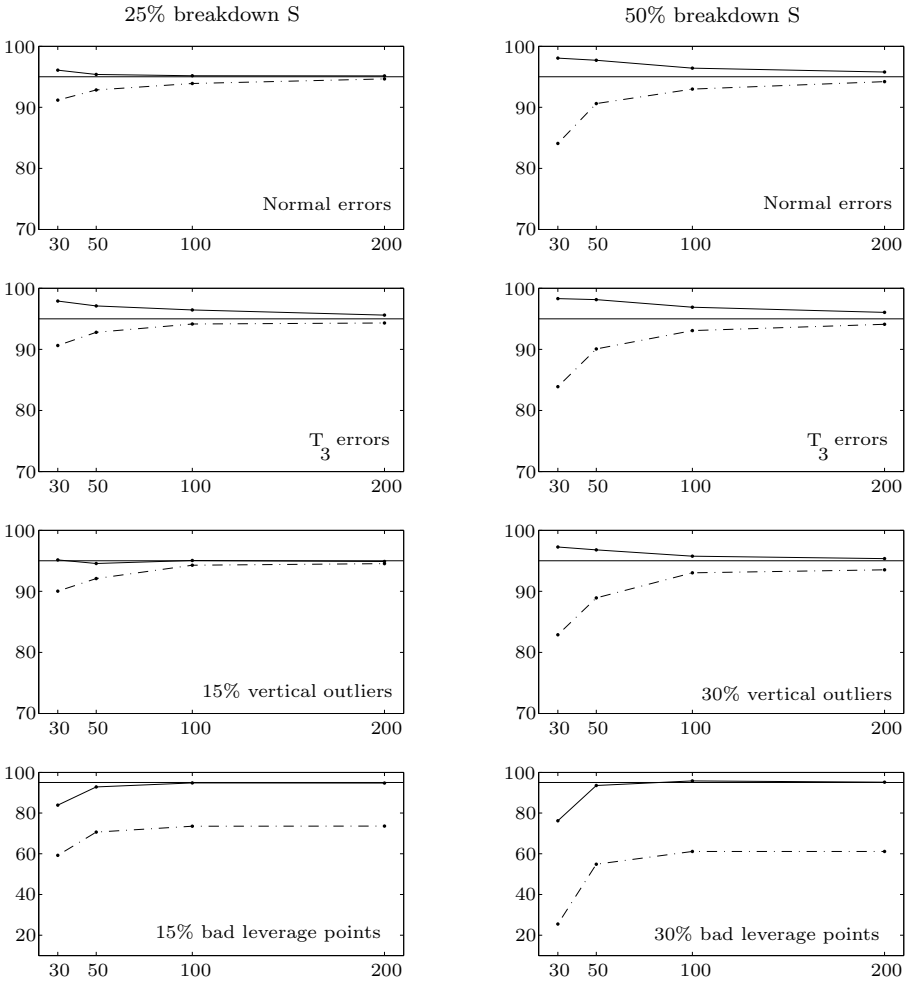

Figure 4.1. Coverage for 95% intervals, for fast bootstrap (solid) and EASV (dash-dotted): $p = 5$, $q = 5$.

multivariate location ($p = 1$, $q = 5$). These figures show the actual percentages of the 95% confidence intervals that were observed to contain the true value of the parameter (i.e., 1). The nominal value of 95% is indicated by the horizontal line. For Figures 4.1 and 4.2 we considered the intervals for the slope parameters (i.e., all parameters contained in $\mathcal{B}$ except for the intercept), while for Figure 4.3 all location parameters were considered (i.e., all parameters contained in $\mathcal{B}$). The left panels contain the results for 25% breakdown while the right panels show the 50% breakdown results. The results for $p = 2$ and $q = 2$ are very similar to the results shown here. Furthermore, results for the T_1 error distribution are omitted since they resemble the results for T_3.

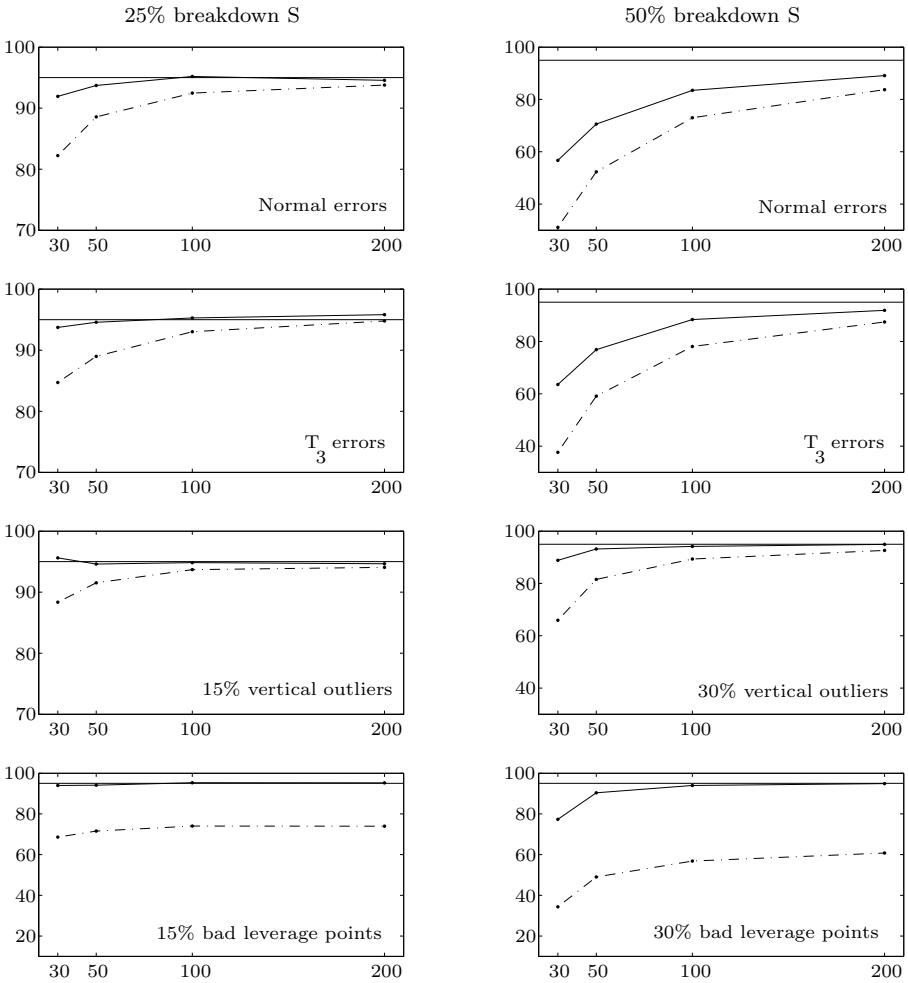

Figure 4.2. Coverage for 95% intervals, for fast bootstrap (solid) and EASV (dash-dotted): $p = 5$, $q = 1$.

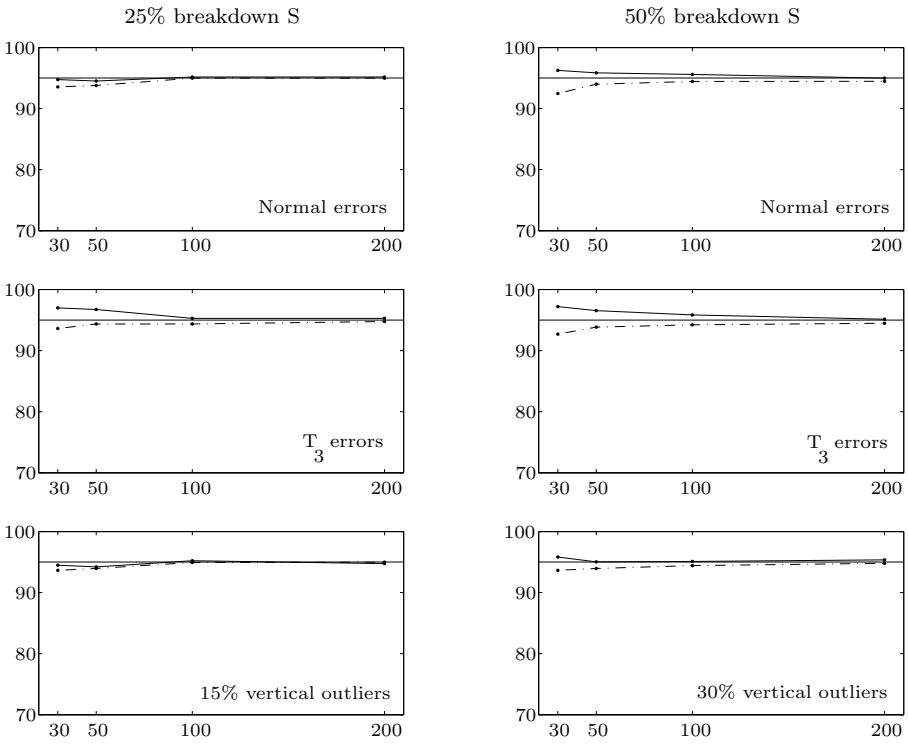

Figure 4.3. Coverage for 95% intervals, for fast bootstrap (solid) and EASV (dash-dotted): $p = 1$, $q = 5$.

In Figure 4.1 it can be seen that the coverage of the intervals based on the EASV (dash-dotted) is generally lower than 95%. It increases however to the nominal value as the sample size grows, except in the case of bad leverage points. In the latter case the EASV is not robust against outliers and hence the intervals are not robust either. The fast bootstrap (solid) outperforms the EASV method, especially in case of bad leverage points. For small sample sizes the fast bootstrap is generally somewhat conservative in this setup. For univariate regression, in Figure 4.2, one notices that for the 50% breakdown estimator the intervals have rather poor coverage in case of normal errors and T_3 errors. Results improve when the sample size grows however. In the other cases for the univariate setup, results for the fast bootstrap are very close to the expected coverage and, in all situations, the bootstrap yields much better results than the EASV method. Finally, for the multivariate location model presented in Figure 4.3, the actual coverage is always very close to the nominal coverage. As in the previous setups, the fast bootstrap is more conservative than the EASV method.

Furthermore, we found that the length of the bootstrap intervals is comparable to the length of the intervals based on the EASV and asymptotic nor-

mality. Table 4.2 shows the average length of the intervals obtained with the 25% breakdown S-estimator, for the cases of normal errors and 15% vertical outliers. Results for the other cases were similar. Naturally, intervals obtained with the 50% breakdown estimator were found to be somewhat longer (results not reported here), especially in the normal case, which is in accordance with the efficiency loss associated with higher breakdown. In general, EASV intervals are shorter than the bootstrap intervals. However, these EASV intervals also have a coverage that is too low and always lower than the bootstrap intervals, as can be seen in Figures 4.1−4.3.

Table 4.2. Average lengths for 95% confidence intervals based on 25% breakdown S-estimator.

	Normal errors				15% vertical outliers			
n	30	50	100	200	30	50	100	200
	$p = 5, q = 5$							
Boot	1.056	0.664	0.431	0.292	1.006	0.694	0.455	0.310
EASV	0.729	0.559	0.397	0.281	0.757	0.598	0.426	0.300
	$p = 5, q = 1$							
Boot	1.246	0.867	0.513	0.339	1.120	0.729	0.473	0.320
EASV	0.717	0.597	0.438	0.313	0.769	0.605	0.431	0.306
	$p = 1, q = 5$							
Boot	0.759	0.577	0.403	0.284	0.786	0.606	0.429	0.303
EASV	0.710	0.554	0.394	0.280	0.754	0.591	0.423	0.300

We can conclude that the fast bootstrap performs well in all situations considered. The method usually outperforms the competing EASV method and, in particular, the fast bootstrap is indeed robust against outliers in the data, as was expected. However, a word of caution is needed here with regard to the level of the fast bootstrap intervals. It should be noted that these confidence intervals are designed to reflect the sampling variability of the S-estimators and do not contain a correction for possible bias due to asymmetric contamination. Therefore, the intervals might not keep the nominal level. Correcting for bias due to contamination is a difficult task. Seminal work in this direction based on the concept of bias bounds has been done by Berrendero and Zamar (2001) and Adrover et al. (2004). Our simulation results indicate that bias is a minor concern for the outlier configurations that were used in this study, but there is no guarantee that this will be the case for any type of contamination.

Several algorithms are available to compute S-estimators, most of which use resampling and are rather slow. Ruppert (1992), however, provided an improved resampling algorithm for univariate regression and multivariate location and scatter which is fast and has good accuracy. In this paper, for the simulations as well

as for the examples in the next section, we used an extension of this algorithm to multivariate regression, similar to Bilodeau and Duchesne (2000).

5. Examples

In this section we present two examples illustrating the use of the multivariate S-estimators and the fast bootstrap method.

5.1. Milk data

First consider the data given by Daudin et al. (1988), which consist of 8 measurements on 85 bottles of milk. We would like to investigate a possible linear dependency of the variables 'Density' and 'Cheese produced' on some or all of the other variables. Figure 5.1 presents the diagnostic plots of the fit of the corresponding multivariate regression model, based on the least squares estimator and on the 25% breakdown S-estimator. These plots show the Mahalanobis or robust distances of the residuals versus the Mahalanobis or robust distances of the explanatory variables (see also Rousseeuw et al. (2004)). The horizontal and vertical lines, respectively, indicate the square roots of the 0.975-quantiles of the χ^2_q and the χ^2_{p-1} distribution. In this example we have $q = 2$ and $p = 7$. The S-estimator detects one extreme bad leverage point (observation 69), some vertical outliers which are less serious, and some good leverage points. The least squares estimator does not reveal observation 69 as an outlier. In fact, this observation with its high leverage pulls the least squares fit toward it.

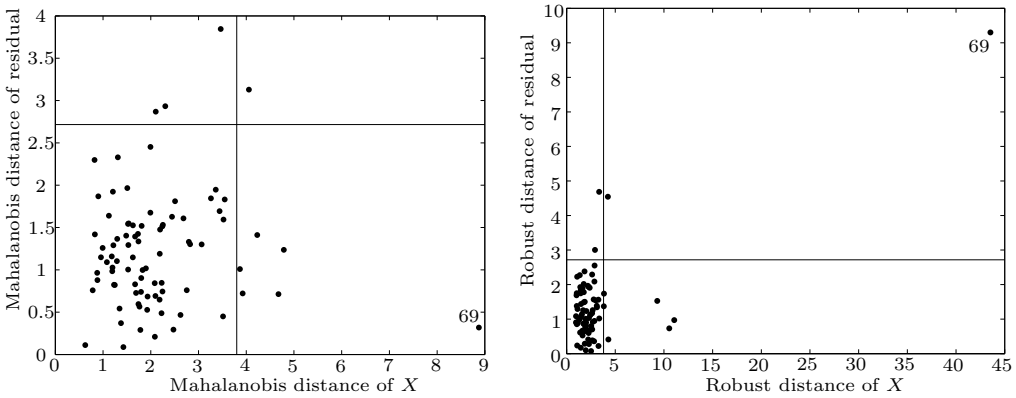

Figure 5.1. Diagnostic plots for the Milk data; LS (left) and 25% breakdown S (right).

To test the significance of the regression fit, we use both the classical and fast bootstrap with $R = 1,000$ on the 25% breakdown S-estimator. An intercept was included so that we have 14 regression coefficients. Nine of these were found

to be not significant by both the classical and the fast bootstrap, based on the fact that their 99% BCA confidence intervals did include 0. For the remaining five coefficients, the results are shown in Table 5.1. We see that for $\mathcal{B}_{22}$ and $\mathcal{B}_{72}$ the classical bootstrap intervals do contain 0, and the fast bootstrap intervals do not. However, these can be considered boundary cases and generally we see that the differences between classical and fast bootstrap results are small. It seems that the one serious outlier did not severely distort the inference results for the classical bootstrap here. Some distortion is present though, as can be seen from Figure 5.2 where we plotted the recalculations from the classical bootstrap versus those from the fast bootstrap, for two of the coefficients. Clearly some bootstrap samples yielded notably different estimates for the fast and the classical procedures, presumably due to the bad leverage point and the other outliers.

Table 5.1. 99% confidence limits for Milk data based on S-estimator.

		fast		classical	
	S-estimate	lower	upper	lower	upper
$\mathcal{B}_{11}$	0.997	0.981	1.008	0.971	1.007
$\mathcal{B}_{71}(\times 10^2)$	0.027	0.012	0.041	0.005	0.042
$\mathcal{B}_{22}$	0.082	0.002	0.162	-0.011	0.194
$\mathcal{B}_{32}$	0.379	0.082	0.765	0.019	0.754
$\mathcal{B}_{72}$	0.050	0.010	0.102	-0.004	0.100

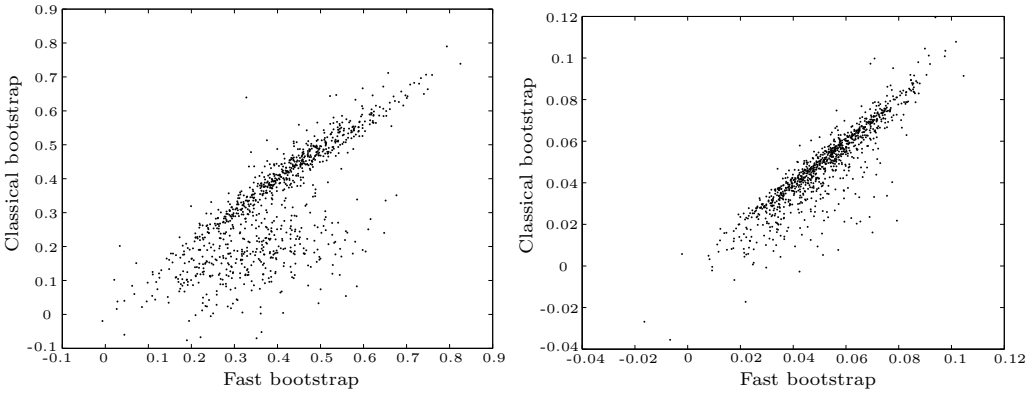

Figure 5.2. Scatter plots for the Milk data: classical bootstrap versus fast bootstrap recalculations; $\mathcal{B}_{32}$ (left) and $\mathcal{B}_{72}$ (right).

Although the lack of robustness of the classical bootstrap was not critical for this application, the fast bootstrap obviously still has the advantage over the classical procedure of being much less time consuming. Performing the classical

bootstrap procedure on the Milk data needed 706 CPU seconds, while the fast bootstrap only took 7.5 CPU seconds.

5.2. Life expectancy data

The following dataset can be found in The World Almanac and Book of Facts 1993 and contains data on the life expectancy in 38 countries, as well as some social characteristics. We consider a multivariate regression model where the female and male life expectancy are the response variables and the two explanatory variables are the number of people per television and the number of people per physician. The latter variables are log-transformed, which yields a better fit. An intercept term is also included. Again we applied both least squares and the 25% breakdown S-estimator. The diagnostic plots for the estimators are shown in Figure 5.3. Two (small) vertical outliers are detected by the S-estimator. The only difference between the LS and the S-estimates here is that the S-estimator more explicitly marks the good leverage points. This example shows that, next to being fast and robust, the fast bootstrap for the multivariate S-estimator also accurately mimics the classical bootstrap when there are only small outliers in the data. The estimates for the standard errors of the S-estimates are given in Table 5.2, respectively obtained through the fast bootstrap without linear correction, the (full) fast bootstrap, the classical bootstrap and the empirical asymptotic variance. We see that the standard errors obtained from the fast and classical bootstrap procedures are very similar. Furthermore, the effect of the linear correction seems to be critical in order for the fast method to approximate the classical bootstrap here. The asymptotic variance yields estimates that are markedly smaller. According to the simulations, the latter presumably are underestimations and the bootstrap estimates should be more accurate.

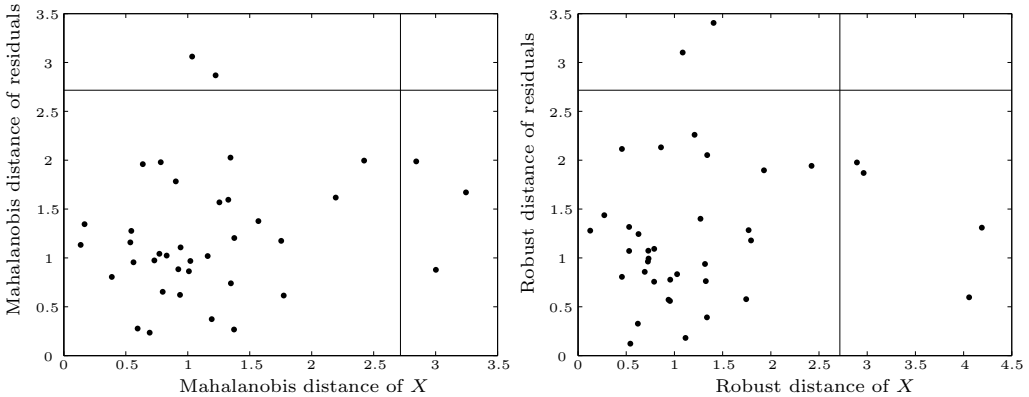

Figure 5.3. Diagnostic plots for the Life expectancy data; LS (left) and 25% breakdown S (right).

Table 5.2. Standard error estimates for Life expectancy data based on S-estimator.

	$\mathcal{B}_{11}$	$\mathcal{B}_{21}$	$\mathcal{B}_{31}$	$\mathcal{B}_{12}$	$\mathcal{B}_{22}$	$\mathcal{B}_{32}$
S-estimate	99.5	-3.03	-3.03	87.0	-2.50	-2.18
uncorrected fast bootstrap	5.27	0.72	0.93	4.66	0.58	0.81
fast bootstrap	8.11	1.08	1.42	7.28	0.81	1.24
classical bootstrap	8.63	1.32	1.55	7.90	0.98	1.38
empirical ASV	4.60	0.62	0.79	4.11	0.56	0.70

6. Remarks

Recently, Tatsuoka and Tyler (2000) introduced multivariate MM-estimators in the location and scatter setting, combining high breakdown with high efficiency in all dimensions. A generalization to multivariate regression is possible and may be worth considering to obtain robust inference through a similar adaptation of the bootstrap method of Salibian-Barrera and Zamar (2002).

References

Adrover, J., Salibian-Barrera, M. and Zamar, R. (2004). Globally robust inference for the location and simple regression models. *J. Statist. Plann. Inference* **119**, 353-375.

Agulló, J., Croux, C. and Van Aelst, S. (2002). The multivariate least trimmed squares estimator. Submitted.

Berrendero, J. R. and Zamar, R. H. (2001). Maximum bias curves for robust regression with non-elliptical regressors. *Ann. Statist.* **29**, 224-251.

Bilodeau, M. and Duchesne, P. (2000). Robust estimation of the SUR model. *Canad. J. Statist.* **28**, 277-288.

Davies, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15**, 1269-1292.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge Univ. Press.

Daudin, J., Duby, C. and Trecourt, P. (1988). Stability of principal component analysis studied by the bootstrap method. *Statistics* **19**, 241-258.

Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich Lehmann* (Edited by P. Bickel, K. Doksum and J. L. Hodges), 157-184. Wadsworth, Belmont, CA.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

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.

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

Johnson, R. A. and Wichern, D. W. (1988). *Applied Multivariate Statistical Analysis*. 2nd Edition. Prentice-Hall.

Koenker, R. and Portnoy, S. (1990). M estimation of multivariate regressions. *J. Amer. Statist. Assoc.* **85**, 1060-1068.

Lopuhaä, H. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17**, 1662-1683.

Lopuhaä, H. and Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19**, 229-248.

Maronna, R. A. and Yohai, V. J. (1997). Robust estimation in simultaneous equation models. *J. Statist. Plann. Inference* **57**, 233-244.

Ollila, E., Hettmansperger, T. P. and Oja, H. (2002). Estimates of regression coefficients based on sign covariance matrix. *J. Roy. Statist. Soc. Ser. B* **64**, 447-466.

Ollila, E., Oja, H. and Koivunen, V. (2003). Estimates of regression coefficients based on lift rank covariance matrix. *J. Amer. Statist. Assoc.* **98**, 90-98.

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

Rousseeuw, P. J., Van Aelst, S., Van Driessen, K. and Agulló, J. (2004). Robust multivariate regression. *Technometrics* **46**, 293-305.

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.

Ruppert, D. (1992). Computing S estimators for regression and multivariate location/dispersion. *J. Comput. Graph. Statist.* **1**, 253-270.

Salibian-Barrera, M. (2003a). The asymptotics of MM-estimators for linear regression with fixed design. Preprint, available at http://mathstat.carleton.ca/~matias/pubs.html

Salibian-Barrera, M. (2003b). Bootstrapping MM-estimators for linear regression with fixed design. Preprint, available at http://mathstat.carleton.ca/~matias/pubs.html

Salibian-Barrera, M. and Zamar, R. (2002). Bootstrapping robust estimates of regression. *Ann. Statist.* **30**, 556-582.

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

Stromberg, A.J. (1997). Robust covariance estimates based on resampling. *J. Statist. Plann. Inference* **57**, 321-334.

Tatsuoka, K. S. and Tyler, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *Ann. Statist.* **28**, 1219-1243.

Van Aelst, S. and Willems, G. (2004). Multivariate regression S-estimators for robust estimation and inference. Technical report, available at http://users.ugent.be/~svaelst/publications.html

Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 S9, B-9000 Ghent, Belgium.

E-mail: Stefan.VanAelst@UGent.be

Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium.

E-mail: Gert.Willems@ua.ac.be

(Received November 2003; accepted October 2004)