ELSEVIER

Statistics & Probability Letters 37 (1998) 341–350

STATISTICS & PROBABILITY LETTERS

A one-step robust estimator for regression based on the weighted likelihood reweighting scheme

Claudio Agostinelli[a], Marianthi Markatou[b,*,1]

[a] *Department of Statistics, University of Padua, 35121 Padua, Italy*
[b] *Department of Statistics, Columbia University, Mail code 4403, 618 Mathematics, New York, NY 10027, USA*

Received 1 April 1997; received in revised form 1 July 1997

Abstract

We propose a one-step estimator for the vector of regression and error-scale parameters in a linear regression model. The estimator is asymptotically normal and fully efficient. Given appropriate initial values it achieves very low bias and high breakdown point. © 1998 Elsevier Science B.V. All rights reserved

Keywords: Breakdown point; Efficiency; Influence; Robust; Weighted likelihood

1. Introduction

We consider the linear regression model defined as $y_i = x_i^T\beta + \varepsilon_i$, $i = 1, 2, \ldots, n$, where y_i is the ith response (dependent) variable, x_i^T is the ith row of the design matrix X which is of dimensions $n \times p$, β is an element of the parameter space Ω which is a subset of the p-dimensional Euclidean space and ε_i is the ith error. We assume that the errors ε_i are independent, identically distributed random variables with distribution function $F(\cdot/\sigma)$, $\sigma > 0$. The interest in this model centers on the estimation of and testing hypotheses about the parameter vector β.

It is well known that not all observations in a set of data play an equal role in determining estimates, tests and other statistics; in certain cases the values of these statistics are determined by only a few data points while most of the data are ignored. These discordant cases are called outliers. In regression there are two types of outliers. Outliers in the response-factor space are points that generate large residuals, while outliers in the factor space are points that generate high leverage. Recall that the leverage of the ith observation is defined as $h_i = x_i^T(X^TX)^{-1}x_i$ and it is the ith diagonal element of the hat matrix.

In this article we develop a one-step estimator for the regression parameters and the scale of the error distribution. The procedure uses a high breakdown point estimator as an initial value and it is analogous

* Corresponding author.

[1] The work of Marianthi Markatou was supported by an RGK CU01610301 grant.

PII S0167-7152(97)00136-3

to the one-step maximum likelihood estimator. The new one-step estimator uses the weighting functions introduced by Markatou et al. (1995) and Markatou (1996). The resulting estimator downweights points that are large residual outliers as well as observations that are both residual outliers and high leverage points. Observations that are only high leverage points are not downweighted. This is in line with the statement in Krasker and Welsch (1982) that "any downweighting in X space that does not include some consideration for how the y values at these outlying observations fit the pattern set by the bulk of the data cannot be efficient". Also note that reweighting has been studied, among others, by He and Portnoy (1992). Another approach that gives full efficiency at the model and high breakdown was given in He (1991).

The new one-step estimator is fully efficient, it is regression equivariant and has a high breakdown point.

The paper is organized as follows. Section 2 provides the background by briefly reviewing the main concepts in the development of the weights that are used. Section 3 presents the estimate and discusses its properties. Section 4 contains asymptotic distributional and robustness results, while in Section 5 examples and Monte Carlo simulations are used to exemplify the properties of the new estimator; the new estimator is also compared with the one-step estimator of Coakley and Hettmansperger (1993) and with the Simpson et al. (1992) estimate in terms of bias and standard error. Finally, Section 6 offers concluding remarks.

2. Background

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the density $m_\beta(x)$ corresponding to the probability measure M_β. Let $u(X;\beta) = \nabla \ln[m_\beta(X)]$ be the score function, where ∇ denotes differentiation with respect to β. Under regularity conditions the maximum likelihood estimator of β is a solution of the likelihood equation $\sum u(X_i;\beta) = 0$.

Given any point x_i in the sample space, Markatou et al. (1997) and Markatou (1996) construct a weight function $w(X_i; M_\beta, \hat{F})$ that depends on x_i, the chosen model distribution M_β, and the sample empirical distribution $\hat{F}$; then estimators for the parameter vector β and the error scale are obtained as solutions to the set of estimating equations $\sum_{i=1}^{n} w(X_i; M_\beta, \hat{F}) u(X_i;\beta) = 0$.

The weight function $w(X_i; M_\beta, \hat{F})$, by construction, takes values in the interval $[0,1]$ and it is defined as

$$w(X_i; M_\beta, \hat{F}) = \min\left\{1, \frac{[A(\delta(X_i))+1]^+}{\delta(X_i)+1}\right\} \quad \text{with} \quad i = 1,2,\ldots,n.$$

The quantity $\delta(x_i)$ is called Pearson residual and it is defined as $\delta(x_i) = [f^*(x_i)/m_\beta^*(x_i)] - 1$, where $f^*(x_i) = \int k(x_i; t, h)\, d\hat{F}(t)$ is a kernel density estimator and $m_\beta^*(x_i) = \int k(x_i; t, h)\, dM_\beta(t)$ is the smoothed model density. The Pearson residual expresses the agreement between the data and the assumed probability model. The function $A(\cdot)$ is a residual adjustment function (Lindsay, 1994) and it operates on Pearson residuals as the Huber ψ-function operates on the structural residuals $y_i - x_i^T\hat{\beta}$. When $A(\delta(x_i)) = \delta(x_i)$ the weight $w(x_i; M_\beta, \hat{F}) = 1$, and this corresponds to maximum likelihood. The function $A(\delta) = 2\{(\delta+1)^{1/2} - 1\}$ corresponds to Hellinger distance. Generally, the weights w use functions $A(\cdot)$ that correspond to a minimum disparity problem. For example the weight $w(\delta) = 1 - \delta^2/(\delta+2)^2$, used in Markatou et al. (1997), corresponds to the symmetric chi-squared distance.

This weighting scheme provides fully efficient and robust estimators, in the sense of breakdown. To calculate the Pearson residuals we need to select the smoothing parameter h. Markatou et al. (1995) select $h^2 = k\sigma^2$, where k is a constant that is independent of the scale of the model and it is selected so that it assigns a very small weight to an outlying observation. In this context, an outlier is an observation which is highly unlikely to occur under the assumed probabilistic model.

This definition generally does not agree with the more common geometric definition of an outlier as being that observation that is far away from the bulk of the data. However, in the normal regression model the two definitions are equivalent.

In the next section we discuss the development of the one-step estimator starting from a high breakdown point initial estimator and using the weighting scheme introduced above.

3. The estimator

Given (x_i, y_i), $i = 1,2,\ldots,n$ the quantities $z_i = (y_i - x_i^{\mathrm{T}}\beta)/\sigma$ are independent, identically distributed random variables if the vector β represents the true parameters. Then the z_is represent a random sample from a distribution with density $f(z)$ which is completely known.

The estimation procedure is then as follows. Start with an initial estimate of $\beta, \hat{\beta}_0$, and calculate the z_is. Treat those as an independent sample and create the Pearson residuals $\delta(z_i)$, $i = 1,2,\ldots,n$.

Consider now the class of estimators defined as solutions to the equations

$$\sum_{i=1}^{n} w(\delta(z_i))u(x_i, y_i; \hat{\beta}_0) = 0, \tag{1}$$

where $u(x_i, y_i; \beta)$ is the usual score function. As an example, when the model is normal the set of estimating equations is given as $\sum_{i=1}^{n} w(\delta(z_i))z_i x_i = 0$ and $\sigma^2 \sum_{i=1}^{n} w(\delta(z_i))\{\sigma^2 - z_i^2\} = 0$. A one-step estimator is based on an initial estimate $\hat{\beta}_0$ and it can be constructed by expanding the left-hand side of (1) about $\hat{\beta}_0$. This leads us to define the one-step estimator as

$$\hat{\beta} = \hat{\beta}_0 - \left[\frac{1}{n}\sum_{i=1}^{n} w(\delta(z_i))(\nabla_\beta u(x_i, y_i; \beta)|_{\beta=\hat{\beta}_0})\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} w(\delta(z_i))u(x_i, y_i; \hat{\beta}_0)\right]. \tag{2}$$

Our definition (2) of the one-step estimator is slightly different from what one would obtain by performing a one-step Taylor expansion in the neighborhood of $\hat{\beta}_0$. The Taylor expansion contains an extra term which is obtained by differentiating the weight with respect to β. This term, when evaluated at the model is exactly 0. Moreover, by defining the estimator as in (2) we insure that the term $1/n \sum_{i=1}^{n} w(\delta(z_i))(\nabla_\beta u(x_i, y_i; \beta)|_{\beta=\hat{\beta}_0})$ is, under the model, always a positive definite matrix. This is because since at least p points follow the model and $n \geqslant 2p + 1$, $\inf_{1 \leqslant j \leqslant p} \omega_j = 1$. Then, to see that the matrix is positive definite follow Simpson et al. (1992, p. 448).

The initial estimate $\hat{\beta}_0$ that we use is a root-n high breakdown point estimate, such as the least trimmed squares (LTS) estimate for the regression parameters. For the scale we use the least median of squares (LMS) as an initial estimate.

Under the assumption that the initial regression estimates are regression equivariant the one-step estimate is also regression equivariant because the weights calculated from the data $\{y_i + x_i^{\mathrm{T}}v, x_i\}$ and those calculated from $\{y_i, x_i\}$ are the same for all $i = 1,2,\ldots,n$.

4. Asymptotic and robustness properties

In this section we discuss the asymptotic and robustness properties of the estimator. The conditions required for consistency and asymptotic normality are as follows:

A1. The weight function $w(\delta)$ is a nonnegative bounded and differentiable function with respect to δ.

A2. The weight function $w(\delta)$ is regular, that is, $w'(\delta)(\delta+1)$ is bounded, $w'(\delta)$ being the derivative of w with respect to δ.

Let $\tilde{u}(x;\beta)=\nabla m_\beta^*(x)/m_\beta^*(x)$ and $u(x;\beta)=\nabla m_\beta(x)/m_\beta(x)$ where $m_\beta^*(x), m_\beta(x)$ are the smoothed version of the model and the true model, respectively.

A3. For every $\beta_0 \varepsilon \Omega$ there is a neighborhood $N(\beta_0)$ such that for $\beta \varepsilon N(\beta_0)$ the quantities $|\tilde{u}(x;\beta)u'(x;\beta)|$, $|\tilde{u}^2(x;\beta)u(x;\beta)|, |\tilde{u}'(x;\beta)u(x;\beta)|$ and $|u''(x;\beta)|$ are bounded by $M_i(x)$, $i=1,2,3,4$, where $E_{\beta_0}[M_i(X)] < \infty$, and the prime denotes derivative.

A4. $E_{\beta_0}[\tilde{u}^2(X;\beta)u^2(X;\beta)] < \infty$.

A5. $I(\beta) = E_\beta[u^2(X;\beta)] < \infty$, that is, the Fisher information is finite.

A6. (i) $\int |\frac{\nabla m_\beta(x)}{m_\beta^*(x)}| \, dx = \int |\frac{u(x;\beta)m_\beta(x)}{m_\beta^*(x)}| \, dx < \infty$,

(ii) $\int |\tilde{u}(x;\beta)u(x;\beta)| \frac{m_\beta(x)}{m_\beta^*(x)} \, dx < \infty$,

(iii) $\int |u'(x;\beta)| \frac{m_\beta(x)}{m_\beta^*(x)} \, dx < \infty$.

A7. The kernel $k(X;t,h)$ is bounded for all x by a finite constant $M(h)$ that may depend on h but not on t or x.

We also need the following lemma from Markatou et al. (1995) which we list here for completeness.

Lemma 1. *Under assumptions* A1–A7

$$n^{1/2}\left|A_n - \frac{1}{n}\sum_{i=1}^{n} u(x_i, y_i; \beta_0)\right| \xrightarrow{P} 0, \quad as\ n \to \infty,$$

$$\left|B_n - \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\beta} u(x_i, y_i; \beta_0)\right| \xrightarrow{P} 0, \quad as\ n \to \infty$$

and

$$C_n = O_p(1)$$

where $A_n = 1/n\sum_{i=1}^n w(\delta(z_i))u(x_i, y_i; \beta_0)$, $B_n = 1/n\sum_{i=1}^n \nabla_\beta\{w(\delta(z_i))u(x_i, y_i; \beta)\}|_{\beta=\beta_0}$, *and* $C_n = 1/n\sum_{i=1}^n \frac{\partial^2}{\partial\beta^2}(w(\delta(z_i))u(x_i, y_i; \beta))|_{\beta=\beta^*}$. *with* β^* *being between the true value* β_0 *and* $\hat{\beta}$.

Theorem 4.1. *Under assumptions* A1–A7 *and assuming that the high breakdown point starting value is a root-n consistent estimate the one-step estimator is consistent and asymptotically normal with mean zero and variance-covariance matrix given by the Fisher information matrix.*

Proof. Expand $\sum_{i=1}^n w(\delta(z_i))u(x_i, y_i; \hat{\beta}_0)$ in the neighborhood of the true value β_0. Now using Lemma 1 and observing that $\int w(\delta(z))(\nabla_\beta u(x, y; \beta))\, dF(x, y)$ is a continuous function in β, we obtain the desired result.

We now discuss the robustness properties of the estimator. Let T be an estimator that is a functional or, asymptotically, can be replaced by a functional. Consider distributions of the form $(1-\varepsilon)F_\beta + \varepsilon\Delta_{u_0}(u)$, where $0<\varepsilon<1$ and $\Delta_{u_0}(u)$ is the distribution that puts mass 1 at the point $u=u_0$. Then, the influence function of T at the distribution F_β is defined by

$$IF(u_0, T, F_\beta) = \lim_{\varepsilon \downarrow 0} \frac{T((1-\varepsilon)F_\beta + \varepsilon\Delta_{u_0}(u)) - T(F_\beta)}{\varepsilon}$$

in those u in the sample space for which the limit exists.

The importance of the influence function lies in its heuristic interpretation; it describes the effect of an infinitesimal contamination at a point u on the estimate, standardized by the mass of the contamination.

Assume the scale parameter is known and equal to one. In practice the scale has to be estimated from the data. Assume $(x, y)\,\varepsilon\,\mathscr{R}^{p+1}$ and let F_β be the probability distribution of $z = y - x^{\mathrm{T}}\beta$. Then, the functional that introduces the one-step estimator is

$$\beta(F_\varepsilon(z)) = \beta_0(F_\varepsilon(z)) - \left[\int w(\delta(z;\beta_0(F_\varepsilon))\nabla_\beta u(y - x^{\mathrm{T}}\beta_0(F_\varepsilon))\,\mathrm{d}F_\varepsilon(z)\right]^{-1}$$

$$\times \left[\int w(\delta(z;\beta_0(F_\varepsilon)))u(y - x^{\mathrm{T}}\beta_0(F_\varepsilon))\,\mathrm{d}F_\varepsilon(z)\right].$$

Theorem 4.2. *The influence function of the one-step estimator is given by*

$$IF(x, y, \beta(F_\varepsilon)) = \left[-\int u'(y - x^{\mathrm{T}}\beta(F))\,\mathrm{d}F\right]^{-1} \cdot u(y - x^{\mathrm{T}}\beta(F)),$$

where u' is the derivative of the score function with respect to β.

Note that when $u(y - x^{\mathrm{T}}\beta(F)) = (y - x^{\mathrm{T}}\beta(F))x$, then the influence function is that of the least-squares estimator. The influence function of the initial estimator does not enter into the calculation. The influence function here clearly depends on the score function and it is unbounded whenever the score is unbounded. Because the one-step estimate has the same influence function as the maximum likelihood, it achieves full asymptotic efficiency. Despite the fact that the influence function is unbounded the estimator is robust as it possesses a high breakdown point.

The breakdown point characterizes the global stability of an estimator. There are both asymptotic (Hampel, 1971; Huber, 1981) and finite sample (Hodges, 1967; Donoho, 1982; Donoho and Huber, 1983) versions of the concept of breakdown. Roughly speaking, the breakdown point of an estimator is the distance from the assumed distribution of the data beyond which the estimator becomes totally uninformative.

Theorem 4.3. *Under assumptions* A1, A2, A5 *and* (i) $\hat{\beta}_0$ *has a breakdown point of* ε^* *less than or equal to* 0.5, (ii) $\hat{\sigma}_0$ *has a breakdown point of at least* ε^*, (iii) $w(\delta(y - x^{\mathrm{T}}\hat{\beta}_0))((\partial/\partial\beta_j)u(y - x^{\mathrm{T}}\beta)|_{\hat{\beta}_0})$ *is bounded by an integrable function of* x, y, *and* (iv) $(1/n)\sum_{i=1}^{n} w(\delta(y_i - x_i^{\mathrm{T}}\beta_0))\nabla_\beta u(y_i - x_i^{\mathrm{T}}\hat{\beta}_0) \xrightarrow{\mathrm{P}} I(\beta)$, *positive definite, the estimator* $\hat{\beta}$ *has a breakdown point of* ε^*.

Note that condition (iii) is satisfied, for example, by the normal model. In fact, it is satisfied by a wide variety of models because since $n \geqslant 2p + 1$, $\sup_z \omega(\delta(z)) = 1$, with $z = y - x^{\mathrm{T}}\hat{\beta}_0$. Moreover, (iv) holds at the model.

5. Examples and simulations

In this section we exemplify the properties of the one-step estimator via examples and simulations. We also compare our one-step estimator with the one-step estimator of Simpson et al. (1992) and that of Coakley and Hettmansperger (1993).

The data sets we use are the soft drink delivery time data in Montgomery and Peck (1982, p. 116), the stack loss data presented by Brownlee (1965), the salinity data (Carroll and Ruppert, 1985) and the artificial data created by Hawkins, Bradu and Kass and reported in Rousseeuw and Leroy (1987, p. 94). Descriptions of these data sets can be found in the above references.

In both, examples and simulations we used the LTS as an initial estimate of regression and the LMS as an initial estimate of scale. Coakley and Hettmansperger (1993) used the same initial estimates to calculate their one-step estimator. To calculate our weights we use the Hellinger distance based weights. We also use chisquare weights (see Section 2). The normal kernel with mean 0 and $h^2 = k\hat{\sigma}_0^2$ where $\hat{\sigma}_0^2$ is the initial high breakdown estimator of the variance is used. The constant k is chosen to correspond to targeted weight of 0.20 when the percentage of contamination is held fixed at $\varepsilon = 20\%$ and an observation is at least three standard deviations away from the mean.

To calculate the Coakley and Hettmansperger (1993) estimator we used the MINITAB program provided in their article, with $\psi_c(t) = \min(c, \max(t, -c))$ and $c = 1.345$. This choice of constant guarantees a 95% efficiency of the location estimate at the standard normal model.

To obtain the Simpson et al. (1992) one-step estimate we used an S-plus code supplied by D. Simpson (personal communication to C. Agostinelli). We use the logistic score function with $c = 1.5$. The initial multivariate location and scatter estimates used were the minimum volume ellipsoid estimates; the exponent in the weights used was 0. This corresponds to a Mallows-type weight. For the calculation of the Simpson, Ruppert, Carroll estimate we performed only one iteration; generally, it is recommended to iterate four or five times in order to obtain the estimates. All computations were performed on a DEC 5000/50 station.

Table 1 presents the resulting one-step estimates associated with the regression parameters of the above data sets, as well as their associated standard errors. The table presents our one-step estimate, the Coakley and Hettmansperger (1993, S1S) and the Simpson et al. (1992, SRC) estimates with their standard errors.

A general inspection of the table shows that the one-step weighted estimator fares very well in terms of efficiency when compared with SRC and S1S estimators as it is indicated from its standard error. The last column of Table 1 presents the ordinary least-squares (OLS) estimators and their associated standard errors; those OLS estimators are calculated after the deletion of the aberrant data points. A comparison of the new one-step weighted estimate with those OLS estimates reveals the general superiority of the new estimate. Welsh and Ronchetti (1993) proved that inferences based on a least-squares analysis on "clean" data, that is data from which outliers are removed, are neither valid nor robust. Our estimates have full asymptotic efficiency, and inferences based on them are robust and valid. For the delivery data the weights associated with observations 1, 9, 22 to 24 are 0.555, 0.168, 0.904 and 0.637 respectively. Coakley and Hettmansperger (1993) downweight points 1, 4, 9, 11, 18, 20, 23, 24 of the delivery data set with point 22 receiving a full weight. We downweight points 1, 9 and 24; point 22 receives almost a full weight. For the stack loss data the weights associated with observations 1, 2, 3, 4 and 21 are 0.000025, 0.374196, 0.000117, 0 and 0.000127, respectively, while for the salinity data the weights associated with observations 3, 5, 16, 23 and 24 are 0.999, 0.007, 0, 0.017, 0.082. Finally, the weights associated with the first ten observations of the artificial data set are all 0, and the weights that observations 11, 12, 13, 14 and 47 receive are 0.998, 0.999, 0.987, 0.998 and 0.239, respectively.

We note that the observations that are both residual outliers and high leverage points are heavily downweighted, whereas the observations that are only high leverage points but fit the model well are not downweighted. The Pearson residual associated with the downweighted points are very large.

We bootstrapped the standard error of the regression estimates for the delivery time data set. The initial scale estimates used were MAD, S_n and Q_n. The last two estimates are given by formulas 2.1 and 3.3 of Rousseeuw and Croux (1993) and they are adjusted for the sample size. The number of bootstrap replications used was 1600. When S_n is used as an initial estimate the bootstrapped standard error of $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ is 1.829, 0.354 and 0.006 respectively, almost double the asymptotic standard error. Analogous results were obtained when the standard error of the coefficients of the SRC estimate was bootstrapped. However, the delivery data set contains only 24 observations. Experimentation with larger samples indicates that the bootstrapped standard error is in fair agreement with the asymptotic standard error.

We now present a small simulation study with the objective to evaluate the performance of the estimators in terms of bias and mean-squared error and relate the large-sample theory to the practical small-sample situations likely to be encountered.

Table 1
Estimates for the parameters of various data sets. The initial estimate of regression is LTS and the scale is LMS. In parenthesis the estimated standard error is reported

	Delivery data							
Parameters	One-step weighted estimate		SRC estimate		S1S estimate		OLS estimate w/o obs. 9	
β_0	3.853	(0.965)	3.559	(0.356)	3.884	(1.870)	4.447	(0.952)
β_1	1.499	(0.139)	1.356	(0.042)	1.465	(0.183)	1.498	(0.130)
β_2	0.013	(0.003)	0.017	(0.001)	0.013	(0.007)	0.010	(0.003)
	Stack loss data							
Parameters	One-step weighted estimate		SRC estimate		S1S estimate		OLS estimate w/o obs. 1,2,3,21	
β_0	−36.488	(4.205)	−43.551	(8.130)	−40.856	(6.835)	−35.484	(4.526)
β_1	0.767	(0.067)	0.990	(0.352)	1.048	(0.367)	0.686	(0.088)
β_2	0.558	(0.146)	0.964	(0.317)	0.684	(0.306)	0.567	(0.153)
β_3	−0.056	(0.056)	−0.211	(0.134)	−0.216	(0.191)	−0.017	(0.063)
	Artificial data							
Parameters	One-step weighted estimate		SRC estimate		S1S estimate		OLS estimate w/o obs. 1-10	
β_0	−0.211	(0.098)	−0.208	(0.127)	−0.142	(0.110)	−0.189	(0.103)
β_1	0.127	(0.064)	0.102	(0.076)	0.093	(0.068)	0.090	(0.065)
β_2	0.030	(0.036)	0.006	(0.031)	−0.076	(0.073)	0.036	(0.038)
β_3	−0.058	(0.032)	−0.012	(0.036)	0.087	(0.072)	0.033	(0.033)
	Salinity data							
Parameters	One-step weighted estimate		SRC estimate		S1S estimate		OLS estimate w/o obs. 5,16	
β_0	29.775	(4.182)	3.691	(14.348)	5.767	(8.539)	23.386	(3.902)
β_1	0.583	(0.070)	0.936	(0.209)	0.890	(0.150)	0.700	(0.067)
β_2	−0.231	(0.112)	−0.110	(0.158)	−0.033	(0.121)	−0.250	(0.131)
β_3	−1.068	(0.157)	−0.087	(0.537)	−0.166	(0.317)	−0.836	(0.148)

The simulated model is given as $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, 2, \ldots, n$. The nominal parameter values are $\beta_0 = \beta_1 = 1$ and $\sigma^2 = 1$. The error distributions considered are $(1-\varepsilon)N(0,1) + \varepsilon N(0,25), (1-\varepsilon)N(0,1) + \varepsilon N(-8,1)$. The first case corresponds to a heavy-tailed distribution; no leverage points are present. The second choice generates a group of residual outliers.

To generate the normal random variables we use the IMSL library. The subroutine DRNNOR was used to generate the standard normal variables. The contaminated samples are generated by taking $100\varepsilon\%$ of the data from $N(0,1)$ and then replace it by $\mu + \sigma X$, $X \sim N(0,1)$. We use sample sizes of 20 and 80 with Hellinger distance and chisquare weight functions. The number of Monte Carlo replications is 250.

An inspection of Table 2 reveals that the one-step weighted estimator performs very well in terms of bias even when the sample size is only 20. We also observe that when the percentage of contamination is very high, the scale estimate breaks down. This phenomenon is independent from the sample size and it is due to the fact that the initial scale estimate breaks down for the high amounts of contamination. We also

Table 2
One-step weighted likelihood and Simpson–Ruppert–Carroll estimates of the regression parameters and scale of the errors. The sample size is 20 and the Hellinger weight is used

		$(1-\varepsilon)N(0,1)+\varepsilon N(0,25)$				$(1-\varepsilon)N(0,1)+\varepsilon N(-8,1)$			
Percent of cont.	Par.	One-step estimate	Standard error	SRC estimate	Standard error	One-step estimate	Standard error	SRC estimate	Standard error
0	β_0	1.024	(0.207)	1.008	(0.252)				
	β_1	1.003	(0.074)	1.003	(0.093)				
	σ^2	0.748		0.656					
5	β_0	1.023	(0.216)	1.019	(0.279)	1.021	(0.213)	0.855	(0.303)
	β_1	1.006	(0.078)	1.004	(0.104)	1.008	(0.077)	1.010	(0.113)
	σ^2	0.796		0.738		0.758		0.795	
10	β_0	1.026	(0.228)	1.025	(0.313)	1.022	(0.223)	0.659	(0.370)
	β_1	1.009	(0.082)	1.004	(0.114)	1.008	(0.08)	1.008	(0.135)
	σ^2	0.875		0.839		0.784		0.954	
15	β_0	1.030	(0.244)	1.032	(0.344)	1.012	(0.232)	0.406	(0.450)
	β_1	1.012	(0.087)	1.001	(0.124)	1.009	(0.082)	1.010	(0.160)
	σ^2	0.985		0.955		0.803		1.133	
20	β_0	1.035	(0.265)	1.017	(0.386)	1.030	(0.240)	0.114	(0.542)
	β_1	1.012	(0.095)	1.007	(0.137)	1.006	(0.085)	1.006	(0.188)
	σ^2	1.145		1.106		0.811		1.366	
30	β_0	1.037	(0.317)	1.015	(0.482)	0.978	(0.288)	−0.773	(0.805)
	β_1	1.015	(0.113)	1.006	(0.168)	1.001	(0.103)	0.997	(0.277)
	σ^2	1.588		1.515		1.019		2.221	
40	β_0	1.049	(0.396)	1.026	(0.582)	0.614	(0.440)	−2.013	(1.135)
	β_1	1.014	(0.14)	1.012	(0.199)	0.995	(0.156)	0.973	(0.392)
	σ^2	2.473		2.182		2.747		4.045	
50	β_0	1.043	(0.514)	1.035	(0.705)	−2.970	(0.959)	−3.019	(1.309)
	β_1	1.023	(0.183)	1.022	(0.245)	0.941	(0.345)	0.980	(0.531)
	σ^2	4.253		3.276		15.424		21.626	

experimented with the scale estimates S_n and Q_n. We used those, with appropriate scale adjustments, as initial scale estimators. The results were very similar to those presented in Table 2, where the MAD was used as the initial scale estimator. For high percentages of contamination we still observe breakdown in the initial scale estimate. However, Q_n is preferable when the contamination is asymmetric, and performs better than S_n when high contamination is present. This holds for both sample sizes 20 and 80.

We did not try S-estimates because on page 208 of Rousseeuw and Leroy (1987) it is stated that ".... preliminary experience (both for simple and multiple regression) indicates that S-estimators do not really perform better than the LMS, at least from a practical point of view ....". However, Ruppert (1992) examined the performance of S-estimators for both real and simulated data and found that S-estimators performed somewhat better than LMS and LTS.

The use of the chisquare RAF provides similar results with those presented here for both sample sizes.

On the other hand, as we can see from Table 2 the bias of the regression estimates is very low even for high amounts of contamination. Similar observations apply for sample size 80 and for the chisquare weight (with the exception of high scale bias for $\varepsilon = 50\%$ when $n = 80$).

We also simulated the case where the data contain high leverage points that are not associated with large residuals. To generate these points we take the error distribution to be $N(0,1)$ and the distribution of the carriers is $(1-\varepsilon)N(0,9)+\varepsilon N(10,1)$. We use sample sizes of 20 and 80 with the same weight functions as before.

In this case both regression estimates and the scale estimate fare well in terms of bias and standard error for both sample sizes. The regression estimates have extremely low bias even for high amounts of contamination, whereas the bias of the scale estimate increases as the amount of contamination increases. When the chisquare weight is used, for both sample sizes, we obtain a dramatic decrease in the bias of the scale estimate. The regression estimates continue to have a very low bias.

The last experiment we performed was to simulate data which include observations that are both, high leverage and large Pearson residual points. To generate such data we took the error distribution to be $(1-\varepsilon)N(0,1)+\varepsilon N(-8,1)$ and the xs were generated from the distribution $(1-\varepsilon)N(0,9)+\varepsilon N(8,1)$.

In this case, the bias of the new regression estimates is very competitive with the bias of the SRC estimate, the bias of the scale estimate is dramatically lower than that of the SRC estimate and generally low when we compare the scale estimate with the true parameter value. The bias is much lower as the sample size increases.

We also simulated data from a t_3. The design we use was the one-way ANOVA with two levels and sample size equal to 60. We assumed that the underlying error distribution was $N(0,1)$ but in reality the error distribution was t_3. To generate t_3 values we used DRNCHI to generate chisquare variables with three degrees of freedom and then we obtained a t_3-variable as the ratio of a $N(0,1)$ variable and the square root of a χ^2_3 variable divided by its degrees of freedom. The nominal value for β_1, β_2 is 1. The initial estimates used were LTS for the regression parameters and LMS for the scale parameters. The estimates obtained when the chisquare weight was used are $\hat{\beta}_0 = 0.99630, \hat{\beta}_1 = 0.98754, \hat{\beta}_2 = 1.00607$ and $\hat{\sigma}^2 = 1.37570$, whereas when the Hellinger distance based weight is used the estimates are $\hat{\beta}_0 = 0.99530, \hat{\beta}_1 = 0.98879, \hat{\beta}_2 = 1.00741$ and $\hat{\sigma}^2 = 1.31844$.

6. Conclusions

We presented a one-step estimator for the regression parameters in a linear model. Additionally, we estimated the scale of the errors simultaneously. The estimators perform very well in terms of bias and standard error. However, for high percentages of contamination the scale estimate breaks down as the initial estimate also breaks down. The weighted likelihood reweighting scheme downweights points that are both high leverage and large residual outliers, and provides estimates that are asymptotically efficient and have high breakdown.

References

Brownlee, K.A., 1965. Statistical Theory and Methodology in Science and Engineering, Wiley, New York.

Carroll, R.J., Ruppert, D., 1985. Transformations in regression; A robust analysis. Technometrics, 27, 1–12.

Coakley, C.W., Hettmansperger, T.P., 1993. A bounded influence, high breakdown, efficient regression estimator. J Amer. Statist. Assoc. 88, 872–880.

Donoho, D.L., 1982. Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Department of Statistics, Harvard University, Cambridge, MA.

Donoho, D.L., Huber, P.J., 1983. The notion of breakdown point, in: Bickel, P.J., Doksum, K., Hodges, J.L. Jr. (Eds.), Festschrift fur Erich L. Lehmann, Wadsworth, Belmont, CA, pp. 157–184.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. Statist. Sci. 1, 54–77.

Hampel, F.R., 1971. A general qualitative definition of robustness. Ann. Math. Statist. 42, 1887–1896.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach based on Influence Functions, Wiley, New York.

He, X., 1991. A local breakdown property of robust tests in linear regression. J. Multivariate Anal. 38, 294–305.

He, X., Portnoy, S., 1992. Reweighted LS estimators converge at the same rate as the initial estimator. The Ann. Statist. 20, 2161–2167.

Hodges, J.L. Jr., 1967. Efficiency in normal samples and tolerance of extreme values for some estimators of location. In: Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, CA, pp. 163–186.

Huber, P.J., 1981. Robust Statistics, Wiley, New York.

Krasker, W.S., Welsch, R.E., 1982. Efficient bounded-influence regression estimation. J. Amer. Statist. Assoc. 77, 595–604.

Lindsay, B.G., 1994. Efficiency versus robustness: The case for minimum Hellinger distance and related methods. Ann. Statist. 22, 1018–1114.

Markatou, M., 1996. Robust statistical inference: weighted likelihoods or usual M-estimation? Commun. Statist. Theory and Methods 25, 2597–2613.

Markatou, M., Basu, A., Lindsay, B.G., 1995. Weighted likelihood estimating equations: the continuous case. Technical Report, Department of Statistics, Columbia University, New York.

Markatou, M., Basu, A., Lindsay, B.G., 1997. Weighted likelihood estimating equations: the discrete case with applications to logistic regression. J. Statist. Planning Inference 57, 215–232.

Montgometry, D.C., Peck, E.A., 1982. Introduction to Linear Regression Analysis, Wiley, New York.

Rousseeuw, P.J., Croux, C., 1993. Alternatives to the median absolute deviation. J. Amer. Statist. Assoc. 88, 1273–1283.

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

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

Ruppert, D., Carroll, R.J., 1980. Trimmed least squares estimation in the linear model. J. Amer. Statist. Assoc. 75, 828–838.

Simpson, D.G., Ruppert, D., Carroll, R.J., 1992. On one-step GM estimates and stability of inferences in linear regression. J. Amer. Statist. Assoc. 87, 439–450.

Welsh, A.H., Ronchetti, E., 1993. A failure of intuition: Naive outlier deletion in linear regression. Preprint.