Journal of Multivariate Analysis 116 (2013) 278–296

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Robust and efficient estimation of the residual scale in linear regression

Stefan Van Aelst [a,*], Gert Willems [a], Ruben H. Zamar [b]

[a] Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 S9, B-9000 Gent, Belgium

[b] Department of Statistics, University of British Columbia, 6356 Agricultural Road, Vancouver, BC, Canada V6T-1Z2

ARTICLE INFO

Article history:
Received 3 October 2011
Available online 7 January 2013

AMS 2010 subject classifications:
62F35
62J05

Keywords:
Robust scale
Maxbias
Influence function
Gross-error sensitivity
Efficiency

ABSTRACT

Robustness and efficiency of the residual scale estimators in the regression model is important for robust inference. We introduce the class of robust *generalized M-scale estimators* for the regression model, derive their influence function and gross-error sensitivity, and study their maxbias behavior. In particular, we find overall minimax bias estimates for the general class and also for well-known subclasses. We pose and solve a Hampel's-like optimality problem: we find generalized M-scale estimators with maximal efficiency subject to a lower bound on the *global* and *local* robustness of the estimators.

1. Introduction

In linear regression, the classical estimators for the regression coefficients and error scale are the well-known least squares estimators. These estimators are optimal under normal errors but extremely sensitive to outliers. This is particularly the case for the residual scale estimator.

Much attention has been paid in the statistical literature to robust and efficient estimation of the regression parameters. In this context, robust residual scale estimators are sometimes proposed as well, but the focus remains on the regression parameters. See e.g. [13,8,2]. For the location-scale model, some attention has been given to the estimation of the scale parameter. See for example [12,14,15,18,17,6].

Robust residual scale estimates play an important role in robust inference for the regression model such as the construction of confidence/prediction intervals, testing of hypotheses and model selection. The properties of such robust inference procedures heavily depend on the parameter estimators involved (see e.g. [9,22]). If the scale is involved in the inference, it is thus desirable to use a highly robust and highly efficient scale estimator to obtain a reliable and effective inference procedure.

Therefore, we study the statistical properties – robustness and efficiency – of a large class of robust residual scale estimators, which we call *generalized M-scales*. This class includes M-scales [10], S-scales [20], and τ-scales [24] as particular cases. We show that the influence function (IF) and breakdown point (BDP) properties do not suffice to characterize the robustness behavior of generalized M-scales because many generalized M-scales can be constructed with the same IF and BDP that still exhibit quite different robustness performance. Therefore, we study the maxbias of generalized M-scales which is a more overall measure of the robustness of an estimator. We investigate the maxbias behavior of generalized M-scales and determine which regression estimators must be used to maximize the robustness of the resulting scale estimator. Moreover, we

* Corresponding author.
E-mail addresses: Stefan.VanAelst@UGent.be (S. Van Aelst), GertllWillems@gmail.com (G. Willems), ruben@stat.ubc.ca (R.H. Zamar).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.12.008

find scale estimators that maximize efficiency under the central model subject to a bound on the gross-error sensitivity and breakdown point.

We now introduce some definitions and notation used throughout this paper. Consider n observations $(y_i, \mathbf{z}_i^t)^t \in \mathbb{R}^p$ and the linear regression model

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta}_0 + \sigma_0 u_i, \quad 1, \ldots, n, \tag{1}$$

where $\mathbf{x}_i = (1, \mathbf{z}_i^t)^t$. Under this central model, the errors u_i are assumed to be independent and identically distributed with a common distribution F_0, which is symmetric around zero and has scale one. Moreover, the errors are assumed to be independent of the predictors $\mathbf{z}_i$. We consider regression models with random predictors which is the standard when studying maxbias properties of regression estimators and allows us to obtain overall results that do not depend on a particular design. The distribution of the predictors under the central model is denoted by G_0 and the common joint distribution of $(y_i, \mathbf{z}_i^t)^t$ is denoted by H_0. To allow for a fraction $0 \le \epsilon \le \frac{1}{2}$ of outliers we assume that the actual underlying distribution H_ϵ of $(y_i, \mathbf{z}_i^t)^t$; $i = 1, \ldots, n$ belongs to the contamination neighborhood

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

where H^* is an arbitrary and unspecified distribution. Robust and efficient estimators for $\boldsymbol{\beta}_0$ or σ_0 are expected to perform relatively well for any $H_\epsilon \in \mathcal{H}_\epsilon$ with $0 \le \epsilon < \epsilon_n^*$, for some ϵ_n^* which would preferably be close or equal to $\frac{1}{2}$. Efficiency of the estimators is measured by their performance at the central model $H = H_0$ (i.e. $\epsilon = 0$).

For any $\mathbf{v} \in \mathbb{R}^p$ let $r_i(\mathbf{v}) = y_i - \mathbf{x}_i^t \mathbf{v}$ denote the corresponding residuals. The most common estimators of the error scale parameter σ_0 are based on the residuals $r_i(\hat{\boldsymbol{\beta}}_n)$ for some regression estimator $\hat{\boldsymbol{\beta}}_n$. We consider the following general class of residual scale estimators that we call *generalized M-scales*:

$$\hat{\sigma}_n(\hat{s}_n, \hat{\boldsymbol{\beta}}_n) = \hat{s}_n \sqrt{\frac{1}{b\,n} \sum_{i=1}^n \rho\left(\frac{r_i(\hat{\boldsymbol{\beta}}_n)}{\hat{s}_n}\right)}. \tag{2}$$

Here $\hat{s}_n$ is an arbitrary *initial scale estimator*, perhaps based on residuals $r_i(\tilde{\boldsymbol{\beta}}_n)$ corresponding to a given regression fit $\tilde{\boldsymbol{\beta}}_n$. The constant b is set equal to $E_{F_0}\{\rho(u)\}$ to obtain a consistent scale estimator at the central model. Often the error distribution F_0 is assumed to be the standard Gaussian distribution to obtain a consistent scale estimator at this model, provided that the initial scale estimator $\hat{s}_n$ is also consistent.

The generalized M-scale estimators considered in this paper are based on loss functions $\rho : \mathbb{R} \to \mathbb{R}$ that satisfy the following conditions:

(A1) ρ is symmetric, bounded and nondecreasing on $[0, \infty]$ with $\rho(0) = 0$. Moreover, ρ is differentiable at all $x \in \mathbb{R}$, except perhaps at a finite number of points.
(A2) The error distribution $F_0(x)$ has a density $f_0(x)$, which is symmetric, continuous, and strictly decreasing for $u \ge 0$.

Some results will require an additional assumption:

(A3) $g(s) = E_{F_0}\left(s^2 \rho(\frac{u}{s})\right)$ is nondecreasing for all $s \ge 0$.

Note that (A3) holds if, for instance, $2\rho(u) - \rho'(u)u \ge 0$, which is true for several well known ρ functions including Tukey's biweight loss function. The symmetry of the error distribution in Assumption (A2) is natural and commonly made in robust regression. Moreover, it is needed to prove our results. Assumption (A1) implies that we can assume without loss of generality that $\lim_{t\to\infty} \rho(t) = 1$ when convenient. The derivative of the loss function $\rho(x)$, which exists according to Assumption (A1), is denoted by $\psi(x)$.

One would expect that a robust choice for the initial estimators $\hat{s}_n$ and $\hat{\boldsymbol{\beta}}_n$ combined with an efficient choice for the loss function ρ would lead to a highly robust and efficient estimator $\hat{\sigma}_n(\hat{s}_n, \hat{\boldsymbol{\beta}}_n)$. In fact, we will show that our class of generalized M-scales includes estimators that can achieve high efficiency at H_0 without compromising their robustness.

To derive the asymptotic properties of the scale estimators, we introduce the generalized M-scale functional corresponding to the generalized M-scale estimator in (2). For any distribution H on $\mathbb{R}^p$ the generalized M-scale functional is defined as

$$\hat{\sigma}(H; \hat{s}, \hat{\boldsymbol{\beta}}) = \hat{s}(H) \sqrt{\frac{1}{b} E_H \rho\left(\frac{y - \mathbf{x}^t \hat{\boldsymbol{\beta}}(H)}{\hat{s}(H)}\right)}, \tag{3}$$

where $\hat{s} = \hat{s}(H)$ and $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(H)$ are scale and regression functionals corresponding to the estimators $\hat{s}_n$ and $\hat{\boldsymbol{\beta}}_n$, respectively.

We now give some examples of special subclasses of generalized M-scale estimators defined by (2)–(3).

Example 1 (*M-scales*). Given a preliminary regression estimator $\tilde{\boldsymbol{\beta}}_n$, the corresponding M-scale estimator $\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$ is implicitly defined as a solution, in s, to the equation

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{r_i(\tilde{\boldsymbol{\beta}}_n)}{s}\right) = b, \tag{4}$$