Inference in robust regression: the impact of scale estimation

Manuel Koller and Werner A. Stahel

[1] Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

Keywords: Inference, Robust Regression, Scale Estimation, MM-estimates, Confidence Intervals

Abstract

M-estimates require an auxiliary scale estimate in order to obtain scale invariance. The robustness properties of estimators based on the framework of M- (and MM-) estimators depend highly on the robustness of this auxiliary scale estimate. This scale estimate is not only required in estimating regression coefficients. It is also an essential part of the covariance matrix estimate of these coefficients, which is finally used to determine confidence limits. We will discuss the need of a robust, but also efficient scale estimate for the estimation of confidence limits.

We focus on Wald type methods, since they lead to confidence intervals in a straightforward manner. They will rely on the asymptotics of M-estimators, since the properties of other efficient procedures are asymptotically equivalent to those of M-estimators.

In the case of MM-estimates this scale is usually taken from an initial S-estimate, ordinarily configured to have maximum (0.5) breakdown point. Unfortunately, this initial estimate has quite low efficiency and also ignores the well-established heteroskedasticity of regression residuals.

For ordinary least squares regression one can show that the variance of the residual for the i-th observation is

$$\mathrm{Var}(r_i) = \sigma^2(1 - h_i) \tag{1}$$

where σ is the error standard deviation and h_i is the leverage of the i-th observation. Hence, even in the simplest case, the variance of regression residuals is not constant and depends on the predictors.

In the case of MM-estimation we write

$$\mathrm{Var}(r_i) = (\sigma\tau_i)^2. \tag{2}$$

We approximate the variance of MM-regression residuals and use this result to devise a correction factor τ_i to adjust the regression residuals. To estimate the scale σ we propose a subsequent M-scale estimator

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i}{\hat{\sigma}\tau_i}\right) = \kappa \tag{3}$$

where ρ is an even function with $\rho(0) = 0$ and monotone increasing for $r > 0$, e.g., $\rho(r) = r^2$. The constant κ is chosen to render the scale estimate consistent.

References

M. Koller (2008) Robust Statistics: Tests for Robust Linear Regression. Master's Thesis, http://stat.ethz.ch/research/dipl_arb/2008/koller.pdf.

P. J. Huber (1973). Robust regression: Asymptotics, conjectures, and Monte Carlo. Ann. Statist. 1, 799-821.