Robust scale estimation in finite samples

J.A. Randal[1]

[1] School of Economics and Finance, Victoria University of Wellington, P.O.Box 600, Wellington, New Zealand

Keywords: Robustness, Scale estimation.

Abstract

In this paper, we consider identification of promising robust estimators of scale in finite samples. The Lax (1985) study is a fundamental one, being the only study to survey the efficiency properties of robust scale estimators. Since the publication of Lax's results, new estimators have been proposed, primarily Rousseeuw and Croux's (1993) S_n and Q_n, and these need to be subjected to the same analysis. Unfortunately, Lax conducted his simulation in an era when computing power was much more limited than what is available today, and as a consequence, his efficiency estimates were very imprecise, and need to be reassessed in light of modern computing developments.

The contribution of this study is threefold: the simulation study we perform dwarfs that of Lax, with the simulation size several orders of magnitude greater; unlike Lax, we benchmark all estimators' efficiencies against the optimal (maximum likelihood) estimators using results of Randal and Thomson (2004); and we consider new estimators of scale in addition to those examined by Lax.

The scale estimators considered are computed for independent samples from Tukey's three corners: the normal, one-wild and slash sampling situations. These were considered by Tukey to reflect the three extreme cases of importance to robust statistics. An estimator's overall quality is assessed by using the triefficiency criterion promoted by Tukey, which is simply the minimum efficiency of the estimator over the three corners, and the "best" estimator will have the largest possible triefficiency. This is the first study to correctly benchmark efficiencies for Tukey's triefficiency for identification of quality general-purpose scale estimators.

As our interest is in finite sample performance, rather than asymptotic results, we conduct simulations for samples of 20 observations, as in Lax (1985). An estimator based on an EM algorithm for t-distributed data is introduced, and the choice of auxiliary scale estimator is considered in detail. The biweight A-estimators remain strong contenders, however the t-estimators provide a simpler alternative which is competitive under the triefficiency criterion. Suitably tuned, the t-estimator is the best estimator we consider under the efficiency measure favoured by Rousseeuw and Croux (1993).

References

D.A. Lax (1985). Robust estimators of scale: finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association*, 80, 736–741.

J.A. Randal, and P.J. Thomson (2004). Maximum likelihood estimation for Tukey's three corners. *Computational Statistics and Data Analysis*, 46, 677–687.

P.J. Rousseeuw and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88, 1273–1283.