Very Robust Regression

Marco Riani
Dipartimento di Economia, Università di Parma, Italy mriani@unipr.it

Anthony C. Atkinson*
The London School of Economics, London WC2A 2AE, UK
a.c.atkinson@lse.ac.uk

Domenico Perrotta
European Commission, Joint Research Centre, Ispra, Italy
domenico.perrotta@ec.europa.eu

There are several very robust estimates of regression parameters. Asymptotically all resist 50% of outliers in the data and so might seem indistinguishable, at least for large samples. However, part of the asymptotic argument assumes that the outliers are infinitely remote from the main body of the regression data. The talk considers what happens if the outliers are not very remote and exhibits differences in behaviour between the methods.

We introduce a parameter λ that defines a parametric path in the space of models and enables us to study, in a systematic way, the properties of estimators as the groups of data move from being far apart to close together. We examine, as a function of λ, the variance and squared bias of five estimators and we also consider their power when used in the detection of outliers. This systematic approach provides tools for gaining knowledge and better understanding of the properties of robust estimators.

The methods compared include Least Trimmed Squares (LTS), the Forward Search (FS) and MM estimation. All estimates are calculated from fits to subsets of the data. The FS is shown to have the best behaviour, both for estimation and for testing. We argue that this is because the estimate compares many subsets to adaptively select the size of the subset used in estimation. The other methods rely on one or two subset sizes. Quantitative differences in the behaviour of these algorithms depend on the distance between the regression data and the outliers. The differences are particularly striking for point contamination.

Key Words: distance of outliers; forward search; least trimmed squares; MM estimate; multiple outliers; overlap index; point contamination; regression diagnostics