Bocconi
Università Commerciale Luigi Bocconi

Department of Decision Sciences

Statistics Seminar

The IV selector: estimation and selection of the variables when the number of variables and instruments can be much larger than the sample size

Eric Gautier
CREST – ENSAE

Thursday, 18 November 2010
12:30pm Room 3-E4-SR03 Via Rontgen 1 Milano

Abstract

We introduce and analyse a new instrumental variables estimation procedure for linear models containing endogenous regressors. It can handle a small sample size n and a large number of regressors, and even larger number of instruments L. Distributional assumptions on the error term or the distribution of the instruments and regressors are very mild. Under proper assumptions on the action of a matrix, relating the variables of the structural model and the instruments, on sparse vectors, and the assumption that the vector of coefficients in the structural model is sparse (or approximately sparse), it is possible to estimate the vector of coefficients even in cases where L is much larger than n. The procedure is a variation on the Dantzig selector of Candµes and Tao (2007). We obtain nonasymptotic upper bounds on the estimation of the vector of coefficients in lp-norms that hold with probability close to 1. The price that is paid in the upper bound, for not knowing which coordinates are non zero, is a power of log(L). Under proper assumptions, a thresholded version of the IV-selector is able to simultaneously select the right nonzero coefficients with a probability close to 1. The estimation reduces to simple linear programming. The method is completely robust to weak instruments and there is no real issue in selecting instruments with the IV -selector.
Taking all possible exogenous instruments (up to an exponential in the sample size!)
Still leads to tight bounds around the true vector of parameters. Adding an instrument which is irrelevant incurs a loss of log(L)-log(L-1) in the constant multiplying the parametric rate. This is totally negligible compared to the potential gain in adding an instrument with some power. It is also robust to heteroscedastic errors.

Department of Decision Sciences

Via Röntgen 1 - 20136Milano

Tel. 02 5836.5632
Fax 02 5836.5630