Data driven tests for homoscedastic linear regression model

Teresa Ledwina
Institute of Mathematics, Polish Academy of Sciences

We describe and discuss new tests for testing the validity of a semiparametric random-design linear regression model.

The construction consists of several steps. First, we follow the classical idea of overfitting and replace the basic problem by a series of auxiliary subproblems. Next, to test whether extra terms are significant we construct a counterpart of classic efficient score statistic. Finally, we combine the solution with model selection methods providing guidelines to choose the right subproblem. This leads to data driven score tests for the initial testing problem. The model selection part of the construction exploits a couterpart of Schwarz selection rule, pertaining to the efficient score statistic. We also apply and investigate new selection rule which combines Schwarz-type and Akaike-type criteria.

Analysis of new solutions consists in several steps: we discuss a useful way of deriving the efficient score vector; we introduce large class of estimators of the efficient score vector and prove that under the null model our constructions are asymptotically distribution free; the limiting distribution of the test statistics under the null hypothesis is given. The proofs adopt and exploit some ideas and results developed in the area of semiparametric estimation.

We also compare the finite sample performace of our tests with the recent solution introduced by Guerre and Lavergne (2005), *Ann. Statist.*, as well as to Cramér-von Mises type construction. The simulation experiment indicates the very good performance of the proposed tests.

The contribution is based on the following papers co-authored by Tadeusz Inglot:

Data driven score tests for a homoscedastic linear regression model: the construction and simulations. *Proc. Prague Stochastics 2006*, 124-137,

Data driven score tests for a homoscedastic linear regression model: asymptotic results. *Probab. Math. Statist.* 2006, 41-61,

Towards data driven selection of a penalty function for data driven Neyman tests. *Linear Algebra and its Appl.* 2006, 579-590.