

# Conditional predictive inference post model selection

Prof. Dr. Hannes Leeb  
(Universität Wien)

November 11, 2015

A finite-sample analysis of predictive inference procedures after model selection in regression with random design is given. The analysis is focused on a statistically challenging scenario where the number of potentially important explanatory variables can be infinite, where no regularity conditions are imposed on unknown parameters, where the number of explanatory variables in a good model can be of the same order as sample size and where the number of candidate models can be of larger order than sample size. The performance of inference procedures is evaluated conditional on the training sample. Under weak conditions on only the number of candidate models and on their complexity, and uniformly over all data-generating processes under consideration, we show that a certain prediction interval is approximately valid and short with high probability in finite samples, in the sense that its actual coverage probability is close to the nominal one and in the sense that its length is close to the length of an infeasible interval that is constructed by actually knowing the best candidate model. Similar results are shown to hold for predictive inference procedures other than prediction intervals like, for example, tests of whether a future response will lie above or below a given threshold.