

GUEST EDITORS' EDITORIAL

RECENT DEVELOPMENTS IN MODEL SELECTION AND RELATED AREAS

HANNES LEEB
Yale University

BENEDIKT M. PÖTSCHER
University of Vienna

Model selection procedures have become an integral part of almost any statistical or econometric analysis of data. Prominent examples include the selection of explanatory variables in (non)linear regression or lag-length selection in time series models. It is safe to say that these methods have become an integral part of the toolbox of modern data analysis. Moreover, research on model selection (or related procedures such as model averaging and shrinkage procedures) and its implications for statistical analysis has intensified in recent years. This is witnessed by the fact that it is difficult to pick up any recent issue of an econometrics or statistics journal and avoid encountering a paper where some form of model selection, or model averaging, or general shrinkage estimation is considered.

The present issue collects a number of papers that represent a selection from the current trends in research on model selection and related methods. We hope that these papers provide a glimpse of some of the important new developments, ideas, and problems that currently drive the research in this area. The topics covered in this special issue range from inference after model selection, to Bayesian model selection methods with rescaled priors, to multiple testing, to estimation and prediction in parametric and nonparametric settings.

The first two papers in this issue deal with inference after model selection or shrinkage. Keith Knight studies the distributional properties of a large class of post-model-selection and shrinkage-type estimators in the case of nearly singular design. In a regression setting, the author assumes that the design matrix is nonsingular but converges to a singular matrix as sample size increases. Knight's analysis covers Lasso-type estimators, including the Lasso, ridge regression, and other types of penalized maximum likelihood estimators with l_p -penalties, in addition to a class of post-model-selection estimators that includes information criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

Address correspondence to Hannes Leeb, Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT 06510, USA; e-mail: hannes.leebe@yale.edu.

Because the finite-sample (and also the asymptotic) distributions of post-model-selection estimators typically depend on unknown parameters in a complicated way, a natural and important question is how these distributions can be estimated. This is studied in the paper by Leeb and Pötscher. It is shown that typically the distribution of post-model-selection estimators can not be estimated with reasonable accuracy even asymptotically. In particular, it is shown that no estimator for such distribution can be uniformly consistent (not even locally). The results also apply to estimators derived from the bootstrap or from similar subsampling methods or randomization methods.

Often, the goal of model selection is to identify the smallest model that correctly describes the data-generating process. In this regard, Tanujit Dey, Hemant Ishwaran, and Sunil Rao study the Bayesian method of selecting the model with highest posterior probability, with particular emphasis on the finite-sample performance of that method. They show that, in finite samples, selecting the highest posterior probability model can result in a marked tendency of underfitting, i.e., in choosing an overly parsimonious model; this is well known for methods like BIC but less so for, say, Bayesian methods based on spike-and-slab priors. Moreover, the selected model can be substantially influenced by the choice of the prior. (These finite-sample phenomena are in stark contrast to the apparently favorable large-sample limit performance of such model selectors, which select the smallest correct model with probability approaching one as sample size increases.) Dey, Ishwaran, and Rao also discuss alternative methods designed to mitigate the undesirable finite-sample effects. In particular, they consider a rescaled version of a spike-and-slab prior and two model selection strategies based on this prior, namely, selecting the model with highest posterior probability and the (posterior) median model.

Closely related to the model selection problem is the problem of multiple testing. Joseph Romano, Azeem Shaikh, and Michael Wolf provide a survey of multiple testing procedures and generalized error rates including the traditional family-wise error rate and variations thereof, in addition to several recent proposals such as the false-discovery rate, the positive false-discovery rate, and the false-discovery proportion. That paper puts particular emphasis on the case where individual test statistics are dependent. The authors give a detailed description of several algorithms for multiple testing that control a generalized error rate, together with simulation examples and empirical applications.

The estimation of a parametric or nonparametric regression function is an area where model selection and related shrinkage methods have been applied with great success. Several papers in this volume report on recent progress in this direction.

Rudy Beran studies the problem of estimating a matrix that is observed with noise. A distinguishing feature of his paper is that the number of unknowns is proportional to sample size. To estimate the mean matrix, Beran considers estimators that minimize the estimated risk over certain (large) classes of linear estimators. It is shown that the asymptotic risk of these estimators is equal to

the risk of an (infeasible) “oracle estimator.” Here, the “oracle estimator” is that which minimizes the risk over the class of estimators under consideration. These results hold uniformly over the set of data-generating processes such that the signal-to-noise ratio and the measurement error variance are bounded. Moreover, the results carry over to errors-in-variables linear regression models.

While traditionally the goal of model selection is to select a single model that performs well in some global sense, Yuhong Yang studies the problem of fitting several models to estimate an unknown regression function such that the corresponding estimator performs well locally, depending on the argument of the regression function. For example, with two models, the first model may fit the regression function well in one region of the regression function’s domain, whereas the second model may be better in another. This results in a larger and more versatile class of possible estimators and thus can potentially outperform methods based on selecting a single model. Yang considers two strategies for local model selection and studies their performance in finite samples and in the large-sample limit.

A somewhat related approach is studied by Gerda Claeskens and Nils Hjort. For model selection in parametric settings, Claeskens and Hjort have recently proposed a so-called focused information criterion (FIC) that aims at selecting a model such that a particular parameter of interest, the focus parameter, is estimated with small risk. Here, the “best” model depends on the particular focus parameter. In this issue, Claeskens and Hjort propose a kind of weighted FIC where a collection of focus parameters is considered and the criterion aims at selecting a model that minimizes the (weighted) average risk, where the average is taken over the collection of focus parameters under consideration.

Next, we have several papers on model selection and shrinkage methods with the aim of finding a “good” model for forecast and prediction out-of-sample.

Ed George and Xinyi Xu consider Bayesian model averaging procedures with the aim of estimating the density of future observations (under Kullback–Leibler risk). They give sufficient conditions for minimaxity and dominance of Bayes procedures. In particular, George and Xu show that the improper Bayes estimator based on a uniform prior, which has constant risk and which is minimax, can in fact be uniformly dominated by an appropriate proper Bayes procedure.

Peter Bartlett gives finite-sample risk bounds and oracle inequalities when the model is selected by minimizing the empirical loss plus a complexity penalty over a (nested) class of candidate models. The results in that paper considerably improve upon existing oracle inequalities in the sense that, for certain situations and with appropriately chosen complexity penalties, the risk of the resulting post-model-selection estimator is shown to approach zero at a faster rate than that delivered by existing oracle inequalities.

The problem of combining (or averaging) several forecasts is studied by Jan Magnus and Andrey Vasnev. Their goal is to combine two microlevel forecasts, where one is based on microlevel data and the other is obtained from macro-

level data. This approach of Magnus and Vasnev is nonstandard in that they consider data from two different levels of aggregation, i.e., micro and macro level, and in that they allow for the case where not all data are available to the forecaster, a situation often encountered in practice. Simulation evidence suggests that the proposed method performs well, and its performance is also demonstrated by forecasting Euro yield based on combining monthly and quarterly data.

Finally, we thank all who contributed to this volume with their research, ideas, opinions, and editorial work: the authors, the referees, and, in particular, Peter Phillips, the editor of *Econometric Theory*. Also, we greatly appreciate Mary Moulder's smooth and efficient secretarial support.