THE LINK BETWEEN STATISTICAL LEARNING THEORY AND ECONOMETRICS: APPLICATIONS IN ECONOMICS, FINANCE, AND MARKETING

Esfandiar Maasoumi[1] and Marcelo C. Medeiros[2]

[1] Department of Economics, Emory University, Atlanta, USA.

[2] Department of Economics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil.

Statistical Learning refers to statistical aspects of automated extraction of regularities (structure) in datasets. It is a broad area which includes neural networks, regression-trees, nonparametric statistics and sieve approximation, boosting, mixtures of models, computational complexity, computational statistics, and nonlinear models in general. Although Statistical Learning Theory and Econometrics are closely related, much of the development in each of the areas is seemingly proceeding independently. This special issue brings together these two areas, and is intended to stimulate new applications and appreciation in Economics, Finance, and Marketing. This special volume contains ten innovative papers covering a broad range of relevant topics.

Keywords: Statistical learning; Forecasting; Model Combination; Bagging; Neural Networks; Nonlinear Models; Support Vector Regression; Regression Trees; Mixture of Models.

JEL Classification: C14; C22; C45; C51; C53

Introduction

Statistical Learning refers to statistical aspects of automated extraction and identification of regularities (structure) in datasets, sometimes large ones. It is a broad area which includes neural networks, regression-trees, nonparametric statistics and sieve approximation, boosting, mixtures of models, computational complexity and dimension reduction, computational statistics, and nonlinear models in general. Two typical applications are forecasting and classification. Although Statistical Learning Theory and Econometrics are closely related, a good deal of the development in each area is seemingly proceeding independently. For example, factor analysis or diffusion indexes (dimension reduction), combination of forecasts (bagging and boosting), sieve approximations and nonparametric econometrics (neural networks, support vector regression, radial basis functions), piecewise linear and nonlinear models in general (regression trees, neural networks) are some examples of areas that have been evolving in both literatures; see, for example, Hastie, Tibshirani, and Friedman (2001) or Vapnik (1995).

This special issue brings together these two areas, and represents our hope to stimulate applications in Economics, Finance, and Marketing that benefit from some new developments in each area.

The present volume contains ten innovative papers by a number of leading researchers. Of the ten articles in this issue, four are specifically concerned with forecasting models with many predictors, one is a large scale forecasting comparison among the most used statistical learning (nonlinear) models, two are related to mixture of models. Support Vector Regression is the main theme of two papers, and, finally, Regression Trees are discussed in one paper.

Overview

Forecasting a macroeconomic or financial variable is challenging in an environment with many potential predictors whose predictive ability can vary over time. In the first four papers in this compendium, this problem is tackled with techniques which have their roots in the Statistical Learning Theory, such as factor models, bagging, and forecast combination.

The first paper in this issue is co-authored by Nii Ayi Armah (Rutgers University) and Norman R. Swanson (Rutgers University) and is entitled “Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Large Scale Macroeconomic Time Series Environments”. The authors consider forecasting from diffusion indexes in situations where common factors are assumed to underlie the co-movements of a set of macroeconomic variables. The authors provide a nice review of the extant literature on diffusion indexes and then, outline a number of approaches to the selection of factor proxies (observed variables that proxy unobserved estimated factors). Their approach to factor proxy selection is examined via a small Monte Carlo experiment, where evidence supporting their proposed methodology is presented, and via a large set of prediction experiments. One of their main empirical findings is that their smoothed approaches to factor proxy selection appear to yield predictions that are often superior not only to a benchmark factor model, but also to simple linear time series models which are generally difficult to beat in forecasting competitions. In some sense, by using their approach to predictive factor proxy selection, one is able to open up the black box often associated with factor analysis, and to identify actual variables that can serve as primitive building blocks for (prediction) models of a host of macroeconomic variables, and that can also serve as policy evaluation instruments. Their findings suggest that important observable variables include: various S&P500 variables, including stock price indices and dividend series; a 1-year Treasury bond rate; various housing activity variables; industrial production; and exchange rates.

In “Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Employment Growth”, David E. Rapach (Saint Louis University) and Jack K. Strauss (Saint Louis University) compare two approaches to forecasting U.S. employment growth with many predictors. Their first approach applies bagging to a general-to-specific procedure based on a general dynamic linear regression model with 30 potential predictors. The second approach considers several methods for combining forecasts from 30 individual autoregressive distributed lag (ARDL) models, where each individual ARDL model contains a potential predictor. The authors analyze bagging and combination forecasts at multiple horizons over four different out-of-sample periods using a mean square forecast error (MSFE) criterion and forecast encompassing tests. They find that bagging forecasts often delivers the lowest mean squared forecast errors. Interestingly, the authors also find that incorporating information from both bagging and combination forecasts based on principal components often leads to further gains in forecasting accuracy.

Forecasting with many predictors is also the subject of the interesting paper by Huiyu Huang (PanAgora Asset Management) and Tae-Hwy Lee (University of California, Riverside). In “To Combine Forecasts or to Combine Information?”, the authors discuss two directions one could follow in a forecasting environment with many predictors: Combination of forecasts (CF) or combination of information (CI). CF combines forecasts generated from simple models each incorporating a part of the whole information set, while CI brings the entire information set into one super model to generate an ultimate forecast. Through linear regression analysis and simulation, the authors show the relative merits of each, particularly the circumstances where forecast by CF can be superior to forecast by CI, when CI model is correctly specified and when it is misspecified, and shed some light on the success of equally weighted CF. In their empirical application on prediction of monthly, quarterly, and annual equity premium, they compare the CF forecasts (with various weighting schemes) to CI forecasts (with principal component approach mitigating the problem of parameter proliferation). The authors find that CF with (close to) equal weights is generally the best and dominates all CI schemes, while also performing substantially better than the historical mean.

The last of the four papers on forecasting with many predictors is “The Benefits of Bagging for Forecast Models of Realized Volatility” by Eric Hillebrand (Louisiana State University) and Marcelo C. Medeiros (Pontifical Catholic University of Rio de Janeiro). In their contribution, the authors show that bagging can improve the forecast accuracy of time series models for realized volatility. They consider 23 stocks from the Dow Jones Industrial Average over the sample period 1995 to 2005 and employ two different forecast models: a log-linear specification in the spirit of the heterogeneous autoregressive model and a nonlinear specification with logistic transitions. Both forecast model types benefit from bagging, in particular in the 1990s part of the sample. The log-linear specification shows larger improvements than the nonlinear model. Bagging the log-linear model yields the highest forecast accuracy on the sample.

The fifth paper of this special issue consists of a very useful forecasting experiment. Nesreen Kamel (Cairo University), Amir F. Atiya (Cairo University), Neamat El Gayar (Cairo University), and Hisham El-Shishiny (IBM Cairo Technology Development Center) in “An Empirical Comparison of Machine Learning Models for Time Series Forecasting” present a large scale comparison study for the major statistical learning models for time series forecasting. Specifically, the authors apply their models to the monthly M3 time series competition data. As claimed by the authors, there have been very few large scale comparison studies for statistical learning models for the regression or the time series forecasting problems. The models considered are the standard feedforward neural network model, Bayesian neural networks, radial basis neural networks, generalized regression neural networks, K-nearest neighbor regression, regression trees, support vector regression, and Gaussian processes. The study reveals significant differences between the different methods. The best two methods turned out to be the simple feedforward neural network model and the Gaussian process regression.

The next two papers in the sequence consider mixture models. The first one is “On Models for Value at Risk” by Philip Yu (The University of Hong Kong), Wai Keung Li (The university of Hong Kong), and Shusong Jin (Fudan University). The authors extend the Conditional Autoregressive Value-at-Risk (CAViaR) model in order to describe possible nonlinearities and structural breaks in Value-at-Risk (VaR) measures. Two different new specifications are described in the paper: the threshold Generalized Autoregressive (TGARCH) and the mixture-GARCH models. The methods are applied to the S&P 500, Hang Seng, Nikkei and Nasdaq indices.

Alexandre X. Carvalho (Institute of Applied Economics Research, IPEA) and Georgios Skoulakis (University of Maryland) in “Time Series Mixtures of Generalized t Experts: ML Estimation and An Application to Stock Return Density Forecasting” propose and analyze a new nonlinear time series model based on local mixtures of linear regressions, referred to as experts, with thick-tailed disturbances. The mean function of each expert is an affine function of covariates that may include lags of the dependent variable and/or lags of external predictors. The mixing of the experts is determined by a latent variable, the distribution which depends on the same covariates used in the regressions. The expert error terms are assumed to follow the generalized t and normal distributions as special cases and allowing separate modeling of scale and kurtosis. The authors show consistency and asymptotic normality of the maximum likelihood estimator, for correctly specified and for misspecified models, and they also provide Monte Carlo evidence on the performance of the standard model selection criteria in selecting the number of experts.

Support Vector Regression is considered in the next two contributions. In “Estimating the Market Share Attraction Model using Support Vector Regressions”, Georgi I. Nalbantov (Erasmus University), Philip Hans Franses (Erasmus University), Patrick J. F. Groenen (Erasmus University), and Jan C. Bioch (Erasmus University) propose to estimate the parameters of the Market Share Attraction Model in a novel way by Support Vector Regressions. Traditionally, the parameters of the Market Share Attraction Model are estimated via a Maximum Likelihood procedure, assuming that the data are drawn from a conditional Gaussian distribution. However, if the distribution is unknown, ordinary least squares may seriously fail. One way to tackle this problem is to introduce a linear loss function over the errors and a penalty on the magnitude of model coefficients. This leads to qualities such as robustness to outliers and avoidance of the problem of overfitting. This kind of estimation forms the basis of the Support Vector Regression and makes it a good candidate for estimating the Market Share Attraction Model. The authors test their approach to predict the evolution of the market shares of 36 car brands simultaneously and report promising results.

André d’Almeida Monteiro (Gávea Investimentos) in “Estimating Interest Rate Curves by Support Vector Regression” proposes a model to estimate an interest rate curve which incorporates the bid-ask spreads of the securities from which it is extracted. The author advocates the use of Support Vector Regression. The motivation for this is that Support Vector Regression features extra capabilities at a low estimation cost. The model is specified by a loss function, a kernel function, and a smoothing parameter. The author models the daily US dollar interest rate swap curves, from 1997 to 2001. Comparing the Support Vector Regression with other models, the former achieved the best cross-validation interpolation performance in controlling the bias-variance trade-off and generating the lowest error considering the desired accuracy fixed by the bid-ask spreads.

Finally, William S. Rea (University of Canterbury), Marco Reale (University of Canterbury), Carmela Cappelli (Universitá di Napoli Federico II), and Jennifer A. Brown (University of Canterbury) in “Identification of Changes in Mean with Regression Trees: An Application to Market Research”, present a computationally efficient method for finding multiple structural breaks at unknown dates based on regression trees. The authors show that the tree-based method performs well in long series which are impractical to analyze with current methods. They apply these methods plus the CUSUM test to the market share of Crest toothpaste between 1958 and 1963.

We hope this collection of contributions by several of the leading experts in the field of statistical learning theory and econometrics will serve to fill part of the gap between these two areas, as well as serve both professionals and readers who might be less familiar with these vibrant areas of research.

Acknowledgments
The authors wish to thank the contributors to this issue for sharing their innovations and superb technical skills in a timely manner, and for participating in the rigorous review process very responsively. Our acknowledgments also go to the numerous referees for their helpful, insightful, and timely reviews. The second author is most grateful for the financial support from CNPq, Brazil.

References
Hastie, T., R. Tibishirani, and J. Friedman (2001). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer.

Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer.