ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence[☆]

Hyun Hak Kim[a,*], Norman R. Swanson[b]

[a] The Bank of Korea, 55 Namdaemunno, Jung-Gu, Seoul 100-794, Republic of Korea

[b] Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C1
C22
C52
C58

Keywords:
Bagging
Bayesian model averaging
Boosting
Diffusion index
Elastic net
Forecasting
Least angle regression
Non-negative garotte
Prediction
Reality check
Ridge regression

ABSTRACT

In this paper, we empirically assess the predictive accuracy of a large group of models that are specified using principle components and other shrinkage techniques, including Bayesian model averaging and various bagging, boosting, least angle regression and related methods. Our results suggest that model averaging does not dominate other well designed prediction model specification methods, and that using "hybrid" combination factor/shrinkage methods often yields superior predictions. More specifically, when using recursive estimation windows, which dominate other "windowing" approaches, "hybrid" models are mean square forecast error "best" around 1/3 of the time, when used to predict 11 key macroeconomic indicators at various forecast horizons. Baseline linear (factor) models also "win" around 1/3 of the time, as do model averaging methods. Interestingly, these broad findings change noticeably when considering different sub-samples. For example, when used to predict only recessionary periods, "hybrid" models "win" in 7 of 11 cases, when condensing findings across all "windowing" approaches, estimation methods, and models, while model averaging does not "win" in a single case. However, in expansions, and during the 1990s, model averaging wins almost 1/2 of the time. Overall, combination factor/shrinkage methods "win" approximately 1/2 of the time in 4 of 6 different sample periods. Ancillary findings based on our forecasting experiments underscore the advantages of using recursive estimation strategies, and provide new evidence of the usefulness of yield and yield-spread variables in nonlinear prediction model specification.

1. Introduction

Technological advances over the last five decades have led to impressive gains in not only computational power, but also in the quantity of available financial and macroeconomic data. Indeed, there has been something of a race going on in recent years, as technology, both computational and theoretical, has been hard pressed to keep up with the ever increasing mountain of (big) data available for empirical use. From a computational perspective, this has helped spur the development of data shrinkage techniques, for example. In economics, one of the most widely applied of these is diffusion index methodology. Diffusion index techniques offer a simple and sensible approach for extracting common factors that underlie the dynamic evolution of large numbers of variables. To be more specific, let Y be a time series vector of dimension $(T \times 1)$, let X be a time-series predictor matrix of dimension $(T \times N)$, and define the following factor model, where F_t denotes a $1 \times r$ vector of unobserved common factors that can be extracted from X_t. Namely, let $X_t = F_t \Lambda' + e_t$, where e_t is an $1 \times N$ vector of disturbances and Λ is an $N \times r$ coefficient matrix. Using common factors extracted from the above model, Stock and Watson (2002a,b) as well as Bai and Ng (2006a) examine linear autoregressive (AR) forecasting models augmented by the inclusion of common factors.

In this paper, we use the forecasting models of Stock and Watson (2002a,b) and Bai and Ng (2006a) as a starting point. In particular, we first estimate unobserved factors, say $\hat{F}_t$, and then forecast a scalar target variable, Y_{t+h}, using observed variables

☆ This paper was prepared for the Sir Clive W.J. Granger Memorial Conference held at Nottingham University in May 2010. We are very grateful to the organizers, Rob Taylor and David Harvey, for hosting this special event. We have benefited from numerous useful comments made on earlier versions of this paper by the editor, Graham Elliott, and by an anonymous referee. In addition to thanking the editor and referee, the authors wish to thank the seminar participants from the conference, as well as at the Bank of Canada and Rutgers University, for useful comments. Additional thanks are owed to Nii Armah, Dongjun Chung, Valentina Corradi, David Hendry, Gary Koop, John Landon-Lane, Fuchun Li, Greg Tkacz, Hiroki Tsurumi, and Hal White for numerous useful suggestions on an earlier version of this paper. The views stated herein are those of the authors and not necessarily those of the Bank of Korea.

* Corresponding author.
E-mail addresses: khdouble@bok.or.kr (H.H. Kim), nswanson@econ.rutgers.edu (N.R. Swanson).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.08.033

and $\hat{F}_t$. We then draw on the fact that even though factor models are now widely used, several issues remain outstanding, such as the determination of the (number of) factors to be used in subsequent prediction model specification (see e.g., Bai and Ng, 2002, 2006b, 2008). In light of this, and in order to add functional flexibility, we implement prediction models where the numbers and functions of factors are selected using a variety of shrinkage methods. In this sense, we add to the recent work of Stock and Watson (2012) as well as Bai and Ng (2008, 2009), who survey several methods for shrinkage in the context of factor augmented autoregression models. Shrinkage methods considered in this paper include bagging, boosting, Bayesian model averaging, simple model averaging, ridge regression, least angle regression, elastic net and the non-negative garotte. We also evaluate various linear models, and hence also add to the recent work of Pesaran et al. (2011), who carry out a broad examination of factor-augmented vector autoregression models.

In summary, the purpose of this paper is to empirically assess the predictive accuracy of various linear models; pure principal component models; principal components models where the factors are constructed using subsets of variables first selected based on shrinkage techniques; principle components models where the factors are first constructed, and are then refined using shrinkage methods; models constructed by directly applying shrinkage methods (other than principle components) to the data; and a number of model averaging methods. The "horse-race" that we carry out allows us to provide new evidence on the usefulness of factors in general as well as on various related issues such as whether model averaging "wins" as often as is usually found to be the case in empirical investigations of this sort.

The variables that we predict include a variety of macroeconomic variables that are useful for evaluating the state of the economy. More specifically, forecasts are constructed for eleven series, including: the unemployment rate, personal income less transfer payments, the 10 year Treasury-bond yield, the consumer price index, the producer price index, non-farm payroll employment, housing starts, industrial production, M2, the S&P 500 index, and gross domestic product. These variables constitute 11 of the 14 variables (for which long data samples are available) that the Federal Reserve takes into account, when formulating the nation's monetary policy. In particular, as has been noted in Armah and Swanson (2011) and on the Federal Reserve Bank of New York's website: "*In formulating the nation's monetary policy, the Federal Reserve considers a number of factors, including the economic and financial indicators which follow, as well as the anecdotal reports compiled in the Beige Book. Real Gross Domestic Product (GDP); Consumer Price Index (CPI); Nonfarm Payroll Employment Housing Starts; Industrial Production/Capacity Utilization; Retail Sales; Business Sales and Inventories; Advance Durable Goods Shipments, New Orders and Unfilled Orders; Lightweight Vehicle Sales; Yield on 10-year Treasury Bond; S&P 500 Stock Index; M2*".

Our finding can be summarized as follows. First, for a number of our target variables, we find that various sophisticated shrinkage methods, such as component-wise boosting, bagging, ridge regression, least angle regression, the elastic net, and the non-negative garotte yield predictions with lower mean square forecast errors (MSFEs) than a variety of benchmark linear autoregressive forecasting models constructed using only observable variables. Moreover, these shrinkage methods, when used in conjunction with diffusion indexes, yield a surprising number of MSFE "best" models, hence suggesting that "hybrid" models that combine diffusion index methodology with other shrinkage techniques offer a convenient way to filter the information contained in large-scale economic datasets, particularly if they are specified using sophisticated shrinkage techniques. More specifically, when using recursive estimation windows, which dominate other "windowing" approaches, "hybrid" models are MSFE "best" around 1/3 of the time, when used to predict 11 key macroeconomic indicators at various forecast horizons. Baseline linear (factor) models also "win" around 1/3 of the time, as do model averaging methods. Interestingly, these broad findings change noticeably when considering different sub-samples. For example, when used to predict only recessionary periods, "hybrid" models "win" in 7 of 11 cases, when condensing findings across all "windowing" approaches, estimation methods, and models, while model averaging does not "win" in a single case. However, in expansions, and during the 1990s, model averaging wins almost 1/2 of the time. Overall, combination factor/shrinkage methods "win" approximately 1/2 of the time in 4 of 6 different sample periods. Ancillary findings based on our forecasting experiments underscore the advantages of using recursive estimation strategies,[1] and provide new evidence of the usefulness of yield and yield-spread variables in nonlinear prediction model specification.

Although we leave many important issues to future research, such as the prevalence of structural breaks other than level shifts, and the use of even more general nonlinear methods for describing the data series that we examine, we believe that results presented in this paper add not only to the diffusion index literature, but also to the extraordinary collection of papers on forecasting that Clive W.J. Granger wrote during his decades long research career. Indeed, as we and others have said many times, we believe that Sir Clive W.J. Granger is in many respects the father of time series forecasting, and we salute his innumerable contributions in areas from predictive accuracy testing, model selection analysis, and forecast combination, to forecast loss function analysis, forecasting using nonstationary data, and nonlinear forecasting model specification.

The rest of the paper is organized as follows. In the next section we provide a brief survey of diffusion index models. In Section 3, we briefly survey the robust shrinkage estimation methods used in our prediction experiments. Data, forecasting methods, and benchmark forecasting models are discussed in Sections 4 and 5, while empirical results are presented in Section 6. Concluding remarks are gathered in Section 7.

2. Diffusion index models

Recent forecasting studies using large-scale datasets and pseudo out-of-sample forecasting include: Artis et al. (2005), Boivin and Ng (2005, 2006), Forni et al. (2005), and Stock and Watson (1999, 2002a,b, 2005, 2006, 2012). Stock and Watson (2006) discuss in some detail the literature on the use of diffusion indices for forecasting. In the following brief discussion of diffusion index methodology, we follow Stock and Watson (2002a).

Let X_{tj} be the observed datum for the j-th cross-sectional unit at time t, for $t = 1, \ldots, T$ and $j = 1, \ldots, N$. We begin with the following model:

$$X_{tj} = F_t \Lambda_j' + e_{tj}, \tag{1}$$

where F_t is a $1 \times r$ vector of common factors, Λ_j is an $1 \times r$ vector of factor loadings associated with F_t, and e_{tj} is the idiosyncratic component of X_{tj}. The product $F_t \Lambda_j'$ is called the common component of X_{tj}. This is a useful dimension reducing factor representation of the data, particularly when $r \ll N$, as is usually assumed to be the case in the empirical literature. Following Bai and Ng (2002), the whole panel of data $X = (X_1, \ldots, X_N)$, where X_i, $i = 1, \ldots, N$, is a $T \times 1$ vector of observations on a single variable, can be represented as in (1). Connor and Korajczyk (1986,

[1] For further discussion of estimation windows and the related issue of structural breaks, see Pesaran et al. (2011).