International Journal of Forecasting 32 (2016) 411–436

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts

W. Jos Jansen[a], Xiaowen Jin[b], Jasper M. de Winter[c,*]

[a] Financial and Economic Policy Department, Ministry of Finance, P.O. Box 20201, The Hague, Netherlands

[b] Munich Graduate School of Economics, Ludwig Maximilians Universität, P.O. Box 1111, München, Germany

[c] Economic Policy and Research Division, De Nederlandsche Bank, P.O. Box 98, 1000 AB, Amsterdam, Netherlands

ARTICLE INFO

Keywords:
Forecasting competitions
Factor models
Professional forecasters
Judgment

ABSTRACT

We conduct a systematic comparison of the short-term forecasting abilities of twelve statistical models and professional analysts in a pseudo-real-time setting, using a large set of monthly indicators. Our analysis covers the euro area and its five largest countries over the years 1996–2011. We find summarizing the available monthly information in a few factors to be a more promising forecasting strategy than averaging a large number of single-indicator-based forecasts. Moreover, it is important to make use of all available monthly observations. The dynamic factor model is the best model overall, particularly for nowcasting and backcasting, due to its ability to incorporate more information (factors). Judgmental forecasts by professional analysts often embody valuable information that could be used to enhance the forecasts derived from purely mechanical procedures.

1. Introduction

Information on economic activity and its short-term prospects is very important for decision makers in governments, central banks, financial markets and non-financial firms. Monetary and economic policy makers and economic agents have to make decisions in real time based on incomplete and inaccurate information on current economic conditions. A key indicator of the state of the economy is the growth rate of real GDP, which is available on a quarterly basis only, and is also subject to substantial publication lags. In many countries, an initial estimate of quarterly real GDP is published around six weeks after the end of the quarter. Moreover, real GDP data are subject to revisions that can be substantial, as more data become available to statistical offices over time.

Fortunately, though, a lot of statistical information related to economic activity is published on a more frequent and timely basis. This information includes data on industrial production, prices of goods and services, expenditures, unemployment, financial market prices, loans, and consumer and business confidence. Recently, the forecasting literature has developed several statistical approaches for exploiting this potentially very large information set in order to improve the assessment of both real GDP growth in the current quarter (nowcast) and its development in the near future. Examples of such approaches include bridge models, factor models, mixed-data sampling models (MIDAS) and mixed-frequency vector-autoregressive (MFVAR) models. These models differ in their solutions to the practical problems of dealing with large information sets and the fact that the auxiliary variables are observed at different frequencies and with different publication lags.

Practitioners now have a wealth of statistical models to choose from; but which one should they use? As each model has its own strengths and weaknesses, it is difficult

* Corresponding author.
E-mail addresses: W.J.Jansen@minfin.nl (W. Jos Jansen), xiaowenjin1115@gmail.com (Xiaowen Jin), J.M.de.Winter@dnb.nl (Jasper M. de Winter).

http://dx.doi.org/10.1016/j.ijforecast.2015.05.008

to make a decision on purely theoretical grounds. The ranking of the models in terms of forecasting abilities, and the extent to which this varies with the prediction horizon or the economic circumstances, has to be determined by empirical analysis. On these issues the jury is still out, however, as large-scale comparative studies are scarce. In many papers, the empirical work refers only to a single country, and usually only limited numbers of models are included. Furthermore, studies differ in the size of the information set and the sample period used.[1]

This paper is motivated by this gap in the empirical literature. We undertake a systematic comparison of a broad range of linear statistical models – twelve models in all – that have been applied in the recent literature. For the sake of comparability and robustness, we include five countries (Germany, France, Italy, Spain and the Netherlands) and the euro area in our analysis, and utilize an information set that is as homogeneous as possible across geographical entities. Moreover, our sample includes the volatile episode of the financial crisis of 2008 and its aftermath, which may make it easier to distinguish between the various models. We contrast the models' forecasting abilities before 2008 with those during the crisis period. This may be of great interest for policy makers, financial analysts and economic agents alike, as information on where the economy stands and where it is headed in the immediate short run is particularly valuable at times of great uncertainty.

The provision of cross-country evidence on the relative performances of twelve different statistical forecasting models is our first contribution to the literature. Model forecasts are the result of purely mechanical recipes, and do not incorporate subjective elements. Our second contribution concerns the potential usefulness of forecasts made by professional analysts (published by Consensus Forecasts on a quarterly basis). From a practical point of view, such forecasts are very cheap and easy to use. Moreover, as an expression of the "wisdom of crowds", they may reflect much more information than the statistical information set, which is inevitably limited. A questionnaire conducted by the European Central Bank (ECB) among the participants of the ECB Survey of Professional Forecasters found that the panelists regard 40% of their short-term GDP forecasts as being judgment-based (Meyler & Rubene, 2009). We investigate the extent to which the subjective forecasts by analysts in our sample contain information beyond that generated by the best mechanical statistical models.

The remainder of the paper is structured as follows. Section 2 describes the various statistical models and discusses how they deal with the challenges posed by large and irregularly shaped datasets. Section 3 describes the data, our pseudo real-time forecast design, and other specification issues. Sections 4 and 5 present the results for the mechanical models and the professional forecasts, respectively. Section 6 summarizes our findings and concludes.

2. Linear statistical models for short-term GDP forecasting

2.1. Overview

In practice, taking advantage of auxiliary information for the forecasting of real GDP in the immediate short run poses several challenges. The first challenge is posed by the large size of the information set. There are countless potentially useful variables for forecasting GDP, and often they are also interrelated. The datasets used in the empirical literature vary greatly in size, and may include more than 300 variables. Moreover, the limited length of the time series involved makes over-parametrization a real issue. The second problem relates to the fact that the indicator variables are observed more frequently (monthly, weekly, daily) than GDP. Moreover, the dating of the most recent observation may vary across indicators because of differences in publication lags. This is known as the "ragged edge" problem; see Wallis (1986).

The various statistical approaches in the literature deal with these challenges in different ways. Broadly speaking, a forecasting procedure involves two transformations of the dataset of indicators in order to produce a final forecast: an aggregation and the application of a forecasting tool, which links auxiliary variables to real GDP growth. These two transformations can be executed in either order, representing two fundamentally different strategies. The first strategy begins by computing an indicator-specific GDP forecast for each variable, which are then aggregated into a single final forecast in the second step. We call this strategy the "pooling forecasts strategy". In this approach, it is necessary to specify the weighting scheme for the individual forecasts. A basic scheme is the simple average, which gives each forecast an equal weight, but weights may also be computed recursively depending on the indicators' (recent) forecasting performances. Examples of the forecast pooling strategy include bridge equations and VAR models. In contrast, the "aggregating information strategy" takes the aggregation step first, by summarizing the large dataset in a small number of series. This strategy exploits the fact that the auxiliary variables are correlated. Factor analysis is used to replace a large number of correlated time series with a limited number of uncorrelated (unobserved) factors representing the common information component of the original data series. The implicit weights (factor loadings) are determined from the correlation patterns in the original dataset. The factors serve as inputs for the forecasting procedure in the next step. Examples of this modeling strategy include dynamic factor models and factor augmented versions of forecasting models that pool forecasts. Finally, a recent development is estimation using Bayesian shrinkage on coefficients, which translates a large set of indicators into a single GDP forecast directly, without a clear

[1] Rünstler et al. (2009) form an important exception, comparing three factor models, a bridge model and a quarterly VAR model for ten European countries; however, their study does not include the financial crisis. Kuzin, Marcellino, and Schumacher (2013) analyzed the relative forecasting performances of MIDAS models and dynamic factor models, including part of the crisis years (2008–2009). Liebermann (2012) analyzed the relative forecasting performances of a range of models over the years 2001–2011, but only for the United States.