International Journal of Forecasting 31 (2015) 188–206

Contents lists available at ScienceDirect

International Journal of Forecasting

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

# Forecasting economic activity with targeted predictors

Guido Bulligan[a], Massimiliano Marcellino[b,c,*], Fabrizio Venditti[a]

[a] Banca d'Italia, Italy
[b] Bocconi University, IGIER, Italy
[c] CEPR, London, United Kingdom

ARTICLE INFO

*Keywords:*
Nowcasting
Forecasting
Mixed frequency data
Variable selection
Bridge models
Principal components
Factor models

ABSTRACT

In this paper we explore the forecasting performances of methods based on a pre-selection of monthly indicators from large panels of time series. After a preliminary data reduction step based on different shrinkage techniques, we compare the accuracy of principal components forecasts with that of parsimonious regressions in which further shrinkage is achieved using the General-To-Specific approach. In an empirical application, we show that the two competing models produce accurate current-quarter forecasts of Italian GDP and of its main demand components, outperforming naïve forecasts and comparing favorably with factor models based on all available information. A robustness check conducted on the GDP growth of the euro area and of its major members confirms these results.


## 1. Introduction

Although a large number of indicators covering all aspects of the economy are usually available at very high frequencies, quarterly national accounts still play a central role in guiding economic decisions and policy analysis. However, the delay with which they are released complicates decision making greatly. Over the past decade, a number of econometric tools have been developed to solve this problem.

One kind of older method, known as bridge models, are single equations in which lower frequency (typically quarterly) target variables are regressed against higher frequency (typically monthly) indicators preliminarily aggregated at the lower frequency, see Baffigi, Golinelli, and Parigi (2004), Barhoumi et al. (2008), Diron (2008) and Hahn and Skudelny (2008). Albeit very simple, these models are still used widely within policy institutions and by private forecasters, for a number of reasons. First, they strike a good compromise between simplicity and accuracy: a small set of indicators appropriately chosen usually guarantees a good forecasting performance. Second, forecasts based on single linear equations are very easy to explain and to communicate to decision makers. Third, dissecting forecast errors is also very easy in a linear context: discrepancies between actual and predicted values of the target variables can be related straightforwardly to those between actual and predicted values in the underlying indicators. However, bridge models present two important drawbacks: (i) they rely on a very parsimonious information set, potentially leaving out informative predictors, (ii) their specification often relies on the judgement and experience of the econometrician.

Their ability to address both of these issues is the reason for the attention that factor models, which have rapidly become the workhorse of short-term forecasting, have attracted in recent years. In these models, the information from a potentially very large dataset is summarized by a small number of linear combinations of the available time series, so that no valuable information is lost. Furthermore, the specification of a factor model requires little judgement: once the number of factors

* Correspondence to: Department of Economics, Bocconi University, Via Roentgen 1, 20136 Milan, Italy.

*E-mail addresses:* guido.bulligan@bancaditalia.it (G. Bulligan), massimiliano.marcellino@unibocconi.it (M. Marcellino), fabrizio.venditti@bancaditalia.it (F. Venditti).

http://dx.doi.org/10.1016/j.ijforecast.2014.03.004
0169-2070/

has been determined on the basis of some information criterion, the common factors can be estimated using various methods (see Forni, Hallin, Lippi, & Reichlin, 2005; Giannone, Reichlin, & Small, 2008; Kapetanios & Marcellino, 2009; Stock & Watson, 2002, for different estimation techniques), and a forecasting equation can be specified easily. In an empirical application, Angelini, Camba-Mendez, Giannone, Reichlin, and Rünstler (2011) find that, on average, factor models score better than bridge models in forecasting euro area GDP.

The issue of variable selection, which is crucial in the context of bridge models, is usually swept under the carpet in the factor model literature, where it seems that all that is needed is a large number of variables that can be used to average out the influence of idiosyncratic components and to estimate the common factors. A recent branch of the literature has questioned the usefulness of "too much information" for factor forecasts. Boivin and Ng (2006), for example, argue that increasing the $N$-dimension of large panels can be detrimental, especially if the errors are strongly cross-correlated and the forecasting power is provided by a factor that is dominant in a small panel but dominated in a larger panel. This problem arises because factors are extracted 'blindly', without taking into consideration the properties of the variable that the researcher is really interested in forecasting. To put it roughly, since principal components maximize the signal to noise ratio of the whole panel, they are well suited for forecasting the variables which load the common factors more strongly, but may perform poorly for other variables. Tailoring the predictors to a specific target variable can then provide substantial gains. Bai and Ng (2008) show that factor forecasts can be improved by identifying useful "targeted" predictors and computing principal components on this restricted dataset. In particular, they show that soft thresholding methods like the least absolute shrinkage selection operator (LASSO) can be used successfully to reduce the size of the information set.

An interesting connection between factor model forecasts and thresholding methods has recently been established by De Mol, Giannone, and Reichlin (2008). They find that, as the panel dimension increases, factor forecasts become more highly correlated with those obtained with LASSO, i.e., with a regression on a few selected predictors. They conclude that "… the result that few selected variables are able to capture the space spanned by the common factors, suggests that small models with accurately selected variables may do as well as methods that use information on large panels and are based on regressions on linear combinations of all variables. This point calls for further research…".

This open question constitutes the main motivation of our paper. However, we go beyond simple Lasso regressions, or more generally regressions of target variables on targeted predictors, by intersecting the targeted predictors argument with the General To Specific (GETS) modeling philosophy (see Hoover & Perez, 1999) that underlies the bridge approach (Krolzig & Hendry, 2001). Our analysis also proceeds in two steps. In the first step we follow Bai and Ng (2008) and use a range of hard- and soft-thresholding methods to reduce the dimension of a large dataset to a limited number of potential regressors. In the second step, information extraction is accomplished through an automatic selection algorithm which picks the most informative variables and specifies parsimonious bridge equations, in order to replicate the process usually followed by the econometricians, guided by their judgement and experience, when setting up bridge models.[1] Hence, our first methodological contribution relates to the specification of bridge equations in the presence of a large set of potentially useful indicators, based on sound statistical procedures rather than simply on either the experience and preferences of the bridge model developer or the use of information criteria and/or testing with only a small set of indicators.

Our second contribution is a comparison, in terms of forecasting performances, of this enhanced bridge approach with simple AR models that do not use any external information, with Diffusion Index models estimated on targeted predictors (as in Bai & Ng, 2008), and with general Diffusion Index models based on all of the information available. We can therefore assess: (i) the accuracy gain associated with monthly timely information, (ii) the "harmfulness" of "too much information", and (iii) the relative gains of two alternative ways of extracting information from targeted predictors (selection based on statistical criteria or information extraction by means of factor estimation).

Our empirical analysis focuses on Italian GDP and on the main demand components. The motivation for looking not only at GDP but also at the demand breakdown is twofold. First, factor models have been employed frequently for forecasting GDP, but seldom if ever for forecasting demand components. However, the business cycle behavior of aggregate GDP is very different from that of its components. For example, investment and trade variables are much more volatile than aggregate GDP, while Private Consumption is typically smoother than total activity, see Artis, Marcellino, and Proietti (2004). Checking how models compare for forecasting variables that behave so differently over the business cycle is an interesting exercise on its own. Second, forecasting demand aggregates is extremely important at the turn of the cycle and in turbulent phases. For example, investment tends to trough before GDP, while consumption only achieves momentum when an expansion is well under way, peaking after the cycle. Having models that complement GDP forecasts with a view on the main drivers of economic activity enables business cycle analysts to provide a much more accurate reading of the cyclical phase.

Our application is to one-step-ahead forecasts of Italian GDP and of the main demand breakdown, that is, Private Consumption, Investment in Construction, Other Investment, Exports and Imports. By *one step ahead*, we mean *the next quarterly release*. Given the delay with which quarterly series are published, this actually amounts to performing a nowcast/backcast exercise. We deliberately limit our forecast horizon to the next quarterly release because we are interested in gauging the relative merits

1 The GETS methodology is implemented using the freeware software GROCER (see http://dubois.ensae.net/grocer.html).