Economic Modelling 44 (2015) 266–272

Contents lists available at ScienceDirect

Economic Modelling

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

Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence

Francisco Dias [a], Maximiano Pinheiro [b], António Rua [a,c,*]

[a] Economics and Research Department, Banco de Portugal, Av. Almirante Reis no 71, 1150-012 Lisboa, Portugal
[b] Banco de Portugal, Av. Almirante Reis no 71, 1150-012 Lisboa, Portugal
[c] Nova School of Business and Economics, Universidade Nova de Lisboa, Lisboa, Portugal

ARTICLE INFO

Article history:
Accepted 17 October 2014
Available online 12 November 2014

Keywords:
Factor models
Diffusion index
Targeted diffusion index
Forecasting
Crisis

ABSTRACT

In this article, we assess the relative performance of factor models to forecast GDP growth in Portugal. A large dataset is compiled for the Portuguese economy and its usefulness for nowcasting and short-term forecasting is investigated. Since, in practice, one has to cope with different publication lags and unbalanced data, we also address the pseudo real-time performance of such models. Furthermore, by considering a relatively long out-of-sample period, we are able to evaluate the behavior of the different models over the pre-crisis period and during the latest economic and financial crisis. As Portugal was one of the hardest hit economies, it is a particularly insightful case to assess the relative performance of factor models during a period of economic stress.

1. Introduction

With the widespread development of the statistical systems, the information set available to policymakers has become progressively larger. Naturally, this poses methodological challenges in terms of how to take on board all the available data, which can involve hundreds of series.

For forecasting purposes, the use of factor models to forecast macroeconomic variables in a data rich environment has become increasingly popular in the literature and among practitioners at central banks and international institutions. See, for example, Stock and Watson (1998, 2002a, 2002b) and Giannone et al. (2008) for the United States, Marcellino et al. (2003) and Angelini et al. (2011) for the euro area, Artis et al. (2005) for the UK, Schumacher (2007, 2010, 2011) and Schumacher and Breitung (2008) for Germany, Barhoumi et al. (2010) for France, de Winter (2011) and den Reijer (2013) for the Netherlands, and for a cross-country study encompassing several European countries see Rünstler et al. (2009).

Factor models allow circumventing the curse of dimensionality when dealing with large datasets by reducing the dimension of the number of series to a manageable scale, which is particularly useful in the case of forecasting. In fact, these models allow one to summarize the information contained in large databases in a set of a handful of unobserved common factors that drive a sizeable fraction of the overall comovement among the whole set of variables in the dataset. However, since it ignores entirely the information content other than the one conveyed by this small set of factors, it may potentially disregard data that can be useful for the variable to be forecasted or the forecast horizon under consideration.

Dias et al. (2010) suggest an alternative procedure to overcome the above mentioned shortfall. In particular, a tailor made targeted diffusion index (TDI) dependent on the variable to be forecasted and the forecast horizon is proposed. This index is simply a weighted average of all the factors of the dataset that take into account both the explanatory power of each factor for the variable to be forecasted and the relative importance of the factor in capturing the total variation of the series. For the US case, such an approach outperforms the standard factor model in forecasting several macroeconomic variables.

Herein, we focus on the Portuguese case which was one of the hardest hit economies as from the latest economic and financial crisis. In particular, we assess the performance of several alternative factor models to forecast GDP growth using a large dataset compiled for Portugal, which encompasses 126 monthly series.

By considering a relatively long out-of-sample period, from 2002 up to 2013, we can assess the relative performance of the different models during the pre-crisis period and during the latest years where pronounced GDP downturns and upswings were observed. This can be particularly useful to assess the robustness of the forecasting performance of factor models in periods of significant economic stress.

Furthermore, since forecasting in real-time typically involves missing observations for some of the variables due to different release lags, we also address how to overcome this issue and evaluate the corresponding pseudo real-time forecasting performance.

The article is organized as follows. In Section 2, an introductory overview of the factor models considered in subsequent analysis is provided.

* Corresponding author at: Economics and Research Department, Banco de Portugal, Av. Almirante Reis no 71, 1150-012 Lisboa, Portugal.
E-mail address: antonio.rua@bportugal.pt (A. Rua).

http://dx.doi.org/10.1016/j.econmod.2014.10.034
0264-9993/

In Section 3, we describe the dataset for Portugal whereas in Section 4 the estimated common factors are discussed. In Section 5, we assess the out-of-sample forecasting performance with balanced data. In Section 6, the issue of how to deal with unbalanced data is addressed whereas in Section 7 the pseudo real-time performance is evaluated. Finally, Section 8 concludes.

2. The factor models

More formally, the static factor model assumes that each and every variable in the data set can be specified as a combination of two terms: one component driven by a small set of latent unobserved static factors common to all variables and an idiosyncratic component specific to each variable, that is

$$X_t = \Lambda F_t + e_t$$

where X_t is the N-dimensional vector time series in the panel, Λ is an (Nxr) matrix of factor loadings, F_t is the vector of r unobserved common factors and e_t is the N-dimensional vector of idiosyncratic terms. The unobserved factors can be estimated relying on the principal components technique which is shown to provide a consistent estimator of the factor space under fairly general conditions.

Dynamic factor models, on the other hand, were originally developed by Geweke (1977), Sargent and Sims (1981), Geweke and Singleton (1977) and Watson and Engle (1983) and applied in the context of a limited number of variables. This type of model has been extended to handle the information conveyed by large data sets. The dynamic factor model has an equivalent static factor model representation, where the r-dimensional static factors comprise both current and lagged values of the q dynamic factors. If the number of static and dynamic factors are the same, that is, $r = q$, then there is no difference between the static and dynamic forms (see Stock and Watson, 2005). Moreover, as pointed out by Bai and Ng (2007), not much is expected to be gained from the distinction between the static factors and the dynamic factors for forecasting purposes.

Typically, the first few top-ranked principal components capture a sizeable share of the collinearity among the series in the dataset. Once the number of factors is selected, the variable to be forecasted y is projected on the set of the r estimated factors and possibly lags of the dependent variable. This results in the following forecasting model

$$y_{t+h} = \beta_0 + \sum_{i=1}^{r} \beta_i \hat{F}_{t,i} + \sum_{j=0}^{p} \delta_j y_{t-j} + v_{t+h}$$

where h refers to the forecast horizon, y_{t-i} are the autoregressive components of the regression and v_{t+h} denotes the forecast error. Such an approach corresponds to the so-called diffusion index (DI) model proposed by Stock and Watson (1998, 2002a, 2002b).

In practice, the above discussed factor model requires a priori the determination of the number of factors and the space spanned by those factors draws on the main principal components. In fact, the factors reflect the top-ranked principal components, that is, the ones that encompass the largest share of the common comovement in the dataset. All other lower-ranked factors are entirely disregarded independently of their possible informational content for forecasting the variable of interest. This can result in an important shortcoming for forecasting purposes as such an approach does not take into account neither the specific variable to be forecasted nor the forecast horizon. This shortfall was circumvented in Dias et al. (2010) where the authors propose a targeted diffusion index (TDI), which reconciles both the spirit of the Stock and Watson approach and the targeting principle discussed by Bai and Ng (2008). Basically, the suggested procedure considers in the forecasting model a synthetic regressor which is computed as a linear combination of all the factors of the dataset, that is

$$y_{t+h} = \beta_0 + \beta_1 \mathring{F}_{(h)t} + \sum_{j=0}^{p} \delta_j y_{t-j} + v_{t+h}$$

$$\mathring{F}_{(h)t} = \sum_{n=1}^{N} \left(\frac{\omega_{(h)n}}{\sum_{i=1}^{N} \omega_{(h)i}} \right) \hat{F}_{(h)t,n}$$

$$\omega_{(h)n} = \left(\frac{1}{T-h} \sum_{t=1}^{T-h} \hat{F}_{(h)t,n} y_{t+h} \right) \left(\frac{\varphi_{(h)n}}{\varphi_{(h)1}} \right).$$

The first equation is the same as in the case of the DI approach but where the top-ranked principal components, i.e. the common factors, are replaced by the synthetic composite indicator. This targeted diffusion index is the convex linear combination of all the factors derived from the database, where the weights attached to each factor take into account both the relative size of the overall variation captured by each factor $\left(\frac{\varphi_{(h)n}}{\varphi_{(h)1}}\right)$ and its correlation with the variable of interest at the relevant forecast horizon $\left(\frac{1}{T-h} \sum_{t=1}^{T-h} \hat{F}_{(h)t,n} y_{t+h}\right)$. The weights attached to each factor are naturally dependent not only on the relative importance of the factor but also on the specific series to be forecasted and corresponding forecast horizon. This modeling strategy avoids discarding potentially relevant information contained in the dataset and tries to obtain a better match between the available data and the variable to be forecasted. As shown in Dias et al. (2010), this approach proved to be quite promising vis-à-vis the diffusion index model, improving considerably the forecast performance for several US macroeconomic variables.

3. Dataset

The monthly dataset compiled for the Portuguese economy comprises 126 series and it includes both hard and soft data.[1] It covers business and consumers surveys (43 series), retail sales (4 series), industrial production (7 series), turnover in industry and services (20 series), employment, hours worked and wage indices in industry and services (24 series), tourism nights spent in Portugal (3 series), car sales (3 series), cement sales, vacancies and registered unemployment (5 series), energy consumption (3 series), goods exports and imports (10 series), real effective exchange rate, Portuguese stock market index and ATM/POS series. Although most series are provided on a seasonally adjusted basis, for those variables that are not but which present a seasonal pattern, a seasonal adjustment was conducted resorting to X12-ARIMA. The sample period runs from the beginning of 1995 up to the end of 2013 (T = 228 monthly observations). Since for some variables the series start later than 1995, we resort to the Expectation–Maximization (EM) algorithm suggested by Stock and Watson (2002a) to balance the dataset at the beginning of the sample period.

Regarding GDP, the series in real terms is available from the Portuguese National Statistics Office (INE) as from the first quarter of 1995 up to the fourth quarter of 2013 on a seasonally adjusted basis.

With the exception of survey data, all series are taken in logarithms. The series are then differenced to obtain stationarity. For GDP we took the first-difference of the quarterly series, which corresponds to the quarter-on-quarter growth rate. For the monthly series we compute a

[1] A detailed list of the series and corresponding source is available from the authors upon request.