MACROECONOMIC FORECASTING: A NON-STANDARD OPTIMISATION APPROACH TO THE CALIBRATION OF DYNAMIC FACTOR MODELS

Fabio Della Marra [1]

[1] Department of Economics, University of Parma, (e-mail: fabio.dellamarra@iispascal.it)

ABSTRACT: In this paper, we present a comparison of the forecasting perfomance of selected static and dynamic factor models on two large monthly data panels. The first dataset contains EU variables, whereas the other contains US variables. These data panels are split into two parts: the first subsample (the *calibration sample*) is used to select the most performing specification for each class of models in a in-sample environment and the second subsample (the *proper sample*) is used to compare the performances of the selected models in an out-of-sample environment. In the calibration sample, genetic algorithms are employed to achieve an efficient exploration of the parameter space. We find that dynamic factor models are globally the most performing methods on both data panels.

KEYWORDS: macreoconomic forecasting, dynamic factor models, genetic algorithms.

1 Introduction

In this paper, a comparative analysis of the forecasting performance of three Large-Dimensional Dynamic Factor Models is presented. As a key feature, Dynamic Factor Models represent each variable in a dataset as the sum of two orthogonal terms: a *common component* χ_t, driven by a reduced (as compared to the number of series in the dataset) number of common factors, and an *idiosyncratic component* ξ_t, which represents measurement errors or local features. Among the different versions of the Dynamic Factor Models we selected:

(i) *SW model*. This time-domain method was introduced in Stock & Watson, (2002a), Stock & Watson, (2002b). The factors are estimated by computing static principal components of the variables in the dataset. Let y_{it} be the variable of the dataset to be forecasted at time t, its h-step-ahead prediction equation (also called *Diffusion Forecast Index*) is obtained by regressing y_{it+h} on the factors and on y_{it} itself. Lags of the factors and of y_{it} may be added.

(ii) *FHLR model*. This frequency-domain method was proposed in Forni *et al.* , (2000), Forni *et al.* , (2005) and requires the computation of two steps. In a first step, the common component χ_t, the idiosyncratic component ξ_t and their covariances are estimated using a frequency-domain method introduced in Forni *et al.* , (2000) named *Dynamic Principal Component*. In the second step, the factors are estimated by computing Generalized Principal Components.

(iii) *FHLZ model*. This frequency-domain method was proposed in Forni *et al.* , (2015), Forni *et al.* , (2016b). Here, the underlying assumption in (i) and (ii) that the common components span a finite-dimensional space as n tends to infinity is relaxed.

There exists some literature comparing the forecasting performances of SW and FHLR, but universal consensus still does not seem to have been reached. Theoretically, time-domain methods consider only relations among the variables at the same time, whereas frequency-domain methods exploit leaded and lagged relations among the variables. However time-domain methods require less parameters to be calibrated. Hence they are more robust to misspecification than frequency-domain methods. Instead, a systematic comparison of the forecasting performances of SW, FHLR and FHLZ can be found only in Forni *et al.* , (2016a), Della Marra, (2017). Forni *et al.* , (2016a) conducted a forecasting exercise on a US macroeconomic dataset, taking an autoregressive process of order 4 as a benchmark. They showed that FHLZ oupterforms SW, FHLR and the benchmark both for Industrial Production and Inflation during the Great Moderation. In the Great Recession, the forecasting performances of the Industrial Production change dramatically: all factor models are outperformed by the benchmark. SW and FHLR outperform FHLZ. Hence, Forni et al. concluded that, due to its more dynamical structure, FHLZ tends to be the best performing method in "stationary periods", but it loses ground during regime changes. Also, they showed that FHLZ tends to be outperforming on nominal variables and FHLR on real variables. Della Marra, (2017) conducted a forecasting exercise on an EU macroeconomic dataset. The global settings of his exercise are basically the same as in Forni *et al.* , (2016a), but also the length of the rolling window is suboptimally selected during the calibration process. He found that, on the proper sample, FHLZ is the most performing for the Inflation. However, mixed evidencies appear over the proper sample for the Industrial Production.

In this paper, the EU dataset is the same employed in Della Marra, (2017).

This dataset is split into two subsamples. To guarantee balancedness, time series with missing data are discarded. The former, from February 1986 to December 2000, is used to calibrate the models, i.e. to produce in-sample forecasts of the variables of the EU dataset for several specifications of SW, FHLR and FHLZ. Then, for each class of models, we select the specification which shows the minimum mean square forecast error (MSFE). These models are then run and compared in the remaining sample, from January 2001 to November 2015. Instead, the US dataset employed in our exercise is accurately described in McCracken & Ng, (2016). This dataset is split into two subsamples. The former, from February 1959 to December 1984, is used to calibrate the models. Then, for each class of models, we select the specification which shows the minimum mean square forecast error (MSFE). These models are then run and compared in the remaining sample, from January 1985 to October 2016.

The paper is structured as follows. In Section 2, the calibration process of the models is described. In Section 3, results are discussed and Section 4 concludes.

2 Description of the two datasets and of the calibration process

Both dataset contain real variables (import/export price indexes, employment, Industrial Production) and nominal variables (money aggregates, consumer price indexes, wages), asset prices (stock prices and exchange rates) and surveys. To achieve stationarity, several series are deseasonalized and transformed. No treatment for outliers is applied. In addition to SW, FHLR, FHLZ, the forecasts of an autoregressive process (AR) are computed. The order p of the AR process is determined in the calibration process. As in Stock & Watson, (2002b), D'Agostino & Giannone, (2012), to assess the forecasting performances, the variables which are taken into account are the level of the logarithm of the Industrial Production (IP) and the yearly change of the logarithm of the Consumer Price Index (CPI). Forecasts are computed h-months ahead, with $h \in \{1,3,6,12,24\}$. For each methods, we employ a rolling-window scheme $[t-l,t]$, whose size l is determined in the calibration sample. To assess the forecasting performance of each model, the mean-square forecast error (MSFE) is employed as a metric.

Since each method is characterized by several parameters, an exhaustive exploration of the parameter space would be computationally infeasbile. Hence, we employ genetic algorithms to explore more efficiently the parameter space in the calibration sample of each dataset. At each epoque, the population of the genetic algorithm is a subset of the strings containing all the possibile configurations of the parameters. We set the fitness as the inverse for its MSFE. For each method, we iterate the genetic algorithm ten times on the calibration sample of the two datasets. The fitness of each individual is stored in a data structure. Eventually, for each method we select as the most performing configuration the one endowed with the greatest fitness. The convergence of each iteration of the genetic algorithms is graphically shown by plotting the boxplot of the results. These plots are not reported here.

3 Results

The forecasting performance of the three dynamic factor models over the IP and CPI are compared on the proper sample of each dataset. As in Forni *et al.* , (2016a) and in Della Marra, (2017), to assess the forecasting performance of each couple of methods locally, each time series of the dataset is smoothed by a centered moving average of length $m = 61$ (with coefficients equal to $1/m$) and then the Fluctuation test (Giacomini & Rossi, (2010)) is run, at 5% significance level. The results for the IP at horizon $h = 12$ are reported in figure1.

As to the EU dataset, all methods outperform AR significatively from the crisis on (which, according to CEPR, starts in April 2008). Globally, FHLR and FHLZ outperforms SW from the crisis on. As to the IP, FHLR tends to outperform FHLZ from the crisis on. Instead, as to the CPI, FHLZ tends to outperform FHLR from the crisis on, but evidencies are less significative. As in Forni *et al.* , (2016a) and in Della Marra, (2017), this exercise has been extended to the other variables in the dataset. The results achieved are omitted here, but it can be seen that FHLR tends globally to outperform the other methods on the real variables and that FHLZ tends globally to outperform the other methods on the nominal variables. On the US dataset, all methods tend, instead, to lose ground against AR significatively during the Great Recession. FHLR tends globally to outperform the other methods on the real variables and that FHLZ tends globally to outperform the other methods on the nominal variables.

4 Conclusions

In this paper, we have shown that FHLR tends globally to outperform the other methods on the real variables and that FHLZ tends globally to outperform the other methods on the nominal variables. As to EU dataset, Della Marra, (2017) found similar results for the CPI, but mixed evidencies appeared for

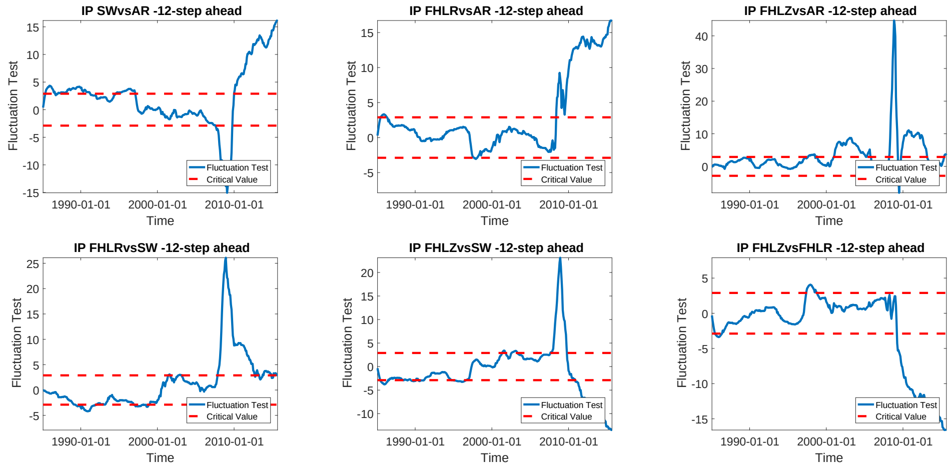

Figure 1: Fluctuation test for the IP.

the IP. As to the US dataset, Forni *et al.* , (2016a) found similar but less significative results. Hence, we have empirically shown that the calibration process plays a crucial role in these applications, since a more efficient exploration of the paramter space allowed us to empirically prove the superiority of frequence-domain dynamic factor models against time-domain factor models in a macroeconomic forecasting setting.

References

BOIVIN, J., & NG, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, **1**(3), 117 – 151.

D'AGOSTINO, A., & GIANNONE, D. (2012). Comparing Alternative Predictors Based on Large-Panel Factor Models. *Oxford bulletin of economics and statistics*, **74**(2), 306–326.

DELLA MARRA, F. (2017). A forecasting performance comparison of dynamic factor models based on static and dynamic methods. *Communications in Applied and Industrial Mathematics*, **8**(1), 44–66.

FORNI, M., & LIPPI, M. (2001). The generalized dynamic factor model: representation theory. *Econometric theory*, **17**(06), 1113–1141.

FORNI, M., HALLIN, M., LIPPI, M., & REICHLIN, L. (2000). The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics*, **82**(4), 540 – 554.

FORNI, M., HALLIN, M., LIPPI, M., & REICHLIN, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, **100**, 830 – 840.

FORNI, M., GIANNONE, D., LIPPI, M., & REICHLIN, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, **25**(05), 1319–1347.

FORNI, M., HALLIN, M., LIPPI, M., & ZAFFARONI, P. (2015). Dynamic factor model with infinite dimensional factor space: representation. *Journal of Econometrics*, **185**, 359–371.

FORNI, M., GIOVANNELLI, A., LIPPI, M., & SOCCORSI, S. (2016a). *Dynamic Factor Model with infinite dimensional factor space: forecasting*. Tech. rept. CEPR Discussion Paper No. DP11161.

FORNI, M., HALLIN, M., LIPPI, M., & ZAFFARONI, P. (2016b). *Dynamic factor model with infinite dimensional factor space: Asymptotic analysis*. Tech. rept. EIEF Working Paper 2016/07, Einaudi Institute for Economics and Finance.

GIACOMINI, R., & ROSSI, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, **25**(4), 595–620.

HALLIN, M., & LIPPI, M. (2013). Factor models in high-dimensional time seriesA time-domain approach. *Stochastic Processes and their Applications*, **123**(7), 2678–2695.

MCCRACKEN, M. W., & NG, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, **34**(4), 574–589.

STOCK, J. H., & WATSON, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**(460), 1167–1179.

STOCK, J. H., & WATSON, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, **20**(2), 147–162.