Short term forecasting of Greek GDP growth using Dynamic Factor Models

Thanassis Kazanas[1]
tkazanas@aueb.gr

Abstract

In recent years, central banks and international organisations have been making ever greater use of factor models to forecast macroeconomic variables. We examine the performance of these models in forecasting Greek GDP growth over short horizons. The factors are extracted from a large data set of around one hundred variables including survey balances as well as real, financial, and international variables.

1. Introduction

Macroeconometricians face a peculiar data structure. On the one hand, the number of years for which there is reliable and relevant data is limited and cannot readily be increased other than by the passage of time. On the other hand, for much of the postwar period statistical agencies have collected monthly or quarterly data on a great many related macroeconomic, financial, and sectoral variables. Thus, macroeconometricians face data sets that have hundreds or even thousands of series, but the number of observations on each series is relatively short, for example 20 to 40 years of quarterly data.

Factor models have received substantial coverage in the literature in recent years (see, e.g, Stock and Watson, 2010; Bai and Ng, 2008b). Central banks and other international organisations are using them increasingly for short-term forecasting of GDP. The models are used in static form (for example at the Federal Reserve [Fed], under the impulse of the studies by Stock and Watson, 1999, 2002a, 2002b) or in dynamic form (at the European Central Bank [ECB], following the studies by Doz, Giannone, and Reichlin, 2011, 2012; Giannone, Reichlin and Small, 2008; at the Bank of Italy with the Eurocoin indicator developed by Altissimo et al., 2001, 2010).

Factor models offer several advantages over classic tools. First, they can incorporate information provided by a large set of variables and summarise it in a small set of factors, which will then serve as explanatory variables in a standard regression model. Second, factor models

[1] Acknowledgments: The author would like to thank Efthymios Tsionas for helpful comments and suggestions.

can be adjusted if observations are missing at the end of a period. This is a valuable property for the short-term economic analyst, who is constrained by the availability of short-term indicators (release times are fairly short for balances of opinion in business and consumer surveys and for financial variables, longer for real variables such as the industrial production index and manufactured-goods consumption). When one uses a factor model does not need to develop auxiliary models to predict missing observations or to use different models depending on the month of the quarter in which the forecast is prepared, that is, depending on the information available to the analyst.

In the recent period, the ECB has been using two concurrent approaches to prepare short-term forecasts of euro area growth in the previous, current and following quarters. Both approaches are used twice a month: mid-month after the release of real indicators such as the IPI; and then at the end of the month after the release of business and consumer tendency surveys and financial data.

The first approach rests on the combination of forecasts drawn from about ten standard calibrations (Rünstler and Sédillot, 2003; Diron, 2008). The second approach is based on dynamic factor models introduced at the ECB (and at the Federal Reserve) in keeping with the method presented by Giannone, Reichlin and Small (2008). Whereas the first approach relies on relatively few monthly indicators (up to 15 in Diron, 2008), the information set in the second approach comprises 85 monthly indicators, real indicators, financial indicators, and indicators derived from business and consumer tendency surveys. A Kalman filter is used to calculate missing factor observations due to the missing months of the monthly indicators. The factor model is estimated using the two-stage estimation method (PCA and Kalman filter) proposed by Doz, Giannone and Reichlin (2011). In this context, Bańbura and Rünstler (2011) measure the variables' contribution to forecasts and apply the results to the short-term GDP forecast for the euro zone.

Our study describes an application of dynamic factor models to the forecasting of Greek GDP growth in the following quarters. We use a database of about one hundred variables such as survey variables, real indicators, monetary and financial variables, and international indicators. An out-of-sample assessment shows that the quality of the forecasts supplied by our factor models is satisfactory, although longer-term forecasts are fragile.

Our paper is organized as follows. Section 2 describes the factor models in their static and dynamic forms, as well as the associated estimation and forecasting methods. Section 3 presents the data used in our study and examines the forecasting performance of factor models tested on their sample base and on an out-of-sample basis.

2. Factor models and their use in forecasting

This section gives a concise description of factor models in their static form and their dynamic extension. We go on to discuss alternative methods for estimating the models. We conclude by reviewing the methods that can be used to construct a forecast based on the prior estimation of a factor model.

2.1. Factor models

2.1.1. Static factor models

Factor models are designed to supply a parsimonious representation of the information provided by a large set of variables when these are correlated. Factor models assume that the observed variables can be described in terms of a small set of latent, unobservable variables called "factors" or "common factors" and that these latent common factors are the source of the correlations between the observed variables. In the static framework, there are two types of factor models: 1) exact factor models, in which the factors explain the entire correlation between variables; and, 2) approximate factor models, which are suited to cases where the number of observed variables tends toward infinity and where the factors explain most of the correlations between variables (the residual portion being negligible).

More formally, with N the number of variables studied, T the number of observations available for each variable and x_{it} the observation of variable i at instant t, the exact model with r factors $\left(f_{jt}\right)_{j=1,\dots,r}$ is written as follows:

$$x_{it} = \mu_i + \lambda_{i1} f_{1t} + \lambda_{i2} f_{2t} + \dots + \lambda_{ir} f_{rt} + e_{it},$$

for $i = 1,\dots,N$, $t = 1,\dots,T$ and $r < N$. That is, in matrix form:

$$x_t = \mu + \Lambda f_t + e_t, \quad t = 1,\dots,T$$

with $x_t = \left(x_{1t},...,x_{Nt}\right)'$ and $e_t = \left(e_{1t},...,e_{Nt}\right)'$ N-dimensional vectors, $f_t = \left(f_{1t},...,f_{rt}\right)'$ a r-dimensional vector and Λ a $\left(N,r\right)$-dimension matrix. The following assumptions hold: $E\left(e_t\right)=0$, $E\left(f_t\right)=0$, $E\left(e_t e_t'\right)=D=diag\left(d_1,...,d_N\right)$, $E\left(f_t f_t'\right)=0$ $\forall$ $\left(t,\tau\right)$, $t \neq \tau$, $E\left(e_t e_t'\right)=0$ $\forall$ $\left(t,\tau\right)$, $t \neq \tau$, I_r representing the r-dimensional identity matrix and $\left(d_1,...,d_N\right)'$ a vector of N positive parameters to be estimated.

In what follows, we shall focus on the case where $\mu = 0$ and work with variables mean-centred beforehand. When r is very small compared with N, the model does indeed yield a parsimonious representation of the covariances between x_{it} variables.

In this static model, the r common factors are not auto-correlated. We can further assume without loss of generality, that they are not correlated with one another and have unit variance. The term, e_{it} called the specific or idiosyncratic component, represents the share of variable x_{it} that is not explained by the common factors. As the e_{it} disturbance terms are uncorrelated two by two, the entire correlation between observed variables is provided by the factors.

The factor weights $\left(\lambda_{ij}\right)$ measure the covariances between the observed variables i and the common factors j. The variance of each variable can thus be written as:

$$V\left(x_{it}\right)=\sum_{j=1}^{r}\lambda_{ij}^2 + d_i$$

The term λ_{ij}^2 represents the share of the variance of x_{it} explained by factor j. The term $\sum_{j=1}^{r}\lambda_{ij}^2$ is the total share of the variance (communality) captured by the r factors. In addition, the variance-covariance matrix of the vector of observed variables is written as $V\left(x_t\right)=\Lambda\Lambda' + D$ and as D is diagonal, the covariances between the observed variables are explicitly expressed in terms of factor loadings. Thus, the variance-covariance matrix of x_t is expressed in terms of the $N(r+1)$ parameters of Λ and D instead of depending on $N(N+1)/2$ parameters if we do not assume the existence of a factor model. Note that the model is invariant to change of scale, so that decomposing the variance-covariance matrix of x_t is equivalent to decomposing its correlation matrix.

In the approximate static model, one no longer assumes that the idiosyncratic terms are uncorrelated two by two. It is merely assumed that in the correlation between the observed variables, the share due to the correlation between the idiosyncratic terms is negligible compared with the share due to the common factors. If one continues to write $E\left(e_t e_t'\right)=D$ (with a non-diagonal matrix *D*), one assumes that when the number *N* of observed variables tends toward infinity, the matrix *D* remains bounded whereas the matrix $\Lambda\Lambda'$ is unbounded. Consequently, as $V\left(x_t\right)=\Lambda\Lambda'+D$, the share of the correlation between variables not explained by the factors can be regarded as negligible.

2.1.2. Dynamic factor models

Dynamic factor models aim to provide a parsimonious description of the common dynamics of the observed variables (or of the co-movements of the observed variables). These models generalize static models (exact or approximate) in two ways. First, the common factors are auto-correlated. Their dynamics are typically modelled in VAR form or in some cases, in vector autoregressive moving average (VARMA) form. Second, the observed variables can be influenced by the factor's contemporary values, but also by their lagged values. In both cases the model can be reduced, via suitable notation changes, to a form close to that of static factor models.

Examining the framework of exact dynamic factor models we may assume that the factor dynamics are correctly represented by a VAR(*p*) model and still using $x_t=\left(x_{1t},\ldots,x_{Nt}\right)'$ to denote the vector of observed variables, one can define an initial class of models in which factors are included only via their contemporary values. These models have the following form:

$$x_t=\Lambda_0 f_t+e_t$$

$$f_t=\sum_{l=1}^{p} A_l f_{t-l}+\varepsilon_t$$

where ε_t is white noise and e_t is a process whose components are uncorrelated two by two and are uncorrelated with the factors.

The factor may operates not only on a contemporary basis but also with its lags, that is, in the context of a model of the form:

$$x_t = \Lambda_0 f_t + ... + \Lambda_s f_{t-s} + e_t$$

$$f_t = \sum_{l=1}^{p} A_l f_{t-l} + \varepsilon_t$$

As with static models, the scope of application of dynamic models may be extended by introducing approximate dynamic factor models when the number N of observable variables tends toward infinity. In this type of model, we allow the components of vector e_t to be correlated with one another, but we assume that the share of the observable variables' dynamics due to the idiosyncratic components is negligible by comparison with the factor-related share.

2.2. Estimation of a dynamic factor models

The framework of approximate dynamic factor models is the standard choice for analyzing macroeconomic data. Various methods for estimating these models have been proposed in the literature. For a full survey of the methods, see Bai and Ng (2008b), and Stock and Watson (2010).

The method most commonly used is principal component analysis (PCA), first proposed by Stock and Watson (2002a). This method is applied to a static factor model (or a dynamic factor model converted to static form). Under the assumptions usually made in the specification of the approximate factor model, PCA is shown to yield convergent estimators of the model's parameters and an approximation of the factors that converges toward their true value when the number N of series studied and the number T of observations tend toward infinity.

However, other estimation methods have been proposed to allow factor dynamics to be taken into account. Forni et al. (2000, 2005) propose an estimation method based on the analysis of the spectral density of observations. Doz, Giannone and Reichlin (2011, 2012) have proposed a pseudo-maximum likelihood estimation method and a two-stage estimation method based on the Kalman filter.

The two-stage estimation method is fairly simple to implement. It has the added advantage of easily adjusting to missing values; one of the main problems faced by short term analysts, as noted earlier. The two-stage method was used, for example by Giannone, Reichlin and Small (2008) to forecast US and euro area GDP, and by Angelini et al. (2008), and Bańbura and Rünstler (2011) to prepare a short-term forecast of euro area GDP.

It is important to stress here that PCA implementation requires a balanced data sample. This imposes a severe constraint on short-term forecasting. If we truncate the sample at the last date for which all the data are available, we deprive ourselves of a part of the existing information.

By contrast, with the two-stage method proposed by Doz, Giannone and Reichlin (2011), we can calculate the best approximations of factor values at each date, taking into account all the information available. Assuming normal disturbances, we know that the Kalman filter and smoother allow us to obtain for a given parameter value, the optimal approximation of the latent variables on the basis of the full information available on the observable variables. The two-stage method seems to be particularly well suited to short-term forecasting.

2.3. Use in forecasting

The estimated factors may be used for forecasting important macroeconomic variables. Assuming that y_t stands for quarterly GDP growth, we may estimate the following model by means of Ordinary Least Squares (OLS):

$$y_t = a + \sum_{i=1}^{n} \beta_i y_{t-i} + \sum_{j=0}^{m} \gamma_j f_{t-j} + e_t$$

The dynamic framework of the model relies on the estimates of factor dynamics obtained when estimating the factor model. If the factors confirm a model of the form

$$f_t = \sum_{l=1}^{p} A_l f_{t-l} + \varepsilon_t$$

we can obtain recursively a forecast $f_{T+h/T}$ at date T using the estimated values of the A_l matrices and the factors. This type of approach is applied by Giannone, Reichlin and Small (2008), Angelini, Bańbura and Rünstler (2008), and Bańbura and Rünstler (2011).

2.4. Choice of model specification

Bai and Ng (2002, 2007) offer criteria for choosing the number of factors. In their 2002 paper, they introduce an initial series of criteria suited to static factor models while in their 2007 paper, they propose a second series of criteria to determine the number of dynamic factors.

In practice, these criteria are used in three stages. First, use one of the six criteria (Bai and Ng, 2002) to determine the optimal number of factors in a static setting. Second, estimate a VAR on these factors and choose the VAR order (p^*) so as to minimize the standard AIC or BIC criterion. Third, apply the Bai and Ng (2007) criteria to the variance-covariance matrix or the correlation matrix of the VAR (p^*) residuals to obtain the optimal number of dynamic factors q^*.

Several studies show that in practice, the use of the Bai and Ng (2002, 2007) criteria can entail the choice of too few factors, undermining forecast quality; see for example, Barhoumi, Darné and Ferrara (2010) for an application to the French GDP forecast, and Schumacher (2007) for an illustration concerning German GDP. A possible explanation is that the choice of factor model specification is totally unrelated to the variable to be forecast.

Schumacher (2007) proposes an alternative to the information criteria so as to compare the results obtained. The alternative consists in choosing the number of factors that minimizes the Root Mean Square Error (RMSE) criterion in the GDP growth regression on the factors. The RMSE criterion also determines the choice of order p of the VAR process on the factors.

3. Use of dynamic factor models to forecast Greek GDP growth

3.1. Data

We use a dataset of 95 variables. Like most studies of GDP forecasts derived from factor models, we have chosen three groups of variables:

● *survey balances*: The main balances of Greek business tendency surveys used to construct the synthetic (or business climate) indicators in manufacturing, services, the building sector and the retail trade, plus the main balances of the consumer tendency survey;

● *real variables*: Real GDP and its main components, household consumption of manufactured goods and its components, new car registrations, building starts and building permits, the industrial production index and its components, labour market variables, tourist arrivals, real effective exchange rate of euro, oil prices;

● *nominal variables* (monetary and financial): Interest rates, yield-curve slope, stock market indexes, monetary aggregates and price indexes;

Many short-term analysts base their forecasts on survey variables and as they become available, on real variables: in particular the industrial production index, household consumption of manufactured goods, building starts, building permits and customs data for foreign trade. The survey balances group includes 31 variables, the real variables group 32 variables and finally, the nominal variables group 32.

Our data set covers the period from the first quarter of 2000 until the second quarter of 2017. The estimates reported in this study were prepared with the series published in early June 2017. All series were downloaded from Eurostat and OECD databases. Some variables that published monthly have been converted to quarterly frequency by taking the mean of each quarter. Some series were seasonally un-adjusted so, using the TRAMO/SEATS filter we proceed to seasonal adjustment of all the series. In order to avoid stationarity issues we log-differentiate the real and nominal variables and take first differences for the survey variables as well as for the interest rates. Finally, we standardize all the variables.

3.2. Estimation of the dynamic factor model

The classification of the variables in three groups allows the estimation of impact of each sector on the whole economy. For this reason we estimate the following dynamic factor model for the Greek GDP:

$$y_t = a + \sum_{i=1}^{n} \beta_i y_{t-i} + \sum_{j=0}^{m} \gamma_j^R f_{t-j}^R + \sum_{j=0}^{m} \gamma_j^N f_{t-j}^N + \sum_{j=0}^{m} \gamma_j^S f_{t-j}^S + e_t \qquad (1)$$

where f^R, f^N and f^S are the factors from the real, the nominal and the survey group of variables correspondingly.

We estimate the factors from each group of variables using the PCA method discussed above. Table 1 shows the cumulative proportion of the variance of each group that explained by a specific number of factors (*k*).

Table 1: Cumulative proportion of variance

Number of factors (k)	Real Sector	Nominal and financial sector	Survey sector
1	31.97%	25.28%	31.52%
2	42.70%	42.75%	45.63%
3	49.70%	51.21%	55.13%
4	56.54%	58.01%	61.45%
5	62.32%	64.08%	66.81%
6	67.05%	69.37%	71.42%
7	71.56%	73.48%	75.13%
8	75.13%	77.30%	78.69%
9	78.60%	80.48%	81.50%
10	81.26%	83.38%	84.01%

Then, we estimate model (1) using various combinations of the estimated factors as well as various lags and compute the Schwarz information criterion. We choose this model with the minimum value of the Schwarz information criterion. According to this value, we use one lag of the real GDP growth, one factor from each group and two lags of the corresponding factors. Namely, the parameters of model (1) are $n=1$, $m=2$, $k^R=1$, $k^N=1$ and $k^S=1$. Using these parameters the estimated model follows (t-statistics are included in brackets):

$$\underset{}{y_t} = -\underset{(0,64)}{0,00054} - \underset{(2,02)}{0,254}\,y_{t-1} + \underset{(7,57)}{0,0097}\,f_t^R + \underset{(0,27)}{0,00041}\,f_t^N + \underset{(0,22)}{0,00024}\,f_t^S + \underset{(2,27)}{0,004}\,f_{t-1}^R + \underset{(1,03)}{0,0018}\,f_{t-1}^N$$

$$+\underset{(1,10)}{0,0011}\,f_{t-1}^S + \underset{(2,38)}{0,0031}\,f_{t-2}^R - \underset{(0,15)}{0,00024}\,f_{t-2}^N - \underset{(0,82)}{0,00078}\,f_{t-2}^S$$

After that, we estimate the following VAR (2) model for the estimated factors:

$$f_t = A_1 f_{t-1} + A_2 f_{t-2} + \varepsilon_t$$

where $f_t = \left(f_t^R, f_t^N, f_t^S\right)'$. So, we can obtain recursively forecast $f_{T+1/T}$, for the third quarter of 2017, at date *T*. Then, we use the estimated model (1) in order to obtain real GDP growth forecast for the third quarter of 2017. Using this forecasted value we estimate the forecasted seasonally adjusted real GDP value for the corresponding quarter. Finally, we may obtain forecast of the seasonal factor of real GDP for the third quarter of 2017 and add this to the forecasted seasonally adjusted real GDP values. So, we have the forecasted value of the seasonally un-adjusted real GDP series which is 0,44%.

4. Conclusion

This study has examined the performance of a tool based on dynamic factor models for forecasting Greek GDP growth over short horizons. Such models allow the inclusion of information provided by a large variable set, summarized into a small set of factors. In their dynamic form, the models allow a time dependence of factors and a dependence of observed variables on contemporary and lagged factor values. If some indicator values are missing, we can adjust the associated estimation methods, avoiding the need for auxiliary models.

Several approaches could be explored for improving these results. The choice of different sets of initial variables seems to yield different forecast qualities. Ahead of factor construction, it might therefore be worth applying the variable selection methods recommended by Boivin and Ng (2006), and more recently Bai and Ng (2008a). The use of these methods by Caggiano et al. (2009) and Schumacher (2010) does show a gain for the GDP forecast, and Charpin's application on French data (2009) of the method proposed by Bai and Ng (2008a) yields encouraging results. Moreover, the introduction of non-linearities in the specification has thus far been relatively little explored in the context of factor models and could also be a major source of improved performance.

References

Altissimo, F. et al. (2001), "A Real Time Coincident Indicator for the Euro Area Business Cycle", *CEPR Discussion Paper Series* No. 3108, 46 pages.

Altissimo, F. et al. (2010), "New Eurocoin: Tracking Economic Growth in Real Time", *The Review of Economics and Statistics*, Vol. 92, No. 4, pp. 1024-1034.

Angelini, E., M. Bańbura and G. Rünstler (2008), "Estimating and Forecasting the Euro Area Monthly National Accounts from a Dynamic Factor Model", *ECB Working Paper* No. 953, October, 29 pages.

Bai, J. and S. Ng (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, Vol. 70, No. 1, pp. 191-221.

Bai, J. and S. Ng (2006), "Confidence Intervals for Diffusion Indexes Forecasts and Inference for Factor-Augmented Regression", *Econometrica*, Vol. 74, No. 4, pp. 1133-1150.

Bai, J. and S. Ng (2007), "Determining the Number of Primitive Shocks in Factor Models", *Journal on Business and Economic Statistics*, Vol. 25, No. 1, pp. 52-60.

Bai, J. and S. Ng (2008a), "Forecasting Economic Time Series Using Targeted Predictors", *Journal of Econometrics*, Vol. 146, No. 2, pp. 304-317.

Bai, J. and S. Ng (2008b), "Large Dimensional Factor Analysis", *Foundations and Trends in Econometrics"*, Vol. 3, No. 2, pp. 89-163.

Bańbura, M. and G. Rünstler (2011), “A Look into the Factor Model Black Box – Publication Lags and the Role of Hard and Soft Data in Forecasting GDP”, *International Journal of Forecasting*, Vol. 27, No. 2, pp. 333-346.

Barhoumi, K., O. Darné and L. Ferrara (2010), “Are Disaggregate Data Useful for Factor Analysis in Forecasting French GDP?”, *Journal of Forecasting*, Vol. 29, No. 1-2, pp. 132-144.

Boivin, J. and S. Ng (2006), “Are More Data Always Better for Factor Analysis”, *Journal of Econometrics*, Vol. 132, No. 1, May, pp. 169-194.

Caggiano, G., G. Kapetanios and V. Labhard (2009), “Are More Data Always Better for Factor Analysis? Results for the Euro Area, the Six Largest Euro Area Countries and the UK”, *Working Paper Series* No. 1051, European Central Bank.

Charpin, F. (2009), “Estimation précoce de la croissance, de la régression LARS au modèle à facteurs”, *Revue de l’OFCE*, Vol. 2009/1, No. 108, pp. 31-48.

Diron, M. (2008), “Short-Term Forecasts of Euro Area Real GDP Growth: An Assessment of Real-Time Performance Based on Vintage Data”, *Journal of Forecasting*, Vol. 25, No. 5, pp. 371-390.

Doz, C., D. Giannone and L. Reichlin (2011), “A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering”, *Journal of Econometrics*, Vol. 164, No. 1, pp 188-205.

Doz, C., D. Giannone and L. Reichlin (2012), “A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models”, *Review of Economics and Statistics*, Vol. 94, No. 4, pp. 1014-1024.

Forni, M. et al. (2000), “The Generalized Dynamic Factor Model: Identification and Estimation”, *Review of Economics and Statistics*, Vol. 82, No. 4, pp. 540-554.

Forni, M. et al. (2005), “The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting”, *Journal of the American Statistical Association*, Vol. 100, No. 471, pp. 830-840.

Giannone, D., L. Reichlin and D. Small (2008), “Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases”, *Journal of Monetary Economics*, Vol. 55, No. 4, pp. 665-676.

Rünstler, G. and F. Sédillot (2003), “Short-Term Estimates of Euro Area Real GDP by Means of Monthly Data”, *Working Paper Series* No. 276, European Central Bank, 28 pages.

Schumacher, C. (2007), “Forecasting German GDP Using Alternative Factor Models Based on Large Datasets”, *Journal of Forecasting*, Vol. 26, No. 4, pp. 271-302.

Schumacher, C. (2010), “Factor Forecasting Using International Targeted Predictors: The Case of German GDP”, *Economic Letters*, Vol. 107, No. 2, pp. 95-98.

Stock, J. and M. Watson (1999), “Forecasting Inflation”, *Journal of Monetary Economics*, Vol. 44, No. 2, pp. 293-335.

Stock, J. and M. Watson (2002a), “Forecasting Using Principal Components from a Large Number of Predictors”, *Journal of the American Statistical Association*, Vol. 97, No. 460, pp. 1167-1179.

Stock, J. and M. Watson (2002b), “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business and Economic Statistics*, Vol. 20, No. 2, pp. 147-162.

Stock, J. and M. Watson (2010), “Dynamic Factor Models”, *in Oxford Handbook of Economic Forecasting*, M.P. Clements and D.F. Hendry (ed.), Oxford University Press, Chapter 2.