DYNAMIC ECONOMETRIC MODELS
Vol. 9 – Nicolaus Copernicus University – Toruń – 2009

*Jarosław Krajewski*
*Nicolaus Copernicus University in Toruń*

# Estimating and Forecasting GDP in Poland with Dynamic Factor Model[†]

A b s t r a c t. Presented paper concerns the dynamic factor models theory and application in the econometric analysis of GDP in Poland. DFMs are used for construction of the economic indicators and in forecasting, in analyses of the monetary policy and international business cycles. In the article we compare the forecast accuracy of DFMs with the forecast accuracy of 2 competitive models: AR model and symptomatic model. We have used 41 quarterly time series from the Polish economy. The results are encouraging. The DFM outperforms other models. The best fitted to empirical data was model with 3 factors.

K e y w o r d s: Dynamic factor models, principal components analysis, GDP.

## 1. Introduction

In recent years, dynamic factor models have become popular in empirical macroeconomics. They are believed to have been pioneered by Geweke (1977) and Sims & Sargent (1977) who applied this type of models to the analysis of small sets of variables. DFMs have a very wide field of applications. Such models are widely used for forecasting, constructing leading indicators of business climate, monetary policy analysis or the analysis of international business cycles.

The purpose of the article is to estimate a dynamic factor model of GDP in Poland in 1997 – 2008.

The second part of the article presents a concept of a dynamic factor model. In the third part the approach to estimating model parameters as well as common factors are discussed. The methods of specifying the number of factors in

[†] Scientific work is financing by European Social Fund and national budget of Poland in The Integrated Regional Operational Programme framework, Measure 2.6 "Regional Innovation Strategies and transfer of knowledge" Kujawsko-Pomorskie voievodship own project "Candidate for doctor's degree grants 2008/2009 – IROP".

the model are also presented. The data used in the study and the empirical results have been described in the fourth part. Final part summarises the whole study.

## 2. Dynamic Factor Model

The concept of factor models bases on the assumption that the behavior of most macroeconomic variables may be well described using a small number of unobserved common factors. These factors are often interpreted as the driving forces in the economy. The particular variables may be then expressed as linear combination of up-to-twenty common factors which usually make it possible to explain a major part of variability of those variables (Kotłowski, 2008).

Let $y_t$ stand for a variable and let $\mathbf{X_t}$ express the vector of N variables containing information that can be useful in modeling and forecasting the future values of $y_t$. In the dynamic factor model we assume that all variables $x_{it}$ contained in vector $\mathbf{X_t}$ may be expressed as a linear combination of current and lagged unobserved factors $f_{it}$.

$$x_{it} = \lambda_i(L)\mathbf{f_t} + e_{it}, \text{for } i = 1,...,N, \tag{1}$$

where $\mathbf{f_t} = [f_{1t}, f_{2t},..., f_{\bar{r}t}]'$ stands for vector $\bar{r}$ of unobserved common factors at moment t, $\lambda_i(L) = \lambda_{i0} + \lambda_{i1} + \lambda_{i2}L^2 + ... + \lambda_{iq}L^q$ represent a lag polynomials and $e_{it}$ express an idiosyncratic errors for variable $x_{it}$ (see Stock, Watson, 1998).

In turn, $y_t$ may be noted as the function of current and lagged common factors contained in vector $\mathbf{f_t}$ and the past values of variable $y_t$, with the following formula

$$y_t = \beta(L)\mathbf{f}_t + \gamma(L)y_t + e_t. \tag{2}$$

The model described with equations (1) and (2) is a dynamic factor model.

## 3. Model Estimation and Selection of the Factor's Number

One of the most widely used methods of parameters and factors estimation in a factor models is the method of principal components. Let us emphasise that both: the factor matrix and the coefficient matrix are unknown. Model (1) is thus equivalent to the model in the matrix form of

$$\mathbf{X} = \mathbf{FHH}^{-1}\mathbf{\Lambda}' + \mathbf{e}, \tag{3}$$

where matrix $\mathbf{H}$ is any non-singular matrix of dimension $r \times r$. It is necessary to carry out the appropriate normalisation of matrix $\mathbf{H}$. Stock & Watson (1998) suggest that for this purpose condition in the form of $(\mathbf{\Lambda}^{'}\mathbf{\Lambda}/N) = \mathbf{I_r}$ may be imposed on the parameters of model which would render matrix **H** orthonormal.

The estimation of matricies **F** and **Λ** using the method of principal component consist in finding such estimates of matrices $\hat{\mathbf{F}}$ and $\hat{\mathbf{\Lambda}}$ that would minimise the residual sum of squares in equation (3) as expressed with the following formula

$$V(\mathbf{F},\mathbf{\Lambda}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \mathbf{\Lambda_i}^{'}\mathbf{F_t})^2. \tag{4}$$

It is necessary, in the first step, to perform a minimisation of function (4) in respect to factor matrix **F** with the assumption that matrix **Λ** is known and fixed. Then we obtain estimate $\hat{\mathbf{F}}$, as function **Λ**, which is subsequently substituted in equation (4) for the true value of **F**. In the second step, we minimise function (4) in respect to matrix **Λ** with a normalisation condition $(\mathbf{\Lambda}^{'}\mathbf{\Lambda}/N) = \mathbf{I_r}$, thus directly obtaining estimate $\hat{\mathbf{\Lambda}}$. It should be emphasised that it is equivalent to maximisation of expression $tr[\mathbf{\Lambda}^{'}(\mathbf{X}^{'}\mathbf{X})\mathbf{\Lambda}]$.

Matrix $\hat{\mathbf{\Lambda}}$ is a matrix whose subsequent columns are eigenvectors of matrix $\mathbf{X}^{'}\mathbf{X}$ multiplied by $\sqrt{N}$ corresponding to highest eigenvalues of the same matrix. In turn, the estimate of matrix **F** is expressed by the formula

$$\hat{\mathbf{F}} = (\mathbf{X}\hat{\mathbf{\Lambda}})/N. \tag{5}$$

Stock & Watson (1998) emphasise that if the number of variables is higher than the number of observations, i. e. N > T, then from the computational point of view it is easier to apply a procedure which determine estimate $\tilde{\mathbf{F}}$ by minimizing concentrated function (4) in respect to matrix F with the condition $\mathbf{F}^{'}\mathbf{F}/T = \mathbf{I_r}$. Matrix $\tilde{\mathbf{F}}$ will then contain eigenvectors of matrix $\mathbf{X}^{'}\mathbf{X}$ corresponding to r highest eigenvalues of this matrix and multiplied by $\sqrt{T}$. In turn, the estimate of matrix $\tilde{\mathbf{\Lambda}}$ will assume the following form

$$\tilde{\mathbf{\Lambda}}^{'} = (\tilde{\mathbf{F}}^{'}\mathbf{X})/T. \tag{6}$$

In practice, the number of factors necessary to represent the correlation among the variables is usually unknown. To determine the number of factors empirically a number of criteria were suggested. Bai and Ng (2002) have suggested information criteria to be used to estimate the number of factors.

$$IC_1(k) = \ln(\hat{V}(k)) + k\left(\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right)\right), \tag{7}$$

$$IC_2(k) = \ln(\hat{V}(k)) + k\left(\left(\frac{N+T}{NT}\right)\ln C_{NT}^2\right), \tag{8}$$

$$IC_2(k) = \ln(\hat{V}(k)) + k\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right), \tag{9}$$

where $\hat{V}(k)$ is residual sum of squares from k – factors model

and $C_{NT} = \min\{\sqrt{N}\sqrt{T}\}$.

## 4. Description of Data and Empirical Results

The data used in the study are macroeconomic quarterly data describing Polish economy and encompass the period from first quarter of 1997 to third quarter of 2008 (47 observations). As explained variable we used polish GDP. All of data were taken from polish Central Statistical Office[1]. Before embarking on the work on factor model specification, the date had to be appropriately modified. In the first step variables were adjusted for the impact of seasonal fluctuations. Next, the series were transformed by taking logarithms and/or differencing so that the transformed series were stationary (Green, 2003). In the final step, all variables were standardized. In total, 41 time series were considering, representing of following macroeconomic categories: output & sales, construction, domestic and foreign trade, prices and labour market, budgetary and monetary policy.

After preliminaries principal components analysis were used to estimate factors. Next Bai and Ng informational criteria were calculated to specify number of factors. Table 1 includes eigenvalues and values of information criteria. The first two criteria reach minimum for the number of factors equal to 3, while the third criteria assumes its lowest value for 10 factors. Due to the fact that two out of three criteria display the same value, we arbitrarily assume that the number of factors in the model is 3. The first three factors explain almost 82% of total variance of GDP.

Some econometricians maintain that factors estimated using the principal component method do not have an economic interpretation. However, in this paper try-out were done. For this reason, it is possible to carry out a regression of particular variables against each of the estimated factors and check which factor explains the behavior of a given variable to the greatest extent. The R-squared values on the regression of particular variables suggest that the first factor primarily affects the variability of labour market and foreign trade. The second factor determines the prices and incomes. The third factor to the greatest extent influences the values of sales.

[1] www.stat.gov.pl

The further stage of study BIC criterion was used to determine the number of GDP and factors delays. The BIC criterion indicated model only with current values of the first three factors. Model estimation results are presented in Table 2. All coefficients are significant under 5% level. Factor model of GDP in Poland estimated in this way has R-squared over 70%. Real values of GDP and values based on the model are shown on the figure 1.

Table 1. Selection of the number of factors in the model

| Number of factors | Eigenvalues | Contribution to variance | Cumulative contribution to variance | IC1 | IC2 | IC3 |
|---|---|---|---|---|---|---|
| 1 | 76.949 | 0.721 | 0.721 | -3.694 | -3.491 | -3.833 |
| 2 | 6.249 | 0.059 | 0.779 | -3.584 | -3.177 | -3.861 |
| 3 | 4.278 | 0.040 | 0.819 | -4.555 | -3.945 | -4.971 |
| 4 | 2.921 | 0.027 | 0.846 | -4.490 | -3.677 | -5.044 |
| 5 | 2.460 | 0.023 | 0.870 | -4.467 | -3.451 | -5.160 |
| 6 | 1.976 | 0.019 | 0.888 | -4.335 | -3.115 | -5.166 |
| 7 | 1.551 | 0.015 | 0.903 | -4.192 | -2.769 | -5.162 |
| 8 | 1.429 | 0.013 | 0.916 | -4.109 | -2.482 | -5.216 |
| 9 | 1.282 | 0.012 | 0.928 | -4.246 | -2.417 | -5.493 |
| 10 | 1.066 | 0.010 | 0.938 | -4.126 | -2.093 | -5.511 |

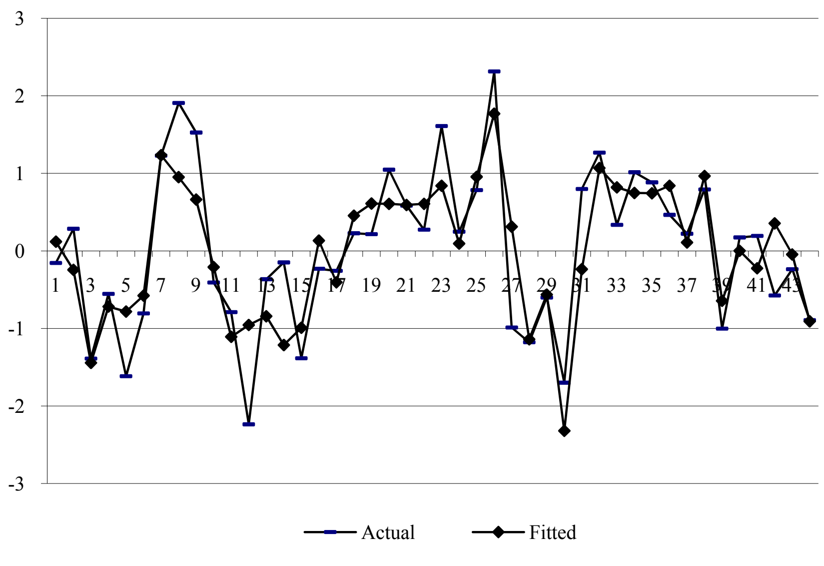


Figure 1. Actual and fitted values of GDP in Poland in 1997 – 2008

Next stage of analysis was to check if lagging some of variables will influence the final result of estimation. This caused increasing the number of factors to four. The resulting model with four factors is presented in Table 3. It is not hard to see that in four factors model R-squared is higher and true significance level of coefficients is lower.

Table 2. Dynamic factor model of GDP in Poland in 1997-2008

| Dependent Variable: GDP | | | | |
|---|---|---|---|---|
| | Coefficient | Std. Error | T - statistic | P - value |
| F1 | -0.0362 | 0.0095 | -3.7985 | 0.0005 |
| F2 | 0.0683 | 0.0335 | 2.0404 | 0.0478 |
| F3 | 0.3707 | 0.0405 | 9.1610 | 0.0000 |
| R-squared | 0.7142 | Akaike Criterion | | 1.7264 |
| Adjusted R-squared | 0.7003 | Schwarz Criterion | | 1.8481 |
| Durbin-Watson | 2.1454 | Hannan-Quinn Criterion | | 1.7715 |

Table 3. Dynamic factor model of GDP in Poland in 1997-2008 – after changes in data set

| Dependent Variable: GDP | | | | |
|---|---|---|---|---|
| | Coefficient | Std. Error. | T - statistic | P - value |
| F21 | 0.1722 | 0.0273 | 6.3160 | 0.0000 |
| F22 | 0.1686 | 0.0329 | 5.1320 | 0.0000 |
| F23 | -0.2548 | 0.0361 | -7.0669 | 0.0000 |
| F24 | 0.2241 | 0.0400 | 5.6090 | 0.0000 |
| R-squared | 0.7909 | Akaike Criterion | | 1.4858 |
| Adjusted R-squared | 0.7748 | Schwarz Criterion | | 1.6497 |
| Durbin-Watson | 2.2511 | Hannan-Quinn Criterion | | 1.5463 |

Table 4. Forecast errors

| | MAPE | RMSE | R-squared |
|---|---|---|---|
| AR | 90.8002 | 0.8116 | 0.1246 |
| DFM | 1.7898 | 0.0160 | 0.7003 |
| DFM2 | 12.8691 | 0.1150 | 0.7748 |
| Causal Model | 4.9267 | 0.0440 | 0.8694 |

In the last stage of our study we generated forecasts and forecast errors. The forecasting performance of the factor models was evaluated by comparing the accuracy of GDP forecasts obtained on the basis of the factor models with the accuracy of GDP forecasts derived from other competitive models. Two competitive models were taken into consideration: an univariate autoregressive model and causal model with two variables. An univariate autoregressive model was adopted as the main benchmark model for evaluating the forecasting performance of the factor models (see Marcellino, Stock, Watson, 2001). The BIC criterion indicated model AR(1). The causal model included industrial production sales and average employment as explanatory variables. Forecasting models were estimated on a shorter sample (up to 4 quarter 2007). The forecast was produced on the one period ahead. First factor model had the best forecast accuracy. Table 4 presents results.

## 5. Summary

The principal component analysis reduced the number of explanatory variables from 41 to 3 factors. The resulting dynamic factor model of GDP in Poland is satisfactory from the statistical point of view.

Changes in data set influenced the final result of model estimation. In this study it brought out increasing number of factors and improvement estimation performance. Unfortunately, it did not improve forecasting performance.

First dynamic factor model of GDP in Poland in 1997 – 2008 gave the best forecasting performance in comparison with three competitive models described above.

## References

Bai, J., Ng, S. (2002), Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191–221.

Geweke, J. (1977), The Dynamic Factor Analysis of Economic Time Series, [in:] Aigner D. J., Goldberger A. S. (ed.), *Latent Variables in Socio–Economic Models*, Amsterdam, North Holland.

Greene, W. H. (2003), *Econometric Analysis*, Pearson Education, New Jersey.

Kotłowski, J. (2008), Forecasting Inflation With Dynamic Factor Model – the Case of Poland, Working Papers, 2-08, SGH, Warszawa.

Marcellino, M., Stock, J. H., Watson, M. W. (2001), Macroeconomic Forecasting in the Euro Area: Country Specific versus Area–Wide Information, Working Paper, 201, Innocenzo Gasparini Institute for Economic Research.

Sargent, T., Sims, C. (1977), *Business Cycle Modelling Without Pretending to Have Too Much A-Priori Economic Theory*, in Sims C. (ed.), *New Methods in Business Cycle Research*, Minneapolis, Federal Reserve Bank of Minneapolis.

Sims, C. A. (1980), Macroeconomics and Reality, *Econometrica*, 48, 1–48.

Stock J., Watson, M. W. (1998), Diffusion Indexes, Working Paper, 6702, National Bureau of Economic Research.

## Zastosowanie dynamicznego modelu czynnikowego do modelowania i prognozowania PKB w Polsce

Z a r y s t r e ś c i. Referat traktuje o podstawach konstrukcji dynamicznych modeli czynnikowych i ich zastosowaniu empirycznym. DFM stosuje się do prognozowania, konstruowania głównych wskaźników koniunktury, analiz polityki monetarnej i badania międzynarodowych cykli koniunkturalnych. W referacie oszacowano dynamiczny model czynnikowy PKB w Polsce w latach 1997–2008, a także oceniono trafność uzyskanych na jego podstawie prognoz w porównaniu do modelu AR i modelu symptomatycznego. Zbiór danych wykorzystanych do badania zawiera 41 zmiennych makroekonomicznych. Najlepszym ze statystycznego punktu widzenia okazał się model z 3 czynnikami.

S ł o w a k l u c z o w e: dynamiczny model czynnikowy, metoda głównych składowych, PKB.