Torun Business Review 15(1) 2016 25-33

APPLICATION OF DYNAMIC FACTOR MODELS FOR INFLATION FORECASTING IN POLAND

JAROSŁAW KRĄJEWSKI[a]
[a] *WSB University in Torun, Poland*

ABSTRACT

The subject of this article is the application of dynamic factor models in modelling and forecasting inflation in Poland. It contains a brief description of the DFM tool. It also provides a glimpse at empirical forms of tools used to determine the forecasts and compares the forecasts using meters normally used for this purpose. The empirical analysis was carried out on the basis of a set of monthly data. The set consists of 70 variables from the period between January 2002 and March 2015.

ARTICLE INFO

Available online 23 March 2016

Keywords:
Dynamic Factor Model,
Method of Principal Components,
inflation,
forecasting

JEL: C38, C53.

Doi: 10.19197/tbr.v15i1.36

INTRODUCTION

Recently, dynamic factor models have been very popular in macroeconomic analyses (Jungbecker and Koopman 2015). Their popularity is undoubtedly influenced by central banks of many countries that stimulate their growth seeing them as an opportunity to discover a tool enabling faster and more accurate forecasts than those obtained with the use of the tools which are normally used for this purpose. Besides forecasting, DFM is used for constructing main indicators of the economic situation, monetary policy analyses, and research of international business cycles.

Geweke (1977) as well as Sims and Sargent (1977), who applied DFM to small data sets, are considered pioneers in this field. Dynamic factor models present the atheoretical approach to econometric modelling (Sims 1980).

The main subject of the article is the evaluation of effectiveness of dynamic factor models in modelling and forecasting inflation in Poland.

The following article introduces the application of the method of principal components for dynamic analysis of macroeconomic time series, which is uncommon in the Polish literature. The methods presented in this article represent a kind of a bridge between statistics and econometrics.

DYNAMIC FACTOR MODEL

The concept of factor models is based on the assumption that the behaviour of most macroeconomic variables can be well defined using a small number of unobservable common factors. These factors are often interpreted as the leading forces in the economy. Particular variables can then be expressed as a linear combination of fewer than 20 factors which explain a significant part of their variability (Kotłowski 2008).

Let y_t represent a certain time series and X_t express the vector N of variables in the form of time series containing information useful in modelling, and forecasting the y_t value. In a dynamic factor model we assume that all x_{it} variables contained in the X_t vector can be expressed as a linear combination of current and delayed unobservable f_{it} factors

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

where $f_t = [f_{1t}, f_{2t}, \ldots, f_{\bar{r}t}]'$ is a $\bar{r}$ vector of unobservable common factors at moment t, $\lambda_i(L) = \lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2 + \ldots + \lambda_{iq}L^q$ is a polynomial of delay operator, whereas e_{it} expresses a kind of x_{it} variable error, responsible for the remaining disturbance of x_{it} variables uncorrelated with factors.

Therefore, y_t can be written as a function of current and delayed common factors contained in f_t vector and delayed values of y_t in the following manner:

$$y_t = \beta(L) f_t + \gamma(L) y_t + e_t. \qquad (2)$$

Therefore, it is possible to say that the dynamic factor model consists of equations (1) and (2).

ESTIMATION OF DFM PARAMETERS AND SPECIFICATION OF NUMBER OF FACTORS

One of the most frequently used methods of estimation of parameters and factors in factor models is the method of principal components. In this method, both matrices of factors and parameters are unknown. A model presented as equation (1) can be written in the following matrix form:

$$X = FHH^{-1}\Lambda' + e, \qquad (3)$$

where H is a non-unit matrix with the dimensions of $r \times r$. It is necessary to perform an appropriate normalisation of the H matrix. Stock and Watson (1998) proposed a condition $(\Lambda'\Lambda / N) = I_r$, which may be imposed on the parameters of the model and will make matrix H orthonormal.

Estimation of F and Λ matrices using the method of principal components consists of finding the estimators of matrices $\hat{F}$ and $\hat{\Lambda}$, which will minimise the sum of squared residuals of equation (3) expressed as follows:

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

In the first step, it is necessary to minimise function (4) with respect to the matrix of factors F under the assumption that the Λ matrix is known and constant. The result will be the $\hat{F}$ estimator as a function Λ, which will subsequently replace the real F values in the above equation. In the second step, function (4) is minimised in relation to the Λ matrix with the normalisation condition $(\Lambda'\Lambda / N) = I_r$, in this way we directly obtain the estimator $\hat{\Lambda}$. It is worth noticing that this is equivalent to maximising the expression $tr[\Lambda'(X'X)\Lambda]$.

Subsequent columns of the $\hat{\Lambda}$ matrix are eigenvectors of X'X matrix multiplied by $\sqrt{N}$, corresponding to the largest eigenvalues of this matrix. In turn, the estimator of the F matrix is expressed as

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

Stock and Watson (1998) stress 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 use a procedure involving estimating $\tilde{F}$ by minimising (4) with regard to the F of $F'F / T = I_r$ condition. The $\tilde{F}$ matrix contains then the eigenvectors of the $X'X$ matrix relating to r largest eigenvalues of this matrix multiplied by $\sqrt{T}$. In turn, the matrix estimator $\tilde{\Lambda}$ takes the following form:

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

Both estimators $\hat{F}$ and $\tilde{F}$ are equivalent.

In practice, the number of factors necessary to demonstrate relationships between variables is usually unknown. However, there are criteria that can be used to determine a number of factors. For this purpose, Bai and Ng (2002) proposed the following information criteria:

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

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

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

In the above formulas, $\hat{V}(k)$ means the sum of squared residuals from the k - factorial model, and $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

DATA AND ANALYSIS OF EMPIRICAL RESULTS

In this study, a monthly data set containing 70 variables in the form of time series with a monthly frequency was used for the construction of factors. The detailed list of variables can be found in the work of Krajewski (2011). The data concerned a period from January 2002 to March 2015 inclusive, so each set consisted of 159 observations. All data were taken from the web page of the National Bank of Poland (NBP)[1] and the Statistical Bulletins of the Central Statistical Office (GUS)[2].

Inflation represented by an increase in the index of prices of consumer goods and services was used in the study as the explained variable.

All data were subjected to appropriate transformations. They were brought to the fixed prices of first periods of each set and purified from seasonal fluctuations using the X-12 ARIMA procedure. In the next step, the data was logarithmised and differentiated according to the owned time series, in order to bring them to stationarity (Greene 2003). The commonly known ADF test (Dickey & Fuller 1979) was used to determine the degree of integration of individual variables. Data concerning sales volume of industrial production in general and its constituent parts, construction in various aspects, domestic and foreign trade, inflation and the labour market in different perspectives, the budgetary sphere, as well as the characteristics of the wider monetary policy, was used for determination of the factors. In addition, data concerning raw materials and domestic and foreign assets was used.

After the initial data preparation, the method of principal components was applied in order to determine the factors. Subsequently, the Bai-Ng information criteria were calculated to specify their number. Table 1 presents values of second information criteria of IC2 for different numbers of factors in each model.

Finally, in each model, one factor was taken into account, since the empirical criterion values indicated so, which largely represents the raw material market.

[1] www.nbp.pl

[2] Statistical Bulletins of the Central Statistical Office form the period between January 2002 and April 2015, ZWS, Warsaw.

Table 1. IC2 values for selecting a number of factors for the models estimated on the basis of monthly data

Number of factors	CPIstacM	CPIstacdyM
1	-6.8651	-7.0502
2	-6.7759	-6.9648
3	-6.6989	-6.8849
4	-6.6512	-6.8461
5	-6.7391	-6.9050
6	-6.6483	-6.8760
7	-6.6177	-6.7899
8	-6.6204	-6.7756
9	-6.5453	-6.8039
10	-6.4532	-6.7171
11	-6.4853	-6.6312
12	-6.4230	-6.6663

Source: own work.

Table 2. Dynamic factor model: CPIstacM

Dependent variable: CPI				
Variable	Factor evaluation	Standard error	t	p
F1	0.0008	0.0002	3.3552	0.0010
F1(-1)	0.0009	0.0004	2.4611	0.0150
F1(-2)	0.0010	0.0004	2.4702	0.0146
F1(-3)	0.0011	0.0003	3.0546	0.0027
F1(-4)	0.0006	0.0002	2.4995	0.0135
CPI_NBP_SA(-1)	0.1319	0.0765	1.7229	0.0870
CPI_NBP_SA(-2)	0.3386	0.0769	4.4058	0.0000
R-square	0.2176	Akaike criterion		-6.1682
Corrected R-square	0.1857	Schwarz criterion		-6.0302

Source: own work.

Then, using the BIC criterion, delays for both the dependent variable and the factors were selected. Results of the estimation procedure in the form of the individual dynamic factor models are contained in Tables 2 and 3.

Table 2 presents a model in which only the current source variables were used (CPIstacM model), and Table 3 presents a model based on both the current and the delayed source variables (CPIstacdyM model).

The estimation resulted in models characterised by statistically significant parameters on a level no higher than 10%. All the estimated models are characterised by a lack of autocorrelation which is not a necessary property in the case of the dynamic factor models, but obviously affects the evaluation preferably. The R-square coefficients do not reach too high values – approximately 22%. However, this is not a key criterion for the assessment of this type of econometric tools. 40% is already considered as a very high level of this meter for the DFM.

Table 3. Dynamic factor model: CPIstacdyM

Dependent variable: CPI				
Variable	Factor evaluation	Standard error	t	p
F1	0.0007	0.0002	3.0705	0.0025
F1(-1)	0.0008	0.0004	2.1935	0.0299
F1(-2)	0.0010	0.0004	2.2406	0.0266
F1(-3)	0.0011	0.0004	3.0068	0.0031
F1(-4)	0.0006	0.0002	2.4835	0.0141
CPI_NBP_SA(-1)	0.1378	0.0773	1.7831	0.0766
CPI_NBP_SA(-2)	0.3454	0.0776	4.4500	0.0000
R-square	0.2194	Akaike criterion		-6.1634
Corrected R-square	0.1873	Schwarz criterion		-6.0248

Source: own work.

The actual values of inflation and the values calculated on the basis of the factor models are presented in Figures 1-2. The diagrams confirm a relatively good matching of values obtained on the basis of the model in relation to the actual values. It is particularly evident in the field of change directions of the analysed economic size.

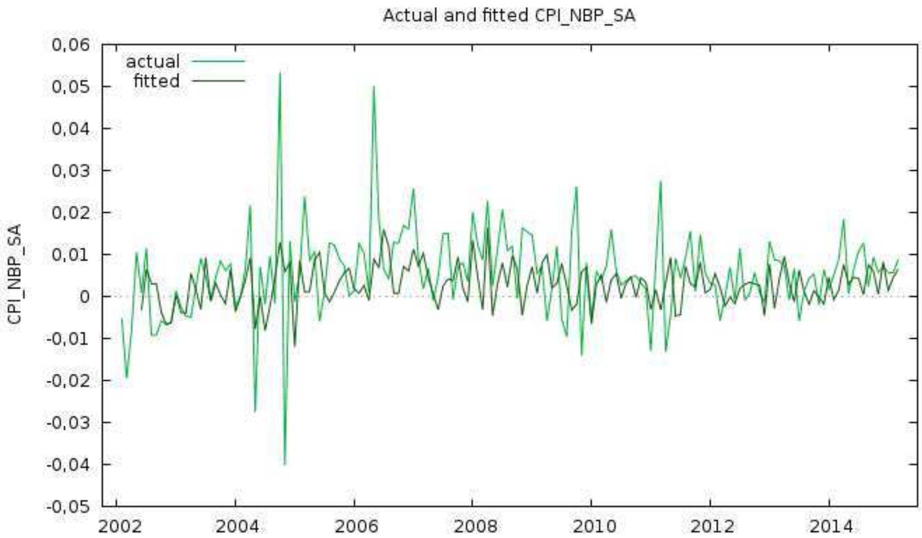

Figure 1. Monthly empirical values of CPI and theoretical values estimated on the basis of the model: CPIstacM
Source: own work.

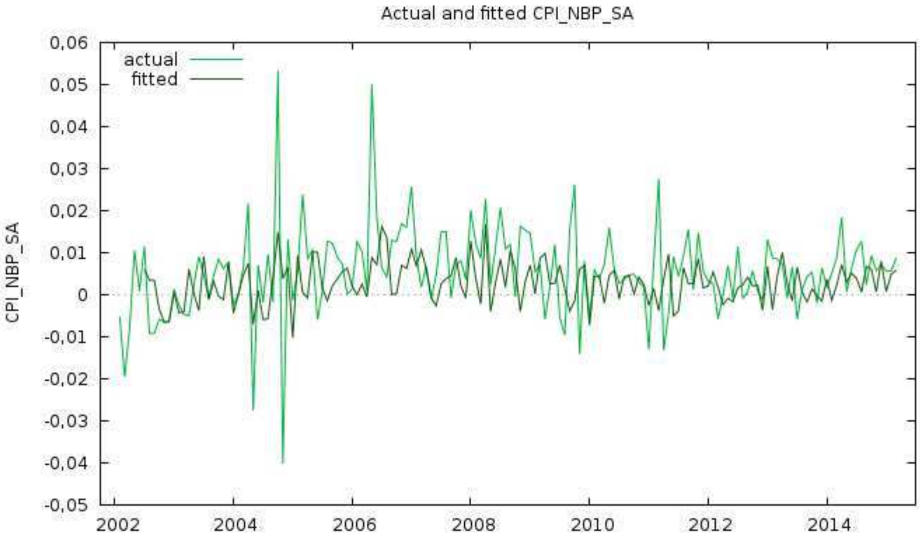

Figure 2. Monthly empirical values of CPI and theoretical values estimated on the basis of the model: CPIstacdyM
Source: own work.

The final stage of the analysis was to determine the forecasts and their errors on the basis of the dynamic factor models. Forecasting based on the DFM is problematic due to the lack of a factor value in a forecast period. One solution is to treat the factors as autoregressive processes and determine their future values based on appropriate AR models.

The resulting forecast errors were then compared with the corresponding ones received from the autoregressive models. In fact, the AR class models represent the most common, but not the only point of reference in the literature. Most frequently, the authors compare forecasts from models with many variables to models with one variable (Marcellino, Stock & Watson 2001). For comparisons, on the basis of the BIC criterion, the stationary AR models (4) were adopted.

For the set of monthly data, the following forecast horizons were adopted: 3, 6, 9, 12 periods. For quality assessment of the obtained forecasts, we use widely known meters of forecasting errors, such as mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) or Theil coefficients.

In the case of forecasting errors, the forecast horizon 6, 9 and 12 turned out to be long enough to allow the performance of the test of Diebold-Mariano (1995), who pointed out that the differences between the lowest forecast errors and errors from the AR models are statistically significant.

Table 4. Forecast errors obtained from models estimated on the basis of the monthly data

H	Model	RMSE	MAE	MAPE	Theil
3	CPIstacM	**0.0071**	**0.0057**	**77.1760**	0.9415
	CPIstacdyM	0.0073	0.0058	82.2690	**0.8958**
	CPI_AR	0.0106	0.0080	89.5270	1.1039
6	CPIstacM	**0.0084**	0.0066	75.5390	0.9560
	CPIstacdyM	0.0085	**0.0066**	**73.0390**	**0.9122**
	CPI_AR	0.0101	0.0083	94.7580	1.1203
9	CPIstacM	**0.0077**	**0.0063**	75.8010	0.9865
	CPIstacdyM	0.0078	0.0064	**75.3250**	**0.9506**
	CPI_AR	0.0093	0.0080	96.5050	1.1520
12	CPIstacM	**0.0072**	**0.0060**	**75.6900**	0.9960
	CPIstacdyM	0.0074	0.0061	77.5330	**0.9643**
	CPI_AR	0.0088	0.0077	97.3790	1.1697

Source: own work.

Presenting the results of forecasting, it should be noted that the dynamic factor models have given lower scores forecast errors in all 16 cases considered. This indicates that they represent an attractive and promising alternative to determining the macroeconomic forecasts of the Polish economy. The CPIstacM proved itself the best model nine times, and on the basis of the CPIstacdyM model, the lowest rate of forecast errors was obtained seven times. It is also worth noticing that the forecast errors from both factor models are similar.

In most cases, the dynamic factor models provide more accurate forecasts than the AR-type models, which suggests that they represent an attractive alternative to be used in the process of forecasting and macroeconomic planning.

CONCLUSION

The analysis led to a reduction in the number of original explanatory variables from 70 factors to 1, which was obtained by applying the method of principal components. As a result, the dynamic factor models describing the Polish economy in terms of inflation in a satisfactory manner from a statistical point of view were obtained.

The dynamic factor models gave lower scores of forecasting errors in all 16 cases under consideration.

REFERENCES

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

Dickey, D. A., & Fuller W. A. (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association, 366*, American Statistical Association, 427 – 431.

Diebold F. X., & Mariano R. S. (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13.

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, North Holland, Amsterdam, 365 – 383.

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

Jungbacker B., Koopman S. J. (2015), Likelihood – based dynamic factor analysis for measurement and forecasting, Econometric Journal, v. 18.

Krajewski, J. (2011), Dynamiczne modele czynnikowe w modelowaniu i prognozowaniu procesów makroekonomicznych na przykładzie inflacji i PKB w Polsce, UMK Toruń.

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, Milano.

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, Federal Reserve Bank of Minneapolis, Minneapolis, 45 – 110.

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

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

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