2. IMPROVING SHORT-TERM FORECASTING OF MACEDONIAN GDP: COMPARING THE FACTOR MODEL WITH THE MACROECONOMIC STRUCTURAL EQUATION MODEL

Dimitar EFTIMOSKI[1]

Abstract

This paper evaluates two different models for short-term forecasting of the Macedonian GDP: (a) the medium-scale static factor model, based on the static principal components analysis, and (b) the small-scale macroeconomic structural equation model. Recursive dynamic pseudo out-of-sample forecasts, based on a panel of quarterly time series, indicate that forecast errors of the factor model are smaller overall in comparison to errors of the structural equation model at all forecast horizons. In line with the existing short-term GDP forecasting practice, our medium-scale factor model (that extracts common factors from a data set of 52 variables) diversifies and strengthens the current macroeconomic forecasting strategy in Macedonia.

Keywords: factor model, macroeconomic structural equation model, forecasting and forecasting evaluation, GDP

JEL Classification: C2, C3, C38, C53

1. Introduction

After the rising public criticism pointed at macroeconomists concerning their failure to predict or warn about the large-scale recession of 2008-2009, writing a paper about macroeconomic modeling and using macroeconomic models for policy analysis and forecasting is somewhat frustrating. Nevertheless, short-term forecasting of quarterly GDP undoubtedly plays an important role when assessing future macroeconomic performance. Its accuracy is a precondition for the sound policy decisions of monetary and fiscal authorities and private sector agents. Inaccurate short-term GDP forecasts may result in macroeconomic instability and highly volatile business cycles.

In general, the short-term forecasting of quarterly GDP is motivated by the considerable delay in official estimates emanating from the national accounts' statistics. In the case of

[1] *Dimitar Eftimoski, St. Clement of Ohrid University, Bitola, and Faculty of Business Economics, Skopje, Macedonia. E-mail: dimitar@ukim.edu.mk*

Macedonia, this delay takes at least two months (the first estimates of GDP are available approximately two months after the end of the reference quarter). Moreover, the short-term forecasts of Macedonian GDP are closely related to the effectiveness of monetary policy in terms of exchange rate targeting and inflation forecasting.

As far as we are aware, there are hardly any published papers that elaborate the problem of Macedonian GDP forecasting. This absence of research papers was crucial to our decision to evaluate the performance of the two diametrically opposed forecasting models: the medium-scale static factor model, based on the static principal components analysis, and the small-scale, predominantly Keynesian macroeconomic structural equation model (SEM).[2]

The intermediate size of our factor model helps us to avoid some specification and information problems immanent for large- and small-scale factor models. The large-scale factor models can violate the assumption of a weak cross-correlation among the idiosyncratic components needed to ensure the consistency of their estimates. The small-scale factor models are relatively exposed to idiosyncratic shocks and suffer an implicit loss of information (Cuevas and Quilis, 2012).

Our SEM is adapted to the structure of the Macedonian economy. It gives insight into the macroeconomic interactions and represents a reliable framework for forecasting the Macedonian GDP. It is worth noting that, in terms of predictive power and accuracy, much simpler and less resource-consuming aggregate supply–aggregate demand models often outperform more sophisticated models. In spite of the development of new methods of modeling, we believe that structural macroeconomic modeling remains the most promising approach for understanding macroeconomic behavior, generally (see Hall, 1995).

In a publication of the Macedonian Central Bank, Jovanovic and Petrovska (2010) evaluate the short-term forecasting performance of six competing models, including the medium-scale static factor model based on the static principal components analysis and the small-scale SEM. After evaluating the forecast accuracy based on 24 out-of-sample one-step-ahead forecasts and employing standard forecast accuracy measures (mean absolute error and root mean squared error criteria), the authors concluded that the static factor model outperforms other alternative models. In doing so, Jovanovic and Petrovska: (a) used static (one-step-ahead) forecasts, (b) extracted factors from a data set of 31 variables, and (c) regressed the real GDP on all the available principal components and retained only the statistically significant ones (common factors). The possibility of improving these three solutions proposed in Jovanovic and Petrovska (2010) — first, by using dynamic (multi-step-ahead) forecasts based on a recursive estimation approach; second, by increasing the number of predictors in the factor model to one third, which implies extracting factors from a data set of 52 variables; and third, by imposing more efficient factor selection procedures based on automatic and "manually fixed factor" selection—reveals the main motivation of

[2] *When forecasting GDP, policymakers are often oriented towards traditional macroeconomic SEMs. However, their shortcomings, usually related to the famous Lucas critique (Lucas, 1976), are the main reason for the frequent use of more sophisticated new generation models, such as vector-autoregressive (VAR) models, factor models, computable general equilibrium (CGE) models, and dynamic stochastic general equilibrium (DSGE) models, as alternatives. Recently, the DSGE models, which are established on solid microeconomic theoretical foundations, were winning the attention of policymakers. However, after the Great Recession of 2008/09, due to their shortcomings related to the insufficient coverage of the banking and financial sector and, therefore, failing to predict an upcoming recession, these structural models are being exposed to heavy criticism.*

our paper. We expect these improvements to be converted into benefits for the macroeconomic policymakers in Macedonia, when using medium-scale factor models.[3]

In general, this paper faces four challenges: (a) to introduce recursive multi-step-ahead forecasts of Macedonian GDP, (b) to find out whether the improvements in the factor selection procedure can ameliorate the Macedonian short-term GDP forecasts - when using an intermediate data-set, (c) to compare the forecasting performance of the factor model relative to the macroeconomic SEM, and (d) to increase the body of research studies concerning the macroeconomic (GDP) forecasting in Macedonia.

The structure of the paper is as follows: Section 2 provides a short literature review. In Section 3, we introduce the forecasting models and accuracy measures. Section 4 describes the data and explains the forecasting methodology used. In Section 5, we provide estimation results and discuss the outcomes of forecasting experiments. Section 6 summarizes the main findings.

2. Literature Review

In this section, we provide a short overview of some empirical studies concerning our paper.

Recently, an increasing number of studies have dealt with forecasting using factor models. Stock and Watson (2002) forecasted a single time series with many predictors using the dynamic forecasting model with static factors obtained by the static principal components analysis. The same approach can be found in Bai and Ng (2002, 2013), Bai (2003), Klein *et al*. (2004) and Bai and Wang (2014, 2016). In its dynamic form, factor models were used as forecasting tools in the following studies: Bernanke and Boivin, 2003; Forni *et al*., 2005; Kapetanios and Marcellino, 2006; Artis *et al*., 2007; Rünstler *et al*., 2009; Schumacher, 2010, 2011; D'Agostino and Giannone, 2012; Bessec, 2013; and den Reijer, 2013. Schumacher (2007) discussed the forecasting performance of alternative factor models in the case of the German economy. His out-of-sample forecast experiments showed that the forecast errors of the factor models are on average smaller than the forecast errors of the parsimonious autoregressive (AR) benchmark model. His study also revealed that the dynamic factor models outperform the static factor model in most cases. Wang (2009) investigated out-of-sample forecasting performance of the dynamic factor model, DSGE model, VAR model, and AR model, for the United States (US) economy. According to his findings, the factor model performs the best, when forecasting GDP growth and inflation, in the short-run. In their papers, Lombardi and Maier (2011) and Winter (2011) confirmed that the dynamic factor model exhibits the best forecast accuracy even during the recent crisis. However, Boivin and Ng (2005) found negligible difference in forecasting between the static and dynamic factor models, just like Eickmeier and Ziegler (2008), using a meta-regression, and reported better forecast performance of static versus dynamic factor models. Nevertheless,

[3] *Note that Jovanovic and Petrovska (2010) concluded that the factor model performs the best while the small-scale SEM and the "foreign demand model" outperform the other three alternative models (ARIMA, Kalman AR, and FAVAR), which are well specified but fail to produce decent forecasts. That is why we have excluded these three alternative models from further analysis. Moreover, we found that the "foreign demand model" (proposed by Jovanovic and Petrovska, 2010) is developed on a simple intuitive premise with extremely strong assumptions, which make this model undesirable for GDP forecasts. Consequently, we have focused our interest on the remaining two best performing models: the static factor model and the SEM.*

as stated in Bai and Ng (2007), not much is expected to be gained from the distinction between the static and dynamic factor models for forecasting purposes. In terms of size, almost all aforementioned studies use either large- or small-scale factor models. From this point of view, it is worth mentioning that Watson (2003) found that the increase of the number of predictors beyond 50 does not substantially improve predictive gains. Boivin and Ng (2006) showed that factors extracted from the 40 series, using sound economic logic, can produce similar results to factors extracted from the 147 series. Furthermore, Bai and Ng (2008) suggested that forecast accuracy does not necessarily increase with the enlargement of the time series number. As a result, the medium-scale factor models have become very popular among researchers and policymakers. To forecast French GDP, along with the small- and large-scale factor models, Barhoumi *et al.* (2008) used medium-scale factor models based on static and dynamic principal components derived from a data set of 51 variables. Banbura and Modugno (2010), in line with the small- and large-scale models, applied a medium-scale dynamic factor model to forecast euro-area GDP. Their composition of predictors consists of the 46 series. To compute short-term forecasts of the euro-area GDP growth, Camacho and Quiros (2010) employed a medium-scale factor model. Bencivell *et al.* (2012) used medium-scale factor model (similar to that proposed by Camacho and Quiros, 2010) to forecast euro-area GDP. In their paper, the information set for each euro-zone country is composed of 20 to 30 variables. Cuevas and Quilis (2012) utilized a medium-scale factor model to forecast the short-term growth rate of the Spanish economy. They used 31 selected economic indicators divided into five large blocks.

Charemza (1994) employed a macroeconomic SEM to simulate and forecast the economic activity in eight Central and Eastern European (CEE) countries. To predict some macroeconomic and monetary variables, Fuhrer and Moore (1995) as well as Rudebusch and Svensson (1999) estimated small-scale macroeconomic models adapted to the structure of the US economy. Basdevant (2000) used macroeconomic SEM for the Russian Federation to predict outcomes of a number of alternative economic policies. Merlevede *et al.* (2003) developed small-scale macroeconomic SEMs for CEE countries to analyze various macroeconomic relationships and adjustments during their path to European Union (EU) integration. The estimated models are subsequently used as inputs for in-sample and out-of-sample forecasting and policy evaluation. Crespo-Cuaresma *et al.* (2009) applied small-scale country-specific macroeconomic SEMs. They consisted of six structural co-integration relationships for modeling: private consumption, investment, exports, imports, nominal exchange rates, and nominal interest rates. Their results, which aimed for a small sample of Central, Eastern, and Southeastern European (CESEE) countries, suggest that the forecasting performance of SEMs is not superior to a parsimonious AR model. Steiner *et al*. (2014) used macroeconomic SEM for semi-annual GDP and imports projections on a small sample of CESEE countries. Their paper represents a modification of the aforementioned small-scale country-specific macroeconomic SEMs proposed by Crespo-Cuaresma *et al.* (2009) in terms of a more precise representation of the economic heterogeneity of observed countries.

When it comes to the macroeconomic forecasting in Macedonia, Jovanovic and Petrovska (2010) employed out-of-sample, one-step-ahead forecasts to evaluate the short-term forecasting performance of six different models. Their results indicate that the static factor model outperforms all other alternative models. Petrovska *et al.* (2016) applied the Qual VAR approach for pseudo out-of-sample forecasting of Macedonian business cycle turning points. They concluded that the economic sentiment indicator (ESI), along with the real GDP and the capacity utilization rate in manufacturing, can fairly predict Macedonian business cycle fluctuations. Recently, Petrovska *et al*. (2017) used four models (a dynamic factor model[4]; three-equation structural model, and two ARIMA models – based on aggregated and disaggregated approach) for short-term forecasting of inflation. They found that the disaggregated ARIMA model performs the best.

3. Forecasting Models and Accuracy Measures

In this section, we present the forecasting models for the out-of-sample forecast experiments. First, we provide a description of the medium-scale static factor model; then, we discuss the small-scale macroeconomic SEM. Both models are kept rather simple and well-specified from a statistical point of view. Finally, the main accuracy measure is introduced.

3.1. The Factor Model

The factor model assumes that each variable in the data set can be represented as a sum of two components: the common component (small number of unobserved factors common to all variables) and the idiosyncratic component (specific to each variable):

$$X_t = w_t(F_t) + \xi_t \tag{1}$$

where: X_t denotes the observed $N \times 1$ dimensional vector of stationary time series with observations for $t = 1, \dots, T$; $w_t(F_t)$ are the common components solely driven by the factors F_t; and ξ_t are the idiosyncratic components for each of the variables, that is, the part of X_t which is not explained by the common components.

Following Stock and Watson (2002), we use the dynamic forecasting model augmented with static factors obtained by the static principal components analysis (PCA). In fact, we are trying to reduce the dimensionality of a data set in which there are a relatively large number of interrelated variables while retaining, as much as possible, the variation present in the data set. This reduction is achieved by transforming into a new set of variables (the principal components) which are uncorrelated and are ordered, so that the first few encompass most of the variation present in all the original variables. The computation of the principal components reduces to the solution of an eigenvalue-eigenvector problem for a positive, semi-definite symmetric matrix (see Jolliffe, 2002). More concretely, we use the PCA to choose the parameters and the number r of common factors F_t in a way to retain, as much as possible, the variation present in the original data set:

$$X_t = \Lambda F_t + \xi_t \tag{2}$$

where: Λ is the $N \times r$ factor loading matrix; F_t is the $r \times 1$ dimensional vector of unobserved common factors; and ξ_t is the $N \times 1$ dimensional vector of idiosyncratic shocks.

The choice of the number of common factors strongly determines the predictive ability of factor models. In order to select an optimal number of common factors, we put in comparison the automatic factor selection procedure, based on information criterion IC_{p1} (proposed by Bai and Ng, 2002), and the "manually fixed procedure," based on manually fixed number of

[4] *Their medium-scale factor model extracts six common factors from a data set of 73 variables. However, the paper does not provide an explanation of the factor selection procedure in details.*

factors — both in terms of the relative mean squared forecast errors.[5] The optimal number of common factors minimizes IC_{p1}:

$$IC_{p1}(r) = \ln(V(r,F)) + r\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right) \qquad (3)$$

The first term on the right-hand side of Eq. 3 shows the goodness-of-fit, that is, the sum of the squared residuals that decreases when the number of factors increases. The second term on the right-hand side (after *r*) stands for the penalty of over-fitting, which can lead to a loss of efficiency and is an increasing function of the cross-section size *N* and time series length *T*. The criterions are evaluated for all values of $r = 1, \dots, r_{max}$, where $r_{max} = 6$.[6]

When the number of common factors is selected, we proceed with the following forecasting model where y (GDP) is projected on a set of estimated common factors and possibly lags of the dependent variable:

$$y_{t+h}^{h} = \beta_0 + \sum_{i=1}^{r} \beta'_i \hat{F}_{t,i} + \sum_{j=1}^{p} \delta_j y_{t-j+1} + \varepsilon_{t+h}^{h} \qquad (4)$$

where: $\hat{F}$ refers to the estimated common factors; β_i denotes the factors' coefficients, which are estimated by OLS for each forecast horizon *h*; y_{t-j+1} refers to the autoregressive components; and ε_{t+h} are the forecast errors.

3.2. The Structural Equation Model

Our second forecasting tool is a small-scale predominantly Keynesian[7] macroeconomic SEM, the core part of which consists of six structural equations modeling private consumption, investment, exports, imports, nominal exchange rate, and nominal interest rate (using an augmented Taylor rule).

Every structural equation represents causal relationships among the different variables in the model, where the dependent variable in one regression equation may appear as an independent variable in another regression equation. The structure of the model is a simple aggregate supply–aggregate demand model, which is given by the following equations:

$$cons = \alpha_0 + \alpha_1 \cdot y + \alpha_2 \cdot (int - cpi) \qquad (5)$$

$$inv = \beta_0 + \beta_1 \cdot (int - ppi) + \beta_2 \cdot y \qquad (6)$$

$$expo = \omega_0 + \omega_1 \cdot err + \omega_2 \cdot expo\_eu + \omega_3 \cdot y\_eu + \omega_4 \cdot y \qquad (7)$$

$$imp = \delta_0 + \delta_1 \cdot y + \delta_2 \cdot err + \delta_3 \cdot expo\_eu \qquad (8)$$

$$ern = \theta_0 + \theta_1 \cdot m4 + \theta_2 \cdot y + \theta_3 \cdot int \qquad (9)$$

$$int = \tau_0 + \tau_1 \cdot cpi + \tau_2 \cdot y + \tau_3 \cdot m4 + \tau_4 \cdot u \qquad (10)$$

In addition to domestic variables, the model includes the EU GDP and exports, so that the specificities of a small open economy have been taken into account. All the GDP

[5] *Note that Bai and Ng (2002) showed that the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) overestimate the number of common factors. This is because the penalty for over-fitting of both criterions is set as an increasing function of the time series length* T *only.*

[6] *We set* $r_{max} = 6$ *since we intend to obtain an appropriate variable-to-factor ratio.*

[7] *The model includes some classical assumptions, such as the dependence of private consumption on interest rates.*

components are modeled as a function of some explanatory variables, except for government consumption (pub), which is taken as an exogenous variable from the government budget projections. Finally, the GDP identity equation is also included.

Following Merlevede *et al.* (2003) and Crespo-Cuaresma *et al.* (2009), private consumption $(cons)$ in Eq. 5 is determined by the output (y) and the nominal interest rates, which are deflated by the consumer price index $(int - cpi)$. The investment (inv) equation is a function of nominal interest rates, which are deflated by the producer price index $(int - ppi)$, and output. In Eq. 7, exports are determined by the real exchange rate (err) as an indicator of price competitiveness, EU exports $(expo\_eu)$ as a proxy for the global trade volume, EU GDP (y_eu) as a proxy of foreign demand, and output. In Eq. 8, the output is assumed to be a proxy for domestic demand. Furthermore, imports are primarily determined by the real exchange rate, but an additional explanation is obtained by the EU exports (as a proxy for the global trade volume). The nominal exchange rate (ern) in Eq. 9 is modeled as a function of the monetary aggregate M4 $(m4)$, output, and nominal interest rates. Finally, Eq. 10 combines elements from Taylor's rule (1993), which includes the rate of inflation, output,[8] and an empirical interest rate rule proposed by Fair (2001), which adds unemployment (u) and money growth to the Taylor rule.

All dependent variables (except private consumption and investment) are assumed to follow an AR process, which helps to reduce the persistence of the error in structural equations. Furthermore, all exogenous variables (that is, all the variables that do not appear on the left-hand side in Eqs. 5-10) are assumed to follow simple AR(1) processes, which presumably is the most efficient way of modeling in terms of lost observations and degrees of freedom.[9] Given the limited sample size, the maximum number of possible lags of the exogenous variables in Eqs. 5-10 is restricted to 4.

3.3. The Benchmark Model

We use an autoregressive model of first order as a simple benchmark model, against which both competing models are compared. For forecasting purposes, the same single equation model (4), estimated with a dynamic (multi-step) estimation approach,[10] is applied:

$$y_{t+h}^{h} = \beta_0 + \sum_{j=1}^{p} \delta_j y_{t-j+1} + \varepsilon_{t+h}^{h} \tag{11}$$

One may see that the only difference between the equations is the lack of factors in Eq. 11. The Bayesian Information Criterion (BIC) is used for the lag length selection.

3.4. Accuracy Measure

The relative mean squared error (relative MSE) is used as a measure for comparing forecasts of the same series between different models. It is defined as:

8 Following Crespo-Cuaresma et al. (2009), output is introduced as an additional term to capture the cyclical stand of the economy, which is traditionally measured in terms of the output gap.

9 The obtained results do not significantly change if the optimal lag length of the AR processes is chosen using the Akaike Information Criterion when, in most of the cases, the optimal lag length is determined to be 1.

10 According to Boivin and Ng (2005), when comparing forecasts of factor models, the direct approach is preferable.

$$relative\ MSE = \frac{n^{-1}\sum_{t=S+1}^{S+n} e_{1t}^2}{n^{-1}\sum_{t=S+1}^{S+n} e_{2t}^2} \tag{12}$$

where: e_{1t}^2 are the squared forecast errors of the alternative model; e_{2t}^2 are the squared forecast errors of the benchmark model; and S and n denote the number of observations in the estimation and validation samples, respectively. For a specified forecast horizon, a relative MSE less than one indicates the superior forecast performance of the alternative model.

The relative MSE is also used for measuring gains (or losses) in the predictive ability of models: gain or loss = 1 – relative MSE, where a score larger than zero indicates that the alternative model exhibits gain in predictive ability.

Finally, to check the robustness of the results obtained by the relative MSE criterion, we introduce two more alternative accuracy measures: the mean absolute error (MAE) and the root squared mean error (RSME).[11]

4. Data and Forecasting Strategy

4.1. Data

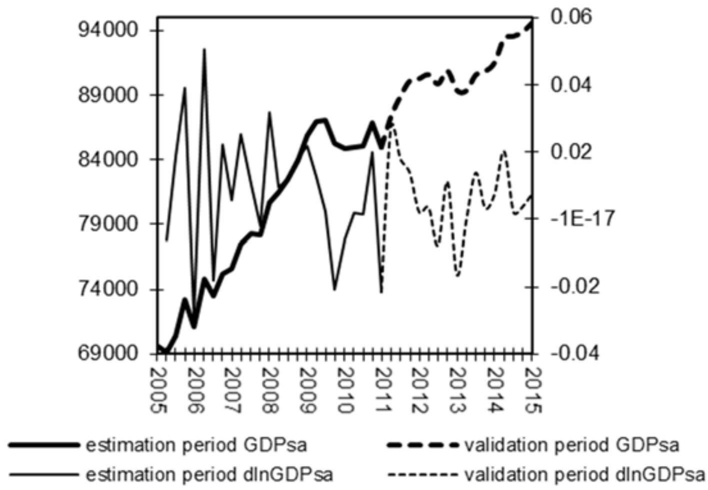
We use quarterly data for the period 2005q1–2015q1. The data for the macroeconomic SEM referring to Macedonia were obtained from the Macedonian Central Bank, Macedonian State Statistical Office, and Jovanovic and Petrovska (2010), while the data concerning the EU were acquired from the Penn World Table. The data for the factor model (52 time series, measuring overall economic activity) were obtained from the Macedonian Central Bank and Jovanovic and Petrovska (2010). They consisted of six types of series: GDP variables (12 series); prices (7 series); labor market variables (5 series); production, foreign direct investment (FDI), and capital stock (6 series); monetary, financial, and fiscal variables (17 series); and miscellaneous variables (5 series). The Macedonian GDP data was provided from the Macedonian State Statistical Office and Jovanovic and Petrovska (2010) (see Appendix, Table A1).

Natural logarithms were taken for all the time series except for the interest rate variables. Stationarity was obtained by appropriately differencing the time series. Seasonal distortions were eliminated using the X11 (Historical) adjustment method. The in-sample period ranges from 2005q1 to 2011q1 while the remaining part, from 2011q2 to 2015q1 (*i.e.,* 16 observations in the validation sample or approximately 40% of the total sample), is used for out-of-sample forecasts (see Figure 1).

[11] The results obtained by the alternative accuracy measures are given in the Appendix. More on alternative accuracy measures can be found in Albu *et al.* (2015).

Figure 1

Estimation and validation periods



Note: Level (left axis) and differenced data (right axis) of the seasonally adjusted GDP; sa denotes seasonal adjustment; dlnGDPsa stands for the first difference of the logarithm of GDPsa.

4.2. Forecasting Strategy

We use the pseudo out-of-sample forecasting method to simulate the real-time performance of forecasting models. The evaluation between the models is performed by recursive forecasts over the validation sample. The recursive validation scheme proceeds as follows: Initially, the models are estimated for the time period 2005q1–2011q1. Forecasts are computed with a forecast horizon of $h = 1,..,4$, and forecast errors are stored. In the next step, the sample size is increased by one period (2011q2 is included in the estimation sample), and the models are re-estimated. Then, forecasts are computed again with a forecast horizon of $h = 1,..,4$, and the forecast errors are stored. This procedure is repeated for the entire validation sample. The forecast series are evaluated using the standard evaluation measure—relative MSE.

5. Estimation and Results

In this section, we provide some estimation details of the competing models, and we discuss the outcomes of the forecasting experiments.

Following the PCA procedure, in order to extract common factors from a data set of 52 variables, we converted correlated variables into uncorrelated factors. The automatic factor selection procedure, based on information criterion $\mathrm{IC_{p1}}$ (proposed by Bai and Ng, 2002), suggests retaining five common factors. It implies that the variable-to-factor ratio is 10.4. The first common factor accounts for 18.99% of the variation in the data, the second for 12.35%, the third for 10.69%, the fourth for 9.79%, and the fifth for 8.60%. In total, all five common factors explain 60.43% of the variation in the data. The minimum value of communalities is 0.72; the maximum value is 0.89, while the average value of communalities is equal to 0.79. Consequently, it is important to mention that if the communalities are high (as in our case), recovery of population factors in sample data is normally very good, almost regardless of sample size, level of over-determination, or presence of model error. Thus, smaller samples are likely sufficient when communalities are high (see MacCallum *et al.*, 1999, 2001). Using variables with high communalities substantially reduces sample size requirements. Therefore, a more manageable design aspect might be the variable-to-factor ratio. Having at least 8 variables per factor is advised, and a ratio of 10 or more is preferred (Pearson and Mundform, 2010).

In order to choose an appropriate number of common factors, we have tried to find out whether the automatic factor selection procedure, based on information criterion IC_{p1}, can systematically outperform the “manually fixed factor” procedure in terms of relative MSEs. Table 1 shows the gains (or losses) in predictive ability of the static factor model (FM) for a different number of estimated common factors over the AR(1) benchmark model. One may see that only one common factor is needed to outperform the AR(1) model and that a number of factors, including the automatic IC selection, can generate superior forecasts to the benchmark model in all forecast horizons (except at $h = 2, 3, 4$ for $r = 2$). In these cases, for different forecast horizons, the factor model yields from 2.8–72% lower relative MSEs than the AR(1) benchmark model. However, since the results indicate that the automatic information criterion-based procedure does not systematically outperform the “manually fixed factor” procedure, the way of selection of an appropriate number of common factors stays ambiguous.

Table 1

Gains (Losses) in predictive ability (in terms of relative MSEs): FM vs AR(1)

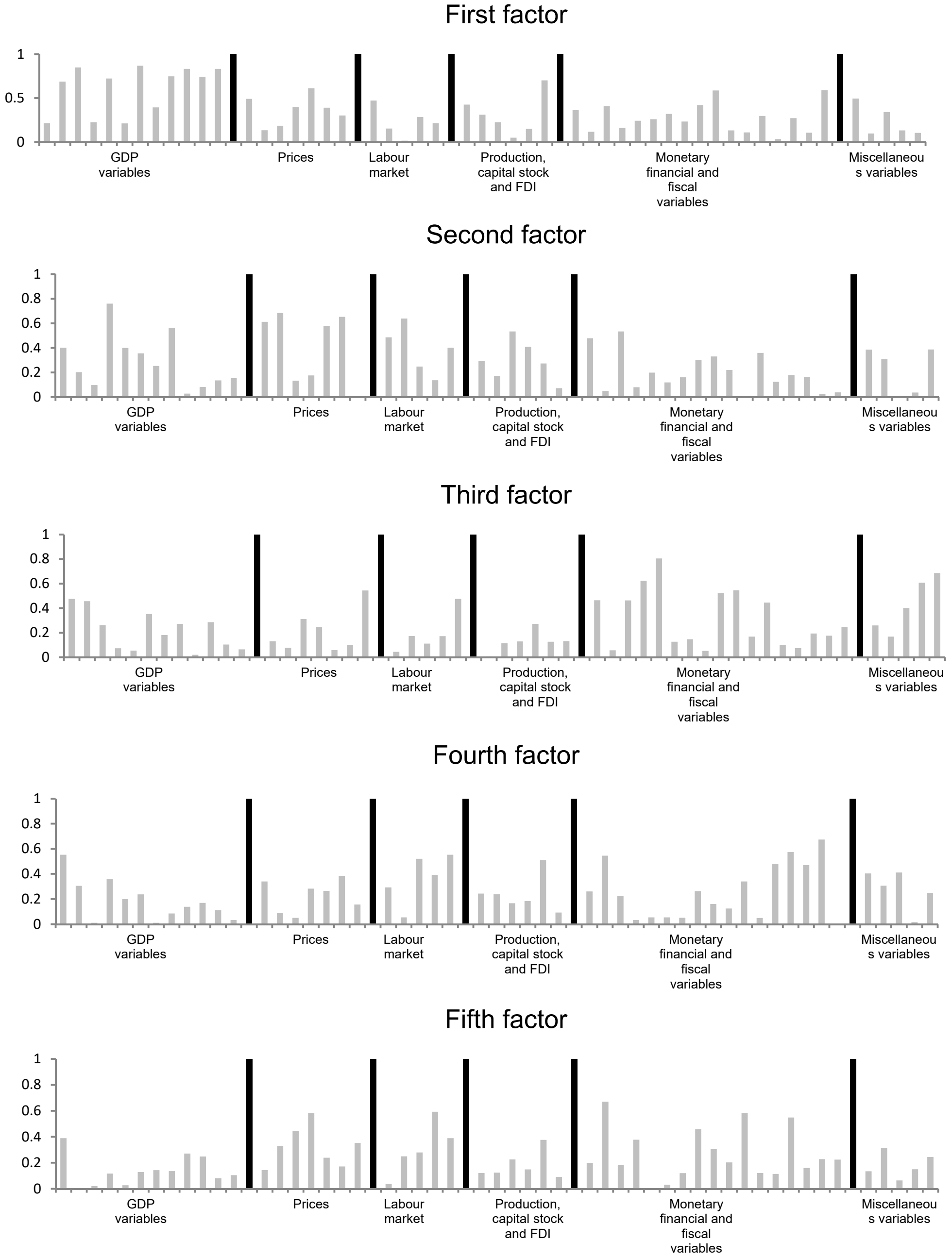
Forecast horizon	Fixed number of factors					
	r = 1	*r* = 2	*r* = 3	*r* = 4	*IC*	*r* = 6
1	0.1838	0.1094	0.5221	0.6151	0.6284	**0.6290**
2	0.0282	-0.1132	0.5456	0.6000	0.6021	0.5938
3	0.0684	-0.2238	0.6068	**0.6635**	0.6440	**0.6596**
4	0.1110	-0.2933	0.6528	**0.7199**	0.7049	0.6909

Note: r denotes the fixed number of factors using manually fixed procedure. IC denotes the automatic factor selection with information criterion IC_{p1} proposed by Bai and Ng (2002). A relative MSE larger than zero indicates that the factor model exhibits gain in predictive ability. Bold values denote higher gains than those obtained by using IC.

Following Stock and Watson (2002) and Dias *et al.* (2015), to characterize the factors we present Figure 2: the R^2 of the regressions of the 52 individual series on each of the five common factors for the entire sample period. One may see that the first common factor is related to the GDP variables. The second common factor describes the prices and the labor market. The third factor reflects the monetary and miscellaneous variables, while the fourth factor is related to the fiscal variables. Finally, the fifth common factor evenly captures variations in all series types and is not related to any particular type. We have estimated the static factor model recursively, using an OLS estimator, regressing the first difference of the logged and seasonally adjusted real GDP on common factors.

Figure 2

R^2 between the individual series and the common factors.



Concerning the macroeconomic SEM, due to the problem of endogeneity, in order to estimate the coefficients of the system, consisting of six structural equations and eight AR processes, we applied the multivariate estimation method—seemingly unrelated regressions (SUR)—that accounts for heteroskedasticity and contemporaneous correlation in the errors across equations (Zellner, 1962). To solve the model, we applied a dynamic-deterministic simulation with the Broyden algorithm as a solver. The results indicate that the real interest rates have almost neutral effect on private consumption as well as on investment; that is, in the short-run, the transmission mechanism from the real interest rates towards both aforementioned variables works slowly, which is in line with our expectations and some previous investigations, for example, Eftimoski (2019). Furthermore, both the real export and the real import display expected positive effect from the increased global trade volume, while in the short-run, the nominal exchange rate is fairly well explained by the monetary variables and output. Finally, the augmented Taylor rule "works properly" when explaining nominal interest rates.

As aforementioned, the forecast accuracy of the models is measured in terms of the relative MSE criterion, computed on the basis of a recursive out-of-sample forecast evaluation. The relative MSEs are computed as ratios between the MSEs of the alternative models and the MSE of the AR(1) benchmark model. A relative MSE less than one indicates superior forecast performance of the alternative model for the chosen forecast horizon $h = 1, .., 4$.

Figure 3 displays forecast evaluation results for all forecast periods. From the one-period-ahead forecasts (pane a), it is obvious that the SEM performs volatile oscillations, but, in general, it follows GDP fluctuations relatively well. For the same forecast period, the factor model and the AR(1) benchmark model exhibit a faster adjustment. The relative MSEs criterion (see Table 2) indicates that, at the one-period-ahead horizon, the forecast errors of the macroeconomic SEM are bigger than the errors of a parsimonious benchmark model. However, if we take into account the SEM's small scale, then we can conclude that it performs quite decently and in line with our expectations.

Table 2

Relative MSE for all forecast horizons

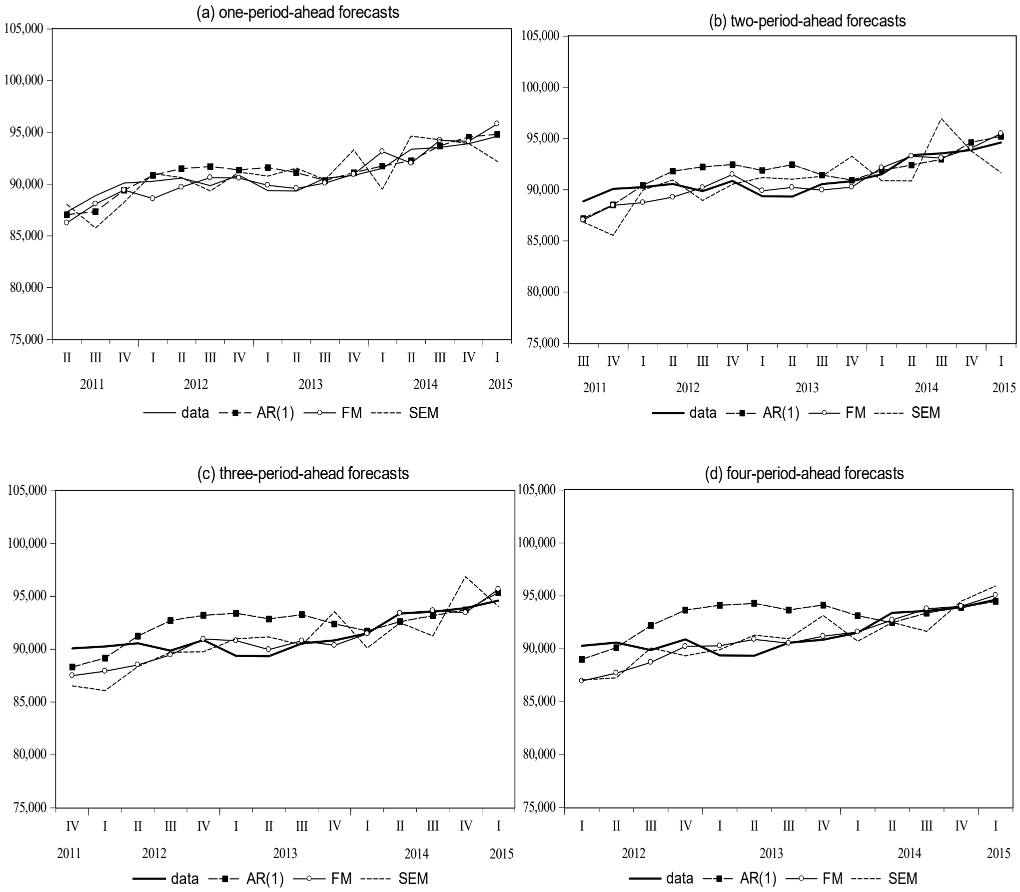
Forecast horizon	Relative MSE					Ranking			
	1	2	3	4		1	2	3	4
SEM	2.2502	1.9238	1.1515	0.4760		3	3	3	2
FM	0.3715	0.3978	0.3560	0.2950		1	1	1	1

Note: The table shows the MSE of the rival models relative to the MSE of the AR(1) model.

According to the relative MSE criterion, the factor model provides smaller forecast errors in comparison with the macroeconomic SEM and AR(1) model at the one-period-ahead horizon. The relevant literature suggests that, in terms of the forecasting performances measured in relative MSE criterion, factor models are very often superior to any other model in the short-run. This superiority is proven in our case as well. In spite of its medium-scale (52 variables included), the factor model produces the best forecast performance at the one-period-ahead horizon as reported in the previously mentioned empirical study of Jovanovic and Petrovska (2010). However, this time the results are reinforced by certain improvements into the factor selection procedure. In this respect, it is worth noting that the results obtained by the two alternative accuracy measures (MAE and RMSE) are quite similar to those reported in Table 2 where the relative MSE criterion is used (see Appendix Tables A3 and A4).

Figure 3

Forecast evaluation results for all forecast horizons



Note: (a) one-period-ahead forecasts, (b) two-period-ahead forecasts, (c) three-period-ahead forecasts, and (d) four-period-ahead forecasts.

At the same time, Figure 3 displays forecast evaluation results at the two-, three-, and four-period-ahead forecast horizons. It gives a slightly different picture of the forecasting performance of alternative models. It is obvious that the SEM adjusts slowly and performs significantly better than at the one-period-ahead horizon (panes b, c, and d). Its forecast ability rises gradually as the forecast horizon expands. On the other hand, after the one-period-ahead horizon, the FM continues to capture the GDP behavior and follows it very well. For the remaining forecast horizons (2-4) the AR(1) model exhibits larger periods of overestimation. It performs best at the two-period-ahead horizon while at the three- and four-period-ahead horizons it exhibits a slower adjustment.

According to the relative MSE criterion, SEM shows notable forecast performance at the four-period-ahead horizon (see Table 2) when it outperforms the AR(1) benchmark model. To a certain extent, these results could be related to the prevailing “wisdom” that structural models perform better in longer forecast horizons.

To assess the robustness of the forecasting performance of alternative models over time, we have conducted a sub-sample analysis. More precisely, first, we shortened the full out-of-sample period by 25% (the first sub-sample-period); then, we shortened it an additional 25% (the second sub-sample-period). The results are reported in Table 3.

Table 3

Relative MSE for all forecast horizons, sub-samples

	Forecast horizon	Relative MSE					Ranking			
		1	2	3	4		1	2	3	4
Out-of-sample period	SEM	2.6519	1.6640	1.2737	0.5824		3	3	3	2
2011Q2-2014Q1	FM	0.4512	0.5358	0.4774	0.2983		1	1	1	1
Out-of-sample period	SEM	1.3167	1.5238	1.1281	1.6916		3	3	3	3
2011Q2-2013Q1	FM	0.3657	0.5007	0.5403	1.5331		1	1	1	2

Note: The table shows the MSE of the rival models relative to the MSE of the AR(1) model.

The results of Table 3 can be directly compared to the results of Table 2. One may see that the results of the full out-of-sample period (Table 2) are in general confirmed by the results of the two sub-out-of-sample periods (Table 3). An additional check of the robustness of the results obtained by the factor model is given in the Appendix.

6. Concluding Remarks

We have employed two alternative models for short-term forecasting of the Macedonian GDP–factor model and macroeconomic SEM. Our findings suggest that quite reliable results can be obtained using a medium-scale static factor model. More specifically, we have found that: (a) the factor model, which exploits an intermediate data-set, shows superior forecasting abilities and does better at the one-period-ahead forecast horizon, which is in line with the results reported by Jovanovic and Petrovska (2010). Moreover, our results show that the factor model exhibits the best forecasting performances at the two-, three-, and four-period-ahead forecast horizons as well; (b) the applied improvements in the factor selection procedure can ameliorate Macedonian short-term GDP forecasts. In this respect, the automatic factor selection procedure suggests retaining five common factors (which implies that the variable-to-factor ratio is equal to 10.4). However, the mode of selecting an appropriate (optimal) number of common factors still stays ambiguous since the automatic factor selection procedure fails to outperform the "manually fixed factor" procedure systematically. Consequently, as an alternative, a more manageable design aspect regarding the optimal number of predictors might be the variable-to-factor ratio. In this respect, at least 10 variables per factor are advised. Nevertheless, the automatic factor selection procedure, based on information criterion proposed by Bai and Ng (2002), seems to be the most adequate solution; (c) it is very plausible that factors extracted from 52 rather than 31 variables can produce more accurate GDP predictions. However, keeping in mind the optimal variable-to-factor ratio, we are not convinced that the extraction of common factors from a larger data set than ours can substantially improve the short-term forecasts of the Macedonian GDP. Of course, we encourage such further investigations; (d) the SEM performs quite well. It produces smaller forecast errors at the four-period-ahead horizon, which is in line with the general perception that the structural models perform better in longer forecast horizons. It also reveals that, in the short run, the transmission mechanism from the real interest rates towards private consumption and investment works slowly, which implies that the real interest rates have almost neutral effect on both variables (private consumption and investment). This particular finding raises some important questions regarding the effectiveness of the current monetary policy, especially in terms of when the Macedonian Central Bank applies a monetary-policy strategy of exchange-rate targeting where the interest rate on Central Bank bills auctions is a basic monetary-policy instrument; and (e) further gains in forecasting accuracy should be expected along with an improvement in data quality and the potential use of longer forecast horizons.

Finally, it is worth noting that, in spite of the prevailing “wisdom" that factor models are useful tools for macroeconomic forecasting, they suffer from serious limitations. For instance, the fact that factor models assume that each variable in the data set can be represented as a sum of two components—the common component (small number of unobserved factors common to all variables) and the idiosyncratic component (specific to each variable)—makes them unable to identify the forces driving the dynamics of the economy. Consequently, our medium-scale factor model faces the same limitations when interpreting the main drivers of forecast accuracy. Nonetheless, there is no doubt that our inherently data-driven model will diversify the existing macroeconomic forecasting strategy in Macedonia (especially when using medium-scale factor models with approximately 50 predictors). More concretely, we expect our factor model to be a very useful device when policymakers are uncertain about the reasons behind some stochastic processes in the economy, that is, when structural models are unable to perform as expected. In such cases, our medium-scale factor model can additionally strengthen the Macedonian real GDP forecast accuracy.

Acknowledgements

I wish to record my gratefulness to Maja Micevska Scharf from Webster University for useful comments, as well as to Branimir Jovanovic from the Macedonian Central Bank for making data available to me.

References

Albu, L.L. Jimenez, C.M. and Simonescu, M., 2015. The Assessment of Some Macroeconomic Forecasts for Spain using Aggregated Accuracy Indicators. Romanian Journal of Economic Forecasting, 18(2), pp.30-47.

Artis, M. Banerjee, A. and Marcellino, M., 2005. Factor forecasts for the UK. Journal of Forecasting, 24(4), pp. 279-298.

Bai, J., 2003. Inferential theory for factor models of large dimension. Econometrica, 71(1), pp. 135-171.

Bai, J. and Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica, 70(1), pp. 191–221.

Bai, J. and Ng, S., 2007. Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics, 25(1), pp. 52-60.

Bai, J. and Ng, S., 2008. Forecasting economic time series using targeting predictors. Journal of Econometrics,146(2), pp. 304-317.

Bai J. and Ng S., 2013. Principal components estimation and identification of static factors. Journal of Econometrics, 176(1), pp. 18–29.

Bai, J. and Wang, P., 2014. Identification theory for high dimensional static and dynamic factor models. *Journal of Econometrics*, *178*(2), pp. 794–804.

Bai, J. and Wang, P., 2016. Econometric analysis of large factor models. *Annual Review of Economics*, *8*, pp. 53–80.

Banbura, M. and Madugno, M., 2010. Maximum likelihood estimation of factor models on data set with arbitrary pattern of missing data. *European Central Bank Working Paper* No. 1189.

Basdevant, O., 2000. An Econometric Model for the Russian Federation. *Economic Modeling*, 17(2), pp. 305–336.

Bernanke, B.S. and Boivin, J., 2003. Monetary policy in a data-rich environment. *Journal of Monetary Economics,* 50(3), pp. 525-546.

Bessec M., 2013. Short-term forecasts of French GDP: a dynamic factor model with targeted predictors. Journal of Forecasting, 32, pp. 500–511.

Boivin, J. and Ng, S., 2005. Understanding and comparing factor - based forecasts. *International Journal of Central Banking*, 1(3), pp. 117-151.

Boivin, J. and Ng, S., 2006. Are more data always better for factor analysis? *Journal of Econometrics*, 132(1), pp. 169-194.

Camacho, M. and Quiros, G.P., 2010. Introducing the euro-sting: Short-term indicator of the euro area growth. *Journal of Applied Econometrics*, 25(4), pp. 663-694.

Charemza, W., 1994. LAM models for East European Economies: General description. University of Gdansk Discussion Paper.

Crespo-Cuaresma, J. Feldkirecher, M. Slacik ,T. and Woerz, J., 2009. Simple but Effective: The OeNB's Forecasting Model for Selected CESEE Countries. *Focus on European Economic Integration,* Q4/09.

Cuevas, A. and Quilis, E.M., 2012. A factor Analysis for the Spanish Economy. SERIEs *Journal of the Spanish Economic Association*, 3, pp. 311-338.

D'Agostino, A. and Giannone, D., 2012. Comparing alternative predictors based on large-Panel factor models. *Oxford Bulletin of Economics and Statistics*, 74 (2), pp. 306-326.

den Reijer, A., 2013. Forecasting Dutch GDP and inflation using alternative factor model specifications based on large and small datasets. *Empirical Economics*, 44, pp. 435–453.

Dias, F., Pinheiro, M. and Rua, A., 2015. Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence. *Economic modeling*, 44, pp. 266-272.

Eftimoski, D., 2019. An assessment of the dynamic effects of monetary policy in Macedonia. *Applied Economics Letters*, 26(10), pp. 823-829.

Eickmeier, S. and Ziegler, C., 2008. How successful are dynamic factor models at forecasting output and inflation? a meta-analytic approach. *Journal of Forecasting*, 27(3), pp. 237-265.

Fair, R.C., 2001. Estimates of the Effectiveness of Monetary Policy. Yale's Cowles *Foundation, Yale University Discussion Paper* No. 1298.

Forni, M. Giannone, D. Lippi, M. and Reichlin, L., 2005. Opening the black box: structural factor models with large cross-sections. *European Central Bank Working Paper*, No. 571.

Fuhrer, J.C. and Moore, G., 1995. Monetary policy trade-offs and the correlation between Nominal interest rates and real output. *The American Economic Review,* 85(1), pp. 219–239.

Hall, S., 1995. Macroeconomics and A Bit More Reality. *The Economic Journal,* 105(431), pp. 974-988.

Harvey, D., Leybourne, S. and Newbold, P., 1997. Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting,* 13(2), pp. 281-91.

Jolliffe, I.T., 2002. *Principal Component Analysis.* Second edition. New York: Springer.

Jovanovic, B. and Petrovska, M., 2010. Forecasting Macedonian GDP: Evaluation of Different models for short-term forecasting. *NBRM Working paper*, August 2010. Available at: <http://www.nbrm.mk/WBStorage/Files/WebBuilder_ForecastingMacedonianGDP.pdf>.

Kapetanios, G. and Marcellino, M., 2006. A Parametric Estimation Method for Dynamic Factors Models of Large Dimensions. *IGIER Working Paper* No.305.

Klein, L. R. Roudoi, A. Eskin, V. Albu, L.L. Stanica, C. Nicolae, M. and Chilian, M.N., 2004. Principal Components Model of the Romanian Economy. GDP – Production Side. *Romanian Journal of Economic Forecasting*, voumel 5 supplement, pp. 52-66.

Lombardi, M. and Maier, P., 2011. Forecasting Economic Growth in the Euro Area during The Great Moderation and the Great Recession. *European Central Bank Working Paper* No. 1379.

Lucas, R., 1976. *Econometric Policy Evaluation: A Critique*. Carnegie - Rochester Conference Series on Public Policy 1: 19–46.

MacCallum, R.C. Widman, K.F. Zhang, S. and Hong, S., 1999. Sample size in factor analysis. *Psychological Methods,* 4, pp. 84-99.

MacCallum, R.C. Widman, K.F. Preacher, K.J. and Hong, S., 2001. Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research,* 36(4), pp. 611-637.

Merlevede, B. Plasmans, J. and Van Aarle, B., 2003. A Small Macroeconomic Model of the EU-Accession Countries. *Open Economies Review,* 14(3), pp. 221–250.

Pearson, R.H. and Mundform, D.J., 2010. Recommended sample size for conducting Exploratory factor analysis on dichotomous data. *Journal of Modern Applied Statistical Methods,* 9(2), pp. 359-368.

Petrovska, M. Krstevska, A. Naumovski, N., 2016. Forecasting Macedonian Business Cycle Turning Points Using Qual VAR. Model. *Journal of Central Banking Theory and Practice*, 5(3), pp. 61-78.

Petrovska, M. Ramadani, G. Naumovski, N. Jovanovuc, B., 2017. Forecasting Macedonian Inflation: Evaluation of different models for short-term forecasting. *NBRM Working paper* No. 6/2017. Available at: <http://www.nbrm.mk/content/Working%20papers/Forecasting_Macedonian_Inflation_Evaluation_of_different_models.pdf>.

Rudebusch, G.D. and Svensson, L.E.O., 1999. Policy Rules for Inflation Targeting. In Taylor, J. B., eds. *Monetary Policy Rules.* Chicago: Chicago U. Press, pp. 203-246.

Rünstler, G. *et al*., 2009. Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of Forecasting*, 28 (7), pp. 595–611.

Schumacher, C., 2007. Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting,* 26(4), pp. 271-302.

Schumacher, C., 2010. Factor forecasting using international targeted predictors: the case of German GDP. *Economic Letters*, 107 (2), pp. 95–98.

Schumacher, C., 2011. Forecasting with factor models estimated on large datasets: a review of the recent literature and evidence for German GDP. *Journal of Economics and Statistics*, 231 (1), pp. 28–49.

Steiner, K. Wörz, J. and Slacík, T., 2014. Can Trade Partners Help Better Forcee the Future? Impact of Trade Linkages on Economic Growth Forecasts in Selected CESEE Countries. *Focus on European Economic Integration,* 1, pp. 36-56.

Stock, J.H. and Watson, M.W., 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association,* 97(460), pp. 1167-1179.

Wang, M.C., 2009. Comparing the DSGE Model with the Factor Model: an Out-of- Sample Forecasting Experiment. *Journal of Forecasting,* 28(2), pp. 167-182.

Watson, M.W., 2003. Macroeconomic forecasting using many predictors. In Dewatripont, M. L. Hansen, M.L. Turnovsky, S., eds., *Advanced Econometrics, Theory and Applications*, Eight World Congress of the Econometric Society, Vol. III, pp. 87-115.

Winter, J., 2011. Forecasting GDP Growth in Times of Crisis: Private Sector Forecasts versus Statistical Models,. *De Nederlandsche Bank Working Paper* No. 320.

Zellner, A., 1962. An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias. *Journal of the American Statistical Association,* 57(298), pp. 348-368.

Appendix

Table A1

Data description

	Variable name	Abbrevi-ation	Units	Type of series	Source
1	GDP	y	mln, quarterly, n.c.	GDP variables	Macedonian State Statistical Office and Jovanovic and Petrovska (2010) paper
2	Government consumption	pub	mln, quarterly, n.c.		
3	Private consumption	cons	mln, quarterly, n.c.		
4	Investment	inv	mln, quarterly, n.c.		
5	Exports of goods		mln, quarterly, n.c.		
6	Exports of goods and services	expo	mln, quarterly, n.c.		
7	Imports of goods		mln, quarterly, n.c.		
8	Imports of goods and services	imp	mln, quarterly, n.c.		
9	Imports of means of production		mln, quarterly, n.c.		
10	Imports of consumption goods		mln, quarterly, n.c.		
11	Gross value added: wholesale trade		mln, quarterly, n.c.		
12	Gross value added: retail trade		mln, quarterly, n.c.		
1	Consumer Price Index	cpi	2005=100	Prices	Macedonian Central Bank
2	Producer Price Index	ppi	2005=100		
3	Export prices		2005=100		
4	Import prices		2005=100		
5	Industrial metal (steel) prices		$ per tone		
6	Oil prices		$ per barrel		
7	Consumer Price Index - European Union		2005=100		
1	Wages - net (all sectors)		period average, n.c.	Labor market	
2	Wages - net (public sector)		period average, n.c.		
3	Wages - net (industry)		period average, n.c.		
4	Wages - net (service sector)		period average, n.c.		
5	Unemployed persons	u	1000		
1	Industrial production		mln, quarterly, n.c.	Production, capital stock and FDI	
2	Production of consumption goods		mln, quarterly, n.c.		
3	Production of capital goods		mln, quarterly, n.c.		
4	Completed construction works		mln, quarterly, n.c.		

5	FDI (Foreign Direct Investments)		period average, euro		and Jovanovic and Petrovska (2010) paper
6	Physical capital		2005=100		
1	Total deposits		mln, quarterly, n.c.	Monetary, financial and fiscal variables	
2	Interest rates on credits	int	percent		
3	M4 monetary aggregate	m4	mln, quarterly, n.c.		
4	Real exchange rate vs. euro	err	period average		
5	Nominal exchange rate vs. euro	ern	period average		
6	Deposits in n.c.		mln, quarterly, n.c.		
7	Deposits in the short-term		mln, quarterly, n.c.		
8	Deposits in the long-term		mln, quarterly, n.c.		
9	Credits in n.c		mln, quarterly, n.c.		
10	Total credits		mln, quarterly, n.c.		
11	Credits to households		mln, quarterly, n.c.		
12	Credits to firms		mln, quarterly, n.c.		
13	Gross foreign reserves		period average, euro		
14	VAT (Value Added Tax)		mln, quarterly, n.c.		
15	Government revenues		mln, quarterly, n.c.		
16	Government expenditures		mln, quarterly, n.c.		
17	Government capital expenditures		mln, quarterly, n.c.		
1	Private transfers		mln, quarterly, n.c.	Miscellaneous variables	
2	Pensions		period average, n.c.		
3	Telecommunications traffic (impulses from mob. tel.)		1000 min		
4	Exports - European Union	expo_eu	mln, quarterly, euro		Penn World Table
5	GDP – European Union	y_eu	mln, quarterly, euro		

All the time series are expressed in real terms, except the telecommunications traffic and unemployed persons that are expressed in natural units. The variables without abbreviation are used in the factor analysis, but are not explicitly mentioned in the text. n.c. stands for the national currency – Denar.

Robustness check of the results obtained by the factor model

In order to check the robustness of the results obtained by the factor model, the information criterion IC_{p2}, proposed by Bai and Ng (2002), is applied:

$$IC_{p2}(r) = \ln\left(V(r,F)\right) + r\left(\frac{N+T}{NT}\right)\ln\left(\min\{N,T\}\right) \tag{13}$$

Table A2 shows the forecast performance of the factor model, in terms of the relative MSE, using IC_{p2} criterion:

Table A2

Relative MSE for all forecast horizons, IC_{p2}

Forecast horizon	Relative MSE				Ranking			
	1	2	3	4	1	2	3	4
FM	0.3710	0.4062	0.3404	0.3091	1	1	1	1

Note: The table shows the MSE of the factor model relative to the MSE of the AR(1) model.

The results of Table A2 can be compared to those reported in Table 2 where the IC_{p1} criterion is used. One can see that the results are clearly confirmed.

Alternative accuracy measures and modified Diebold-Mariano (MDM) test statistic

The results acquired from the alternative accuracy measures (MAE and RMSE) are reported in Tables A3 and A4:

Table A3

MAE for all forecast horizons

Forecast horizon	Mean absolute error (MAE)				Ranking			
	1	2	3	4	1	2	3	4
AR(1)	814.9	1233.4	1629.6	1979.9	2	2	2	3
SEM	1250.2	1651.5	1836.7	1468.0	3	3	3	2
FM	558.4	803.2	852.4	946.9	1	1	1	1

Table A4

RMSE for all forecast horizons

Forecast horizon	Root mean squared error (RMSE)				Ranking			
	1	2	3	4	1	2	3	4
AR(1)	1049.9	1508.7	2037.9	2561.7	2	2	2	3
SEM	1574.9	2092.7	2186.8	1767.4	3	3	3	2
FM	639.9	951.6	1215.9	1391.4	1	1	1	1

The results are quite similar to those reported in Table 2.

It is well known that accuracy measures ranking sometimes cannot be taken as a convenient tool for model comparison since the differences in measures may not be systematic. Consequently, the modified Diebold-Mariano (MDM) test (proposed by Harvey, Leybourne, and Newbold, 1997) has been applied. The MDM statistic tests the null hypothesis to verify that the predictive abilities of two different forecasting models, based on a specified loss function $\mathrm{E}[\mathrm{g}(\mathrm{e}_t') - \mathrm{g}(\mathrm{e}_t'')] = 0$, are equal. Table A5 displays the results:

Table A5

MDM test statistics

Forecast horizon		SEM
1	FM	0.006
2	FM	0.010
3	FM	0.003
4	FM	0.005

Note: The entries are p-values of pairwise tests of equal predictive ability.

The MDM test confirms our previous findings that the static factor model systematically outperforms the macroeconomic SEM at all forecast horizons.