LATVIJAS BANKA

ISBN 9984–676–39–0

VIKTORS AJEVSKIS
GUNDARS DĀVIDSONS

DYNAMIC FACTOR MODELS IN FORECASTING LATVIA'S GROSS DOMESTIC PRODUCT

2 • 2008

WORKING PAPER

CONTENTS

Abstract	2
Introduction	3
1. Stock–Watson Factor Model	4
2. GDFM of M. Forni, M. Hallin, M. Lippi and L. Reichlin	6
3. Comparative Analysis of Stock–Watson Factor Model and GDFM	7
4. Data Description	7
5. Empirical Results	8
Conclusions	12
Appendices	13
Bibliography	23

ABBREVIATIONS

ADL model – Autoregressive Distributed Lags Model
APT – Arbitrage Pricing Theory
BIC – Bayesian Information Criterion
CSB – Central Statistical Bureau of Latvia
DSGE model – Dynamic Stochastic General Equilibrium Model
GDFM – Generalised Dynamic Factor Model
GDP – gross domestic product
RMSFE – root mean squared forecast error
US – United States of America
VAR – vector autoregression

ABSTRACT

The study aims at evaluating how useful the application of models using large panels of data in forecasting Latvia's GDP is. Two factor models have been used: the Stock–Watson factor model and the generalised dynamic factor model. The forecast findings by the two models have been compared with the results obtained by the benchmark autoregressive model. The results suggest that compared with simpler autoregressive models both the Stock–Watson factor model and the generalised dynamic factor model ensure forecast improvement, which, however, has not been statistically significant if statistical tests are used.

Keywords: *forecasting, factor models, large cross section*

JEL classification: *C32, C33, C53*

INTRODUCTION

Forecasting is among the most important activities of the decision making process at national central banks. In order to produce forecasts as accurately as possible, central banks improve their forecasting methodology on a constant basis using up-to-date forecasting methods. In the forecasting process, information from a large number of series and various econometric models is used. This approach has an advantage: potentially significant information is not neglected.

The models used by central banks differ notably in the employed variables, econometric techniques and economic theories underpinning them.

The majority of central banks employ traditional structural macroeconomic models as the core models. Recently, they are replaced by DSGE models with a stronger theoretical foundation. In short-term forecasting, central banks, additionally to structural models, often use different time series econometric models that do not incorporate economic theory. Univariate time series models and VAR models with a small number of variables are currently used as standard short-term forecasting models.

From a theoretical point of view, the conditional mathematical expectation based on all available information is considered to be an optimal forecast or a forecast with a minimum root mean squared forecast error. From practical positions, information related to a variable's forecast can be very broad. Central banks closely monitor tens or even hundreds of macroeconomic indicators, each of which can provide useful information. B. S. Bernanke and J. Boivin called it "looking at everything".(3) In regression type (e.g. VAR) models, employing a large number of variables impairs the estimation efficiency and results in a forecast of lower quality (with the number of variables exceeding the number of observations, estimation is not possible at all). Such conditions are the drivers behind the macroeconomic research trend of building statistical models using information from a large number of variables. The latter is a trend called the dynamic factor analysis. The idea underpinning it is based on an assumption that the dynamics of macroeconomic variables is determined by a few unobservable factors that can be estimated using broad panel data. In such a way, information inherent in a large number of variables can be used in constructing forecasting models of smaller dimension.

According to the factor model structure, the dynamics of variables can be described by two mutually orthogonal components: 1) the common component, which is a linear combination of common factors and hence closely correlates with all panel variables and 2) idiosyncratic components, which comprise specific information on each variable and are weakly correlated with broad panel dimension variables. The features of the common and specific components of factor models distinguish one type of the factor model from another. According to the classical factor model, idiosyncratic (specific) components of variables are mutually orthogonal. The approximate factor model of G. Chamberlain and M. Rothschild (7) as well as the model of G. Connor and R. A. Korajczyk (9) or the generalised static factor model allow for a moderate correlation between idiosyncratic components. These models are usually used in financial econometrics and the arbitrage pricing theory (APT). Another improvement of the classical factor model is the assumption regarding dynamics.(15; 23) The respective research trend is usually referred to as the dynamic factor analysis.

J. Stock and M. W. Watson introduced dynamics into the approximate factor model.(28) They assumed that common factors affect the observed variable with a finite number of lags. The dynamics results from the inclusion of lagged factor values in the model specification. Simplicity of estimation techniques is an advantage of the static representation of the dynamic factor model. In a generalised factor model, static factors can be estimated using the principal component method.

GDFM proposed by M. Forni, M. Hallin, M. Lippi and L. Reichlin (12; 13) uses the data dynamics structure, estimating factors as the dynamic key components of the spectral density matrix.

Factor models are used in forecasting inflation and GDP in the US (14; 27; 28), such euro area countries as Germany, the Netherlands, Austria, Belgium and France (2; 6; 12; 13; 17; 20; 22; 24; 25; 26, etc), New Zealand, Canada and Australia (8; 19) and the United Kingdom (1; 2). The findings of the respective research papers show that usually the root mean squared forecast error of factor models is lower than that obtained by smaller benchmark models, autoregressive models as an example.

The aim of this paper is to find out to what extent, compared with simpler models, the abovementioned models improve forecasting results when Latvia's data are used. Although in other countries the models yield satisfactory results, regarding Latvia such specific features as structural adjustments and short time series should be accounted for. Notwithstanding availability of the major data as of the mid-1990s, a large share of indicators that would be appropriate to characterise the situation and useful in forecasting and nowcasting (where factor models yield potentially largest benefits) are available starting with later periods.

Sections 1 and 2 give an overview of the two models used in the paper. Section 3 provides comparative analysis of both models. Section 4 deals with the research data. Section 5 presents empirical results, whereas the concluding part comprises inferences regarding model application and suggestions regarding further studies in this area.

1. STOCK–WATSON FACTOR MODEL

Two models are used in the study. The first is the Stock–Watson factor model.(28) It is assumed that a scalar series y_t is to be forecast. X_t as a vector of n-dimensional indicators (stationary and with zero mean) with $t = 1, 2, \ldots T$ is analysed.

J. Stock and M. W. Watson (27) assume that the following factor model describes the dynamics of (y_t, X_t):

$$y_{t+j} = \beta(L) f_t + \gamma(L) y_t + u_{t+j} \qquad [1],$$

$$X_{it} = \lambda_i(L) f_t + \xi_{i,t}, \; i = \overline{1,n} \qquad [2]$$

where

$\xi_{i,t}$ are idiosyncratic component errors,

$\beta(L)$, $\gamma(L)$ and $\lambda_i(L)$ are lag polynomials,

f_t is the vector of m-dimensional common factors.

It is additionally assumed that

$$E[u_t | X_t, f_{t-1}, y_{t-1}, X_{t-2}, f_{t-2}, y_{t-2}, ...] = 0.$$

It implies that $\hat{y}_{t+1} = \beta(L) f_t + \gamma(L) y_t$ with information available at time period t is the best forecast (in terms of RMSFE) for y_{t+1}. $\beta(L)$, $\gamma(L)$ and $\lambda_i(L)$ are finite degree polynomials defined as

$$\beta(L) = \sum_{j=0}^{q} \beta_j L^j,$$

$$\gamma(L) = \sum_{j=0}^{q} \gamma_j L^j,$$

$$\lambda(L) = \sum_{j=0}^{q} \lambda_j L^j.$$

On the basis of this assumption, models [1] and [2] can be rewritten in static form:

$$y_{t+1} = \beta' F_t + \gamma(L) y_t + u_{t+1}$$

$$X_t = \Lambda F_t + \xi_t \qquad [3]$$

where

$F_t = (f_t', f_{t-1}', ..., f_{t-q}')$ is the r-dimensional vector with $r \le m(q+1)$, i-th row of matrix Λ is $(\lambda_{i0}, \lambda_{i1}, ..., \lambda_{iq})$ and $\beta = (\beta_0, \beta_1, ..., \beta_q)'$.

The model written in this form will simplify computation, for it allows for an easier estimation of model parameters with principal components. The following target function is considered:

$$V(F, \Lambda) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(x_{it} - \lambda_t' F_t \right)^2.$$

The minimisation problem of function $V(F, \Lambda)$ is equivalent to maximisation problem of $tr(\Lambda' X'X\Lambda)$ subject to condition $\Lambda' \Lambda = I$. The presented minimisation problem can be solved as a principal component problem where $\hat{\Lambda}$ matrices are eigenvectors corresponding to the largest r eigenvalues of matrix $X'X$ (arranged in diminishing order), and the resulting matrix is multiplied by $\sqrt{n}$. Consequently, the estimation of principal components is

$$\hat{F} = \frac{X\hat{\Lambda}}{n}.$$

In the case of $n>T$, F can be even more easily estimated using $T \times T$ matrix XX'. A non-parametric estimation of $\tilde{F}$ is obtained as an eigenvector matrix associated with the largest r eigenvalue of matrix XX' divided by $\sqrt{T}$. As the spaces formed by $\hat{F}$ and $\tilde{F}$ coincide, both $\hat{F}$ and $\tilde{F}$ can be estimated.

2. GDFM OF M. FORNI, M. HALLIN, M. LIPPI AND L. REICHLIN

The second model employed in the study is based on the research by M. Forni, M. Hallin, M. Lippi and L. Reichlin.(13) Similar to the Stock–Watson factor model, it is assumed that the vector-valued process X_t is the sum of two unobservable components: the common component χ_t and the idiosyncratic component ξ_t:

$$X_t = \chi_t + \xi_t \qquad [4].$$

The similarity between this model and the Stock–Watson factor model should be noted; equation [4] can be rewritten in the same way as equation [2]:

$$X_t = \lambda(L) f_t + \xi_t .$$

Similar to the Stock–Watson factor model, we need to find out how much of X_t variance Γ_{nk}^{T} (T is the number of sample periods, n is the number of series, k is the number of lags) is explained by the common variance component Γ_{nk}^{χ} and how much by the idiosyncratic variance component (Γ_{nk}^{ξ}).

The estimation of common and idiosyncratic components is conducted in two steps.

First, we calculate the spectral density matrix for different frequencies. As a part of these calculations, we first determine sample autocorrelation matrices Γ_{nk}^{T}; afterwards, on the basis of Fourier transformation, we compute spectral density matrices for different frequencies using the Bartlett lag-window estimator. From it, we can obtain estimations for the common and idiosyncratic covariance matrices Γ_{nk}^{χ} and Γ_{nk}^{ξ}.

Second, from these covariance matrices we can construct a linear combination for the current period, which provides the smallest ratio of the common variance and idiosyncratic variance using the generalised principal component method.

A more detailed description of the model can be found in Annex 2 (see also (12) and (13)).

3. COMPARATIVE ANALYSIS OF STOCK–WATSON FACTOR MODEL AND GDFM

There are three principal distinctions between the Stock–Watson factor model and GDFM (the discussion hereinafter is based on (12) and (13)). First, the weights differ when common factors are calculated. The Stock–Watson factor model uses the standard principal component method for the purpose of obtaining common factors. GDFM estimation, in turn, is based on the method of generalised principal components. Intuitively, the generalised principal component method is the standard principle component method estimated on the basis of weighted data where weights are inversely proportional to idiosyncratic component variance. This weighting scheme is a more efficient estimation method.

Second, the common factor calculation is different (in projection). The Stock–Watson factor model uses the least squares method, while the GDFM employs a non-parametric regression accounting for the differences between dynamic factors and their lagged values, imposing rank reduction to the spectral frequency matrix.

Third, the methods differ in the way they are used to forecast the idiosyncratic component. The Stock–Watson factor model uses lagged values in the forecast calculation, while the GDFM forecasts the idiosyncratic component on the basis of the assumption about orthogonality of the common and idiosyncratic components.

From the point of view of computation and daily application, the advantage of the static representation of the Stock–Watson factor model is the simplicity of its estimation techniques.

4. DATA DESCRIPTION

Quarterly data of 126 economic indicators (from the second quarter of 2000 to the fourth quarter of 2006; total of 27 quarters) have been used in GDP forecasting (see Chart 1 for the GDP dynamics).

This set of economic indicators comprises information on the following statistical data groups:

– GDP in the breakdown by sector (A–O; total of 16 indicators).

– Year-on-year volume index of industrial output (manufacturing, manufacture of food and beverages, wood and articles of wood, wearing apparel, printing and other industries (total of 21 indicators).

– GDP expenditure (consumption, investment, demand, imports and exports, etc; total of 8 indicators).

– Wage statistics (statistical data in the breakdown by sector: agriculture, construction, transport, education and other sectors; total of 16 indicators).

– New orders in the Latvian market and for exports (statistical data in the breakdown by sector; total of 26 indicators).

– Inflation indicators (in the breakdown by goods and services group: food and non-alcoholic beverages, alcoholic beverages and tobacco, health, transport, communication, education, recreation, etc; total of 13 indicators).

– Deflators (11 indicators).

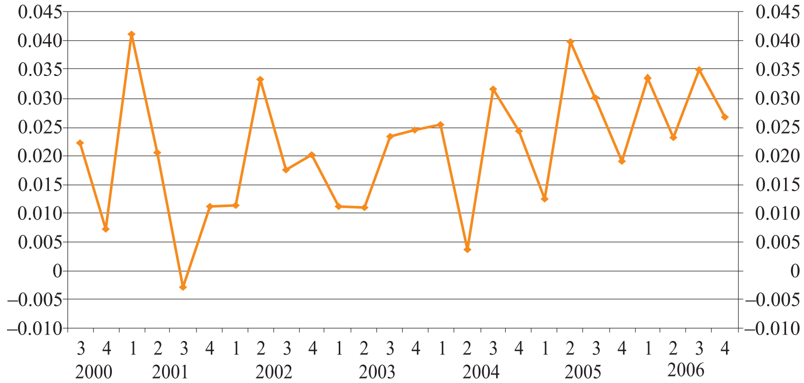
– Employment indicators (number of unemployed, unemployment rate, job vacancies, etc; total of 7 indicators).

– Other indicators (currency in circulation, interest rates, etc; total of 8 indicators).

(The list of all data is available in Appendix 4.)

All time series were first presented in logarithmic form; then they were seasonally adjusted and, if necessary, differentiated to obtain stationary time series. Hence the first order differences were mostly used. When the calculation is based on GDFM, the mean is subtracted from the data prior to calculation (making a forecast, it is added up). Data of different time series become available at different periods. For instance, when the GDP data is published in Latvia the respective quarterly data for a part of indicators (e.g. volume index of industrial output, wage statistics, new orders, employment and monetary indicators) are already available. Both abovementioned forecasting methods foresee the application of complete matrices. The solution to this problem is as follows. Those time series that lack the last period have been moved forward by one period, i.e. if the time series are arranged in columns with the last period down at the bottom, the respective row in this column is also pushed down.

Chart 1
Latvia's GDP growth



5. EMPIRICAL RESULTS

As the set of data used in this paper has a small number of time series observations (only 27), restrictions associated with the number of factors and maximum lags arise.

Four forecasting strategies were used. Within each of them, GDP forecasting for 0 (nowcasting), one and two steps ahead was conducted. The 2-steps ahead forecast has been viewed as y_{t+2} projection to explanatory variables determined for time period t. The out-of-sample forecasting period is from the first quarter of 2005 to the third quarter of 2006. In fact, a time series consisting of 19 periods has been used in the calculation. Traditionally, the forecasting is estimated by RMSFE. The smaller the error, the better the result is.

In the first approach (S–W), the Stock–Watson methodology is used. A model that produces the best forecast (in terms of the minimal RMSFE) for the selected out-of-sample interval is selected from all combinations of factors and lags (the maximum number of factors and lags is 12 and 3 respectively). In addition, the factors and model coefficients were recalculated after each forecasting step, taking into account the new available information, while maintaining the model specification unchanged.

In the second approach (S–W *fbic*), the Stock–Watson methodology is used as well. The search for the best model [1] in compliance with the BIC for the estimation interval from the third quarter of 2000 to the fourth quarter of 2004, using different combinations of the first 12 factors (without lags), is conducted. The factors and model coefficients were recalculated after each forecasting step, taking into account the new available information, while maintaining the model specification unchanged.

The third approach (S–W *flagbic*) is also based on the Stock–Watson methodology. In this case, the focus is on the models with combinations of the first four factors and their two time lags for the estimation interval from the third quarter of 2000 to the fourth quarter of 2004; the best of them is selected in compliance with the BIC. The rest of the analysis is analogous to the second approach.

The fourth approach (GDFM) uses the methodology of M. Forni, M. Hallin, M. Lippi and L. Reichlin. The number of dynamic and static factors was selected as follows. As this is an exercise in pseudo-real time[1], the same forecasting strategy was used for all periods (but not the same model), i.e. as if a hypothetical forecaster invariably uses one and the same principle to select the best model for each period. In each period "the forecaster":
a) finds such a combination of dynamic and static factors, which, if used in forecasting, produces the least root mean squared forecast error for the past (in this case, from the start of the database in 2000);
b) uses this combination in the forecasting of the next period.

The method of dynamic factor number determination based on the jump of the spectral frequency matrix's eigenvalues was also used in the calculation. The method has been described by M. Forni *et al.*; it, however, yielded worse forecasting results than the one described above (the number of static factors is still selected using a) method).

Publicly accessible codes created by M. Forni *et al.* (29) and Matlab software codes created by the authors of this paper are used.

---

[1] The real-time forecast means that information available at the given period is only included; in this case, however, the word "pseudo" is added due to regular updates carried out by, e.g. CSB, being ignored and a revised data set is used.

All factor models are compared with a benchmark model chosen from all autoregressive models on the basis of BIC (the model with the fifth lag and a constant proved to be the best).

It should be noted that the forecasts obtained by models using the first approach shall not be compared with the forecasts obtained by any of the other models due to *ex post* forecast basis used in the S–W approach. Hence it is quite natural that these models have the best forecast. Good forecasting qualities for a fixed out-of-sample interval serve as a precondition for their further application in forecasting. However, no guaranties exist as to the selected model forecasting qualities remaining as good in the future as well. From all autoregressive models, the same model with the fifth time lag and a constant, which was used as a benchmark model, proved to be the best (by *ex post* forecasts).

Out-of-sample forecasts and actual GDP series charts are given in Appendix 3. Table 1 shows RMSFE ratios of the factor to benchmark regression models. If the value of the ratio is less than unity, the forecast quality of this model is better than that of the benchmark model.

Table 1
RMSFE for different models

Forecasting horizon		S–W	S–W *flagbic*	S–W *fbic*	GDFM
0 (nowcasting)					
	RMSFE	0.0036	0.0086	0.0132	0.0100
	D–M *p*-values	0.1411	0.5282	0.3239	0.7072
	Relative versus benchmark model RMSFE	0.3293	0.7911	1.2110	0.9107
1 step					
	RMSFE	0.0052	0.0102	0.0104	0.0096
	D–M *p*-values	0.0777	0.7206	0.7812	0.5587
	Relative versus benchmark model RMSFE	0.4797	0.9281	0.9518	0.8818
2 steps					
	RMSFE	0.0038	0.0101	0.0093	0.0110
	D–M *p*-values	0.1109	0.7516	0.6190	0.9730
	Relative versus benchmark model RMSFE	0.3478	0.9277	0.8509	1.0060

S–W – 12 factors, three time lags, *ex post* RMSFE best model;
S–W *flagbic* – 4 factors, two time lags, estimation period from the third quarter of 2003 to the fourth quarter of 2004, the best ADL model using BIC;
S–W *flbic* – 12 factors, without time lags, estimation period from the third quarter of 2003 to the fourth quarter of 2004, the best ADL model using BIC;
D–M *p*-values – *p*-values of the Diebold–Mariano test.

If the comparison is conducted on the basis of the relative RMSFE value, the performance of the benchmark model can be improved by using factor models. The best nowcasting is for the S–W *flagbic* model. The best 1-step-ahead forecast can be obtained by using GDFM strategy, and for 2 steps ahead for the model selected within the S–W *fbic* strategy.

The Diebold–Mariano test (11; for test description see Appendix 1) indicates, however, that these differences cannot be considered as statistically significant (an unlikely surprising conclusion as forecasts are made for only 7 periods). The best *ex post* model produces around three times smaller RMSFE than the benchmark model, notwithstanding that the Diebold–Mariano test suggests that the improvement is only marginally significant.

CONCLUSIONS

The study aims at evaluating how useful the application of models using large panel of data in forecasting Latvia's GDP is. Two factor models have been used: the Stock–Watson factor model and the generalised dynamic factor model. The forecasts of the two models have been compared with the forecasts of the benchmark autoregressive model.

The findings confirm that at this point, compared with smaller autoregressive models, both the Stock–Watson factor model and the GDFM can improve the forecast of the benchmark autoregressive model in most cases; however, the obtained improvement is not statistically significant, if statistical tests (the Diebold–Mariano test in this case) are used. Currently, the data for a very short time span are available, and the comparison of only 7 periods was conducted, with the time series on the basis of which the model is derived covering only 20 periods; consequently, the results obtained cannot be taken as ultimate.

Quarterly data (monthly data, if used, grouped into quarters) are used in this study, and the entire available database (126 time series) was involved in forecasting. As pointed out by J. Boivin and S. Ng (4) and according to empirical evidence in other works (24), the use of a more extensive database is not always advantageous for improving the outcome of forecasting, particularly if the idiosyncratic part is serially correlated. A change in selecting the database would be useful for further research. The use of a balanced data set (only quarterly data) is another aspect of the research. When the Stock–Watson factor model is used, the outcome improvement is possible via employing unbalanced time series as well (using both the monthly and quarterly data).

APPENDICES

Appendix 1
The Diebold–Mariano test

F. Diebold and R. S. Mariano test the null hypothesis to verify that RMSFEs of both forecasts are equal. It is assumed that e'_t and e''_t are two competing forecast errors, and $d_t = e'^2_t - e''^2_t$, where $t = 1, 2, ..., P$. The Diebold-Mariano statistic is as follows:

$$DM = \frac{\bar{d}}{\sigma_d} \sim t(P-1) \qquad \text{[A1]}$$

where $\bar{d} = \frac{1}{P}\sum_{t=1}^{P} d_t$,

$$\sigma_d = \sqrt{\frac{\sum_{t=1}^{P}(e'^2_t - e''^2_t)^2 - [\sum_{t=1}^{P}(e'^2_t - e''^2_t)^2]^2 / P}{P(P-1)}} .$$

Test statistic is presented in Student's distribution with $P - 1$ freedom degrees. If the null hypothesis is rejected, the sign of the statistic points to the best forecast model. If the sign is positive, the first forecast is better than the second one. If the statistic value is negative, the second forecast is better than the first one.

Appendix 2

Approach of M. Forni, M. Hallin, M. Lippi and L. Reichlin (13)

The following steps are made (for the sake of lucidity, the narration below follows the vectorised representation used in the MATLAB codes as close as possible). First, autocovariance matrices of order k Γ_{nk}^T are computed:

$$\Gamma_{nk}^T = \left(\frac{1}{(T-k)}\right)\sum_{t=k+1}^{T} X_{n,t}X'_{n,t-k} \qquad \text{[A2]},$$

in such a way obtaining matrices within the range $\Gamma_{n(-k)}^T, \ldots, \Gamma_{n0}^T, \ldots, \Gamma_{nk}^T$;

where $X_{n,t} = (x_{1t}, \ldots, x_{nt})'$; n is the number of time series used in the computation, and T is the number of periods in the data set.

Then spectral density matrix is computed

$$\Sigma_n^T(\theta_s) = \sum_{k=-M}^{M} w_k \Gamma_{nk}^T e^{-i\theta_s k} \qquad \text{[A3]}$$

where $\Sigma_n^T(\theta_s)$ is (n x n) matrices for the spectrum $\theta_s = \dfrac{2s\pi}{2M+1}$ where $s = -M, -M+1, \ldots, M$; $M = round(\sqrt{T})$ and $w_k = 1 - \dfrac{|k|}{M+1}$.

Technically these matrices are obtained by multiplying

$$\begin{bmatrix} \Gamma_{n(-k)}^T(1,1) & \ldots & \Gamma_0^T(1,1) & \ldots & \Gamma_{nk}^T(1,1) \\ \Gamma_{n(-k)}^{nT}(1,2) & \ldots & \Gamma_0^T(1,2) & \ldots & \Gamma_{nk}^T(1,2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \Gamma_{n(-k)}^T(1,n) & \ldots & \Gamma_0^T(1,n) & \ldots & \Gamma_{nk}^T(1,n) \end{bmatrix} \begin{bmatrix} 1 & e^{ik\theta_0} & e^{ik\theta_1} & \ldots & e^{-ik\theta_{2M}} \\ 1 & \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 1 & 1 & 1 \\ 1 & \ldots & \ldots & \ldots & \ldots \\ 1 & e^{-ik\theta_0} & e^{-ik\theta_1} & \ldots & e^{ik\theta_{2M}} \end{bmatrix},$$

$$\begin{bmatrix} \Gamma_{n(-k)}^T(1,1)+\ldots+\Gamma_{n0}^T(1,1)+\ldots+\Gamma_{nk}^T(1,1) & \Gamma_{n(-k)}^T(1,1)e^{i2\theta_1}+\ldots+\Gamma_{n0}^T(1,1)+\ldots+\Gamma_{nk}^T(1,1)e^{-i2\theta_1} & \ldots & \ldots & \ldots \\ \Gamma_{n(-k)}^T(1,2)+\ldots+\Gamma_{n0}^T(1,2)+\ldots+\Gamma_{nk}^T(1,2) & \Gamma_{n(-k)}^T(1,2)e^{i2\theta_1}+\ldots+\Gamma_{n0}^T(1,2)+\ldots+\Gamma_{nk}^T(1,2)e^{-i2\theta_1} & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}.$$

The first column produces $\Sigma_n^T(\theta_s)$ $s=0$, and the second $s=1$ respectively. Thus for grid of frequencies $s = 0, \ldots, 2M$ spectral density matrices $\Sigma_n^T(\theta_s)$ are obtained.

From these matrices the first q eigenvalues are computed (q is the number of dynamic factors which is assumed as known) and $\Sigma_n^{\chi T}(\theta_s)$ matrix for $s = 0,...,2M$ is derived

$$\Sigma_n^{\chi T}(\theta_s) = \lambda_{n1}^T(\theta)\widetilde{p}_{n1}^T p_{n1}^T + \quad .... \quad + \lambda_{nq}^T(\theta)\widetilde{p}_{nq}^T p_{nq}^T, \tag{A4}$$

where $\lambda_{nj}^T(\theta)$ is the *j*-th largest eigenvalue of the respective $\Sigma_n^T(\theta_s)$ matrix, but p_{nj}^T is the *j*-th corresponding eigenvector, with $\widetilde{p}_{nj}^T$ denoting that the vector has been transposed and conjugated. Thus 2M+1 matrices are again obtained.

From $\Sigma_n^{\chi T}(\theta_s)$ matrices, using the inverse Fourier transformation, we can already derive $\Gamma_{nh}^{\chi T}$ matrix:

$$\Gamma_{nh}^{\chi T} = \left(\frac{1}{(2M+1)}\right)\sum_{h=-M}^{M}\Sigma_n^{\chi T}(\theta_s)e^{i\theta_h k} \tag{A5}$$

$Z_n^T = (Z_{n1}^T{}'...Z_{nr}^T{}')$, Z_{nj}^T is the generalised eigenvectors matrix for the pair of matrices $(\Gamma_{n0}^{\chi T}, \Gamma_{n0}^{\xi T})$, r is the number of static factors, hence Z_n^T is

(r x n) matrix where $\Gamma_{n0}^{\xi T} = \Gamma_{n0}^T - \Gamma_{n0}^{\chi T}$.

The obtained variables allow for computing matrix

$$K_n^{Th} = \Gamma_{nh}^{\chi T} Z_n^T (Z_n^T{}'\Gamma_{n0}^T Z_n^T)^{-1} Z_n^T{}' \tag{A6}$$

and the corresponding projection

$$\chi_{i,T+h|T}^{nT} = \sum_{j=1}^{n} K_{n,ij}^{Th} x_{jT} \tag{A7}$$

where $K_{n,ij}^{Th}$ is the K_n^{Th} matrix element in the *i*-th row and *j*-th column respectively.

As a result, $\chi_{i,T+h|T}^{nT}$ is obtained, which is the matrix of (1 x n) row and produces a forecast using the part explained by the common factors.

Appendix 3
GDP forecast plot

Chart A1
Preliminary assessment (nowcasting) (GDFM model)

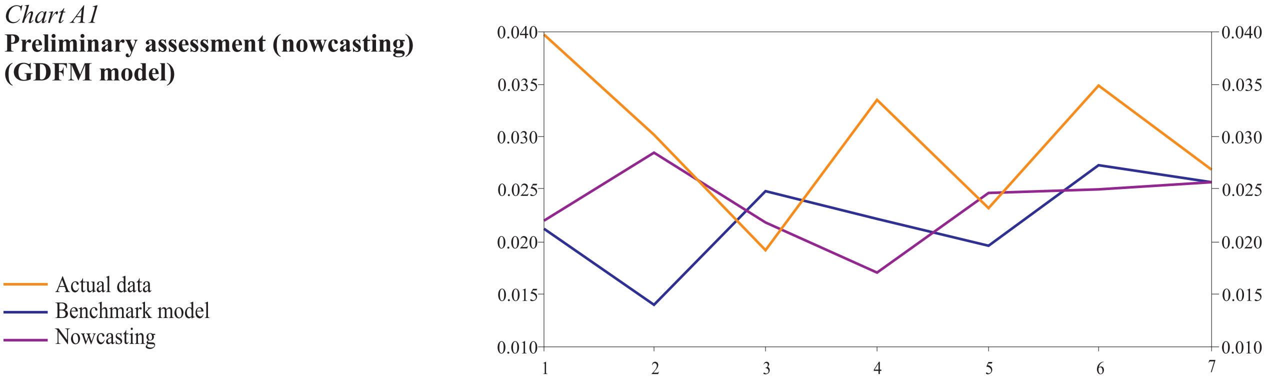


Chart A2
1-step-ahead forecast (GDFM model)

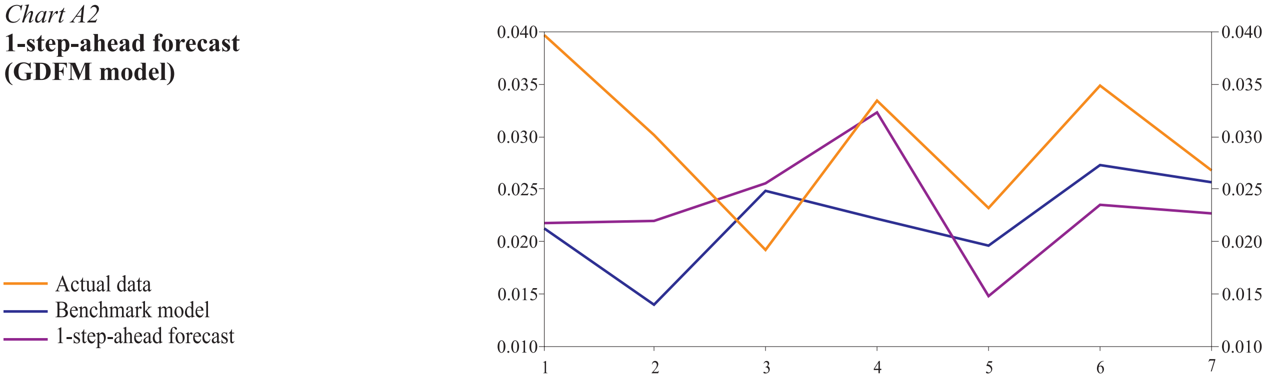


Chart A3
2-steps-ahead forecast (GDFM model)

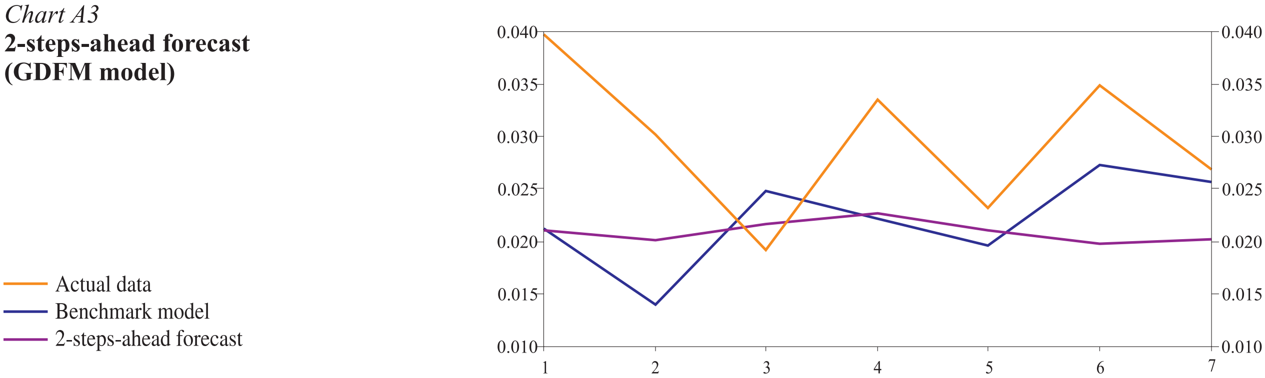


Chart A4
Nowcasting (S–W *flagbic* modelis)

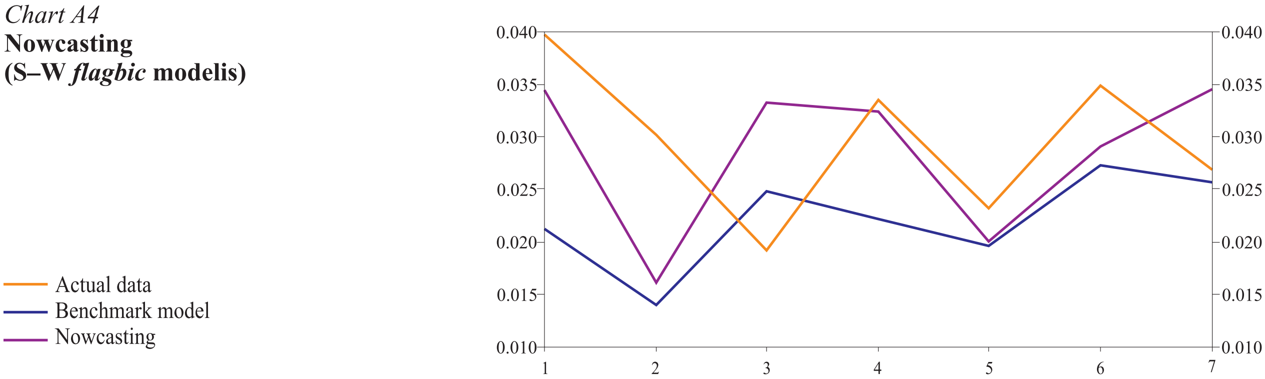


Chart A5
1-step-ahead forecast
(S–W *flagbic* model)

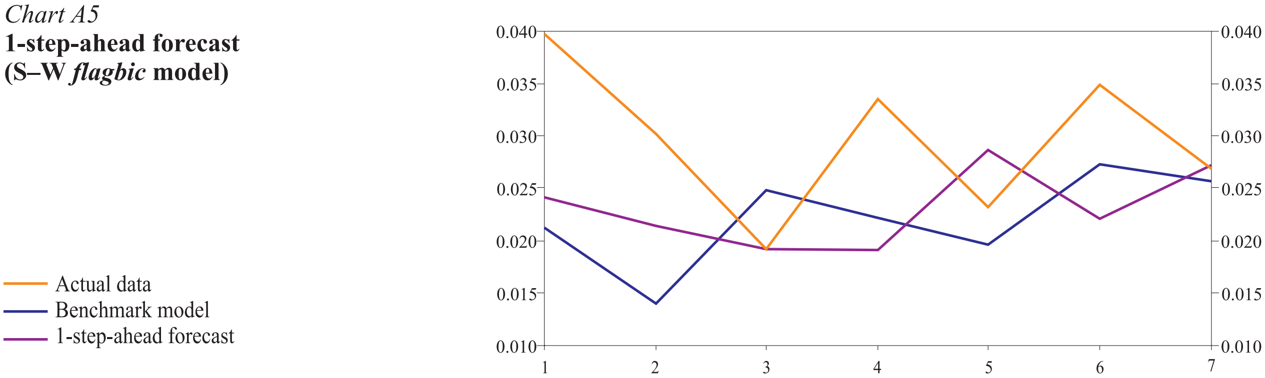


Chart A6
2-steps-ahead forecast
(S–W *flagbic* model)

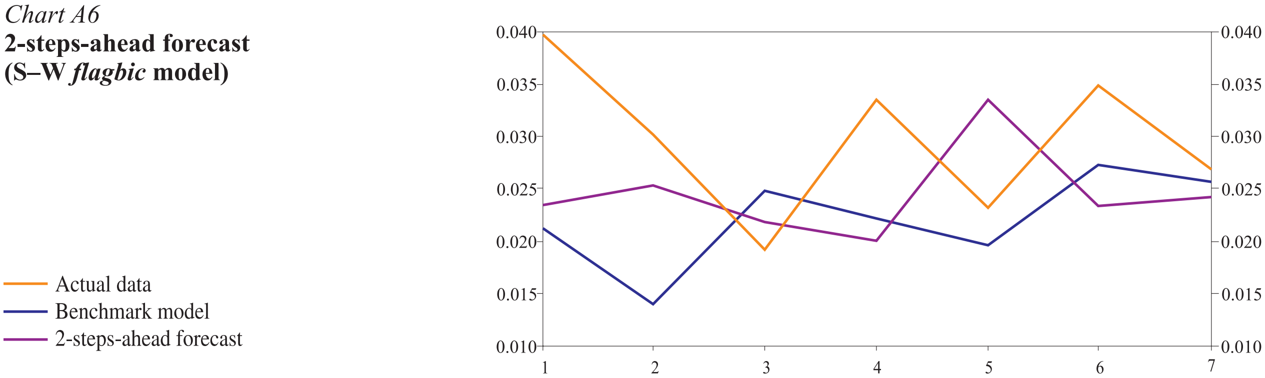


Chart A7
Nowcasting
(S–W *fbic* model)

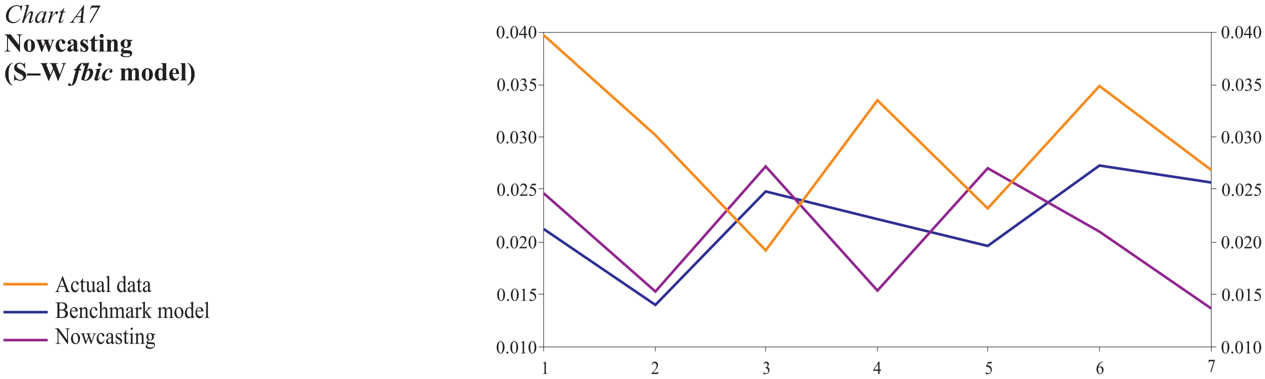


Chart A8
1-step-ahead forecast
(S–W *fbic* model)

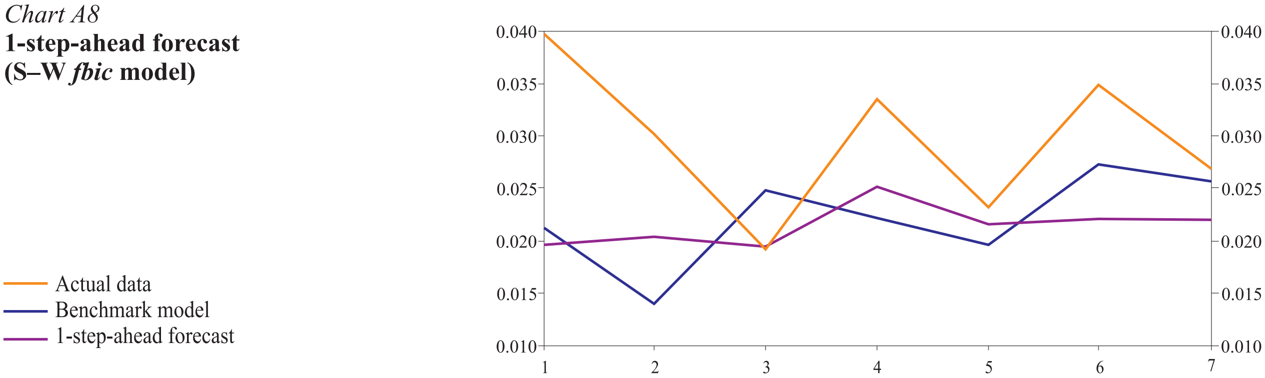
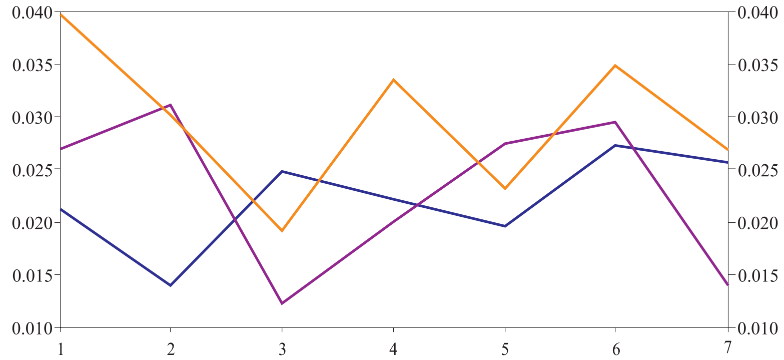
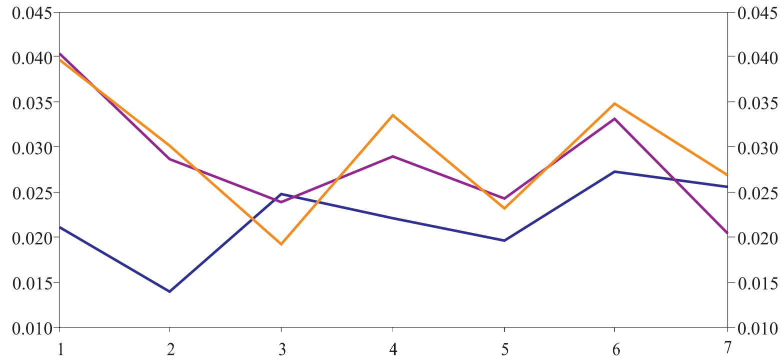


Chart A9
2-steps-ahead forecast
(S–W *fbic* model)



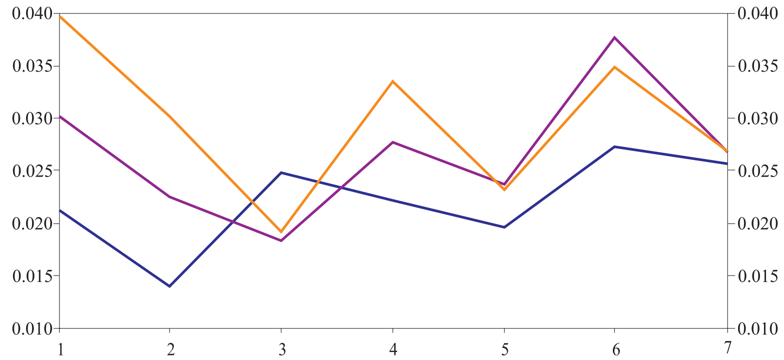
Actual data
Benchmark model
2-steps-ahead forecast

Chart A10
Nowcasting
(S–W model)



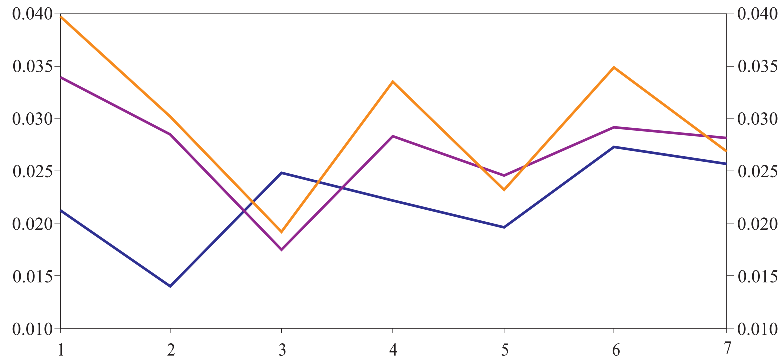
Actual data
Benchmark model
Nowcasting

Chart A11
1-step-ahead forecast
(S–W model)



Actual data
Benchmark model
1-step-ahead forecast

Chart A12
2-steps-ahead forecast
(S–W model)



Actual data
Benchmark model
2-steps-ahead forecast

Appendix 4
Data list

GDP growth at constant prices and by sector

1 GDP – gross domestic product
2 A – agriculture, hunting, forestry
3 B – fishing
4 C – mining and quarrying
5 D – manufacturing
6 E – electricity, gas and water supply
7 F – construction
8 G – wholesale and retail sale; repair of motor vehicles, motorcycles and personal and household goods
9 H – hotels and restaurants
10 I – transport, storage and communication
11 J – financial intermediation
12 K – real estate, renting and business activities
13 L – public administration and defence; compulsory social security
14 M – education
15 N – health and social work
16 O – taxes (less subsidies) on products

Year-on-year volume index of industrial output (overall and by sector)

17 Total industry
18 Mining
19 Manufacturing
20 Manufacture of food products and beverages
21 Manufacture of textiles
22 Manufacture of wearing apparel
23 Dressing of leather and manufacture of leather products
24 Manufacture of wood and of products of wood
25 Manufacture of pulp, paper and paper products
26 Printing industry
27 Manufacture of chemicals, chemical products and man-made fibres
28 Manufacture of rubber and plastic products
29 Manufacture of other non-metallic mineral products
30 Manufacture of fabricated metal products, except machinery and equipment
31 Manufacture of machinery and equipment n.e.c.
32 Manufacture of electrical machinery and apparatus
33 Manufacture of radio, television and communication equipment and apparatus
34 Manufacture of other transport equipment
35 Manufacture of furniture
36 Electricity, gas and water supply
37 Recycling

GDP expenditure (at 2000 prices)

38 Consumption
39 Household consumption

40 Government consumption
41 Investment
42 Fixed investment
43 Domestic demand
44 Exports of goods and services
45 Imports of goods and services

Nominal wages and salaries in the economy overall and by sector (in lats)

46 Economy overall
47 Agriculture, hunting and related services
48 Fishing
49 Industry
50 Mining
51 Manufacturing
52 Industry, electricity, gas and water
53 Construction
54 Wholesale and retail trade
55 Hotels and restaurants
56 Transport, storage and communication
57 Financial intermediation
58 Real estate
59 Public administration and defence; compulsory social security
60 Education
61 Health and social work

New orders (in thousands of lats)

62 Textiles, domestic market
63 Dressmaking/sewing of clothing, domestic market
64 Paper, domestic market
65 Chemicals, domestic market
66 Metal, domestic market
67 Metal products, domestic market
68 Machinery and equipment, domestic market
69 Office equipment and computers, domestic market
70 Electrical machinery, domestic market
71 Radio and television, domestic market
72 Watches, domestic market
73 Motor vehicles, domestic market
74 Other transport equipment, domestic market
75 Textiles, non-domestic market
76 Dressmaking/sewing of clothing, non-domestic market
77 Paper, non-domestic market
78 Chemicals, non-domestic market
79 Metal, non-domestic market
80 Metal products, non-domestic market
81 Machinery and equipment, non-domestic market
82 Office equipment and computers, non-domestic market
83 Electrical machinery, non-domestic market

84 Radio and television, non-domestic market
85 Watches, non-domestic market
86 Motor vehicles, non-domestic market
87 Other transport equipment, non-domestic market

Inflation indicators

88 Food products and non-alcoholic beverages
89 Alcoholic beverages and tobacco
90 Wearing apparel and footwear
91 Housing, water, electricity, gas and other fuel
92 Furnishings, household equipment and routine maintenance of the house
93 Health
94 Transport
95 Communication
96 Recreation and culture
97 Education
98 Hotels, cafes and restaurants
99 Miscellaneous goods and services
100 Total

Deflators

101 GDP
102 Total consumption
103 Household consumption
104 Public sector consumption
105 Investment
106 Fixed investment
107 Absorption
108 Domestic demand
109 Exports
110 Imports
111 Foreign trade

Employment

112 Unemployed persons at end-month
113 Unemployment rate (%)
114 Benefit recipients (number)
115 Long-term unemployed (% of total number of unemployed persons)
116 Short-term unemployed (% of total number of unemployed persons)
117 Vacancies (number)
118 Loading coefficient (%)

Other indicators

119 Currency in circulation (in millions of lats)
120 Average weighted interest rates on loans in lats to non-financial corporations (%)
121 Average weighted interest rates on loans in lats to households (%)
122 Births (number)

123 Non-adjusted construction production indices (%; 2000 = 100)
124 Non-adjusted new orders in construction (in thousands of lats)
125 Non-adjusted retail trade turnover indices (%; 2000 = 100)
126 Non-adjusted turnover indices of motor vehicle sales and automotive fuel retailing (%; 2000 = 100)

BIBLIOGRAPHY

1. ARTIS, Michael J., BANERJEE, Anindya, and MARCELLINO, Massimiliano. *Factor Forecasts for the UK*. CEPR Discussion Paper, No. 3119, January 2002.

2. BANERJEE, Anindya, MARCELLINO, Massimiliano, and MASTEN, Igor. *Forecasting Macroeconomic Variables for the New Member States of the European Union*. ECB Working Paper, No. 482, May 2005.

3. BERNANKE, Ben S., BOIVIN, Jean. Monetary Policy in a Data-Rich Environment. *Journal of Monetary Economics*, vol. 50, April 2003, pp. 525–546.

4. BOIVIN, Jean, NG, Serena. *Are More Data Always Better for Factor Analysis?* NBER Working Paper, No. 9829, July 2003.

5. BREITUNG, Jörg, EICKMEIER, Sandra. *Dynamic Factor Models*. Deutsche Bundesbank Discussion Paper Series 1: Economic Studies, No. 38, 2005.

6. CAMBA-MÉNDEZ, Gonzalo, KAPETANIOS, George. *Forecasting Euro Area Inflation Using Dynamic Factor Measures of Underlying Inflation*. ECB Working Paper, No. 402, November 2004.

7. CHAMBERLAIN, Gary, ROTHSCHILD, Michael. Arbitrage, Factor Structure, and Mean- Variance Analysis on Large Asset Markets. *Econometrica*, vol. 51, issue 5, September 1983, pp. 1281–1304.

8. CHEUNG, Calista, DEMERS, Frédérick. *Evaluating Forecasts from Factor Models for Canadian GDP Growth and Core Inflation*. Bank of Canada Working Paper, No. 07-8, 2007.

9. CONNOR, Gregory, KORAJCZYK, Robert A. Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis. *Journal of Financial Economics*, vol. 15, issue 3, March 1986, pp. 373–394.

10. D'AGOSTINO, Antonello, GIANNONE, Domenico. *Comparing Alternative Predictors Based on Large-Panel Factor Models*. Central Bank and Financial Services Authority of Ireland Research Technical Paper, No. 14/RT/06, December 2006.

11. DIEBOLD, Francis X., MARIANO, Roberto S. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, vol. 13, No. 3, July 1995, pp. 253–263.

12. FORNI, Mario, HALLIN, Marc, LIPPI, Marco, and REICHLIN, Lucrezia. The Generalised Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics*, vol. 82, issue 4, November 2000, pp. 540–554.

13. FORNI, Mario, HALLIN, Marc, LIPPI, Marco, and REICHLIN, Lucrezia. The Generalised Dynamic-Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association*, vol. 100, No. 471, September 2005, pp. 830–840.

14. GAVIN, William T., KLIESEN, Kevin L. *Forecasting Inflation and Output: Comparing Data-Rich Models with Simple Rules*. Federal Reserve Bank of St. Louis Working Paper, No. 2006-054A, September 2006.

15. GEWEKE, John F. The Dynamic Factor Analysis of Economic Time Series Models. ***In:*** *Latent Variables in Socio-economic Models*. Edited by Dennis J. Aigner and Arthur Stanely Goldberger. North Holland, 1977, pp. 365–282.

16. GILLITZER, Christian, KEARNS, Jonathan. *Forecasting with Factors: the Accuracy of Timeliness*. Reserve Bank of Australia Research Discussion Paper, No. 2007-03, April 2007.

17. MARCELLINO, Massimiliano, STOCK, James H., and WATSON, Mark W. Macroeconomic Forecasting in the Euro Area: Country Specific Versus Area-Wide Information. *European Economic Review*, vol. 47, issue 1, February 2003, pp. 1–18.

18. MARINI, Marco, ROSSI, Alessandro. FLASH Version 2.1 User's quide. December 2002, 41 p.

19. MATHESON, Troy. *Factor Model Forecasts for New Zealand*. Reserve Bank of New Zealand Discussion Paper, No. DP2005/01, May 2005.

20. NIEUWENHUYZE, Christophe van. *A Generalised Dynamic Factor Model for the Belgian Economy. Useful Business Cycle Indicators and GDP Growth Forecasts*. National Bank of Belgium Working Paper, No. 80, February 2006.

21. QUAH, Danny, SARGENT, Thomas J. A Dynamic Index Model for Large Cross Sections. ***In:*** *Business Cycles, Indicators, and Forecasting*. Edited by James H. Stock and Mark W. Watson. Chicago : NBER and University of Chicago Press, 1993.

22. REIJER, Ard D. *Forecasting Dutch GDP Using Large Scale Factor Models*. De Nederlandsche Bank Working Paper, No. 028, February 2005.

23. SARGENT, Thomas J., SIMS, Christopher A. Business Cycle Modeling Without Pretending to Have Too Much a Priori Economic Theory. ***In:*** *New Methods in Business Cycle Research*. Edited by Christopher A. Sims. Minneapolis : Federal Reserve Bank of Minneapolis, 1977.

24. SCHNEIDER, Martin, SPITZER, Martin. *Forecasting Austrian GDP Using the Generalised Dynamic Factor Model*. Oesterreichische National Bank Working Paper, No. 89, August 2004.

25. SCHUMACHER, Christian, DREGER, Christian. *Estimating Large-Scale Factor Models for Economic Activity in Germany: Do They Outperform Simpler Models?* Jahrbücher für Nationalökonomie und Statistik, No. 224, 2004, pp. 732–750.

26. SCHUMACHER, Christian. *Forecasting German GDP Using Alternative Factor Models Based on Large Datasets*. Deutsche Bundesbank Discussion Paper Series 1: Economic Studies, No. 24, 2005.

27. STOCK, James H., WATSON, Mark W. *Diffusion Indexes*. NBER Working Paper, No. 6702, August 1998.

28. STOCK, James H., WATSON, Mark W. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, vol. 20, No. 2, April 2002, pp. 147–162.

29. Available: *http://www.economia.unimore.it/forni_mario/matlab.htm*.