DISCUSSION PAPERS

DP/86/2011

Short-Term Forecasting of Bulgarian GDP Using a Generalized Dynamic Factor Model

Petra Rogleva

BULGARIAN NATIONAL BANK

DISCUSSION PAPERS

DP/86/2011

BULGARIAN
NATIONAL
BANK

Short-Term Forecasting of Bulgarian GDP Using a Generalized Dynamic Factor Model

Petra Rogleva

October 2011

ISBN: 978–954–8579–44–5

Printed in the BNB Printing Centre.

Views expressed in materials are those of the authors and do not necessarily reflect BNB policy.

Elements of the 1999 banknote with a nominal value of 50 levs are used in cover design.

Send your comments and opinions to:
Publications Division
Bulgarian National Bank
1, Knyaz Alexander I Square
1000 Sofia, Bulgaria
Tel.: (+359 2) 9145 1351, 9145 1978
Fax: (+359 2) 980 2425
e-mail: Dimova.L@bnbank.org
Website: www.bnb.bg

Contents

SUMMARY: In the present paper we utilize the Generalized Dynamic Factor Model proposed by Forni *et al.* [2002] to construct a short-term forecast of the real GDP in Bulgaria using a large data set of 140 indicators from all sectors of the economy. A simple method is described for selecting a smaller subset of data series that provides better forecasts. Even in this case the number of variables used for the forecasts is still much higher than that in the clasical multivariate time series models. The latest available observations of the series with smaller publication delay are also exploited to capture the most recent developments in the economy.

Petra Rogleva, Bulgarian National Bank, Economic Research and Forecasting Directorate, Rogleva.P@bnbank.org

1 Introduction

The assessment of the current economic situation together with forming an expectation for the next few quarters ahead is an important input in the creation of medium to long-term forecasts and in the decision making process as a whole. Whereas mainly structural models are exploited for medium to long-term forecasts, researchers and practitioners put more efforts in extracting and summarizing information from the great variety of sources available nowadays for the short-run. Both nowcasting and short-term forecasting require focusing on time series data that can provide information on the current state of the economy and the number of potentially useful series can be very large. Dynamic factor models have been successfully applied in a number of papers to forecasting US, euro area and country specific macroeconomic variables, including Stock and Watson (2002), Marcellino *et al.* (2003), Schneider and Spitzer (2004) and Artis *et al.* (2007). The main characteristic of factor models is that they can exploit information from large data sets, while in classical time series models the number of parameters increases quickly with the number of variables modeled, so that usually only univariate or small-scale multivariate models are considered.

One of the specifics of modeling Bulgarian macroeconomic variables is that the number of observations is very small because Bulgaria's transition to market economy started in the beginning of 90's. The short history hinders the use of traditional parametric econometric models and can give additional advantage to dynamic factor models.

The dynamic factor approach is not new for the Bulgarian economy but is still relatively little explored. In Vassileva (2006) a dynamic factor model is used to produce two indicators for the Bulgarian economy. The first one is developed to forecast the nominal GDP and the second – to summarize the current tendencies in the economy on monthly basis. The approach in the present paper differs from the one in Vassileva (2006) in that it is focused on nowcasting and short-term forecasting of the real GDP and in the methodology for selection of variables and model specification.

The basic idea underlying dynamic factor analysis (as in Sargent and Sims (1977) and Geweke (1977)) is that the comovements in a large set of time series can be explained by a small number of unobserved variables, called common factors or common shocks. The models assume that each time series in the panel can be expressed as the sum of two orthogonal components. The first one is a common component, which should explain the main part of the variance of the time series and is driven by the common factors. The second one is a variable specific idiosyncratic component.

In the present paper, we use the Generalized Dynamic Factor Model (GDFM) of Forni *et al.* (2002) to produce a short-term forecast of real Bulgarian GDP from one to four quarters ahead. As the first estimate of the GDP for the last quarter is available almost at the end of the next one, the one-step ahead forecast of the output is practically equivalent to a nowcast — a forecast of the present. Even if the flash estimate of GDP were used one still will have to wait until the middle of the next quarter when all short-term business statistics, surveys and financial data for the last quarter are already published. We have collected 140 real-activity, survey and financial series with monthly and quarterly frequencies. The latest observations of series with a smaller publication delay which are available for the current quarter are used to produce a nowcast of GDP. The latest available information is expected to be beneficial to the forecasts for further periods as well. To deal with a dataset with end-of-sample unbalancedness, *i.e.* when the latest observations for some of the variables in the data set are missing, we use an approach similar to the one used by Altissimo *et al.* (2001).

Forni *et al.* (2002) and Forni and Lippi (2001) prove that their method gives a forecast that is a consistent estimator of the optimal h-step ahead forecast as both the size of cross-section and the series length go to infinity. However, there are some results which indicate that increasing the number of variables over a certain size does not improve, or even worsens forecasting results (Schneider and Spitzer (2004), Bai and Ng (2002)). In this paper we use a simple method for selecting a subset of all the variables collected. The criterion for inclusion of a variable is based on the relative performance of the model in a pseudo-real time forecasting exercise with and without the given variable. We achieve better forecasting performance with a data set smaller than the full data set of 140 variables.

The paper is organized as follows. The next two sections briefly describe the GDFM and our data set. Section Four explains the empirical application of the model. Section Five describes how the latest available observations are exploited. Section Six concludes and outlines the directions for future work. Appendices A and B give technical details and a complete list of variables.

2 The Generalized Dynamic Factor Model

Dynamic factor analysis is based on the idea that the comovements in a large set of related time series can be explained by a small number of common factors and each time series can be decomposed as a sum of two orthogonal components. The common factors of the dataset drive the common component of each series and explain the major part of the variance in the series. The second component is a variable specific idiosyncratic term. In traditional factor analysis (Sargent and Sims (1977) and Geweke (1977)), referred to also as exact factor models, it is assumed that there is no cross-correlation among the idiosyncratic components at any lead and lag. This assumption allows for identification of the models, but is unrealistic for most applications. As an example how idiosyncratic terms can interact with each other, Schneider and Spitzer (2004) describe the following situation. There are two industries and each of their productions consists of common and idiosyncratic components. If both of the sectors are connected with an input-output relationship then an idiosyncratic shock in one of them can affect the idiosyncratic term in the output of the other industry and this can happen with a lag. Thus, the assumption for uncorrelated idiosyncratic components appears unsuitable in our context.

Forni *et al.* (2000a) provide a generalization to these models which allows for some contemporaneous and lagged correlation between the idiosyncratic terms. Stock and Watson (2002) also propose a model allowing for mutual correlation in the idiosyncratic components. One of the advantages of the latter is its simpler estimator of the factors and that it requires only one parameter to be set — the number of the factors. Thus, the chance for the model to be misspecified decreases. However, the former model allows for richer dynamics when the factors are extracted.

This is a very desirable feature of the model and is exploited here not only with a view to capturing the lagged relationship between the variables but also for treatment of the end-of-sample unbalancedness, *i.e.* when the latest observations for some of the variables in the data set are missing.

The model utilized to develop a framework for short-term forecasting of real GDP in Bulgaria is the Generalized Dynamic Factor Model of Forni *et al.* (2002). This model elaborates the work of Forni *et al.* (2000a) and Forni and Lippi (2001), where the GDFM is introduced. Their estimator of the components is based on a two-sided filtering of the observable variables and exploits past, present and future observations. This feature causes problems for the estimation at the end of the sample and is highly undesirable in forecasting. Forni *et al.* (2002) exploit the dynamic techniques developed in Forni *et al.* (2000a) to obtain an estimate of the cross-covariance matrices of common and idiosyncratic components at all leads and lags and then, under some additional restrictions for the structure of the model, they obtain a consistent estimator (for both cross-section and time dimensions going to infinity) of the optimal linear predictor of the common component. This estimator exploits only past and present observations. In Section 2.1 we present the model with some of its main features and in Section 2.2 we briefly describe the estimation procedure.

2.1 The Model

Let $X_n^T = \{x_{it}, i = 1, \ldots, n, t = 1, \ldots, T\}$ denote the observations in a data set of n time series at periods of time $t = 1, \ldots, T$. We assume that X_n^T is a finite realization of a real-valued stochastic process $X = \{x_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$, where for any $n \in \mathbb{N}$, the n-dimensional vector process $\{\mathbf{x}_n = (x_{1t}, \ldots, x_{nt})', t \in \mathbb{Z}\}, n \in \mathbb{N}$ is weakly-stationary, with mean $\mathbf{0}_n$ and finite second order moments $\Gamma_{nk} = \mathbf{E}[\mathbf{x}_{nt}, \mathbf{x}'_{n,t-k}]$, for $k \in \mathbb{N}$. In practice this assumption cannot be verified for a particular dataset. The common way to try to fulfill it is by transforming the data so that the trend and the seasonality be removed.

The GDFM suggests that each variable $x_{it}, i = 1, \ldots, n$ can be represented as the sum of two, mutually orthogonal, unobserved components: the common component, χ_{it}, and the idiosyncratic component, ξ_{it}. The common component is driven by q-dimensional vector of common factors (q is typically much smaller than n). These factors are the same for all variables in the panel but loaded with different coefficients and lag structure.

More precisely,

$$x_{it} = \chi_{it} + \xi_{it} \tag{1}$$

$$= b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \ldots + b_{iq}(L)u_{qt} + \xi_{it}$$

$$= \mathbf{b}_i(L)\mathbf{u}_t + \xi_{it},$$

where L is the lag operator, and $\mathbf{u}_t = \{(u_{1t}, \ldots, u_{qt})', t \in$

DISCUSSION PAPERS

$\mathbb{Z}\}$ is the vector of common factors and is orthonormal white noise, and the processes $\{\mathbf{u}_t, t \in \mathbb{Z}\}$ and $\{\xi_{it}, i = 1, \ldots, n, t \in \mathbb{Z}\}$ are mutually orthogonal, hence the common and the idiosyncratic components are mutually orthogonal; and $b_{ij}(L) = \sum_{k=0}^{\infty} b_{ij,k} L^k$, $i = 1, \ldots, n$, $j = 1, \ldots, q$ are $n \times q$ square-summable filters, called factor loadings.

The eigenvalues of the spectral density matrix of the variables show what share of the total variation in the data set can be explained by a particular factor. If $\lambda_{nk}(\theta)$ is the k-th largest eigenvalue at frequency θ of the spectral matrix $\Sigma_n(\theta)$ of $\{\mathbf{x}_{nt}, t \in \mathbb{Z}\}$, then

$$\frac{\lambda_{ni}(\theta)}{\sum_{j=1}^{n} \lambda_{nj}(\theta)}$$

gives the contribution of the i-th common factor to the total variance at frequency θ. Forni and Lippi (2001) prove that there exists a q-dimensional white noise process $\{\mathbf{u_t}, t \in \mathbb{Z}\}$, satisfying the assumptions of the model if and only if it holds that

> $\lambda_{nq}(\theta) \to \infty$, as $n \to \infty, \theta$-a.e. in $[-\pi, \pi]$ and there exists $\Lambda \in \mathbb{R}$ such that $\lambda_{n,q+1}(\theta) \leq \Lambda$, for any $\theta \in [-\pi, \pi]$, and any $n \in \mathbb{N}$.

With some oversimplification this condition ensures that: 1) each common factor affects infinitely many items of the cross-section as n increases; and 2) a limited amount of cross-correlation among the various idiosyncratic components is allowed. So, this condition relieves the condition for orthogonality of the idiosyncratic components in the exact dynamic factor model.

The same condition also suggests a criterion for choice of the number q of the common factors. If we set

$$\lambda_{in} = \int_{\theta \in [-\pi,\pi]} \lambda_{in}(\theta) d\theta, \qquad i = 1, \ldots, n,$$

to be the aggregated eigenvalues over all frequencies, then for large n, we expect that the first q aggregated eigenvalues are sizably bigger then the $(q+1)$-th one, *i.e.* there is a substantial gap between λ_{qn} and $\lambda_{q+1,n}$.

Forni *et al.* (2000a) prove that under certain assumptions, the model (1) is identified in the sense that the components χ_{it} and ξ_{it} are uniquely determined. They also construct a consistent finite-sample estimator of the components but using a two-sided filter. The modification proposed in Forni *et al.* (2002) makes the model more appropriate for end-of-sample estimation and forecasting, and requires a more restrictive structure of the factor loadings. It is assumed that the common components $\boldsymbol{\chi}_{nt}$ follow a $VARMA(S, s)$ structure of the form

$$\boldsymbol{\chi}_{nt} = \mathbf{B}_n(L)[\mathbf{A}(L)]^{-1}\mathbf{u}_t,$$

where $\mathbf{B}_n(L) = B_0^n + B_1^n L + \ldots + B_s^n L^s$ is a $n \times q$ matrix polynomial [1] in the lag operator L of order s, with $B_s^m \neq \mathbf{0}$, for some m; and $\mathbf{A}(L) = \mathbf{I} - A_1 L - \ldots - A_S L^S$ is a $q \times q$ polynomial of order S and $S \leq s+1$.

For fixed k, B_k^n are nested for $n \in \mathbb{N}$, so if we add more series to $\mathbf{x}_{nt}$ and the number of series n increases, the factor loadings of the old series won't be changed. This together with $B_s^m \neq \mathbf{0}$, for some m implies that $B_s^n \neq \mathbf{0}$, for $n \geq m$ and guarantees that at least one of the factors is loaded with its s-th lag at some of the common components.

Actually the dynamic model (1) can be rewritten in a static form: for

$$\mathbf{f}_t := (f_{1t}, \ldots, f_{qt})' = [\mathbf{A}(L)]^{-1}\mathbf{u}_t$$

and $r := q(s+1)$, $\mathbf{F}_t = (\mathbf{f}_t', \ldots, \mathbf{f}_{t-s}')'$ is a $r \times 1$ vector of, as they are referred to later, static factors. The model for $\mathbf{x}_{nt}$ becomes

$$\mathbf{x}_{nt} = \mathbf{B}_n(L)\mathbf{f}_t + \boldsymbol{\xi}_{nt} = B_0^n \mathbf{f}_t' + \ldots + B_s^n \mathbf{f}_{t-s}' + \boldsymbol{\xi}_{nt} = \mathbf{C}_n \mathbf{F}_t + \boldsymbol{\xi}_{nt},$$

with $\mathbf{C}_n = (B_0^n B_1^n \ldots B_s^n)$ - $n \times r$ matrix of coefficients.

2.2 The Estimation of Components

The procedure for estimation of common and idiosyncratic components of the variables consists of two main

[1] Each element of $\mathbf{B}_n(L)$ is a scalar polynomial in L and B_k^n is $n \times q$ matrix of coefficients premultiplying the variables in lag k.

parts[2]. Firstly, using the dynamic techniques in Forni *et al.* (2000a), estimates $\Sigma_n^{\chi T}(\theta)$ and $\Sigma_n^{\xi T}(\theta)$ of the spectral density matrices of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$ are constructed. From these spectral matrice estimators, the matrices $\Gamma_{nk}^{\chi T}$ and $\Gamma_{nk}^{\xi T}$ of the common and idiosyncratic auto-covariance matrices for all lags k are obtained by applying an inverse Fourier transformation. The estimates of $\Gamma_{nk}^{\xi T}$ have a non-zero off-diagonal elements, as the different idiosyncratic terms are allowed to be correlated in all lags k. However Forni *et al.* (2002) gives some empirical evidence that forcing to zero the off-diagonal entries of the estimated $\Gamma_{n0}^{\xi T}$ improves forecasting performance when the number of the series n is large with respect to the number of observations T. The explanation provided for their results is that the computation of $\Gamma_{n0}^{\xi T}$ brings some spurious large covariances, even when the true covariance is zero. They argue that when n increases, the error obtained estimating $\Gamma_{n0}^{\xi T}$ increases much faster than the error of ignoring the true covariances and setting them to zero. Moreover, they show that replacing $\Gamma_{n0}^{\xi T}$ with any symmetric positive semi-definite matrix with bounded eigenvalues does not affect the consistency obtained. In our application of the model we also set to zero the off-diagonal entries of $\Gamma_{n0}^{\xi T}$ before computing the eigenvectors.

Secondly, the r-dimensional unobserved space spanned by the static factors is estimated by the space spanned by linear combinations of the observations. These lin-

[2] For further details see Appendix A.

ear combinations $\mathbf{W}_{nt}^{T} = (W_{nt}^{1T}, \ldots W_{nt}^{rT})$ are such that the variance in their common part is the biggest for a given idiosyncratic variance. The weights for these combinations are achieved as a solution of the generalized eigenvalue problem of $\Gamma_{n0}^{\chi T}$ and $\Gamma_{n0}^{\xi T}$. Thus, time series are weighted according to their signal-to-noise ratios, *i.e.* the ratio between the variance in their common and idiosyncratic terms. This gives additional advantages of the model in a situation when there are sizable differences in the common-to-idiosyncratic variance ratios, which we take to be the case. For example, we expect the idiosyncratic variances in survey data to be much higher than in structural business statistics because of the source of data. The same holds for the index of industrial production in 'Mining and quarrying' compared to 'Manufacturing' because of the sizes of the sectors. However, the survey data is expected to contribute to forecast accuracy because it is designed to be leading for the rest of the indicators and this justifies its use.

The space spanned by $\mathbf{W}_{nt}^{T}$ is a consistent estimator of the space spanned by the static factors, as $n, T \to \infty$. Hence $\mathbf{W}_{nt}^{T}$ can consistently approximate any point in the common-factor space, including the common components $\boldsymbol{\chi}_{nt}$, as $n, T \to \infty$.

Since the idiosyncratic components are orthogonal to the common components, $\boldsymbol{\chi}_{nt}$ can be approximated by projecting the data X_{n}^{T} on $\mathbf{W}_{nt}^{T}$. Similarly, forecasts of the common components (k-leads ahead) are obtained by projecting $\boldsymbol{\chi}_{n,t+k}$ on $\mathbf{W}_{nt}^T$.

Forni *et al.* (2002) argue that, under certain conditions, the frequency domain approach by Forni *et al.* (2000a), which they use for their predictor, can provide a substantial improvement over the predictor proposed in Stock and Watson (2002) based on static principal components. The improvement can be expected when the various cross-sectional items differ significantly in the lag structure of the loadings. The methodology, which we use to deal with a dataset with end-of-sample unbalancedness (described in Section 5) results in applying the model to a dataset with a significant difference in the timing of the variables.

It should be stressed that $\mathbf{W}_{nt}^T$ is a consistent estimator for some basis of the space spanned by the factors, not for the factors themselves. The identification of the common factors and the coefficients of the filter $\mathbf{B}_n(L)$ calls for further informational and economic assumptions and is thoroughly discussed in Forni *et al.* (2007).

3 The Data

From a theoretical point of view, all available relevant variables should in principle be included in the data set as the GDFM is proved to give a forecast that is a consistent estimator of the optimal h-step ahead forecast as both the size of cross-section and the series length go to infinity. However, as explained in Forni *et al.* (2000b), in practice it is not recommended to include variables which have small commonality and large idiosyncratic components, since the latter could survive aggregation and be wrongly interpreted as additional common factors. Schneider and Spitzer (2004) and Bai and Ng (2002) give evidence that forecasts based on a small dataset can outperform such based on a bigger one.

To find the balance between the large data set that can be beneficial for the forecast and the inclusion of variables which can worsen the results we employ a two-fold strategy. First, we collect a comprehensive data set. It combines more reliable variables like national account data and price indices with indicators with possibly bigger measurement errors and smaller common-to-idiosyncratic ratio. These could be variables from business surveys and production indices and they usually have smaller publication delay and give earlier indications on the latest economic developments. Second, we apply a criterion for selecting a smaller or optimal subset of the full dataset. We base our criteria on the relative performance of the model in a pseudo-real time forecasting exercise, described in Section 4.3.

Thus, our data set for Bulgaria comprises 140 monthly and quarterly series. We include national accounts data (real and nominal GDP and its components), labour market variables, industrial production and industrial and trade turnover, price indices (CPI, HICP and raw material prices), survey data (consumer and business surveys), external trade data, monetary aggregates and interest rates, among others. To make the frequencies of the series into the data set equal, we choose to aggregate the monthly series to a quarterly frequency. In order to achieve stationary series the following transformations have been made. A fourth-seasonal difference was applied to series with pronounced seasonality and then for all series a first difference was taken. The initial range of the data set is from 2000 Q1 to 2009 Q2 and after this treatment the sample includes 33 observations. Finally, the series were taken in deviation from the mean and divided by the standard deviation to obtain results independent of the units of measurement. A complete list of the variables and their transformations is reported in Appendix B.

4 Forecast Evaluation

4.1 Overview

In this section we describe the application and the performance of the model presented in Section 2 for forecasting the Bulgarian real GDP. We also test the relative performance of the model with full data set of 140 variables and the selected smaller subset in our pseudo-real time forecasting exercise.

Let $\tilde{Y}$ be GDP at constant prices from 2000 Q1 to 2009 Q2. After the transformations described in Section 3 we obtain seasonally adjusted series in first differences

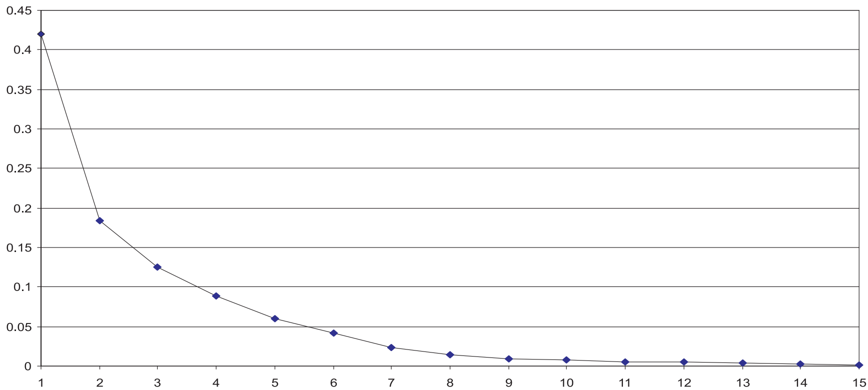
$Y = (1-L)(1-L^4)\tilde{Y}$ and normalized series $y = \frac{Y-\hat{\mu}}{\hat{\sigma}}$, where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Y_t$, $\hat{\sigma} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(Y_t - \hat{\mu})^2}$ and T denotes the sample size. The series y is the variable we are modeling together with the other normalized series.

Since the common and the idiosyncratic components, say χ_y and ξ_y, in the decomposition of the transformed GDP y are orthogonal at any lead and lag, they can be forecast separately. In view of the orthogonality or only weak cross-correlation between the idiosyncratic components of distinct cross-sectional items, Forni *et al.* (2002) propose that the forecast of ξ_y be based on standard univariate or low-dimensional multivariate methods. However, the results from the pseudo out-of-sample exercise in D'Agostino and Giannone (2006) suggest that the idiosyncratic components are actually highly unforecastable. The authors conclude that *the common factors, constructed to explain the maximum amount of cross-sectional variance of the panel, are also able to capture all the predictable dynamics of the key aggregated variables.* We choose to ignore the idiosyncratic terms and our forecast of y is obtained as a forecast of the common component χ_y.

In addition, the parameters specifying the model have to be chosen. These parameters are the number of common factors q and the order s of the factor loading $\mathbf{B}_n(L)$. One possible test for choice of the number of factors is based on the aggregated eigenvalues in decreasing order $\bar{\lambda}_{jn}$ of the spectral density matrix of the panel[3]. It sug-

[3] $\bar{\lambda}_{jn}$ is the sum of the j-th eigenvalues of the spectral density

Figure 1: Share of the total variance explained by first 15 common factors



gests that q should be chosen so that the gap between $\bar{\lambda}_{qn}$ and $\bar{\lambda}_{q+1,n}$ is more substantial than the differences between the other consecutive eigenvalues. Figure 1 displays the aggregated eigenvalues $\bar{\lambda}_{jn}$. We cannot find a substantial gap between any two consecutive eigenvalues except perhaps for the first and the second. Forni *et al.* (2000a) remark that setting the number q bigger than its true value cannot have dramatic effects on estimation. As Table 1 shows, the first three common factors explain more than 70 per cent of the total variability and the fourth common factor would add only 9 percentage points to the total variance explained. Furthermore, this suggests that if the model is estimated with, for example, 3 common factors,

matrices calculated in 13 frequencies, equally spaced in the interval $[0, 2\pi]$ (see Appendix A).

then one can expect the variance of χ_y to be around 70 per cent of the variance of y. So a forecast for Y to be derived, the predicted value for χ_y should be scaled with a coefficient ρ bigger than the standard deviation of Y, $\hat{\sigma}$. In our application the model is specified by the parameters q, s and ρ and they are also determined according to the out-of-sample forecasting performance of the models.

Table 1: Variance explained by first 15 common factors

Common factor	Share	Cumulated variance
1	0.42	0.42
2	0.18	0.60
3	0.12	0.73
4	0.09	0.82
5	0.06	0.88
6	0.04	0.92
7	0.02	0.94
8	0.01	0.96
9	0.01	0.97
10	0.01	0.97
11	0.01	0.98
12	0.00	0.98
13	0.00	0.99
14	0.00	0.99
15	0.00	0.99

4.2 Forecasting Exercise

We evaluated the forecasting performance of a number of models which differ in their specifying parameters —

the number of dynamic common factors q, the maximal order of factor loadings s and the scaling parameter ρ. We consider four forecasting horizons h from one to four steps ahead and for each of them we choose the optimal set of parameters (q_h^*, s_h^*, ρ_h^*). We set the parameter q to vary in the range from 1 to 4. That means that up to 79 per cent of the total variability (for the full data set) is explained by the common factors. For s we test values from 1 to 4. The upper bound of 4 corresponds to the assumption that none of the common factors is leading with more than one year for any of the observable economic indicators. We search the value for our scaling parameter ρ in a wide grid of values R around an approximate value[4] of the ratio of the standard deviation of Y and the common component of y.

For each different combination of variables we estimate the forecasting performance in 17 rolling windows. First, we estimate the model specified by given values of q and s using observations from 2001 Q1 to 2005 Q1 (2005 Q1 is nearly the middle of the sample) and compute $h = 1, 2, 3, 4$-step ahead forecasts $y_{\tau+h|\tau}$ at time τ=2005 Q1 of $y_{\tau+h}$. We then reestimate the model using one additional observation and compute $y_{\tau+h|\tau}$ for τ=2005 Q2 and $h = 1, 2, 3, 4$. This is repeated 17 times until all available observations are added, from 2001 Q1 to 2009 Q2. As a result we achieve one vector of out-of-sample forecasts $y_{\tau+h|\tau}$

[4] We take R as a grid of values around $\hat{\rho} = \sqrt{\frac{\sum_{j=1}^{n} \lambda_j}{\sum_{j=1}^{q} \lambda_j}}\hat{\sigma}$.

for each combination of forecasting horizon h, number of factors q and maximal lag of the factor loadings s.

Furthermore, we add back the variance and the mean to the forecast $y_{\tau+h|\tau}$, $\tau = 2005Q1, \ldots, 2009Q2$ to produce a forecast for our non-normalized variable $Y_{\tau+h}$:

$$Y_{\tau+h|\tau} = y_{\tau+h|\tau}\rho + \hat{\mu}.$$

The forecast $Y_{\tau+h|\tau}$ already depends on the choice of the scaling parameter ρ. The forecast accuracy of the model is measured by $RMSE(h, q, s, \rho)$ which is the root mean squared errors between the observed transformed GDP $Y_{\tau+h}$ and its forecast $Y_{\tau+h|\tau}$:

$$RMSE(h, q, s, \rho) = \frac{1}{17-h}\sqrt{\sum_{\tau=2005:Q1}^{2009:Q2}(Y_{\tau+h} - Y_{\tau+h|\tau})^2}.$$

For a given data set and forecasting horizon we choose the model with minimal RMSE. It is specified by the parameters (q_h^*, s_h^*, ρ_h^*) which are such that:

$$RMSE(h, q_h^*, s_h^*, \rho_h^*) =$$

$$min\Big\{RMSE(h, q, s, \rho) : q, s = \{1, 2, 3, 4\}, \rho \in R\Big\},$$
$$\text{for } h = 1, 2, 3, 4.$$

The measure $RMSE_h^* := RMSE(h, q_h^*, s_h^*, \rho_h^*)$ is used to assess the forecasting performance for a given data set.

4.3 Selecting an Optimal Subset of the Data Set

The criterion which we use to compare the different subsets of the data set is their corresponding $RMSE^*_h$ defined in the previous subsection. For each forecasting horizon, we would idealy be able to compare $RMSE^*_h$ for all possible subsets of the full data set of 140 variables and to choose the one with the smallest $RMSE^*_h$. Unfortunately, this calls for too many calculations and is not feasible. Instead we proceed as follows.

In the first step we sort the data set according to the contributions of the variables to the performance of the model. Let $X_n = (y, x_1, \ldots, x_n)$ be our complete data set, so $n = 139$, and let $X^j_n = (y, x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)$ denote the set without the j-th variable. We estimate $RMSE^*_h$ for X^j_n, say $RMSE^j_h$, for $j = 1, \ldots, n$ and sort the variables $x_j, j = 1, \ldots, n$ in decreasing order according to $RMSE^j_h$. Thus we use the performance of the set X^j_n as a measure of the importance of the variable x_j for our forecast. Obviously the bigger $RMSE^j_h$ is, the more important x_j is.

Once we have sorted the variables in the set, $X^s_{nh} = (y, x^h_{s1}, \ldots, x^h_{sn})$, in the second step we start to exclude variables from the end of the sorted data set one by one. We calculate $RMSE^*_h$, say $RMSE^\nu_h$, for each subset $X^s_{nh\nu} = (y, x^h_{s1}, \ldots, x^h_{s\nu})$, $\nu = 20, \ldots, n$. We choose ν to vary from 20 to n, because there are some limitations for the minimal number of variables in the set[5]. Finally our optimal subset we choose is the set $X^s_{nh\nu}$ with minimal $RMSE^\nu_h$, denoted, respectively, by X^{opt}_{nh} and $RMSE^{opt}_h$.

We repeat this procedure for every forecasting horizon $h = 1, 2, 3, 4$. In Table 2 we report, for each h, the number of variables in the optimal subset X^{opt}_{nh}, the optimal set of parameters and $RMSE^{opt}_h$, as well as $RMSE^*_h$ for the full data set. In this table we also show the relative improvement of the optimal subset on the full data set.

As Table 2 shows, the size of the optimal subsets varies widely among the horizons from 31 variables for forecasts one step ahead to almost the full data set of 138 variables for two steps ahead forecasts. However, a significant improvement in the performance is seen only in the one step ahead forecast where the $RMSE^{opt}$ is 33 per cent less than the $RMSE^*$ for the full data set. Although the $RMSE^{opt}$ for the other horizons is still smaller than $RMSE^*$, the improvement there does not seem significant.

The optimal subsets for the different forecasting horizons are indicated in Appendix B. The conclusion that can be made is that the subsets are quite different in respect to the included variables and a clear pattern for including or excluding of variables cannot be drawn. However, it should be noted that for all horizons there are representatives of the main groups of indicators like labour market, price indices, production and turnover indices, external

[5]For example if we want to estimate $r = q(s + 1)$ static factors we need at least r variables in the set.

Table 2: Forecasting performance, using a balanced data set

	h = 1	h = 2	h = 3	h = 4
Number of variables	31	138	62	86
Parameters				
q^*	2	2	1	2
s^*	2	4	1	4
ρ^*	$2.6\hat{\sigma}$	$3.6\hat{\sigma}$	$7.4\hat{\sigma}$	$2.3\hat{\sigma}$
$RMSE^{opt}$				
in levels	151 986	208 230	254 894	258 075
as a part of $RMSE^*$-full data set	0.67	0.99	0.91	0.96
$RMSE^*-$full data set				
in levels	225 909	210 818	279 072	269 587

trade and consumer and business surveys from the four different sectors (construction, industry, retail trade and services). This provides one justification of the use of factor models and shows that potentially useful information is contained in a very wide range of data.

Figure 2 and Figure 3 show correspondingly the normalized and non-normalized output (y_τ and Y_τ) and their best predictions $y_{\tau+h|\tau}$ and $Y_{\tau+h|\tau}$, for $h = 1, 2, 3, 4$. As it can be expected the fit of the series is better for the forecasts with shorter horizons. However, the graphs show that the direction of the changes in the predicted series is correctly matched in many of the data points even in the longer forecasting horizons.

In order to obtain forecast $\tilde{Y}_{\tau+h|\tau}$ for the real GDP $\tilde{Y}_{\tau+h}$, we use the predictions for $Y_{\tau+1}, \ldots, Y_{\tau+h}$:

$$\tilde{Y}_{\tau+h|\tau} = \tilde{Y}_\tau - \tilde{Y}_{\tau-4} + Y_{\tau+1|\tau} + \ldots + Y_{\tau+h|\tau} + \tilde{Y}_{\tau+h-4}.$$

Thus, the forecast for the level of GDP at constant prices in the second quarter of 2010, for instance, depends not only on the four-steps ahead forecast for the transformed GDP $Y_{\tau+4|\tau}$ in 2010 Q2 and the last observation of GDP in 2009 Q2 but also on the forecasts three-steps ahead for 2009 Q3, two-steps ahead for 2009 Q4 and one-step ahead for 2010 Q1. As a consequence the results for the longer horizons worsen not only because of the smaller predictive power of the model at these horizons but also because of the accumulation of errors of the forecasts for the previous quarters. The results are shown in Figure 4. As a next step growth rates could be calculated. However, again for the longer horizons the accumulated error worsens the results. A possible way to sidestep this problem is to model directly the growth rates or the first differences of the logarithms of the variables. Although additional work on this topic would be worthwhile, the preliminary results show much smaller commonality among the series in the data set when such transformations of the series are done.

Figure 2: Transformed and normalized real output and its best predictions 1, 2, 3 and 4 steps ahead, using a balanced data set

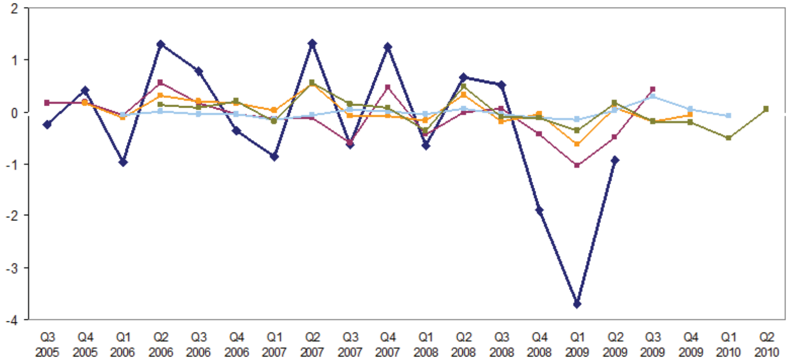


Figure 3: Transformed real output and its best predictions 1, 2, 3 and 4 steps ahead, using a balanced data set

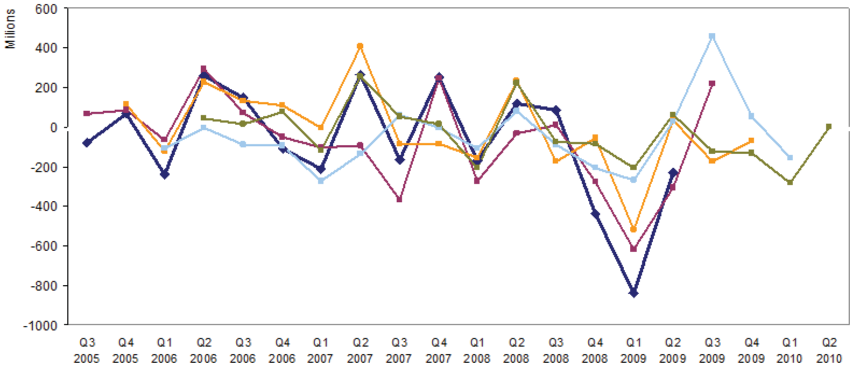
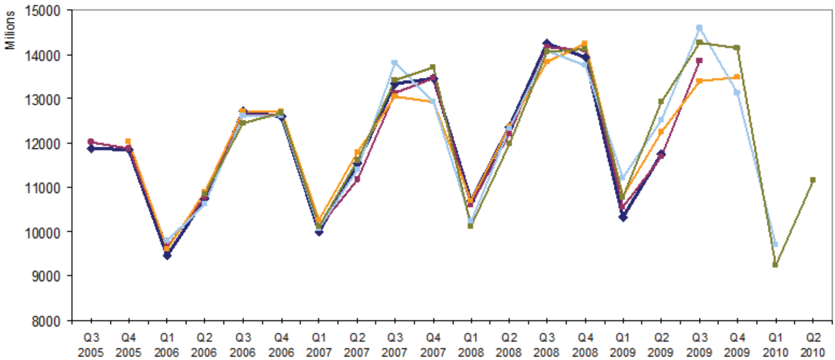


Figure 4: Real output and its best predictions 1, 2, 3 and 4 steps ahead, using a balanced data set



5 Using the Last Available Observations

The method described in Section 2 and applied in Section 4 has the feature that if at time T we want to produce a forecast for time $T + h$, then we need observations for all variables until time T, *i.e.* we need a balanced data set. In the context of forecasting GDP this means that one-quarter ahead prediction for $T + 1$ can be made no sooner than 6 weeks after the end of the period T, when the flash estimate for Bulgarian GDP is published for T, or 10 weeks if we want to use the first estimate of GDP for period T. So, because of the publication delay, our forecast for $T + 1$ is made in the middle or almost at the end of period $T + 1$. This, in combination with the need for a balanced data set, creates the disadvantage that information published during the first 6 (or 10) weeks after the end of period T cannot be exploited. Since almost all of the variables have smaller publication delay (the surveys are published at the end of the same month, inflation is available half month later, *etc.*) the effects of this delay can be very substantial.

Altissimo *et al.* (2001) propose a method for treatment of an unbalanced dataset. In this section we exploit a similar idea for a dataset with observations available up to 2 weeks after the first release of the GDP for 2009 Q2. The method exploits the property of the GDFM to take into account the leading and lagging relations across series by means of principal components in the frequency domain. We shift series with available observations for the next period in the data matrix backward so that the latest available observation for each series can now be obtained at time T. For example, if a given variable is available up to period $T+1$, it will be shifted back one period so that the last observation to be for period T and the first observation will be deleted. So, instead of the variable itself we will use its lead. For variables with monthly frequency which are available for only one or two months of the next quarter this transposition is done before their aggregation to quarterly frequency. The GDFM is designed to be able to capture correctly the time structure and relations in the data set reconstructed this way. The rest of the procedure is the same as the one described in Section 4. It should be noted that all the results are dependent on the cut off date for the data, which determines exactly which additional observations are included in the data set.

The one step ahead forecast in this case is a prediction for the GDP for the current quarter taking in account all available new information and is practically equivalent to a nowcast. Table 3 reports the number of the variables in the optimal subsets at the different horizons and the optimal model specifying parameters. Surprisingly the maximal lag of the factor loadings s for forecasts one and two steps ahead is smaller for the unbalanced data set, which includes the last available observations, than for the balanced one. One possible explanation could be that some

Table 3: Forecasting performance, using the last available observations

	h = 1	h = 2	h = 3	h = 4
Number of variables	107	24	56	62
Parameters				
q^*	4	3	2	1
s^*	1	1	2	4
ρ^*	$2.8\hat{\sigma}$	$3.2\hat{\sigma}$	$6\hat{\sigma}$	$4.2\hat{\sigma}$
$RMSE^{opt}$				
in levels	112 588	170 256	194 158	198 068
as a part of $RMSE^*$ for full data set	0.72	0.68	0.72	0.72
as a part of $RMSE^{opt}$ for balanced data set	0.74	0.82	0.76	0.77
as a part of $RMSE^*$ for full balanced data set	0.50	0.81	0.70	0.73
$RMSE^*$ - full data set				
in levels	157 298	249 302	268 663	275 614
as a part of $RMSE^*$ for full balanced data set	0.70	1.18	0.96	1.02

of the variables which have been shifted backwards in the reconstruction of the data set are in fact leading for GDP and this increases the commonality between variables in 0 or ± 1 lag.

Considering the results of the model with the full data set, the inclusion of the last available observations reduces the root mean squared error significantly but only for one-step ahead forecast. The improvement for this horizon is 30 per cent (measured as $RMSE^*$ – full data set as a part of $RMSE^*$ for full balanced dataset), while for the other horizons the performance remains broadly the same. According to the selection of variables the results seem significantly better than the results from the balanced data set. The difference is that now improvement (measured as $RMSE^{opt}$ as a part of $RMSE^*$ for the full data set) is observed not only for the first horizon but for all of them and is around 30 per cent.

When selection of variables is done for both data sets, the one with included last available observations surpasses the one with balanced data set in all forecasting horizons. The root mean squared error improvements vary from 18 per cent for two-steps ahead forecast to 26 per cent for one-step ahead forecast. An interesting finding is that while for a forecast one-step ahead, the method for inclusion of the last available observations contributes for a better performance with or without selection of an optimal subset, for longer forecasting horizons an improvement is not observed for the full data set but only after selection of variables.

Predictions are shown in Figures 5, 6 and 7. A comparison between these figures and the analogous figures for the predictions with balanced data set (Figures 2, 3 and 4) confirms the conclusion for improvement in the accuracy of the forecasts.

As a whole both methods applied here — first for selection of variables, and second for dealing with unbalanced data set which includes the last available information, bring an accumulated reduction of the root mean squared error in the forecasts up to 50 per cent for the forecast one-step ahead. For the longer horizons the improvement is smaller but still significant — 19 per cent for two steps ahead, 30 per cent for three steps ahead and 27 per cent for four steps ahead (Table 3, $RMSE^{opt}$ as a part of $RMSE^{*}$ for full balanced data set).

Figure 5: Transformed and normalized real output and its best predictions 1, 2, 3 and 4 steps ahead, using the last available observations

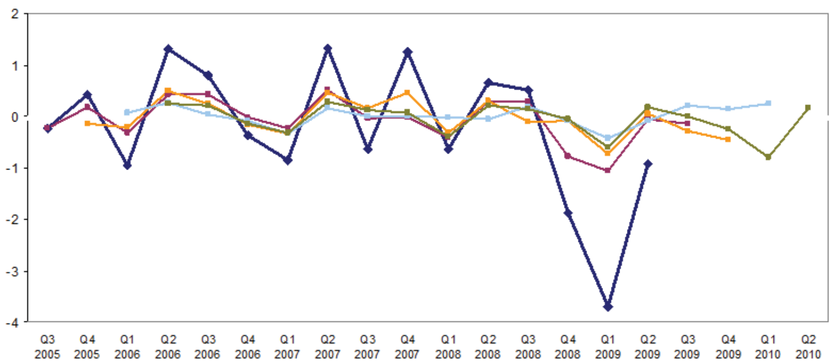


Figure 6: Transformed real output and its best predictions 1, 2, 3 and 4 steps ahead, using the last available observations

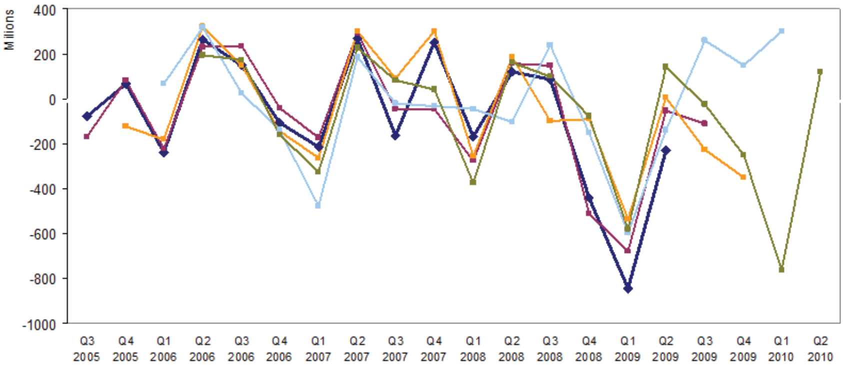
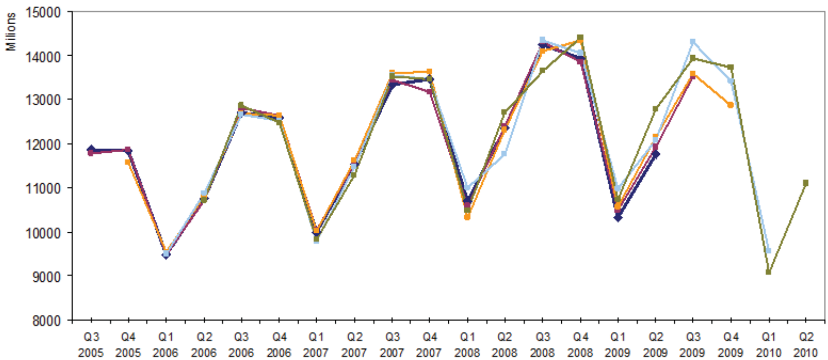


Figure 7: Real output and its best predictions 1, 2, 3 and 4 steps ahead, using the last available observations



DISCUSSION PAPERS

6 Conclusions

In the present paper we apply the GDFM to produce nowcasts and short-term forecasts of real Bulgarian GDP from large data set. A method for selection of model specification and a dataset based on the pseudo out-of-sample performance of the model is proposed. The selected subsets for the different forecast horizons still contain variables from all economic sectors and the forecast remains based on a very wide range of information.

Also a simple method, which exploits the dynamic properties of the GDFM, is used for including the latest available observations of series with smaller publication delay. The results show improvement of the chosen measure of the model performance and we can conclude that our methodology successfully exploits information in the recent observations of the series. Thanks to both data subset selection and the inclusion of the latest available information, a reduction of up to 50 per cent of the root mean squared error of the forecasts is obtained.

The specifics of the Bulgarian economy and in particular the short time span of the dataset have a negative impact on the stability of the model specification and variable selection over time. However, there is a scope for a further examination of the properties and the results of the model over time. The transformation of the data series also is a potential area for improvement of the results, although the initial examination of some of the alternatives does not give encouraging results.

7 Appendix A: Technical Details

Our estimation starts with computation of the sample covariance matrix Γ^T_{nk} of X^T_n and X^T_{n-k}, for $k = 0, 1, \ldots, M$. We use it to estimate the spectral-density matrix of X^T_n *via* discrete Fourier transformation of the two-sided sequence $\Gamma^T_{n,-M}, \ldots, \Gamma^T_{n,0}, \ldots, \Gamma^T_{n,M}$, where $\Gamma^T_{n,-k} = \Gamma^{T\prime}_{n,k}$. Precisely,

$$\Sigma^T_n(\theta_h) = \frac{1}{2\pi} \sum_{k=-M}^{M} \Gamma^T_{nk} w_k e^{-ik\theta_h},$$

where the frequencies are equally spaced in the interval $[0, 2\pi]$, *i.e.* $\theta_h = \frac{2\pi h}{2M+1}$, $h = 0, \ldots, 2M$, and $w_k = 1 - \frac{|k|}{M+1}$, $k = -M, \ldots, M$ are the weights corresponding to the Bartlett lag window of size M. We choose $M = 6$, which is close to the proposed in Forni *et al.* (2002) value of $[T^{1/3}] + 1$.

Further we perform the dynamic principal decomposition. For each frequency of the grid (*i.e.* for $h = 0, 1, \ldots, 2M$) we compute the eigenvalues $\lambda^T_{nj}(\theta_h)$ and their associated eigenvectors $\mathbf{p}^T_{nj}(\theta_h)$, $j = 1, \ldots, n$ of $\Sigma^T_n(\theta_h)$, where $\lambda^T_{nj}(\theta_h)$ are sorted in decreasing order, *i.e.* $\lambda^T_{n1}(\theta_h) \geq \ldots \geq \lambda^T_{nn}(\theta_h)$. We put $\mathbf{P}^T_n(\theta_h) = (\mathbf{p}^{T\prime}_{n1}(\theta_h), \ldots \mathbf{p}^{T\prime}_{nq}(\theta_h))'$ and $\mathbf{Q}^T_n(\theta_h) = (\mathbf{p}^{T\prime}_{n,q+1}(\theta_h), \ldots \mathbf{p}^{T\prime}_{nn}(\theta_h))'$. The spectral-density matrices of the common and idiosyncratic components are defined as

$$\Sigma^{\chi T}_n(\theta_h) = \tilde{\mathbf{P}}^T_n(\theta_h) \left\{ \begin{array}{ccc} \lambda^T_{n1}(\theta_h) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda^T_{nq}(\theta_h) \end{array} \right\} \mathbf{P}^T_n(\theta_h)$$

and

$$\Sigma_n^{\xi T}(\theta_h) = \tilde{\mathbf{Q}}_n^T(\theta_h) \left\{ \begin{array}{ccc} \lambda_{n,q+1}^T(\theta_h) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_{nn}^T(\theta_h) \end{array} \right\} \mathbf{Q}_n^T(\theta_h),$$

where the tilde denotes complex conjugation and transposition. We obtain the covariance matrices $\Gamma_{nk}^{\chi T}$ and $\Gamma_{nk}^{\xi T}$ of the common and idiosyncratic components *via* the inverse discrete Fourier transformation:

$$\Gamma_{nk}^{\chi T} = \frac{2\pi}{2M+1} \sum_{h=-M}^{M} \Sigma_n^{\chi T}(\theta_h) e^{i\theta_h k}$$

$$\Gamma_{nk}^{\xi T} = \frac{2\pi}{2M+1} \sum_{h=-M}^{M} \Sigma_n^{\xi T}(\theta_h) e^{i\theta_h k}$$

The aggregates which we use to approximate the common-factor space are of the form $W_{nt}^{kT} = Z_{nk}^T X_n^t$, where the weights Z_{nk}^T are defined as the solution, for $1 \leq k \leq r$ of the generalized eigenvalue problem

$$\begin{array}{rcl} Z_{nk}^T & = & Arg\ max\ \mathbf{a}\Gamma_{n0}^{\chi T}\tilde{\mathbf{a}} \\ \text{s.t.} & & \mathbf{a}\Gamma_{n0}^{\xi T}\tilde{\mathbf{a}} = 1 \\ & & \mathbf{a}\Gamma_{n0}^{\chi T}\tilde{Z}_{nm}^T = 0 \qquad \text{for } 1 \leq m \leq k, 1 \leq k \leq n \end{array}$$

Thus the vectors Z_{nj}^T, $j = 1, \ldots, n$ are the generalized eigenvectors, associated with the generalized eigenvalues η_{nj}^T of the couple of matrices $(\Gamma_{n0}^{\chi T}, \Gamma_{n0}^{\xi T})$.

Our estimate for $\chi_{i,T+h}$ is

$$\chi_{i,T+h}^{nT} = \left(\Gamma_{nh}^{\chi T} Z_n^T (\tilde{Z}_n^T \Gamma_{n0}^T Z_n^T)^{-1} \tilde{Z}_n^T X_n^T\right)_i,$$

where $Z_n^T = (Z_{n1}^{T\prime}, \ldots, Z_{nr}^{T\prime})'$.

Analogously,

$$\chi_{it}^{nT} = \left(\Gamma_{n0}^{\chi T} Z_n^T (\tilde{Z}_n^T \Gamma_{n0}^T Z_n^T)^{-1} \tilde{Z}_n^T X_n^t\right)_i, \qquad \text{for } t \leq T$$

is our estimate of the common components χ_{it}.

The model is implemented in Matlab on the basis of codes published by M. Forni

(http://www.economia.unimore.it/forni_mario/matlab.l

8 Appendix B: Data List

	Variables	Transf.	Variables used for h steps ahead prediction			
			$h=1$	$h=2$	$h=3$	$h=4$
	Quarterly data					
1	Gross domestic product, value at constant prices	2	+	+	+	+
2	Gross domestic product, value at current prices	2	-	+	+	+
3	Consumer survey - financial situation relative to 12 months ago	1	-	+	-	+
4	Consumer survey - financial situation over the next 12 months	1	-	+	-	+
5	Consumer survey - general economic situation over the last 12 months	1	+	+	-	+
6	Consumer survey - general economic situation over the next 12 months	1	-	+	-	+
7	Consumer survey - consumer price developments over the last 12 months	1	-	+	-	+
8	Consumer survey - inflation expectations over the next 12 months	1	-	+	-	-
9	Consumer survey - unemployment expectations over the next 12 months	1	-	+	-	+
10	Consumer survey - preference for making major purchases (for durable goods, etc.) - assessment of the present situation	1	-	+	+	+
11	Consumer survey - expectations for making major purchases (for durable goods, etc.) over the next 12 months	1	-	+	-	+
12	Consumer survey - possibility to save at present	1	-	+	-	-
13	Consumer survey - expectations for saving over the next 12 months	1	-	+	+	+
14	Consumer survey - current financial situation of the household	1	-	+	+	+
15	Consumer survey - expectations for car purchases over the next 2 years	1	-	+	+	+

	Variables	Transf.	$h=1$	$h=2$	$h=3$	$h=4$
16	Consumer survey - expectations for house purchases over the next 2 years	1	-	+	-	+
17	Consumer survey - expectations for purchases related to home improvements over the next 12 months	1	-	+	-	-
18	Consumer survey - confidence indicator	1	-	+	-	+
19	LFS-employment	2	+	+	+	+
20	LFS-unemployed	1	-	+	-	+
21	Labour productivity	2	+	+	-	-
22	Unit labour cost, nominal	2	-	+	-	+
23	Final individual consumption, value at current prices	2	-	+	+	-
24	Final collective consumption, value at current prices	2	-	+	-	-
25	Exports of goods and services, value at current prices	2	-	+	+	+
26	Imports of goods and services, value at current prices	2	+	+	-	-
27	Gross fixed capital formation, value at current prices	2	+	+	+	-
28	Change in inventories, value at current prices	1	-	-	-	-
29	Final individual consumption, value at constant prices	2	-	+	-	-
30	Final collective consumption, value at constant prices	2	-	+	-	-
31	Exports of goods and services, value at constant prices	2	-	+	+	-
32	Imports of goods and services, value at constant prices	2	-	+	-	-
33	Gross fixed capital formation, value at constant prices	2	-	+	-	-
34	Change in inventories, value at constant prices	1	-	-	+	-
35	Exports of goods, EUR	2	-	+	+	-
36	Imports of goods, EUR	2	-	+	-	-
37	Exports of services, EUR	2	-	+	+	-
38	Imports of services, EUR	1	-	+	-	+
39	Current account, EUR	2	-	+	-	-
40	Net inflow of FDI	1	-	+	+	-
41	Government Expenditures, cash basis	1	-	+	-	-
42	Government Revenue, cash basis	1	+	+	-	+
43	Value added tax, cash basis	1	-	+	+	+

	Variables	Transf.	$h=1$	$h=2$	$h=3$	$h=4$
44	Capital expenditure, cash basis	1	-	+	-	-
45	BNB official reserves	1	-	+	+	-
	Monthly data					
46	Unemployment rate (%)	2	-	+	+	+
47	Job vacancies over the last month	1	-	+	+	+
48	Share of long-term unemployed in the total number of unemployed	1	-	+	+	+
49	Business climate in the euro zone	1	-	+	-	+
50	Index of Fuel and Non Fuel Commodities (1995=100)	1	-	+	+	-
51	Index of Non-Fuel Primary Commodities (1995=100)	2	+	+	-	+
52	Edibles Index (1995=100)	1	+	+	+	+
53	Food Index: Cereals, vegetable oils, protein meals, meats, seafood, sugar, bananas and oranges	1	-	+	+	+
54	Index of Beverages, Coffee, Cocoa, and Tea	1	-	+	+	+
55	Index of Industrial Inputs (1995=100)	1	-	+	+	+
56	Index of Agricultural Raw Materials (1995=100)	1	-	+	+	-
57	Metals index	1	-	+	+	+
58	Commodity Fuel (energy) Index, 2005=100, includes Crude oil (petroleum), Natural Gas, and Coal Price Indices	1	-	+	+	-
59	Average Petroleum Spot index of UK Brent, Dubai, amd West Texas	1	-	+	+	-
60	Turnover Indices in 'Wholesale and retail trade; repair of motor vehicles and motorcycles'r	1	-	+	-	-
61	Turnover Indices in 'Wholesale and retail trade and repair of motor vehicles and motorcycles'	1	+	+	+	+
62	Turnover Indices in 'Wholesale trade, except of motor vehicles and motorcycles'r	1	-	+	+	+
63	Turnover Indices in 'Retail trade, except of motor vehicles and motorcycles'r	1	-	+	-	-
64	Industrial Production Indices	2	+	+	+	+
65	Production Indices in 'Mining and quarrying'	1	-	+	-	-
66	Production Indices in 'Manufacturing'	2	+	+	-	+

	Variables	Transf.	$h=1$	$h=2$	$h=3$	$h=4$
67	Production Indices in 'Electricity, gas, steam and air conditioning supply'	2	+	+	-	-
68	Industrial Turnover Indices	2	-	+	+	+
69	Turnover Indices in 'Mining and quarrying'	1	-	+	-	-
70	Turnover Indices in 'Manufacturing'	2	-	+	+	-
71	Turnover Indices in 'Electricity, gas, steam and air conditioning supply'	2	+	+	+	+
72	Construction survey - current business situation	1	+	+	-	+
73	Construction survey - current situation of production activity	1	+	+	-	+
74	Construction survey - order-books level, total	1	-	+	+	+
75	Construction survey - orders received over the last month compared to the previous month, total	1	-	+	+	+
76	Construction survey - orders received over the last month compared to the previous month, civil engineering	1	-	+	-	+
77	Construction survey - orders received over the last month compared to the previous month, building construction	1	-	+	+	-
78	Construction survey - number of clients with delayed payments, total	1	-	+	+	+
79	Construction survey - expected business situation over the next 6 months	2	-	+	+	+
80	Construction survey - Future production tendency	2	-	+	+	+
81	Construction survey - Selling price expectations	1	-	+	+	+
82	Construction survey - employment expectations	1	-	+	-	+
83	Construction survey - composite business climate	2	-	+	+	+
84	Industry survey - present business situation	1	-	+	+	+
85	Industry survey - present production tendency	1	-	+	-	-
86	Industry survey - order-books level - total	2	-	+	-	-
87	Industry survey - order-books level - dometic	2	-	+	-	-
88	Industry survey - order-books level - export	1	+	+	-	-

	Variables	Transf.	$h = 1$	$h = 2$	$h = 3$	$h = 4$
89	Industry survey - stocks of finished goods	1	-	+	-	-
90	Industry survey - future business situation	1	-	+	-	+
91	Industry survey - future production tendency	1	-	+	-	-
92	Industry survey - price expectations	1	-	+	-	-
93	Industry survey - price increase	1	-	+	-	+
94	Industry survey - composite business climate	1	-	+	+	+
95	Industry survey - composite confidence indicator	1	-	+	-	+
96	Industry survey - insufficient domestic demand	2	+	+	-	-
97	Industry survey - insufficient foreign demand	1	-	+	-	-
98	Industry survey - competitive imports	1	-	+	-	+
99	Industry survey - shortage of labour, including skilled	1	-	+	+	+
100	Industry survey - lack of appropriate equipment	1	-	+	+	-
101	Industry survey - shortage of energy	2	-	+	+	+
102	Industry survey - finacial problems	2	-	+	+	+
103	Industry survey - unclear economic laws	1	-	+	-	-
104	Industry survey - uncertainty of the economic environment	1	-	+	-	+
105	Industry survey - others constraints	1	-	+	+	-
106	Industry survey - no constraints	1	-	+	+	+
107	Retail trade survey - current business situation	1	-	+	-	-
108	Retail trade survey - current situation of the competitive pressures in the sector	1	-	+	+	+
109	Retail trade survey - current situation of the volume of stock of finished goods	1	-	+	+	+
110	Retail trade survey - current situation of selling prices	1	+	+	+	+
111	Retail trade survey - current increase of selling prices	1	+	+	+	-
112	Retail trade survey - expected business situation over the next 6 months	1	-	+	-	+
113	Retail trade survey - expected volume of orders to suppliers, total	1	-	+	-	+

	Variables	Transf.	$h=1$	$h=2$	$h=3$	$h=4$
114	Retail trade survey - expected volume of orders to domestic suppliers	1	-	+	-	+
115	Retail trade survey - expected volume of orders to suppliers abroad	1	-	+	-	+
116	Retail trade survey - sales expectations over the next 3 months	1	-	+	-	+
117	Retail trade survey - selling price expectations over the next 3 months	1	+	+	-	+
118	Retail trade survey - selling price increase over the next 3 months	1	+	+	-	+
119	Retail trade survey - employment expectations	1	-	+	-	+
120	Retail trade survey - demand constraints	1	-	+	-	+
121	Retail trade survey - supply constraints	1	+	+	+	-
122	Retail trade survey - constraints related to competition	1	+	+	+	+
123	Retail trade survey - constraints related to uncertainty	1	-	+	+	-
124	Retail trade survey - constraints related to the work-force	1	-	+	+	+
125	Retail trade survey - other constraints	1	+	+	+	+
126	Retail trade survey - no constraints	2	-	+	+	+
127	Total income of households - in pecuniary form	2	-	+	-	+
128	Total income of households - total	2	-	+	-	+
129	Total expenditures of households - in pecuniary form	2	-	+	-	+
130	Total expenditures of households - total	2	+	+	-	+
131	Headline CPI	1	-	+	-	+
132	Headline HICP	1	-	+	-	+
133	Employees	2	+	+	+	+
134	Wages	1	+	+	-	-
135	Claims on non-government sector	1	+	+	-	+
136	Monetary aggregate M3	1	-	+	-	-
137	Monetary aggregate M1	1	+	+	-	-
138	Currency outside MFIs	2	+	+	-	+
139	Monetary aggregate M2 (M1+Quasi-money)	1	-	+	-	-
140	Interest rate on the inter-bank money market, BGN	2	-	+	-	+

Transformation code:

1 $x_{it} = (1-L)X_{it}$

2 $x_{it} = (1-L)(1-L^4)X_{it}$

References

F. Altissimo, A. Bassanetti, R. Cristadoro, M. Forni, M. Lippi, L. Reichlin, and G. Veronese. A real time coincident indicator of the Euro area business cycle, Augist 2001. CEPR Discussion Paper no. 3108.

M. Artis, A. Banerjee, and M. Marcellino. Factor forecasts for the UK, January 2007.

J. Bai and S. Ng. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221, January 2002.

A. D'Agostino and D. Giannone. Comparing alternative predictors based on large-panel factor models, 2006. ECB working paper No 680.

M. Forni and M. Lippi. The Generalized Dynamic Factor Model: representation theory., 2001.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Econometrics and Statistics*, 82(4):540–554, November 2000a.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. Coincident and leading indicators for the euro area, September 2000b.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The Generalized Dynamic Factor Model: one-sided estimation and forecasting, 2002.

M. Forni, D. Giannone, M. Lippi, and L. Reichlin. Opening the black box: Structural factor models with large cross-sections, 2007. ECB working paper No 712.

J. Geweke. The dynamic factor analysis of economic times series, 1977. ed. by D. J. Aigner and A. S. Goldberger.

M. Marcellino, J. H. Stock, and M. W. Watson. Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review*, 47:1–18, 2003.

T. Sargent and C. Sims. Business cycle modeling without pretending to have too much a priori economic theory, 1977. in New Methods in Business Cycle Research, ed. by C. Sims.

M. Schneider and M. Spitzer. Forecasting austrian GDP using the generalized dynamic factor model, September 2004. working paper 89.

J. H. Stock and M. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20:147–162, 2002.

I. Vassileva. Construction of indicators for the bulgarian economy with generalized dynamic factor models, 2006. No 42006bg, Working paper series from Agency for Economic Analysis and Forecasting (in Bulgarian).

DISCUSSION PAPERS

DP/1/1998	The First Year of the Currency Board in Bulgaria Victor Yotzov, Nikolay Nenovsky, Kalin Hristov, Iva Petrova, Boris Petrov
DP/2/1998	Financial Repression and Credit Rationing under Currency Board Arrangement for Bulgaria Nikolay Nenovsky, Kalin Hristov
DP/3/1999	Investment Incentives in Bulgaria: Assessment of the Net Tax Effect on the State Budget Dobrislav Dobrev, Boyko Tzenov, Peter Dobrev, John Ayerst
DP/4/1999	Two Approaches to Fixed Exchange Rate Crises Nikolay Nenovsky, Kalin Hristov, Boris Petrov
DP/5/1999	Monetary Sector Modeling in Bulgaria, 1913–1945 Nikolay Nenovsky, Boris Petrov
DP/6/1999	The Role of a Currency Board in Financial Crises: The Case of Bulgaria Roumen Avramov
DP/7/1999	The Bulgarian Financial Crisis of 1996–1997 Zdravko Balyozov
DP/8/1999	The Economic Philosophy of Friedrich Hayek (The Centenary of his Birth) Nikolay Nenovsky
DP/9/1999	The Currency Board in Bulgaria: Design, Peculiarities and Management of Foreign Exchange Cover Dobrislav Dobrev
DP/10/1999	Monetary Regimes and the Real Economy (Empirical Tests before and after the Introduction of the Currency Board in Bulgaria) Nikolay Nenovsky, Kalin Hristov
DP/11/1999	The Currency Board in Bulgaria: The First Two Years Jeffrey B. Miller
DP/12/2000	Fundamentals in Bulgarian Brady Bonds: Price Dynamics Nina Budina, Tzvetan Manchev
DP/13/2000	Currency Circulation after Currency Board Introduction in Bulgaria (Transactions Demand, Hoarding, Shadow Economy) Nikolay Nenovsky, Kalin Hristov
DP/14/2000	Macroeconomic Models of the International Monetary Fund and the World Bank (Analysis of Theoretical Approaches and Evaluation of Their Effective Implementation in Bulgaria) Victor Yotzov
DP/15/2000	Bank Reserve Dynamics under Currency Board Arrangement for Bulgaria Boris Petrov

DP/16/2000	A Possible Approach to Simulate Macroeconomic Development of Bulgaria Victor Yotzov
DP/17/2001	Banking Supervision on Consolidated Basis *(in Bulgarian only)* Margarita Prandzheva
DP/18/2001	Real Wage Rigidity and the Monetary Regime Choice Nikolay Nenovsky, Darina Koleva
DP/19/2001	The Financial System in the Bulgarian Economy Jeffrey Miller, Stefan Petranov
DP/20/2002	Forecasting Inflation via Electronic Markets Results from a Prototype Experiment Michael Berlemann
DP/21/2002	Corporate Image of Commercial Banks (1996–1997) *(in Bulgarian only)* Miroslav Nedelchev
DP/22/2002	Fundamental Equilibrium Exchange Rates and Currency Boards: Evidence from Argentina and Estonia in the 90's Kalin Hristov
DP/23/2002	Credit Activity of Commercial Banks and Rationing in the Credit Market in Bulgaria *(in Bulgarian only)* Kalin Hristov, Mihail Mihailov
DP/24/2002	Balassa – Samuelson Effect in Bulgaria *(in Bulgarian only)* Georgi Choukalev
DP/25/2002	Money and Monetary Obligations: Nature, Stipulation, Fulfilment Stanislav Natzev, Nachko Staykov, Filko Rosov
DP/26/2002	Regarding the Unilateral Euroization of Bulgaria Ivan Kostov, Jana Kostova
DP/27/2002	Shadowing the Euro: Bulgaria's Monetary Policy Five Years on Martin Zaimov, Kalin Hristov
DP/28/2002	Improving Monetary Theory in Post-communist Countries – Looking Back to Cantillon Nikolay Nenovsky
DP/29/2003	Dual Inflation under the Currency Board: The Challenges of Bulgarian EU Accession *(in Bulgarian only)* Nikolay Nenovsky, Kalina Dimitrova
DP/30/2003	Exchange Rate Arrangements, Economic Policy and Inflation: Empirical Evidence for Latin America Andreas Freytag
DP/31/2003	Inflation and the Bulgarian Currency Board Stacie Beck, Jeffrey B. Miller, Mohsen Saad

DP/32/2003	Banks – Firms Nexus under the Currency Board: Empirical Evidence from Bulgaria Nikolay Nenovsky, Evgeni Peev, Todor Yalamov
DP/33/2003	Modelling Inflation in Bulgaria: Markup Model *(in Bulgarian only)* Kalin Hristov, Mihail Mihailov
DP/34/2003	Competitiveness of the Bulgarian Economy Konstantin Pashev
DP/35/2003	Exploring the Currency Board Mechanics: a Basic Formal Model Jean-Baptiste Desquilbet, Nikolay Nenovsky
DP/36/2003	A Composite Tendency Indicator for Bulgaria's Industry *(in Bulgarian only)* Tsvetan Tsalinsky
DP/37/2003	The Demand for Euro Cash: A Theoretical Model and Monetary Policy Implications Franz Seitz
DP/38/2004	Credibility Level of the Bulgarian Exchange Rate Regime, 1991–2003: First Attempt at Calibration *(in Bulgarian only)* Georgi Ganev
DP/39/2004	Credibility and Adjustment: Gold Standards Versus Currency Boards Jean-Baptiste Desquilbet, Nikolay Nenovsky
DP/40/2004	The Currency Board: "The only game in town" *(in Bulgarian only)* Kalin Hristov
DP/41/2004	The Relationship between Real Convergence and the Real Exchange Rate: the Case of Bulgaria Mariella Nenova
DP/42/2004	Effective Taxation of Labor, Capital and Consumption in Bulgaria *(in Bulgarian only)* Plamen Kaloyanchev
DP/43/2004	The 1911 Balance of Payments of the Kingdom of Bulgaria *(in Bulgarian only)* Martin Ivanov
DP/44/2004	Beliefs about Exchange Rate Stability: Survey Evidence from the Currency Board in Bulgaria Neven T. Valev, John A. Carlson
DP/45/2005	Opportunities of Designing and Using the Money Circulation Balance *(in Bulgarian only)* Metodi Hristov
DP/46/2005	The Microeconomic Impact of Financial Crises: The Case of Bulgaria Jonathon Adams-Kane, Jamus Jerome Lim

DP/47/2005	Interest Rate Spreads of Commercial Banks in Bulgaria *(in Bulgarian only)* Mihail Mihailov
DP/48/2005	Total Factor Productivity Measurement: Accounting of Economic Growth in Bulgaria *(in Bulgarian only)* Kaloyan Ganev
DP/49/2005	An Attempt at Measurement of Core Inflation in Bulgaria *(in Bulgarian only)* Kalina Dimitrova
DP/50/2005	Economic and Monetary Union on the Horizon Dr Tsvetan Manchev, Mincho Karavastev
DP/51/2005	The Brady Story of Bulgaria *(in Bulgarian only)* Garabed Minassian
DP/52/2005	General Equilibrium View on the Trade Balance Dynamics in Bulgaria Hristo Valev
DP/53/2006	The Balkan Railways, International Capital and Banking from the End of the 19th Century until the Outbreak of the First World War Peter Hertner
DP/54/2006	Bulgarian National Income between 1892 and 1924 Martin Ivanov
DP/55/2006	The Role of Securities Investor Compensation Schemes for the Development of the Capital Market *(in Bulgarian only)* Mileti Mladenov, Irina Kazandzhieva
DP/56/2006	The Optimal Monetary Policy under Conditions of Indefiniteness *(in Bulgarian only)* Nedyalka Dimitrova
DP/57/2007	Two Approaches to Estimating the Potential Output of Bulgaria Tsvetan Tsalinski
DP/58/2007	Informal Sources of Credit and the "Soft" Information Market (Evidence from Sofia) Luc Tardieu
DP/59/2007	Do Common Currencies Reduce Exchange Rate Pass-through? Implications for Bulgaria's Currency Board Slavi T. Slavov
DP/60/2007	The Bulgarian Economy on Its Way to the EMU: Economic Policy Results from a Small-scale Dynamic Stochastic General Equilibrium Framework Jochen Blessing
DP/61/2007	Exchange Rate Control in Bulgaria in the Interwar Period: History and Theoretical Reflections Nikolay Nenovsky, Kalina Dimitrova

DP/62/2007	Different Methodologies for National Income Accounting in Central and Eastern European Countries, 1950–1990 Rossitsa Rangelova
DP/63/2008	A Small Open Economy Model with a Currency Board Feature: the Case of Bulgaria Iordan Iordanov, Andrey Vassilev
DP/64/2008	Potential Output Estimation Using Penalized Splines: the Case of Bulgaria Mohamad Khaled
DP/65/2008	Bank Lending and Asset Prices: Evidence from Bulgaria Michael Frömmel, Kristina Karagyozova
DP/66/2008	Views from the Trenches: Interviewing Bank Officials in the Midst of a Credit Boom Neven Valev
DP/67/2008	Monetary Policy Transmission: Old Evidence and Some New Facts from Bulgaria Alexandru Minea, Christophe Rault
DP/68/2008	The Banking Sector and the Great Depression in Bulgaria, 1924–1938: Interlocking and Financial Sector Profitability Kiril Danailov Kossev
DP/69/2008	The Labour Market and Output in the UK – Does Okun's Law Still Stand? Boris Petkov
DP/70/2008	Empirical Analisys of Inflation Persistence and Price Dynamics in Bulgaria Zornitsa Vladova, Svilen Pachedjiev
DP/71/2009	Testing the Weak-form Efficiency of the Bulgarian Stock Market Nikolay Angelov
DP/72/2009	Financial Development and Economic Growth In Bulgaria (1991–2006). An Econometric Analysis Based on the Logic of the Production Function) Statty Stattev
DP/73/2009	Autonomy vs. Stability: the Relationship between Internal and External Money in Bulgaria (1879–1912) Luca Fantacci
DP/74/2009	The Size of the Shadow Economy in Bulgaria: A Measurement Using the Monetary Method Hildegart Ahumada, Facundo Alvarado, Alfredo Canavese, Nicolàs Grosman
DP/75/2009	Efficiency of commercial banks in Bulgaria in the wake of EU accession Kiril Tochkov, Nikolay Nenovsky
DP/76/2009	Structural Current Account Imbalances: Fixed Versus Flexible Exchange Rates? Slavi T. Slavov

DP/77/2009	Econometric Forecasting of Bulgaria's Export and Import Flows Grigor Stoevsky
DP/78/2009	Explanations for the Real Exchange Rate Development in the New EU Member States in Transition Galina Boeva
DP/79/2009	The Great Depression in the Eyes of Bulgaria's Inter-war Economists (How History of Economic Thought Could Matter for Today's Policy Advice) Stefan Kolev
DP/80/2010	Modeling Interest Rates on Corporate Loans in Bulgaria *(in Bulgarian only)* Mihail Mihailov
DP/81/2010	A Small Open Economy Model with Financial Accelerator for Bulgaria: The Role of Fiscal Policy and the Currency Board Ivan Lozev
DP/82/2010	The Impact of the Global Economic Crisis on Bulgaria's Accession to the Euro Area *(in Bulgarian only)* Tsvetelina Marinova
DP/83/2011	Are Long-term Inflation Expectations Well-anchored? Evidence from the Euro Area and the United States Tsvetomira Tsenova
DP/84/2011	Relative Inflation Dynamics in the EU Accession Countries of Central and Eastern Europe Hiranya K Nath Kiril Tochkov
DP/85/2011	Trade, Convergence and Exchange Rate Regime: Evidence from Bulgaria and Romania Emilia Penkova-Pearson