JEL Classification: C22, C32, C38, C52, C53, E23, E27
Keywords: GDP forecasting, bridge models, principal components, dynamic factor models, real-time evaluation

Short-Term Forecasting of Czech Quarterly GDP Using Monthly Indicators*

Kateřina ARNOŠTOVÁ – Macroeconomic Forecasting Division, Czech National Bank (katerina.arnostova@cnb.cz)
David HAVRLANT – Macroeconomic Forecasting Division, Czech National Bank (david.havrlant@cnb.cz)
Luboš RŮŽIČKA – Macroeconomic Forecasting Division, Czech National Bank (lubos.ruzicka@cnb.cz)
Peter TÓTH – Macroeconomic Forecasting Division, Czech National Bank, and Anglo-americká vysoká škola, o.p.s. (peter.toth@cnb.cz) – *corresponding author*

Abstract
We evaluate the out-of-sample forecasting performance of six competing models at horizons of up to three quarters ahead in a pseudo-real time setup. All the models use information in monthly indicators released ahead of quarterly GDP. We estimate two models – averaged vector autoregressions and bridge equations – relying on just a few monthly indicators. The remaining four models condition the forecast on a large set of monthly series. These models comprise two standard principal components models, a dynamic factor model based on the Kalman smoother, and a generalized dynamic factor model. We benchmark our results to the performance of a naive model and the historical near-term forecasts of the Czech National Bank's staff. The findings are also compared with a related study conducted by ECB staff (Barhoumi et al., 2008). In the Czech case, standard principal components is the most precise model overall up to three quarters ahead. However, the CNB staff's historical forecasts were the most accurate one quarter ahead.

1. Introduction

Short-term forecasting of quarterly GDP using monthly indicators is well motivated by the substantial delay in the publication of National Accounts statistics. This delay is ten weeks after the end of the quarter in the case of the Czech Republic. However, weeks before the appearance of the latest GDP figure, several monthly indicators gradually become available for the full quarter. These indicators are usually closely related to real economic activity in the broad sense and include industrial production, construction output, domestic or foreign sales, and labor market data. Other, alternative monthly indicators that provide timely information on recent developments on the aggregate level, such as price statistics, financial variables, and survey data, can also be considered. The availability of early information in monthly series therefore offers potential for improving the accuracy of GDP forecasts. Accordingly, GDP forecast accuracy is of key importance for monetary policy in particular, as well as in the general context of macroeconomic forecasting.

* The authors would like to thank Gerhard Rünstler (at the time affiliated with the ECB) for a consultation on factor models at the CNB in October 2007 and for providing program codes for the estimation and evaluation of the models considered in the present paper. We are grateful for the valuable comments from the three referees of the working paper – Michal Franta, Gabriel Perez-Quiros, and Petr Zemčík – and also from our colleagues at the CNB during presentations of earlier results of this project. We acknowledge the support provided by CNB Research Project B6/09.

Using Czech data, we compare the out-of-sample GDP forecasting performance of six models relying on information from monthly statistics. Out of the six models using monthly indicators, averaged bivariate VARs and bridge equations consider just a few important monthly series. The two models differ mainly in how they bridge the monthly data to quarterly GDP. The remaining four models rely on a relatively large number of indicators (up to 98), which are summarized using recent techniques, such as static and dynamic factor models. In the latter group we include two static factor models estimated by principal components. One dynamic factor model is estimated based on the Kalman smoother and the other one by generalized principal components in the frequency domain. In most of the six approaches outlined we closely follow a related study carried out by ECB staff and described in Barhoumi et al. (2008).[1] In addition, we test the performance of the average forecast obtained from the above models. This is motivated by studies (for example Kapetanios et al., 2007) that have found pooled forecasts outperform individually well-performing but differently mis-specified models.

What we do slightly differently from the aforementioned ECB study is to narrow down the large set of 98 monthly indicators. We apply a simple rule of thumb so as to include only those variables which are "well correlated" with GDP and to exclude those indicators which are mutually "too correlated". By doing this we aim to reduce variable-specific noise within the panel of monthly series by excluding somewhat irrelevant indicators. The described variable pre-selection results in better performance, especially in the case of the large-scale factor models.

We benchmark the above results to the performance of a naive univariate model of GDP and the historical near-term forecasts of the Czech National Bank's staff. The bank's near-term forecasting framework is multivariate, although it does not explicitly rely on monthly data. In addition, the bank incorporates a considerable amount of expert judgment into its near-term forecast. To summarize, the current project enables us to assess, in the context of Czech data, the forecasting abilities of monthly indicators, various ways of bridging monthly data to quarterly GDP, different versions of factor models, and the role of expert judgment.

We evaluate the competing GDP forecasts in a pseudo-real time setup with monthly updating of the forecasts. The availability of the different monthly indicators at the end of every month is therefore reflected in the evaluation. This results in an unbalanced monthly panel with ragged edges around the latest observations, given the variable-specific delays in data releases. Accordingly, the ragged edges are dealt with in various ways by the models considered. Regarding the described real-time setup we follow the ECB study of Barhoumi et al. (2009). We are also aware of an improved approach to real-time simulations by Camacho and Perez-Quiros (2008, 2009) for the euro area and Spain. The authors use real-time data vintages for their evaluation of a dynamic factor model, which means unrevised versions of both GDP and the monthly indicators exactly as they were available in every period. We have left this possible extension of our application for future research.

The paper is organized in the following way. First, we describe the competing models in detail, grouping them as non-factor models and factor models. Second, we

[1] The ECB study considers the euro area as whole, six countries from the euro area (Belgium, Germany, France, Italy, the Netherlands, and Portugal) and three new EU member states (Hungary, Poland, and Lithuania).

provide some details on our dataset, consisting of 98 monthly indicators. Third, we evaluate our models in terms of out-of-sample root-mean-squared errors and compare the results to the aforementioned ECB study. In the same part we also provide an illustrative example of monthly updates of a forecast as new data becomes available. The last section concludes.

2. Competing Models

We evaluate the out-of-sample forecast accuracies of the eight competing models.[2] The list can be divided into two groups, namely, four non-factor models and four factor models. The first four include the naive model of four-quarter moving averages of GDP, the core near-term forecasting (NTF) framework used at the bank, the averaged bivariate VAR models (VAR), and bridge equations (BEQ). The naive model is univariate and the NTF approach does not explicitly rely on information from monthly indicators. The VAR and BEQ models directly link GDP to selected monthly series that are considered good predictors.

The second group is composed of four factor models which link GDP growth to a set of a few estimated factors that are common to a larger set of monthly indicators. Here we include two versions of standard principal components,[3] where quarterly GDP is bridged with static factors. The first version of the principal components model (PC) follows the ECB's cross-country study by Barhoumi et al. (2008). We include a second, slightly different version (PC-Q) as a reference to an earlier work with leading indicators by Benda and Růžička (2007) at the Czech National Bank.

The group of factor models under our evaluation further includes two models with an explicitly dynamic structure of factors. In the dynamic factor model of Doz et al. (2007), dynamic factors are estimated by Kalman filtering (DFM). The approach of Forni et al. (2000, 2005), referred to as the generalized dynamic factor model (GDFM), estimates dynamic factors in the frequency domain.

As an additional benchmark we test the out-of-sample forecasting performance of the average forecast coming from the five main models listed above (VAR, BEQ, PC, DFM, and GDFM). This approach is motivated by evidence in the literature that pooling forecasts from individually well-performing but differently mis-specified models can improve forecast accuracy. For theoretical studies with related findings see, for example, Fischer and Harvey (1999) and Hendry (2004), while Kapetanios et al. (2007) provide empirical evidence in favor of forecast combination in the context of UK GDP and inflation.

2.1 Non-Factor Models

As a benchmark for the out-of-sample forecast precisions of our models we use the naive forecasts from the four-quarter moving averages of GDP.[4] We selected

[2] We replicate the majority of the models listed below from the cross-country study of EU member states conducted by ECB staff in Barhoumi et al. (2008).

[3] For a seminal work on forecasting with factors obtained from principal components, see Stock and Watson (2002)

[4] In the related ECB study by Barhoumi et al. (2007) a simple AR(1) model and the full-sample historical average growth are used as a benchmark. We use the four-quarter average growth instead, as its out-of-sample forecasts turned out to be more precise in the Czech case compared to the naive models used in the aforementioned study.

four quarters as the averaging interval based on a rule of thumb. The out-of-sample RMSEs of this model, however, were not very sensitive to slightly altering the averaging window. This result was also robust to changing the start and the end of the evaluation interval. The h-quarters-ahead forecast of the naive model can be expressed as:

$$\hat{y}_{t+h} = \tfrac{1}{4}\sum_{i=1}^{4} y_{t+h-i}$$

where for $h > i$ the forecasted values of y_{t+h-i} are used.

Next, we test the accuracy of the historical near-term forecasts (NTF) produced at the CNB. The bank's core NTF method can be characterized as a set of linear regression equations describing most of the expenditure components and deflators of GDP. Based on the situation and a subjective evaluation of the quality of the various equations, the forecasts were complemented with expert judgment. The year-on-year GDP growth forecast was computed as the weighted average of the growth rates of the expenditure components using the nominal weights from the same period of the preceding year. When compiling the GDP figure, further expert judgment was applied in order to produce a sensible and sufficiently smooth trajectory for GDP. The aforementioned equations were estimated at a quarterly frequency and the information in the monthly indicators was not reflected explicitly. However, the evolution of the most important leading indicators was usually reflected in the form of expert judgment. Since the set of the aforementioned equations and the amount of expert judgment in the historical forecasts have been changing over time, we will not specify the equations in further detail.

The remaining two non-factor models – averaged bivariate VARs (VAR) and bridge equations (BEQ) – already use the information in the monthly indicators explicitly. Hence, in these two cases we encounter the problem of mixed frequency data. The VAR and BEQ approaches differ mainly in how they bridge the monthly variables to quarterly GDP, including how incomplete quarters, i.e., the ragged edges of the panel of indicators, are treated.

In the case of the averaged VAR models,[5] the N monthly indicators $x_{i,\tau}$ are first aggregated to quarterly frequency (1), where τ indexes time at the monthly frequency and t denotes quarterly time periods. Incomplete quarters are simply omitted. Next, pairs of the GDP series (y_t^Q) and each of the quarterly aggregate indicators $x_{i,t}^Q$ are formed (2). For each of the N pairs a VAR(2,p_i) model is estimated (3), where the lag length p_i is set on the basis of the Akaike information criterion (AIC). Finally, we get the GDP forecast as the average of the forecasts from all the pairwise models (4).

VAR:

$$x_{i,t}^Q = \frac{1}{3}\sum_{s=0}^{2} x_{i,\tau-s} \tag{1}$$

$$\mathbf{z}_{i,t}^Q = \left\{ y_t^Q , x_{i,t}^Q \right\} \tag{2}$$

[5] The averaged VAR approach has been also applied to data from the United Kingdom by Camba-Mendez et al. (2001). For a general treatment of vector autoregressions, see, for example, Hamilton (1994).

$$\mathbf{z}_{\mathbf{i,t}}^{\mathbf{Q}} = \mu_i + \sum_{s=1}^{p_i} \mathbf{A}_{\mathbf{s}} \, \mathbf{z}_{\mathbf{i,t\text{-}s}}^{\mathbf{Q}} + \boldsymbol{\varepsilon}_{\mathbf{i,t}}^{\mathbf{Q}} \quad (3)$$

$$\hat{y}_{t+h|t}^{Q} = N^{-1} \sum_{i=1}^{N} \hat{y}_{i,t+h|t}^{Q} \quad (4)$$

In the case of the bridge equations[6] model (BEQ), the monthly indicators are first forecasted at the horizon of interest, which is H = 3h months ahead of the latest available GDP figure. The forecast is performed individually for each monthly series based on an AR(q_i) process. The lag length is again selected on the basis of the AIC. Next, each monthly series and its forecast are aggregated to quarterly frequency (5). In the following step, similarly to the VAR($2,p_i$) approach above, pairs of GDP and each of the indicators are formed (6). For every pair i bridge equation (7) is estimated by OLS, where the lag length of the indicator series is set on the basis of the AIC. The final GDP forecast is obtained as the arithmetic average of the forecasts from the pairwise models (8).

BEQ:

$$x_{i,t+h}^{Q} = \frac{1}{3} \sum_{s=0}^{2} x_{i,\tau+H-s} \quad (5)$$

$$\left\{ y_t^Q \,,\, x_{i,t}^Q \right\} \quad (6)$$

$$y_t^Q = \mu_i + \sum_{s=0}^{q_i} \beta_{i,s} \, x_{i,t-s}^Q + \varepsilon_{i,t} \quad (7)$$

$$\hat{y}_{t+h|t}^{Q} = N^{-1} \sum_{i=1}^{N} \hat{y}_{i,t+h|t}^{Q} \quad (8)$$

2.2 Factor Models

In the second group, containing factor models, we consider two versions of static principal components (PC and PC-Q) similar to the approach of Stock and Watson (2002) and two models assuming dynamic factors (DFM and GDFM). The PC model is estimated as follows. To get a balanced panel of the monthly indicators, incomplete quarters are filled in by forecasts from an autoregressive model (lag lengths set via AIC). Factors are estimated by principal components from the panel of N monthly series x_t. The number of static factors, $r<<N$ to be included in the model are determined by the Bai and Ng (2002) information criterion.[7] Every monthly series can thus be decomposed into a linear combination of r factors common to all series and an idiosyncratic term specific to the particular series (9). Next, the estimated factors are aggregated to quarterly frequency (10) and bridged with GDP. The h-quarters-ahead forecast of GDP then follows from the OLS-estimated bridge equation (11).

[6] The outlined bridge equations procedure closely follows the features of Baffigi et al. (2004), Kitchen and Monaco (2003), and Mariano and Murasawa (2003).

[7] In our data, the Bai and Ng (2002) criterion usually suggested only one static factor to be included in the model.

The outlined procedure is thus very standard in estimating principal components, as in Stock and Watson (2002), and bridging the quarterly aggregates of factors with GDP, as in Giannone, Reichlin, and Sala (2004) and Giannone, Reichlin, and Small (2005).

PC:

$$x_{i,\tau} = \sum_{j=1}^{r} \lambda_{i,j} F_{j,\tau} + \upsilon_{i,\tau} \tag{9}$$

$$F_{j,t}^{Q} = \frac{1}{3} \sum_{s=0}^{2} F_{j,\tau-s} \tag{10}$$

$$y_{t+h|t}^{Q} = \rho\, y_t^{Q} + \boldsymbol{\beta}\, \mathbf{F_t^Q} + \varepsilon_t \tag{11}$$

Notice that similarly to Stock and Watson (2002), Barhoumi et al. (2007), and others, the h-periods-ahead forecasts are obtained using the projection of GDP growth y_{t+h} on y_t and $\mathbf{F_t}$ in (11). As an alternative we could have considered vector autoregression of the (y_t, $\mathbf{F_t}$) pair and acquired h-steps-ahead forecasts by iterating forward by one period h times. However, as Stock and Watson (2002) argue, this alternative may imply estimating more parameters than necessary and therefore worse forecast performance.

We also evaluate an alternative version of the model based on static principal components, namely, PC-Q. PC-Q differs from PC in three ways. First, in PC-Q the principal components are estimated on the quarterly aggregates of monthly indicators (12). Although quarterly aggregation cuts the number of observations for estimating principal components, this approach was followed by Benda and Růžička (2007) in order to allow for potential leading indicators that are available only at quarterly frequency. However, in our evaluation exercise we estimate PC-Q on the same set of monthly indicators as PC. Second, the number of static factors is set by the Kaiser criterion.[8] The last aspect in which PC-Q differs from PC is the treatment of ragged edges in the monthly series. PC-Q uses a balanced panel of indicators by simply omitting incomplete quarters. To sum up, by comparing PC-Q to PC we mainly try to assess whether the information in timely monthly data matters for forecast precision. The h-quarters-ahead forecast of PC-Q (13) is estimated by OLS.

PC-Q:

$$x_{i,t}^{Q} = \sum_{j=1}^{r} \lambda_{i,j} F_{j,t}^{Q} + \upsilon_{i,t} \tag{12}$$

$$y_{t+h|t}^{Q} = \rho\, y_t^{Q} + \boldsymbol{\beta}\, \mathbf{F_t^Q} + \varepsilon_t \tag{13}$$

The dynamic factor model (DFM) due to Doz et al. (2007) is estimated in the steps described below. First, static factors are retrieved by the same procedure as in PC above, which includes balancing the panel of monthly indicators and estimating principal components. Next, we assume a VAR(q,p) process (14) driving the dynamics of factors, where q and p are set based on the information criterion developed

[8] According to the Kaiser criterion, principal components with eigenvalues greater than 1 are included in the model. In our data, the Kaiser criterion usually suggested about ten static factors.

by Bai and Ng(2007).[9] The coefficients in A_k and B are estimated by OLS using static factors obtained by principal components. In the next step, dynamic factors are estimated by the Kalman smoother taking the OLS estimates of the coefficients from above as given. Then, forecasts of the dynamic factors at the relevant horizons are produced by the Kalman filter.[10] The ragged edges in the panel of indicators are automatically taken care of by the Kalman filter, which uses all the available information in the monthly series. In the final stage, the factors are aggregated to quarterly frequency (15) and bridged with GDP via the OLS equation (16). For more details on the estimator and the estimation procedure, see Doz et al. (2007).

DFM:

$$\mathbf{f_\tau} = \sum_{k=1}^{p} \mathbf{A_k} \mathbf{f_{\tau\text{-}k}} + \mathbf{B}\omega_{i,\tau} \tag{14}$$

where $\mathbf{f_\tau}$ is a [q x 1] vector of dynamic factors, p is the lag length and $\boldsymbol{\omega_{i,\tau}}$ is a vector of innovations [q x 1].

$$f^{Q}_{j,t+h} = \frac{1}{3}\sum_{s=0}^{2} f_{j,\tau+H-s} \tag{15}$$

$$y^{Q}_{t+h|t} = \mu + \boldsymbol{\beta}'\mathbf{f^{Q}_{t+h|t}} + \varepsilon_{i,t} \tag{16}$$

The last, generalized dynamic factor model (***GDFM***) is based on Forni et al. (2000, 2005). This method allows both dynamic impacts of factors on particular indicators and some degree of cross-sectional and intertemporal correlation among the idiosyncratic shocks of the indicators. The model is estimated by generalized principal components in the frequency domain instead of the time domain, as suggested by the authors. In this approach, individual factors are not estimated directly, therefore the GDP forecast can be obtained only by including GDP in the set of indicators. The GDP forecast is then represented by the forecast of the common components of GDP and the indicators. We follow the ECB study by Barhoumi et al. (2008) and estimate GDFM at quarterly frequency. For more details on the estimation procedure, see also Nieuwenhuyze (2006). Ragged edges of the monthly indicators are equalized by forecasts from autoregressive models, similarly to PC. The number of dynamic factors q and the lag length p are selected based on information criteria.

3. Data

The full database of our monthly indicators consists of 98 time series describing the development of the main areas of the Czech economy and the external environment (see also *Appendix A*). The supply side is represented by indicators depicting the development of industry (29 series), construction (5 series), and the main branches of services (9 series). Several indicators represent the labor market (5). The list also contains foreign trade indicators (4), real and nominal exchange rates (5), monetary aggregates (2), and interest rates (12). We also add surveyed confidence indicators (6) and

[9] In our data, the Bai and Ng (2007) criterion usually suggested both q and p to be equal to 1.

[10] For a general treatment of Kalman filtering and smoothing in the context of time series analysis, see, for example, Hamilton (1994).

consumer and producer prices (11). The development of foreign demand is captured by leading indicators for Germany and Europe (10), such as the IFO business climate indicators, the composite leading indicators compiled by the OECD, and new car registrations in both Germany and the eurozone. The beginning of the time period examined was set at January 2001, as longer time series are not available for many of the indicators listed above.

The indicators under consideration were pre-adjusted in the following way. The first step is seasonal adjustment. Next, the three-month ($x_t - x_{t-3}$) growth rates are calculated and some series were further differenced to achieve stationarity. Finally, missing values and outliers were identified in the transformed time series and were replaced by the largest admissible values. As for aggregate output, we considered quarterly growth rates of GDP at constant prices of 2000 seasonally adjusted by the Czech Statistical Office and published on June 9, 2010. The series in levels was observed from Q1 1996 until Q1 2010.

To make sure that only the most relevant monthly indicators are considered when estimating the models, we narrowed down the full list of 98 series by a simple rule of thumb. We aggregated the series to quarterly frequency and computed the quarterly growth rates. Next, we calculated the correlation coefficient for every series and the quarterly growth rate of GDP. All variables with a correlation coefficient of less than 0.5 in absolute terms were excluded. In the next step we eliminated further series if the absolute value of their correlation with any other series was above 0.9. From such a pair, the one more correlated with GDP was kept in each case. The final set of indicators consists of 27 series, as indicated in *Appendix A*, including correlation coefficients with GDP growth. We are aware that some properties of the estimators we consider in our factor models are based on the assumption of a "large" number of indicators. Motivated by the resulting improvement in the forecasting performance of the factor models, we still opted for reducing the initial set of monthly series for estimating those models.

4. Evaluation

In what follows, the evaluation of forecast accuracies is described. First, the out-of-sample root-mean-squared errors (RMSEs) of the forecasts for the set of methods are compared both in absolute terms and relative to the naive model. Next, an example of the forecast for Q4 2009 is presented based on all the examined methods. This is aimed at illustrating how forecasts are updated and accuracies improve as additional information becomes available. Finally, the outcomes are compared to an ECB study that covers other countries of the European Union.

The monthly dataset used was available for the period Q1 2001–Q4 2009. The pseudo-real time forecast simulation was carried out for the evaluation period of Q1 2005–Q4 2009. The RMSEs of the out-of-sample forecasts were computed compared to the actual quarterly growth rates of GDP as observed at the time of the exercise (with GDP available up to Q1 2010). Within the simulation the time series are cut to the beginning of the evaluation period, also reflecting the publication lags of the monthly variables. As subsequent months of observations are gradually added, the models are re-estimated and the forecasts are updated on a monthly basis. The exercise continues this way until the end of the evaluation interval.

Table 1 Out-of-Sample RMSEs (in terms of quarterly GDP growth)

RMSE	+1Q	+2Q	+3Q	Average
NTF	*0.99*	*1.43*	*1.69*	*1.37*
VAR	1.15	1.45	1.63	1.41
BEQ	0.82	1.20	1.46	1.16
PC	**0.82**	**0.89**	**1.24**	**0.98**
PC-Q	0.94	1.42	1.75	1.37
DFM	0.89	1.03	1.36	1.09
GDFM	1.23	1.21	1.35	1.26
Average forecast	0.97	1.12	1.30	1.13
PC – full panel	1.09	1.07	1.31	1.15
DFM – full panel	1.26	1.43	1.43	1.37
GDFM – full panel	1.29	1.28	1.39	1.32
AR(1)	1.31	1.49	1.50	1.43
historical mean	1.33	1.33	1.33	1.33
4Q averages	1.19	1.31	1.38	1.29

Note: Forecast errors for NTF were computed on unrevised GDP data, therefore RMSE-s of NTF are not comparable to the rest.

4.1 Out-of-Sample Forecast Errors

Table 1 [11] presents the out-of-sample RMSEs 1Q to 3Q ahead, where the six models in the upper part of the table are the main focus of our study and the ones below the line are included as benchmarks. In the second part of the table, the forecast errors of the PC, DFM, and GDFM factor models in their full panel versions are presented. Here, all 98 monthly indicators were used for the estimation.[12] In addition, we report the performance of the average forecast, which is the simple average of the forecasts from the five main models, namely, VAR, BEQ, PC, DFM, and GDFM. Finally we report the performances of three simple univariate models – AR(1), historical average, and four-quarter average. The first two were also evaluated in the ECB study by Barhoumi et al. (2007), where the second one was taken as the naive model. We henceforth refer to the four-quarter average as our naive model.

In *Table 1*, the forecast errors tend to widen with longer horizons. Overall, the best performance is shown by standard principal components (PC). DFM and BEQ are quite close to the top ranked model. The remaining positions are taken by GDFM, PC-Q, and VAR. We can argue that the 1Q-ahead horizon is in most cases a nowcast or a backcast, which makes the availability of monthly indicators for the same period most promising for forecasting GDP. We find some evidence in favor of this argument in our results, as the RMSEs for nearly all the models tend to worsen substantially from 2Q onwards. Note that the RMSEs of the NTF model are not comparable to the rest of the data, as the CNB's historical forecasts were made on unrevised data while our model evaluation is performed on the latest data vintage.

Our results in *Table 1* further suggest that the average forecast is by far not the most precise forecasting tool in our case. This is not what one would expect given

[11] We have omitted the CNB's NTF framework from the table, as the results are not comparable with the rest of the models. This is because the NTF forecasts were performed on unrevised data.

[12] As we described in the Data section, we reduced the initial panel of 98 monthly indicators considered for estimating the PC, DFM, and GDFM factor models to a narrower set of 27 series by a simple rule of thumb in order to improve the forecast accuracy. For a list of the full and narrow sets of monthly series, see *Appendix A*.

Table 2 RMSEs Relative to the Naive Model of Four Quarter Averages

Relat. RMSE	+1Q	+2Q	+3Q	Average
NTF	**0.67**	0.80	0.91	0.81
VAR	0.97	1.11	1.18	1.09
BEQ	0.69	0.92	1.06	0.90
PC	0.69	**0.68**	**0.90**	**0.76**
PC-Q	0.80	1.09	1.27	1.06
DFM	0.75	0.79	0.99	0.85
GDFM	1.04	0.93	0.98	0.98
Average forecast	0.81	0.86	0.95	0.88
PC – full panel	0.92	0.82	0.95	0.89
DFM – full panel	1.06	1.10	1.04	1.07
GDFM – full panel	1.09	0.98	1.01	1.02
AR(1)	1.10	1.14	1.09	1.11
historical mean	1.13	1.02	0.97	1.03
4Q averages	1.00	1.00	1.00	1.00

Note: RMSE-s are presented relative to the naive model of 4Q averages. The RMSE-s of NTF are expressed relative to the naive model run on unrevised GDP data.

the experience of previous studies, such as Kapetanios et al. (2007). Moreover, we find that the full panel versions of the PC, DFM, and GDFM factor models do much worse than those estimated on the narrower set of indicators, as seen in *Table 1*. This suggests that smaller models containing only the most relevant indicators make more precise forecasts, in line with the conclusions of Camacho and Perez-Quiros (2008). Finally, we find that, overall up to three quarters ahead, the four-quarter average is the best performing naive model compared to AR(1) and the historical mean. Therefore, our naive benchmark is somewhat stricter than that of Barhoumi et al. (2007).

All the methods considered were estimated and evaluated on revised GDP data as available at the time of the evaluation exercise (GDP observed until Q1 2010). This is not the case for the NTF approach, since the historical forecasts had been performed on unrevised time series. Moreover, the evaluation interval of the NTF approach was somewhat shorter than in the rest of the models, as it started from Q3 2006 instead of Q1 2005 owing to the limited availability of historical records of the CNB's short-term forecasts. As a result, the absolute RMSEs are not comparable between the NTF approach and the rest of the models. Therefore, we computed the RMSEs for the naive model of four-quarter average quarterly GDP growth on both revised and unrevised GDP vintages. Next, we used the unrevised version to compute the relative RMSEs of the NTF and the revised version for the relative RMSEs of the remaining models, as shown in *Table 2*.

Apparently, the NTF framework does well compared to models using monthly data, and at the shortest horizon of 1Q ahead the NTF performs best.[13] We can hypothesize that the relative success of the NTF comes first from expert judgment, which is likely to take into account developments in the monthly indicators. The second

[13] Please note again that the precision of the NTF framework was computed on unrevised historical vintages of GDP data, unlike the rest of the models in the horserace, which were evaluated on the revised version of the GDP series. To allow better comparability, in *Table 2* we report RMSEs relative to the naive model's performance on the respective versions of GDP series (revised or unrevised). Reporting RMSEs relative to a naive model's performance is a standard approach in the literature to improve the comparability of results across different datasets or countries.

Table 3 Diebold-Mariano Test Statistics for the H0 of Equal Squared Forecast Errors

	VAR	BEQ	PC	DFM	GDFM	4Q average
VAR		2.01*	3.81**	3.28**	1.87*	2.05*
BEQ	-2.01*		2.95**	1.44	-0.41	-0.41
PC	-3.81**	-2.95**		-2.63**	-3.25**	-3.21**
DFM	-3.28**	-1.44	2.63**		-3.17**	-3.03**
GDFM	-1.87*	0.41	3.25*	3.17*		-0.02
4Q average	-2.05*	0.41	3.21**	3.03**	0.02	

Note: Note: negative statistics indicate smaller forecast errors for the model in the row. * and ** denote significance at the 95 % and 99 % levels. Degrees of freedom equals 159.

possible explanation stems from the consistency of the NTF forecasts in the structure of the expenditure components of GDP. The structural consistency perhaps better captures turning points in GDP. However, the top ranking of principal components is maintained when all horizons are considered, where the NTF proves to be the second best.

Not all the models, however, are able to beat the naive model in terms of forecast precision. At all the horizons considered, VAR and PC-Q do worse than the naive model. In addition, GDFM does not outperform the naive model at the shortest horizon of 1Q ahead.

To test whether the differences between the forecasting performances of the reported models are statistically significant, we computed pairwise Diebold-Mariano test statistics[14] on squared errors. This statistic tests the zero hypothesis of equal forecast errors and has a *t*-distribution with H–1 degrees of freedom, where H is the number of forecast errors observed. Since our forecasts were updated on a monthly basis up to eight months before every quarter of GDP growth observation and the evaluation interval consists of 20 quarters, we have 160 forecast errors available for every model. In *Table 3* we report the test statistics for all horizons together. The listed pairwise test statistics consider the five main models and the naive forecast. Negative numbers indicate smaller forecast errors for the model in the row. One and two stars denote significance at 95% and 99%, respectively. According to *Table 3*, most of the pairwise tests reject the hypothesis of equal forecast errors. The tests performed at particular horizons provide a slightly smaller number of statistically significant outcomes, though at substantially smaller degrees of freedom.

4.2 An Example: Monthly Updates of the Forecast for Q4 2009

The forecast for Q4 2009 is shown in *Figures 1–5* in order to illustrate the updating of the forecasts and their standard errors on a monthly basis. The forecast for q-o-q GDP growth depending on the number of months remaining to the actual GDP data release is shown together with the standard deviations. The first estimate is evaluated eight months before the GDP data release (the 3Q-ahead forecast), and it continues until the forecast made shortly before the official figures are published (the 1Q-ahead forecast). The last value in point 0 represents the actual GDP growth rate.

[14] The test was developed in Diebold and Mariano (1995). For an applied treatment in the context of ARIMA models, see, for example, Enders (2004).

Figure 1 VAR (q-o-q, in %, ± 1 std. dev.)

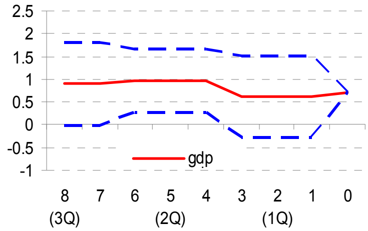


Figure 2 BEQ (q-o-q, in %, ± 1 std. dev.)

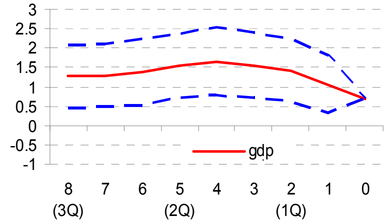


Figure 3 PC (q-o-q, in %, ± 1 std. dev.)

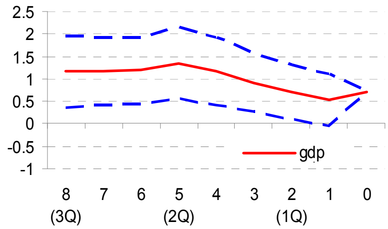


Figure 4 DFM (q-o-q, in %, ± 1 std. dev.)

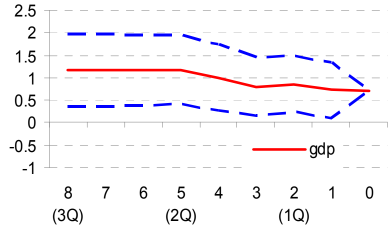
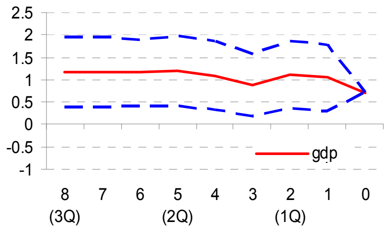


Figure 5 GDFM (q-o-q, in %, ± 1 std. dev.)



*Note:*The *gdp* line represents the evolution of the GDP forecast depending on the number of months remaining to the actual data release.

4.3 Comparable Results from Other EU Countries

A similar comparison of the methods discussed was carried out by the ECB (Barhoumi et al., 2008) for EU countries. The list of countries included the euro area as a whole and six members of the euro area – Belgium, Germany, France, Italy, the Netherlands, and Portugal. In addition, three new EU member states were considered, namely, Hungary, Poland, and Lithuania. This study yields different conclusions for the euro area members on the one hand and for the new EU members on the other hand in terms of the forecast errors of quarterly GDP growth. A fundamental difference emerges right at the beginning and is related to the quantity and quality of the available data. For the euro area countries the methods could be applied to a wide database from Q1 1991 to Q3 2005, whereas in the case of the new member states the time interval started only in Q1 1995. At the same time, the number of monthly indicators, ranging from 76 to 393, was often larger than in the case of the new member states (80 to 103 series).[15] The individual methods therefore in general show

[15] Note that Barhoumi et al. (2008) do not use a procedure to exclude monthly indicators less correlated with GDP. Therefore, all the reported results of the PC, DFM and GDFM models are based on the full dataset of available monthly series.

Table 4 RMSEs for Eurozone Countries Relative to the Naive Model of the Historical Mean

Relat. RMSE	+1Q	+2Q	+3Q	Average
AR	0.99	0.99	1.00	0.99
VAR	0.97	0.99	1.01	0.99
BEQ	0.93	0.96	1.00	0.97
PC	0.86	0.92	**0.94**	0.91
DFM	**0.83**	**0.90**	0.95	**0.89**
GDFM	0.91	0.92	0.98	0.94

Note: Simple averages across different countries were computed based on the detailed tables of Barhoumi et al. (2007).

Table 5 RMSEs for New EU Member-States Relative to the Naive Model of the Historical Mean

Relat. RMSE	+1Q	+2Q	+3Q	Average
AR	0.91	0.95	0.99	0.95
VAR	0.95	**0.94**	**0.95**	**0.90**
BEQ	0.94	0.98	0.98	0.96
PC	1.24	1.06	1.07	1.09
DFM	1.14	1.06	1.01	1.05
GDFM	**0.90**	**0.94**	0.99	0.94

Note: Simple averages across different countries were computed based on the detailed tables of Barhoumi et al. (2007).

better forecasting precision for the euro area countries than for the new EU members.[16] *Table 4* reports the average RMSEs of the evaluated methods for six countries of the euro area and the euro area aggregate relative to the naive model of the historical mean.

The models using additional information from monthly indicators apparently surpass the naive forecast. At the same time, the approaches using a large set of monthly indicators (PC, DFM, and GDFM) show a better forecasting ability than the simpler AR,[17] VAR, and BEQ methods, some of which use only a limited amount of monthly information.

From the new EU member states, Lithuania, Hungary, and Poland were considered. As *Table 5* shows, their average RMSEs relative to the naive model are considerably worse for some models compared to the euro area results. This is true especially for the approaches that use all the monthly indicators at once, namely, PC and DFM (but not GDFM), which perform worse than the simpler AR, VAR, and BEQ methods. At the same time their RMSEs are higher in comparison to the naive forecast (with the exception of GDFM).

5. Conclusion

We evaluated the out-of-sample GDP forecasting performance of six models based on monthly indicators. We compared those results to the average forecast, various naive univariate models, and the historical near-term forecasts made by the Czech

[16] In our opinion, the results in *Tables 4* and *5* reflect the general picture of the cross-country comparison, though simple averages may conceal some heterogeneity of the results. For details we refer the reader to Barhoumi et al. (2008).

[17] AR stands for the standard autoregressive model of GDP.

National Bank's staff. The CNB staff forecasts do not explicitly rely on monthly data, but incorporate a significant amount of expert judgment. Our results suggest that at all the horizons considered (up to three quarters ahead), most of the competing models did better than the naive univariate model of four quarter averages. This is not the case for averaged bivariate VARs, which could only beat the naive model one quarter ahead. A version of principal components where incomplete quarters of monthly indicators were not taken into account also did worse than the naive benchmark on average.

Overall, at all the horizons considered, the best performing model was standard principal components, which conditions the forecast on a large number of monthly indicators. The next positions were taken by the following models: the Czech National Bank's historical near-term forecasts, the dynamic factor model using the Kalman smoother, bridge equations, the generalized dynamic factor model, the alternative version of principal components (PC-Q), and averaged bivariate VARs. The top ranking of principal components was maintained for the horizons of two and three quarters ahead. At the shortest horizon, however, the CNB's historical forecasts provided the best results.[18] Quite surprisingly, the average forecast combining the forecasts from the five main models that we focus on was rather poor compared to the individual models. Finally, some of the large-scale factor models, such as DFM and GDFM, did worse than the naive benchmark when estimated on the full set of 98 monthly indicators instead of a narrowed set of 27. This finding is in line with some other studies suggesting that smaller factor models tend to have better forecasting performance. Our model rankings based on forecast precision were found to be statistically significant in most cases according to pairwise Diebold-Mariano tests of the equality of the squared forecast errors.

In comparison to the ECB study, we got similar results to the euro area countries. Principal components and the dynamic factor model based on the Kalman smoother did well on data from both the Czech Republic and the euro area countries. However, we obtained quite different results in terms of model rankings compared to the new EU member states (Hungary, Poland, and Lithuania). At this stage, therefore, it is difficult to draw general conclusions from our results. Nevertheless, we need to interpret such comparisons with substantial care, as the forecast errors are relatively sensitive to the selection of the evaluation interval and the set of monthly indicators. Both these parameters of our evaluation design are difficult to relax. This is first because the sample size of the Czech data is relatively small and therefore we used the longest possible time interval for the evaluation, and second because the set of available indicators differs substantially across countries too.

Concerning the implications for GDP forecasting practices at the Czech National Bank, we recommend using all the models evaluated in the present study. The average forecast of the above models could serve as a good benchmark for the bank's core near-term forecasting framework, at least at the one-quarter horizon. For further quarters ahead, greater weight could be given to the principal components model, which ranked as clearly the best in our evaluation.

[18] Note that the precision of the CNB's forecasts was computed on historical vintages of GDP data and was reported relative to the performance of the naive model run on the same historical vintages. The relative forecast errors thus allow us to compare the CNB's results with the other models, which were evaluated on the revised version of the GDP series.

APPENDIX

Monthly Indicators Considered for Estimation

Serie s No.	Name	Correlation with GDP*	Included in factor models?	Number of log differences**
1	Industrial Production Index (IPI)	0.57	N	1
2	IPI mining	-0.03	N	1
3	IPI manufacturing	0.57	Y	1
4	IPI electricity, gas, water	-0.09	N	1
5	IPI food processing	-0.18	N	1
6	IPI beverages	0.47	N	1
7	IPI textile	0.29	N	1
8	IPI clothing	0.08	N	1
9	IPI leather	0.51	Y	1
10	IPI wood	0.00	N	1
11	IPI paper	0.03	N	2
12	IPI printing	0.49	N	1
13	IPI coke, oil refinery products	0.08	N	1
14	IPI chemicals	0.01	N	1
15	IPI pharmaceutics	0.34	N	1
16	IPI plastics	0.31	N	1
17	IPI non-metal mineral products	0.41	N	2
18	IPI metals	0.22	N	1
19	IPI fabricated metal products	0.49	N	1
20	IPI computers, electronic and optical devices	0.32	N	1
21	IPI electric devices	0.08	N	1
22	IPI machinery	0.63	Y	2
23	IPI motor vehicles excl. motorcycles	0.51	Y	1
24	IPI other vehicles	0.01	N	1
25	IPI furniture	0.15	N	2
26	IPI other manufacturing	0.34	N	1
27	IPI service and maintanance of machines	-0.29	N	1
28	Industry sales	0.55	Y	1
29	New orders in industry	0.32	N	1
30	Construction output	0.20	N	1
31	Building construction	0.28	N	2
32	Civil engineering	-0.16	N	1
33	Bulding permits issued	-0.08	N	1
34	Approximate value of permitted constructions	-0.06	N	1
35	Sales – wholes., retail, service and maint. of motor vehicles	0.63	Y	2
36	Sales – retail (incl. fuels)	0.42	N	2
37	Sales – services total	0.86	Y	2
38	Sales – transport and warehousing	0.65	N	2
39	Sales – accommodation, catering and hospitality	0.52	Y	2
40	Sales – information and communication services	0.51	Y	2

41	Sales – real estate services	0.29	N	2
42	Sales – professional, scientific and technical services	0.58	Y	2
43	Sales – administrative and complementary activities	0.63	Y	2
44	Number of unemployed	-0.66	N	2
45	Free vacancies	0.73	Y	1
46	Newly registered unemployed (inflows)	-0.77	Y	2
47	Unemployed removed from the register (outflows)	0.02	N	2
48	Unemployment rate (total)	-0.72	Y	2
49	CPI total	0.02	N	2
50	CPI net (excl. regulated prices and indirect taxes)	0.21	N	2
51	CPI adjusted (CPI net, excl. food and fuels)	0.21	N	1
52	Export (current prices)	0.53	Y	1
53	Import (current prices)	0.57	Y	1
54	Export price index	-0.11	N	1
55	Import price index	0.15	N	1
56	Oil price (Brent)	0.32	N	1
57	Earth gas price (Russian, price in Germany)	0.41	N	1
58	Gasoline price (ARA exchange)	0.23	N	1
59	Eurozone PPI (effective)	0.64	Y	1
60	PPI minerals	0.13	N	1
61	PPI manufacturing	0.57	Y	1
62	PPI electricity, gas, steam	-0.30	N	1
63	CZK/EUR	-0.44	N	1
64	CZK/USD	-0.28	N	1
65	NEER	0.31	N	1
66	REER (defl. by CPI)	0.33	N	1
67	REER (defl. by PPI)	0.27	N	1
68	M1	0.03	N	2
69	M2	-0.05	N	2
70	3M PRIBOR	0.52	Y	2
71	1Y PRIBOR	0.44	N	2
72	ECB 3M rate	0.75	N	2
73	ECB 1Y rate	0.75	Y	1
74	PX stock index (Prague)	0.35	N	2
75	Czech government bond yield (5Y)	0.11	N	1
76	Czech government bond yield (10Y)	-0.01	N	1
77	FRA 3*6 (forward rate)	0.49	N	2
78	FRA 3*9 (forward rate)	0.46	N	2
79	FRA 6*9 (forward rate)	0.43	N	2
80	FRA 6*12 (forward rate)	0.41	N	2
81	FRA 9*12 (forward rate)	0.38	N	2
82	Confidence indicator index (entrepreneurs)	0.61	Y	1
83	Confidence indicator index (consumers)	0.32	N	1

84	Confidence indicator index (total)	0.60	N	1
85	Industry survey – overall economic situation	0.71	Y	2
86	Industry survey – demand	0.60	Y	1
87	Industry survey – expected economic situation in 3 months	0.28	N	1
88	The Ifo Business Climate for Germany – Business Climate	0.38	N	1
89	The Ifo Business Climate for Germany – Business Situation	0.67	Y	2
90	The Ifo Business Climate for Germany – Business Expectations	0.00	N	1
91	OECD Composite Leading Indicator – Czech Republic	0.46	N	1
92	OECD Composite Leading Indicator – Germany	0.81	Y	1
93	OECD Composite Leading Indicator – Europe	0.76	N	1
94	OECD Composite Leading Indicator – Total	0.75	N	1
95	New car registrations – EU	-0.12	N	1
96	New car registrations – Germany	-0.54	Y	1
97	Euro area Business Climate Indicator	0.52	Y	1
98	Electricity consumption in the Cz. Rep. (adj. for temperatures)	0.37	N	1

*Notes:** Correlation coefficients were calculated from q-o-q growth rates of the quarterly aggregates.
** Monthly indicators were log-differenced before estimation to achieve stationarity.

REFERENCES

Baffigi A, Golinelli R, Parigi G (2004): Bridge Models to Forecast the Euro Area GDP. *International Journal of Forecasting,* 20(3):447–460.

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

Bai J, Ng S (2007): Determining the Number of Primitive Shocks in Factor Models. *Journal of Business & Economic Statistics*, 21(1):52–60.

Barhoumi K, Benk S, Cristadoro R, Reijer A den, Jakaitiene A, Jelonek P, Rua A, Rünstler G, Ruth K, Nieuwenhuyze C van (2008): Short-Term Forecasting of GDP Using Large Monthly Datasets. A Pseudo Real-Time Forecast Evaluation Exercise. *ECB Occasional Paper Series*, no. 84.

Benda V, Růžička L (2007): Short-term Forecasting Methods Based on the LEI Approach: The Case of the Czech Republic. *Czech National Bank, Research and Policy Notes*, no. 1/2007.

Camacho M, Perez-Quiros G (2008): Introducing the Euro-STING: Short-Term Indicator of Euro Area Growth. *Bank of Spain, Documentos de Trabajo*, no. 0807.

Camacho M, Perez-Quiros G (2009): N-Sting: Espana Short Term Indicator of Growth. *Bank of Spain, Documentos de Trabajo*, no. 0912.

Camba-Mendez G, Kapetanios G, Smith M, Weale R (2001): An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries. *Econometrics Journal*, 4(1):56–80.

Diebold F, Mariano R (1995): Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13:253–263.

Doz C, Giannone D, Reichlin L (2007): A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering. *CEPR Discussion Paper*, no. 6043.

Enders W (2004): *Applied Econometric Time Series*. 2nd ed. Wiley Series in Probability and Statistics, John Wiley & Sons.

Fischer I, Harvey N (1999): Combining Forecasts: What Information do Judges Need to Outperform the Simple Average? *International Journal of Forecasting*, 15(3):227–246.

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

Forni M, Hallin M, Lippi M, Reichlin L (2005): The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association*, 100:830–840.

Giannone D, Reichlin L, Sala L (2004): Monetary Policy in Real Time. In: Gertler M, Rogoff K (Eds): *NBER Macroeconomics Annual 2004*, pp.161–200 (MIT Press).

Giannone D, Reichlin L, Small DH (2005): Nowcasting GDP and Inflation. The Real Time Informational Content of Macroeconomic Data Releases. *ECB Working Paper Series*, no. 633.

Hamilton JD (1994): *Time Series Analysis*. Princeton University Press.

Hendry DF, Clements MP (2004): Pooling of Forecasts. The *Econometrics Journal*, 7(1):1–31.

Kapetanios G, Labhard V, Price S (2007): Forecast Combination and the Bank of England's Suite of Statistical Forecasting Models. *Bank of England Working Paper Series*, no. 323.

Kitchen J, Monaco R (2003): Real-Time Forecasting in Practice: The U.S. Treasury Staff's Real-Time Forecast System. *Business Economics*, 38(4):10–28.

Mariano R, Murasawa Y (2003): A New Coincident Index of Business Cycles Based on Monthly and Quarterly Data. *Journal of Applied Econometrics*, 18:427–443.

Nieuwenhuyze C van (2006): A Generalized Dynamic Factor Model for the Belgian Economy – Useful Business Cycle Indicators and GDP Growth Forecasts. *National Bank of Belgium Working Paper Research*, no. 80.

Stock J, Watson MW (2002): Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.