Nowcasting GDP Growth:

Statistical Models versus Professional Analysts

Jasper de Winter

Nowcasting GDP Growth:
statistical models versus professional analysts

Jasper M. de Winter

ISBN: 978-94-91602-73-3

Nowcasting GDP Growth: statistical models versus professional analysts

Nowcasten van de bbp-groei: statistische modellen versus professionele analisten

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 23 december 2016 om 13.30 uur
door

Jasper Matthijs de Winter

geboren te Amersfoort.

Erasmus University Rotterdam

Promotiecommissie

Promotor:
Prof. dr. J. Swank

Overige leden:
Prof. dr. D. van Dijk
Dr. J.P.A.M. Jacobs
Prof. dr. S.J. Koopman

Acknowledgements

Like a little boy exploring the world on his own two feet for the first time, writing this thesis gave me a chance to discover new grounds. It was an experience I thoroughly enjoyed. This would not have been possible without De Nederlandsche Bank (DNB) who provided me with the opportunity, time and all the necessary resources to complete this work.

Several people have contributed to my research and I would like to extend my thanks to them. First of all, I would like to thank my supervisor Job Swank for having confidence in me from the very start, and for his constructive criticism on my writings. Secondly, I would like to thank my co-authors. Without your help this thesis would not be the same. Siem Jan Koopman, thank you for sharing your econometric expertise and your infectious enthusiasm for the profession. It has been an honor to work with you. Jos Jansen, I admire your determination, perfectionism and originality when doing research. You truly are one of a kind. Irma Hindrayanto, for your patience when explaining tough econometric concepts again and again, and for your impeccable programming skills. Xiaowen Jin for our short, but very productive, collaboration. The amount of work you can do in a short few hours is impressive. A more exhaustive list of people who contributed to my research are mentioned on the first page of each chapter.

Furthermore, I would like to thank a few others who have helped me with this thesis in one way or another. Peter Berkhout, for sparking my enthusiasm for applied econometrics and for our endless discussions on music, science and life in general. Ide Kearney, thank you for your careful edit of parts of my manuscript. My roommate Robert-Paul Berben, for putting up with me over the last couple of years. You may run faster, but I will try to catch up! I am grateful to my parents, who have provided me with both the genes and the means necessary to be educated at university and eventually obtain my PhD degree. Finally, I would like to thank my lovely wife Sabine and our two daughters, Puck and Fieke. I feel blessed we share our life's paths, and sure wish we can continue sharing them for many, many years to come.

Jasper de Winter
November 2016

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation contains four essays on short-term forecasting. The main goal is to evaluate whether the monthly flow of statistical information and predictions by professional analysts are helpful to accurately forecast the quarterly growth rate of real Gross Domestic Product (GDP). This topic has regained interest amongst practitioners and academics since the onset of the financial crisis of 2008–2009, when most models failed to forecast the large and abrupt contraction of GDP.

Nowadays, practitioners face a potentially very large information set of monthly indicators. This information includes data on industrial production, unemployment, consumer confidence, stock markets and prices of goods and services. It is still an open question what strategy policy makers and economic agents should use to transform this information set to an understanding of where the economy currently stands and where it is heading in the near-term. A possible approach would be to purely rely on mechanical statistical models. Recently, the forecasting literature has developed several of these models to exploit large datasets. Examples include dynamic factor models, mixed-data sampling models and Bayesian vector autoregressive models. These models differ in their approach to the practical problems of how to handle a large-scale information set and the fact that the auxiliary variables are observed at different frequencies and publication lags. Apart from model-based forecasts, practitioners can also take advantage of published forecasts made by professional analysts. From a practical point of view, such forecasts are cheap and easy to use. Moreover, they may, as an expression of the "wisdom of crowds", reflect much more information than statistical information, which is inevitably limited.

The four chapters in this thesis contribute to the ongoing debate on the forecasting accuracy of short-term forecasting models, the usefulness of forecasts of professional analysts and the merits of combining the forecasts of professional analysts and statistical models. New light is shed on the issue as to which model features are especially valuable for short-term forecasting. Key to these results is the collection of new data on the quarterly GDP forecasts of professional analysts and snapshots ("vintages") of the available data when the analysts made their forecasts. The remainder of this introduction provides a short overview of each chapter.

Chapter 2 examines whether short-term forecasting models were helpful in forecasting Dutch real GDP growth during the dot-com recession of 2001–2002 and the financial crisis of 2008–2009. Two forecasting strategies based on linear statistical models are evaluated. The first strategy extracts information from the dataset of monthly indicators by averaging the quarterly GDP growth forecasts from quarterly indicator models, i.e.: quarterly vector autoregressive (QVAR) models and quarterly bridge equations (BEQ). The second strategy extracts principal components (or factors) from the set of monthly indicators and uses these factors to forecast quarterly GDP growth. These strategies are compared to a naïve autoregressive model and a strategy that simply takes the average quarterly GDP forecasts of professional analysts. The latter were collected from paper copies of the monthly publication *Consensus forecasts*, published by the private sector firm Consensus Economics. The different strategies are evaluated in a so-called pseudo real-time setup. This setup takes into account the publication delays of GDP and the monthly indicators but does *not* take into account the possibility of data revisions. The empirical results provide compelling evidence that using linear statistical models pays off: the forecasting errors are much smaller than the errors from a naïve autoregressive model. The dynamic factor model (DFM) has the edge over the other statistical models, especially when forecasting the adjacent quarters. The main message from the comparison of the mechanical statistical models and the predictions of professional analysts is that the latter seem to embody information that mechanical models fail to pick up.

Chapter 3 extends the analysis in Chapter 2 in several directions. Firstly, the range of candidate linear statistical models is extended. Besides the QVAR, BEQ and DFM Chapter 3 also examines the forecasting accuracy of the Bayesian quarterly vector autoregressive (BVAR) model, the mixed-frequency vector autoregressive (MFVAR) model and the mixed-data sampling regression (MIDAS) model. In addition, factor-augmented versions of the QVAR, MFVAR and BEQ are evaluated, and it is tested whether allowing for autoregressive terms (GDP's own past) in MIDAS and BEQ models enhances their forecasting quality. In total, twelve models are analyzed. Secondly, the analysis is enriched by analyzing the forecasting quality of the various models in the euro area and its five largest countries (Germany, France, Italy, Spain and the Netherlands), adding robustness to the outcomes. The analysis sheds new light on the performance of forecasting models during periods of heightened volatility, by examining the forecasting accuracy of the different forecasting strategies before as well as during a after the financial crisis of 2008–2009. The main results can be summarized as follows: The dynamic factor model displays the best forecasting capabilities overall, confirming the results in Chapter 2 for a broader set of models and countries. The ability of the DFM to incorporate more than one factor is key to this result. However, factor-augmented MFVARs and MIDAS models produce slightly better one-quarter ahead predictions after the start of the recent financial crisis due to their richer dynamic specification. The BVAR is the best quarterly model. It performs quite well for Germany, the Netherlands and Spain in the more stable period of the Great Moderation. Remarkably, all other models, including the dynamic factor model, perform (very) poorly in the case of Spain during the Great Moderation. This finding suggests that Bayesian estimation is a fundamentally different way of extracting information from a large dataset, which may deliver benefits, even if the model makes inefficient use of the available monthly information. Regarding crucial model features, summarizing the available monthly information in one or more factors clearly delivers better results than the alternative of pooling single-indicator-based forecasts. Allowing for autoregressive terms in forecasting equations leads to improvements in forecast reliability, but the improvement is relatively small compared to the other model features. The scope for improving GDP forecasts by combining the "views" of various models is rather limited in economic terms. Lastly, the (partially) subjective forecasts of private sector analysts embody valuable information that sophisticated mechanical forecasting procedures fail to pick up.

Chapter 4 builds on the results of the previous two chapters, and addresses a new question within the empirical forecasting literature: can predictions by analysts improve GDP forecasts generated by statistical procedures in a *truly* real-time context? This setup differs from the *pseudo* real-time setup in the previous chapters in two respects. Firstly, it mimics a practitioner deciding on his forecasting-strategy at a certain point in time, based on the forecasting performance of the mechanical models and professional analysts in the recent past. Secondly, the real-time setup respects the preliminary nature of many series, i.e. GDP and many of the monthly indicators are revised regularly as more information becomes available to the statistical office over time. This implies exact copies –or "snaphots"– of the macro-economic indicators available when professional analysts made their predictions had to be collected. These so-called real-time datasets were constructed for each of the G7 countries (United States, United Kingdom, Canada, Japan, Germany, France and Italy). The forecasts of professional analysts are compared to the DFM, which turned out to have the best overall near-term forecasting performance in the previous chapters. One of the main insights is that since 2008, Consensus forecasts are a tough competitor for the mechanical DFM for most countries. In the post-crisis period newly released Consensus nowcasts and forecasts have a higher forecast accuracy than the DFM. Another insight is that the difference in forecasting performance between professional analysts and the DFM tends to be greatest for a fresh Consensus nowcast. In the stable pre-crisis period, the professional analysts do worse or at most marginally better than the dynamic factor model across the board.

This pattern suggests that analysts pay more attention and devote more effort to forecasting in volatile times. The value added of the predictions of professional analysts is largest when they know at least some hard and soft data pertaining to the quarter of interest. Another finding is that the relative forecasting advantage of professional analysts declines as their forecasts age. This is related to the fact that the DFM is able to fully exploit all newly released monthly data. However, combining the forecasts of professional analysts and the DFM delivers sizable gains in forecasting ability of statistical models for most countries, even when the forecasts of analysts are somewhat dated. A final insight is that determining the optimal combination scheme in real-time is infeasible. Overall, using a simple average of the various combination rules provides the best hedge against misspecification and instability.

Based on the relatively good forecasting performance of the dynamic factor model in the previous chapters and the widespread use of this nowcasting model amongst practitioners, Chapter 5 presents an analysis on the most appropriate specification of the dynamic factor model. The analysis concentrates on four factor models: the canonical factor model of Stock and Watson (2002b) who initiated the current literature on factor models, the widely used dynamic factor model of Bańbura and Rünstler (2011), its modification proposed by Bańbura and Modugno (2014), and the recently proposed factor model specification of Bräuning and Koopman (2014). The forecasting accuracy of the four models is compared in a large-scale forecasting 'horse-race' for the euro area and its five largest countries (Germany, France, Italy, Spain and the Netherlands). To examine whether periods of high volatility favor a different factor model structure, the forecasting accuracy is examined before as well as during and after the financial crisis of 2008–2009. Furthermore, two modifications to the factor models based on the outcome of Chapter 3 are proposed and tested. Firstly, the inclusion of one or more lags of the targeted variable (GDP's own past) is introduced in the dynamic factor models. Secondly, a simple alternative for handling the non-synchronous nature of the monthly data-releases in the Bräuning and Koopman (2014) model is proposed. Both modifications clearly improve the forecasting accuracy of the factor models. The main conclusion is that the modified factor model specification of Bräuning and Koopman (2014) has the edge over the other factor models for most countries and most forecasting horizons. This conclusion holds before as well as during and after the recent financial crisis.

Chapter 6 summarizes and concludes.

Chapter 2

Nowcasting real GDP growth in the Netherlands

This chapter examines the forecasting accuracy of linear statistical models tailored to forecast GDP growth in the adjacent quarters, with a special focus on the dot-com recession of 2001–2002 and the financial crisis of 2008–2009. The forecasting accuracy of the mechanical models is confronted with the forecasting accuracy of professional analysts, and it is evaluated whether a combination of the two can improve forecasting performance. The analysis covers the Netherlands over the years 1995–2010. Overall, the recently proposed dynamic factor model showed the highest forecasting accuracy of all models considered. During the recent financial crisis, the forecasting accuracy of all models deteriorated. The dynamic factor model shows the smallest deterioration of all models considered. Interestingly, enhancing the forecasts of the dynamic factor model with the judgmental forecasts of professional analysts can increase the forecasting accuracy, but only during the financial crisis.[1]

KEYWORDS: Factor models; Professional analysts; Pseudo real-time data.

2.1 Introduction

The recession of 2008–2009 marked the largest fall in Dutch real Gross Domestic Product (GDP) since the Second World War. After peaking in the fourth quarter of 2007, output declined for six consecutive quarters, by 4.2 percentage point in total. Although some indicators clearly revealed the build up of imbalances that usually precede a crisis (Reinhart and Rogoff, 2008), most forecasters across the world failed to forecast the

[1] Comments and suggestions by Marta Bańbura, Jos Jansen, Pierre Lafourcade, Job Swank and seminar participants at the ECB Expert Meeting on Activity Forecasting (2011, Frankfurt) are gratefully acknowledged. An early version of this chapter was circulated as DNB Working Paper 320 under the title "Forecasting GDP growth in times of crisis: private sector forecasts versus statistical models".

depth and duration of the crisis. This also holds for the Netherlands (Roscam Abbing et al., 2010).

This chapter examines if forecasting models that are specifically tailored to forecast real GDP growth in the adjacent quarters could have been helpful in forecasting the dynamics of the financial crisis in the Netherlands. The forecasting accuracy of these so called “nowcasting” models is a highly relevant issue for policy makers and economic agents. Especially in times of heightened uncertainty, surrounding recessions and crises. The recent literature on nowcasting GDP (see e.g Rünstler et al., 2009, Bańbura et al., 2011 and Baffigi et al., 2004) has not explicitly analyzed these periods. This chapter aims to partially fill this gap in the literature by conducting an in-depth analysis of the forecasting performance of linear nowcasting models in the Netherlands. The sample covers the quarters 1995.I–2010.IV and allows comparison of the forecasting quality of the models during both the dot-com recession (2001.I–2003.III) and the financial crisis (2008.I–2010.IV). Furthermore, the forecasting accuracy of the mechanical linear models is confronted with the forecasting accuracy of professional analysts, using a new dataset of the quarterly GDP forecasts of professional analysts. For policy makers and market participant the forecasts of professionals can be a cheap and easy to use alternative for using statistical models. Moreover, they could reflect much more information than the information contained in the monthly indicators alone, which is inevitable limited. Lastly, it is evaluated whether the forecasts of professional analysts could have enhanced the forecasts of the mechanical models.

The main conclusion is that the dynamic factor model has the highest forecasting accuracy of all linear nowcasting models considered. Moreover, the forecasting accuracy of the dynamic factor model is higher than the forecasts of professional analysts both during periods of heightened volatility (financial crisis, dot-com recession) as well as during more tranquil times. Interestingly, the forecasts of mechanical linear models can be enhanced with the forecasts of the professional analysts, but the advantage is limited to the period of the financial crisis.

The remainder of this chapter is structured as follows. Section 2.2 describes the quarterly forecasts of professional analysts and their accuracy during the financial crisis and the dot-com recession. Section 2.3 discusses the mechanical linear nowcasting models, while Section 2.4 outlines the empirical setup of the analysis. Section 2.5 reports the empirical results. Section 2.6 concludes.

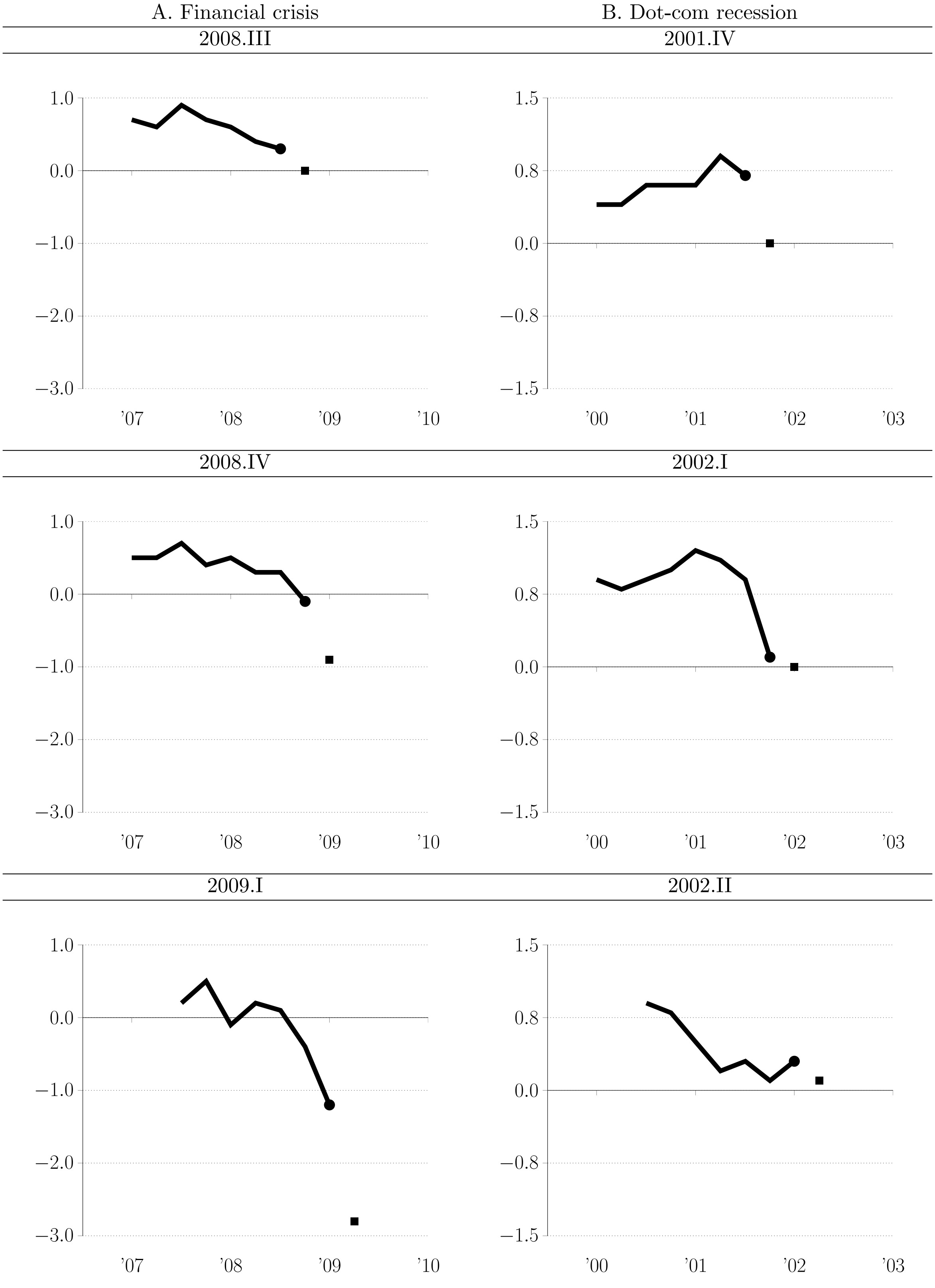
2.2 Quarterly forecasts of professional analysts

The private firm Consensus Economics has been collecting and publishing forecasts on a monthly basis under the name of *Consensus forecasts*. The Consensus forecasts provide the forecasts of professional analysts for a set of key macroeconomic variables for a broad range of countries. Consensus forecasts is best known for the *annual* GDP growth projections for the current and next year, which have been analyzed in several papers (e.g. Ager et al. 2009; Loungani and Rodriguez 2008). However, once a quarter Consensus Economics also provides the averaged forecasts of professional analysts for the *quarterly* real GDP growth rate over a horizon of one up to six quarters, starting with the nearest quarter for which no officially released GDP figure is available. The survey data (deadline for respondents) is typically the last Monday of the third month of a quarter. The number of respondents varies somewhat over time, but on average there are nine institutions participating in the poll for the Netherlands, mainly banks. The quarterly real GDP growth forecasts for the Netherlands have been published since December 1994.

The black lines in Figure 2.1 show the evolving mean quarterly GDP growth forecast of the panelists in the quarters leading up to the financial crisis (2008.III–2009.I; panel A) and the dot-com recession (2001.IV–2002.II; panel B). The black dots indicate the last Consensus forecasts before the first (or "flash") release of GDP growth. The black squares indicate the flash GDP release. The horizontal axes show the dating of the forecasts, whilst the vertical axes show the mean Consensus forecasts. Figure 2.1 clearly indicates that the collapse of Lehman Brothers in September 2008 started a sequence of downward adjustments to the quarterly real GDP growth forecasts of professional analysts, but these adjustments ultimately underestimated the size of the downturn in real GDP growth, i.e. the black dots all indicate a higher real GDP growth rate than the black squares. Panel B of Figure 2.1 shows panelists also overestimated real GDP growth in the quarters leading up to the dot-com recession, although the overestimation was somewhat smaller.

The apparent sluggish adjustment of the *quarterly* Consensus forecasts is in line with parts of the literature analyzing the *annual* Consensus forecasts. Some authors claim that professional analysts do not revise their forecasts promptly and sufficiently to reflect incoming (foreign) news because they are excessively cautious or because they want to smooth their forecasts (Loungani et al. 2013; Isiklar et al. 2006). The seemingly slow adjustment of the quarterly forecasts might be partly traced back to the difficult task of aggregating the vast amount of monthly incoming information in a consistent manner. The main difficulty is that the monthly indicators can, and often do, provide conflicting signals, and it is no easy task to transform these signals into a forecast of real GDP growth. Moreover, most monthly indicators have sizable publication lags, which cause "ragged edges" in the dataset (Bańbura et al., 2011). For instance, in July 2008 analysts only had information on real GDP growth until 2008.I. They did have more recent information on industrial production and retail trade, but the latest monthly figure was for May 2008. The most recent information stemmed from financial markets, which is available on a daily basis, and survey data, which are promptly

Figure 2.1: Quarterly GDP forecasts of Consensus forecasts, 2008.III–2009.I



Notes: the black lines denote the evolving Consensus forecasts. The black dots denote the last Consensus forecasts prior to the flash release of GDP growth. The black squares indicate the flash release of GDP growth.

available at the end of the month.

2.3 Mechanical linear nowcasting models

An alternative to using the forecasts of professional analysts is to use pure mechanical statistical models to forecast quarterly GDP growth. The main difference with the professional analysts is that the mechanical models do not contain any subjective interpretation, but just let the data speak. From an econometric viewpoint, combining all monthly information to forecasts GDP growth in the current and adjacent quarters is not a straightforward problem. Regression by ordinary least squares is not viable since the number of monthly indicators is too large to estimate all model coefficients ("curse of dimensionality"). Moreover, ordinary least squares is not able to handle the "ragged edged" structure of the monthly dataset. In the literature, there are basically two approaches to overcome these estimation problems. In the first approach, the information in the monthly dataset is summarized in a limited number of series. This approach exploits the fact that the auxiliary variables are correlated. Principle components analysis is used to replace a large number of correlated time series with a limited number of uncorrelated (unobserved) factors representing the common information component of the original data series. The factors serve as inputs for the forecasting procedure in the next step. This so-called factor model approach has been shown to provide relatively accurate forecast in the United States (Giannone et al., 2008), the euro area (see Bańbura et al., 2011; Rünstler et al., 2009) and the Netherlands (den Reijer, 2013). The second approach starts with the computation of indicator-specific forecast of GDP growth, which are then aggregated into a single final GDP growth forecast in the second step (Timmermann, 2006). To benchmark both approaches a AR(1) model of GDP is estimated. Within the second approach a variety of specifications is possible. In this chapter the quarterly bridge equation and the quarterly vector autoregressive model are considered. Chapter 3 presents a more comprehensive comparison of model specifications.

2.3.1 Extracting information through factors

Recently, the use of factor models to forecast near-term GDP growth has become quite popular among academics and practitioners at central banks. See, for example, Stock and Watson (2002a) and Giannone et al. (2008) for the United States, Angelini et al. (2011) for the euro area, Schneider and Spitzer (2004) for Austria, Schumacher and Breitung (2008) for Germany, Barhoumi et al. (2010) for France and den Reijer (2013) for the Netherlands. In this chapter the so-called dynamic factor model proposed by Bańbura and Rünstler (2011) is used. A key feature of this models is the use of the

Kalman filter, which allows for an efficient handling of the unbalancedness of the dataset and the different frequencies of the data.[2] The Kalman filter replaces any missing monthly indicator observations with optimal predictions and also generates estimates of unobserved monthly real GDP subject to a temporal aggregation constraint for the quarterly observation. The first equation of the model is:

$$x_m = \Lambda f_m + \xi_m, \qquad \xi_m \sim N(0, \Sigma_\xi) \tag{2.1}$$

which relates the n monthly indicators $x_m = (x_{1,m}, \ldots, x_{n,m})'$ to r monthly static factors $f_m = (f_{1,m}, \ldots, f_{r,m})'$ via a matrix of factor loadings Λ and an idiosyncratic component $\xi_m = (\xi_{1,m}, \ldots, \xi_{n,m})'$, where $r << n$ and m is a monthly time index. The dynamic factor model assumes that the idiosyncratic components are a multivariate white noise process, hence the covariance matrix Σ_ξ is diagonal. Furthermore, the dynamic factor model assumes that the factors follow a vector-autoregressive process of order p:

$$f_m = \sum_{s=1}^{p} A_s f_{m-s} + \zeta_m, \qquad \zeta_m \sim (0, \Sigma_\zeta) \tag{2.2}$$

where A is a square $r \times r$ matrix. Moreover, the covariance matrix of the VAR (Σ_ζ) is driven by a q dimensional standardized white noise process η_m:

$$\zeta_m = B\eta_m, \qquad \eta_m \sim N(0, I_q) \tag{2.3}$$

where B is a $r \times q$ matrix and $q \leq r$ by definition. The final equation is a forecasting equation linking the factors to (unobserved) mean-adjusted real GDP growth:

$$y_m = \beta' f_m + \varepsilon_m, \qquad \varepsilon_m \sim N(0, \sigma_\varepsilon^2) \tag{2.4}$$

where y_m denotes the (unobserved) three-month growth rate of monthly real GDP, i.e. the growth rate vis-à-vis the same month of the previous quarter. Quarterly real GDP growth in quarter t, y_t^Q, is assigned to month $3t$ on the monthly time scale. The relation between the quarterly and monthly GDP growth rates is given by $y_t^Q = \frac{1}{3}(y_{3t} + y_{3t-1} + y_{3t-2})$.

The model is estimated in four steps. In the first step the factors loadings Λ and the estimated static factors $\hat{f}_m$ are obtained. In the second step the coefficient matrices A_s in Eq. (2.2) and β in Eq. (2.4) are estimated by OLS using $\hat{f}_m$. In the third step, ζ_m and its covariance matrix Σ_ζ are computed, and an estimate of the matrix B is obtained by principal components analysis. In the final step, the model is cast in state space form and the Kalman filter and smoother are used to re-estimate the estimated factors ($\hat{f}_m$) and monthly GDP growth.

[2] See Durbin and Koopman (2012) for a comprehensive treatment of state space models and the use of the Kalman filter and smoother. See Chapter 5 for a comparison of the forecasting accuracy of different factor model specifications.

To estimate the model, the number of static and dynamic common factors need to specified, denoted by r and q respectively. The largest possible value of r is set at 8, based on the scree test of Cattell (1966). The maximum value of p is set at 3. The specification that was used in the main text is r=5, q=4 and p=2, using the root mean squared forecast error as a selection criterion, following Matheson (2013) and Rünstler et al. (2009). Alternatively, one could choose the number of factors r and q on the basis of in-sample criteria, as described in Bai and Ng (2002, 2007). However, preliminary estimates indicated that these criteria tend to indicate a relatively large number of factors, in line with the outcome in Bańbura and Rünstler (2011), leading to volatile and less accurate forecasts.

2.3.2 Extracting information by pooling

This section describes two approaches that pool the monthly information, i.e. quarterly bridge equations (BEQ) and quarterly bivariate vector autoregressive models (QVAR).

Quarterly bridge equation (BEQ)

The quarterly bridge equation is a widely used method for forecasting GDP growth using all available observations of monthly indicators; for applications see Kitchen and Monaco (2003) and Baffigi et al. (2004). Bridge equations are linear regressions that "bridge" monthly variables, such as industrial confidence and retail sales, to quarterly GDP. Various specifications are possible within this approach. Here, a simple version of the bridge equation is proposed, proceeding in two steps. Firstly, predictions of the necessary monthly values of indicator x_i are obtained over the forecasting horizon with the help of univariate autoregressive models and aggregated to appropriate quarterly values x_i^Q. Secondly, these quarterly aggregates are used to predict GDP. The bridge model for x_i is:

$$y_t^Q = \alpha + \sum_{s=0}^{p} \beta_s x_{i,t-s}^Q + \varepsilon_{i,t}^Q, \qquad \varepsilon_{i,t}^Q \sim N(0, \sigma^2_{\varepsilon^Q}) \tag{2.5}$$

where α is a constant, p denotes the number of lags in the bridge equation and ε_i^Q is a normally distributed error-term. Eq. (2.5) is estimated for each of the n indicators. The final forecast is then calculated by weighting the n indicator-specific forecasts for each horizon. The lag parameter p in Eq. (2.5) is determined recursively by the Schwartz information criterion (SIC).

Quarterly vector autoregressive model (QVAR)

The VAR approach is very similar to the bridge equation approach. Unlike bridge equations, VAR models use the information content of GDP itself to produce forecasts of

GDP (e.g. Camba-Mendez et al., 2001). Moreover, it is a system approach, attempting to exploit the interdependence of indicator and real GDP dynamics. However, misspecification anywhere in the system may affect the accuracy of the GDP predictions. More importantly, the QVAR model only uses monthly observations that correspond to a full quarter. Consequently, it does not fully exploit the available monthly information. In total, n quarterly bivariate VAR models that include one of the indicators and GDP growth were estimated:

$$z_{i,t}^{Q} = \alpha + \sum_{s=1}^{p_i} A_s z_{i,t-s}^{Q} + \varepsilon_{i,t}^{Q}, \qquad \varepsilon_{i,t}^{Q} \sim N(0, \Sigma_{\varepsilon^Q}) \tag{2.6}$$

where $z_{i,t}^{Q} = (y_t^Q, x_{i,t}^Q)'$. From each bivariate VAR an indicator-specific quarterly GDP forecast for quarter $t+h$ at time t is obtained, denoted as $y_{t+h|t}^{Q}$. As in the case of the bridge equations, the final forecast is constructed in the second stage, as a weighted average of the individual indicator model forecasts. The lag parameter p is determined recursively by the SIC.

The BEQ and QVAR models construct a large number of different indicator specific forecasts in the first stage described above, which have to be aggregated in the second stage. The weights are inversely proportional to the root mean squared forecast error (RMSFE), and calculated recursively; i.e. from the start of the sample period until the previous quarter.

2.4 Data and forecast design

This section describes the dataset (Section 2.4.1) and the pseudo real-time design (Section 2.4.2).

2.4.1 Dataset

The dataset consists of eighty monthly time-series variables that are spread over four groups: hard, quantitative information (30), financial variables (11), prices (11) and soft, qualitative information (28). Most of the series refer to the Netherlands, but also included are series referring to important trading partners of the Netherlands, to take into account the important role of exports for explaining growth in the small open Dutch economy. Appendix 2.A provides details on the sources, availability and the applied transformations of the data series. The available monthly data are usually already adjusted for seasonality (and calendar effects). When necessary, raw data series are seasonally adjusted using the US Census X-12 method. All monthly series are made stationary by differencing or log-differencing (in the case of trending data, such as industrial production, retail sales and monetary aggregates). Finally, each variable is standardized by subtracting the mean and dividing by the standard deviation. This

Table 2.1: Timing of forecasting exercise for third quarter GDP growth

Nr.	Forecast type	Month	Forecast made in middle of
1	Two-quarter ahead	1	January
2		2	February
3		3	March
4	One-quarter ahead	1	April
5		2	May
6		3	June
7	Nowcast	1	July
8		2	August
9		3	September
10	Backcast	1	October
11		2	November

normalization is standard practice in order to avoid the overweighting of series with large-variances series in the extraction of common factors.

2.4.2 Pseudo real-time design

The aim is to replicate the availability of the data at the time the forecast were made in order to mimic as closely as possible the real-time flow of information. To this end, a dataset that was downloaded on May 10th, 2011 was combined with the typical data release calendar to reconstruct the available dataset on the 10th of each month during the period January 1995 until December 2010. All monthly series start in January 1985, whilst the quarterly GDP series start in 1985.I. This approach is called pseudo real-time, since it takes into account the publication delays in the data, but does not take into account the possibility of data revisions. Abstracting from data revision may affect the comparison of mechanical models and forecast by analysts to a certain extent. The expectations of professional analysts necessarily reflect the inaccurate initial estimates of GDPs recent past and this puts them at a disadvantage vis-à-vis mechanical models in a pseudo real-time setting as the latter can take data revisions on board.[3]

The parameters of all models are estimated recursively using only the information available at the time of the forecast. This approach is regularly used in empirical studies (see Rünstler et al., 2009; Giannone et al., 2008; Garratt et al., 2008). More precisely, a sequence of eleven forecasts for GDP growth in a given quarter is considered, obtained in consecutive months. Table 2.1 explains the timing of the forecasting exercise, taking the forecast for the third quarter of 2008 as an example. The first forecast is made in January 2008, which is called the two-quarter ahead forecast in month one. Subsequently, monthly forecasts are produced for the next ten months through Novem-

[3] See Roodenburg and den Reijer (2006) for a discussion of revisions in Dutch GDP figures and Croushore (2011) and Chapter 4 of this thesis for a discussion of the effects of data revisions on forecast accuracy.

ber. The last forecast is made just before the first release of GDP in mid-November. Following the conventional terminology, *forecasts* refer to one or two-quarter ahead forecasts, *nowcasts* refer to current quarter forecasts and *backcasts* refer to forecasts for the preceding quarter, before official GDP figures become available. In case of the current example, i.e. 2008.III, the models produce two-quarter ahead forecasts from January to March, one-quarter ahead forecasts from April to June, nowcasts from July to September, and backcasts in October and November.

2.5 Empirical results

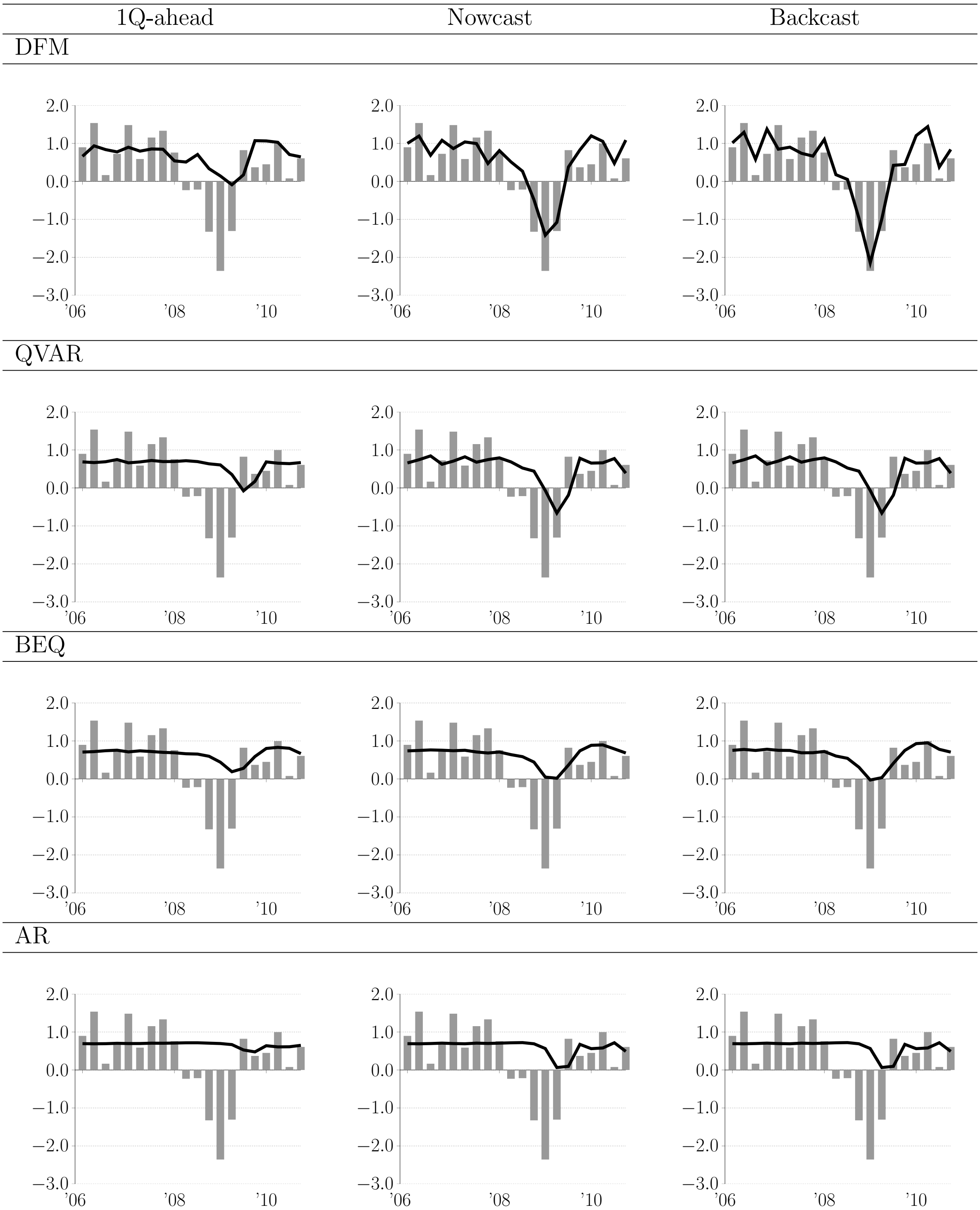
This section presents the outcome of two forecasting "horse races". The first race is between the mechanical linear nowcasting models. Its aim is to determine the model with the highest forecasting accuracy before and during the financial crisis. The second race is between the best performing mechanical linear model and the average Consensus forecast.

2.5.1 A first look at the outcomes

Figure 2.2 shows the forecasts of the mechanical linear models (black lines) against the GDP realizations (grey bars). Moving from left to right the graphs show the forecasting performance of the one-quarter ahead forecasts, nowcasts and backcasts in the period surrounding the financial crisis.[4] The horizontal axis shows which quarter the model is forecasting, the vertical axis shows the GDP realizations. The outcomes in Figure 2.2 can be summarized in three points. Firstly, on the one-quarter ahead forecasting horizon, all models missed the abrupt slowdown at the end of 2008. As the forecast horizon shortens the forecasting performance of the models steadily increases. Secondly, starting from the nowcast, the dynamic factor model is the only model that clearly indicates an abrupt slowdown at the end of 2008, whilst the backcast was relatively close to the published GDP growth rate. Thirdly, the forecasts of the BEQ and QVAR show very little dynamics: most forecasts fluctuate only slightly around trend growth (about 0.5 percent quarter-on-quarter over the period 1995.I–2010.IV). Overall, the forecasts of the BEQ and QVAR as well as the naïve benchmark were only revised downward following the publication of the (very) negative GDP growth rate for 2008.IV, in February 2009. In contrast, the dynamic factor model was much quicker in acknowledging the quick deterioration in the economy. Overall, the graphical analysis indicates that the dynamic factor model is more efficient in translating the flow of monthly information to an accurate GDP growth forecast.

[4] Figure 2.2 only shows the forecasting performance for the one-quarter ahead forecasts, nowcasts and backcasts in the third month (second month for backcasts) to keep the presentation parsimonious.

Figure 2.2: Quarterly GDP forecasts of mechanical linear models, 2006.I–2010.IV



Notes: the black lines denote the 1Q-ahead forecasts, nowcasts and backcasts of the mechanical linear models on month 3, 3 and 2, respectively. The grey bars denote GDP growth realisations. DFM: dynamic factor model, QVAR: quarterly vector autoregressive model, BEQ: quarterly bridge equation, AR: AR(1) model.

2.5.2 Forecasting performance linear models

Table 2.2 shows the out-of sample forecasting performance of the linear models. The forecast accuracy is measured by the root mean squared forecast error (RMSFE). The Diebold-Mariano test (Diebold and Mariano, 1995) indicates the statistical significance of the differences in forecasting accuracy. Entries in bold indicate the model with the highest forecasting accuracy, i.e. the lowest RMSFE. Grey cells indicate that the forecasting accuracy of the best performing model is significantly better according to the Diebold-Mariano test (Diebold and Mariano, 1995). Panel A shows the performance for the complete sample period 1995.I–2010.IV. Panel B shows the performance for the whole sample period, excluding the financial crisis (1995.I–2007.IV). Panel C shows the performance for the whole sample period excluding both the dot-com recession and the financial crisis (1995.I–2000.IV and 2003.IV–2007.IV).

The outcomes in Table 2.2 point to several interesting results. First, the dynamic factor is almost unequivocally the best forecasting model across forecast horizons when considering the whole sample. The dynamic factor model reduces the RMSFE against a naïve autoregressive model considerably, by up to 50% for the backcast in month 2 (panel A in Table 2.2). The forecasting advantage of the dynamic factor decreases as the forecasting horizon increases. Consequently, the difference in forecasting accuracy between the dynamic factor model and the other models is very small for the two-quarter ahead forecasts. This outcome is an indication that the mechanical nowcasting models considered are less suited for forecasting purposes, when no monthly data on the pertaining quarter are available. Second, the dynamic factor model is relatively well equipped to translate sharp drops in the monthly variables to sharp drops in GDP growth. The other models clearly had more difficulty forecasting the depth and timing of the crisis, as can be seen by the much higher RMSFEs (panel A in Table 2.2). Thirdly, the dynamic factor model seems especially advantageous in periods of abrupt changes, as can be seen from the relatively favorable forecasting performance during the financial crisis and the dot-com recession. Excluding both the financial crisis and the dot-com recession, the forecasting accuracy of the dynamic factor is only superior to the other models when backcasting (panel C in Table 2.2).

The fact that the dynamic factor model outperforms the other models does not necessarily mean the other model forecasts do not contain any additional information. To test if the models are complementary, i.e. if pooling of two model types increases forecasting accuracy, encompassing tests were conducted. The test regression, proposed by Granger and Ramanathan (1984) and used by among others, Fair and Shiller (1990) and Liebermann (2014), is:

$$y^Q_{t+h} = \alpha + \beta \hat{y}^Q_{\mathrm{b}(t+h|t)} + \gamma \hat{y}^Q_{\mathrm{a}(t+h|t)} + \varepsilon_t \tag{2.7}$$

where y^Q_{t+h} is the GDP growth in quarter $t+h$, $\hat{y}^Q_{\mathrm{b}(t+h|t)}$ and $\hat{y}^Q_{\mathrm{a}(t+h|t)}$ are the predictions

Table 2.2: Forecasting performance of mechanical statistical models, 1995.I–2010.IV

	RMSFE														
	A. Whole sample					B. Sample without financial crisis					C. Sample without financial crisis and dot-com recession				
	Month	AR	BEQ	QVAR	DFM	Month	AR	BEQ	QVAR	DFM	Month	AR	BEQ	QVAR	DFM
2Q-ahead	1	0.73	0.72	0.73*	**0.72**	1	0.52	0.52	0.53*	**0.52**	1	**0.44**	0.45	0.46	0.45
	2	0.73*	0.71	0.73*	**0.71**	2	0.52	0.52	0.53*	**0.51**	2	**0.44**	0.45	0.46	0.45
	3	0.73*	0.71	0.72*	**0.70**	3	0.52	0.51	0.52	**0.51**	3	**0.44**	0.45	0.45	0.46
1Q-ahead	1	0.73*	0.70	0.72*	**0.69**	1	0.52	0.51	0.52	**0.50**	1	**0.44**	0.45	0.45	0.46
	2	0.73*	0.69	0.72*	**0.67**	2	0.52	0.50	0.52	**0.49**	2	**0.44**	0.45	0.45	0.46
	3	0.72*	0.68**	0.71**	**0.62**	3	0.52*	0.50*	0.52**	**0.46**	3	**0.44**	0.45	0.46	0.44
Nowcast	1	0.72*	0.67*	0.71*	**0.59**	1	0.52	0.50	0.52	**0.47**	1	**0.44**	0.45	0.46	0.45
	2	0.72*	0.66*	0.71*	**0.53**	2	0.52	0.49	0.52*	**0.46**	2	**0.44**	0.45	0.46	0.45
	3	0.71*	0.63*	0.62*	**0.48**	3	0.54*	0.49	0.50**	**0.45**	3	0.45	0.44	0.46	**0.44**
Backcast	1	0.71**	0.62*	0.62**	**0.41**	1	0.54***	0.48***	0.50***	**0.41**	1	0.45***	0.44	0.46*	**0.40**
	2	0.71**	0.61**	0.62**	**0.39**	2	0.54***	0.48***	0.50***	**0.39**	2	0.45***	0.43**	0.46**	**0.38**

Notes: entries denote RMSFEs. Entries in bold indicate the model with the lowest RMSFE. Grey cells indicate that the forecasting accuracy of the best performing model is significantly better according to the Diebold-Mariano test. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. AR: AR(1) model, BEQ: quarterly bridge equation. QVAR: quarterly vector autoregressive model. DFM: dynamic factor model.

Table 2.3: Forecast encompassing test of dynamic factor model versus alternative models, 1995.I–2010.IV

A.Coefficient dynamic factor model (β)

	Whole Sample				Without financial crisis				Sample without financial crisis and dot-com recession			
	Month	AR	BEQ	QVAR	Month	AR	BEQ	QVAR	Month	AR	BEQ	QVAR
	1	1.26	1.13	1.28	1	0.47	0.41	0.48	1	-0.18	-0.19	-0.18
2Q-ahead	2	1.53**	1.29	1.52**	2	0.59	0.57	0.60	2	-0.25	-0.25	-0.24
	3	1.61**	1.32	1.61**	3	0.95	0.80	0.95	3	-0.57	-0.58	-0.57
	1	1.59***	1.09	1.57***	1	1.14	1.00	1.19*	1	-0.53	-0.63	-0.57
1Q-ahead	2	1.69***	1.15*	1.68***	2	1.33**	1.29*	1.41**	2	-0.44	-0.50	-0.50
	3	2.12***	1.93***	2.18***	3	2.01***	1.92***	2.04***	3	0.37	0.38	0.35
	1	1.59***	1.27***	1.61***	1	1.19***	1.02**	1.20***	1	-0.16	-0.13	1.20***
Nowcast	2	1.24***	1.02***	1.24***	2	0.93***	0.77**	0.93***	2	0.24	0.24	0.24
	3	1.23***	1.03***	1.22***	3	0.90***	0.72**	0.92***	3	0.37	0.31	0.37
	1	1.08***	1.08***	1.08***	2	1.01***	0.91***	1.03***	2	0.67**	0.66**	0.68**
Backcast	2	1.06***	1.00***	1.03***	1	0.97***	0.85***	0.98***	1	0.71***	0.72***	0.72***

B.Coefficient alternative model (γ)

	Whole Sample				Without financial crisis				Sample without financial crisis and dot-com recession			
	Month	AR	BEQ	QVAR	Month	AR	BEQ	QVAR	Month	AR	BEQ	QVAR
	1	-0.70	0.27	-0.74	1	-3.55*	-1.12	-2.86*	1	1.83	0.85	1.40
2Q-ahead	2	-0.45	0.67	-0.46	2	-3.50*	-0.54	-2.81*	2	1.83	0.67	1.39
	3	0.10	0.66	-0.01	3	-2.81	0.29	-1.42	3	2.15	0.87	1.85
	1	0.67	1.27	0.35	1	-2.63	0.74	-1.38	1	2.06	0.97	1.84
1Q-ahead	2	1.34	1.34	0.77	2	-2.20	0.72	-1.16	2	2.01	0.68	1.83
	3	1.03	0.51	-0.25	3	-0.27	0.46	-0.71	3	1.04	0.32	0.55
	1	1.25	1.01	-0.07	1	-0.23	1.14	-0.43	1	1.16	0.36	0.74
Nowcast	2	1.35	0.92	0.08	2	-0.18	1.17	-0.36	2	1.00	0.22	0.54
	3	-0.24	0.62	-0.03	3	-0.81	1.74*	0.61	3	0.48	0.70	0.04
	1	-0.18	-0.05	-0.05	2	-0.60	0.97	0.47	2	0.45	0.11	-0.12
Backcast	2	-0.14	0.20	0.07	1	-0.82	1.07	0.57	1	0.28	0.01	-0.12

Notes: entries denote the estimated coefficients of the dynamic factor model (β) and the alternative model (γ), respectively. *, ** or *** denotes that the estimated coefficient is statistically different from 0 at the 10%, 5% or 1% significance level, respectively. Grey cells indicate β is significantly different from zero and γ is *not* significantly different from zero. AR: AR(1) model, BEQ: quarterly bridge equation. QVAR: quarterly vector autoregressive model.

for quarter $t+h$ on time t of the best (b) and alternative (a) model respectively. The alternative model does *not* contain any additional information with respect to the best model *if and only if* γ is not significantly different from zero and β is significantly different from zero. Both models contain useful information if β as well as γ are significantly different from zero. Table 2.3 shows the coefficients of the dynamic factor model (β, panel A) and the alternative model (γ, panel B) in the encompassing test. Grey cells indicate the alternative model contains no extra information with regards to the dynamic factor model (β significantly different from zero, γ *not* significantly different from zero). In other words, the most accurate forecasts are made by using the dynamic factor model alone instead of a combination of the dynamic factor model and one of the other statistical models.

The main message from Table 2.3 is that there is *no* gain in forecasting accuracy from combining the dynamic factor model with any of the other models when (severe) recessions are included in the sample, except for the (first month(s)) of the two-quarter ahead forecasts. This conclusion also holds when the period of the financial crisis is excluded. When excluding both the dot-com recession and the financial crisis, there seems to be an advantage form combining the dynamic factor model and the alternative models. However, since both β and γ are *not* significantly different from zero it actually means that the models have no value added compared to a forecast that equals the mean growth rate of GDP over the sample period (α in Eq. (2.7)). Overall, the results indicate that there is little gain in forecasting accuracy from combining the dynamic factor model with any of the other models.

2.5.3 Forecasting performance professional analysts

This section compares the forecasting accuracy of the dynamic factor model with the forecasting accuracy of professional analysts. Professional analysts can use much more information than just the monthly indicators in the mechanical models, and could add a –possibly sizable– judgmental element to their near-term forecasts. The evidence on the size and frequency of these judgmental adjustment is scarce however. One of the few pieces of information on judgmental adjustment in the near-term forecasts of analysts can be found in a recent questionnaire conducted by the European Central Bank (ECB) among the participants of the ECB Survey of Professional Forecasters (Meyler and Rubene, 2009). The main finding of Meyler and Rubene (2009) is that panelists regard approximately 40% of their short-term GDP forecasts as being judgment-based. Table 2.4 presents the forecasting performance of the Consensus forecasts and the dynamic factor model. The comparison is limited to forecasts in the third month of the quarter, when Consensus Economics releases a *fresh* Consensus forecasts. It also means that the comparison does not include the backcasts, since GDP is released approximately two weeks prior the release of the Consensus forecasts.

Table 2.4: Forecasting performance of dynamic factor model versus Consensus forecasts, 1995.I–2010.IV

	A. Root Mean Squared Error		
	2Q-ahead	1Q-ahead	Nowcast
Whole sample			
DFM	0.70	0.62	0.48**
CF	0.67	0.62	0.56
Without financial crisis			
DFM	0.51**	0.46***	0.45**
CF	0.52	0.52	0.49
Without fin. crisis and dot-com recession			
DFM	0.46**	0.44**	0.44**
CF	0.51	0.52	0.52

Notes: entries denote RMSFEs. Entries in bold indicate the approach (DFM or CF) with the lowest RMSFE. Grey cells indicate that the forecasting accuracy of the best performing approach (DFM of CF) is significantly better according to the Diebold-Mariano test. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. DFM: dynamic factor model, CF: Consensus forecasts.

	B. Encompassing test		
	2Q-ahead	1Q-ahead	Nowcast
Whole sample			
β	1.22	1.60**	1.11***
γ	0.55**	0.38*	0.33*
Without financial crisis			
β	1.55**	1.88***	0.80**
γ	0.04	0.08	0.26
Without fin. crisis and dot-com recession			
β	0.39	0.97	1.01
γ	0.07	0.27	0.21

Notes: entries denote the estimated coefficients of the dynamic factor model (β) and the Consensus forecasts (γ), respectively. *, ** or *** denotes that the estimated coefficient is statistically different from 0 at the 10%, 5% or 1% significance level, respectively. Grey cells indicate β is significantly different from zero and γ is *not* significantly different from zero.

Panel A of Table 2.4 compares the RMSFE of the Consensus forecasts with the RMSFE of the dynamic factor model over the whole sample period, the sample period without the financial crisis and the sample period without both the financial crisis and the dot-com recession. Over the whole sample period, the dynamic factor model only beats the Consensus nowcasts. This seems to indicate the professional analysts are a tough competitor for the dynamic factor model at the one and two-quarter forecasting horizon. However, closer inspection reveals that this relative good performance of the Consensus forecasts is limited to the period excluding the financial crisis. The results for the sample period excluding the financial crisis, clearly reveals the dynamic factor model beats the Consensus forecasts on all forecasting horizons.

Panel B of Table 2.4 presents the outcome of the encompassing test of the dynamic factor model against the Consensus forecasts. The results indicate that there is a clear gain from combining the dynamic factor model and Consensus forecasts as both β and γ are significantly different from zero. However, again, this gain in forecasting accuracy is limited to the period of the financial crisis. Without this period the forecasting accuracy is higher when only using the dynamic factor model. In line with the outcome of the encompassing test of the mechanical models in Table 2.3, taking out both the financial crisis and the dot-com recession drastically diminishes the value added of using either the dynamic factor model or the Consensus forecasts (both β and γ are insignificant). Overall, combining the Consensus forecasts with the dynamic factor model could have increased the forecasting accuracy during the financial crisis, when it really counts.

2.6 Conclusion

This chapter makes two contributes to the empirical literature on forecasting real GDP in the short run. The first contribution is a systematic comparison of the forecasting accuracy of three popular statistical linear models for the Netherlands. The sample period (1995.I–2010.IV) allows making a comparison between the forecasting accuracy of the models during the financial crisis and the pre-crisis period. The main finding is that, in the analyzed sample, the dynamic factor model has a higher forecasting accuracy than the alternative forecasting models considered, especially surrounding periods of increased volatility. Enhancing the forecasts of the dynamic factor model with the forecasts of the other linear models does not (significantly) increase the forecasting accuracy. The second contribution concerns the potential usefulness of (subjective) forecasts made by professional analysts. Interestingly, forecasts by professional analysts appear to be quite different from mechanical models, at least in the analyzed sample. During periods of heightened volatility, such as the financial crisis, the near-term forecasts for Dutch GDP growth by professional analysts seem to embody information ("judgment") that the analyzed mechanical models fail to incorporate.

Appendix

2.A Dataset

The main source of the monthly data is the statline database of Statistics Netherlands (CBS). Data on world trade are from CPB Netherlands Bureau for Economic Policy Analysis (CPB). Commodity prices and most financial market variables are taken from Thomson Reuters datastream. Table 2.5 provides an overviews of all monthly series and the applied transformations. The data can be classified into four categories: hard, quantitative information (hard), consumer and producer prices (price), financial and monetary variables (financial) and soft, qualitative information (soft).

Table 2.5: Description monthly dataset

No.	Variable	Type	Transformation			Lag
			ln.	dif.	filter	
1	Ind. prod. - industry (total)	hard	1	1	3	2
2	Ind. prod. - manufacturing (total)	hard	1	1	3	2
3	Ind. prod. - manufacture of wearing apparel	hard	1	1	3	3
4	Ind. prod. - manufacture of motor vehicles and (semi) trailers	hard	1	1	3	3
5	Ind. prod. - manufacture of other transport equipment	hard	1	1	3	3
6	Ind. prod. - manufacture of basic metals and metal products	hard	1	1	3	2
7	Ind. prod. - treatment and coating of metals and machines	hard	1	1	3	3
8	Ind. prod. - electricity, gas, steam, water and air conditioning	hard	1	1	3	2
9	Ind. prod. - manufacture of textiles	hard	1	1	3	3
10	Ind. prod. - printing and reproduction of recorded media	hard	1	1	3	3
11	Ind. prod. - construction	hard	1	1	3	2
12	Ind. prod. - manufacture of food products and beverages	hard	1	1	3	2
13	Ind. prod. - mig capital goods industry	hard	1	1	3	2
14	Ind. prod. - mig durable consumer goods industry	hard	1	1	3	2
15	Ind. prod. - mig non-durable consumer goods industry	hard	1	1	3	2
16	Ind. prod. - consumer goods industry	hard	1	1	3	2
17	Ind. prod. - Belgium (total)	hard	1	1	3	3
18	Ind. prod. - Germany (total)	hard	1	1	3	2
19	Ind. prod. - Spain (total)	hard	1	1	3	2
20	Ind. prod. - France (total)	hard	1	1	3	2
21	Ind. prod. - United Kingdom (total)	hard	1	1	3	2
22	Ind. prod. - Italy (total)	hard	1	1	3	2
23	Car registration - new commercial vehicles	hard	1	1	3	2
24	Car registration - new passenger car	hard	1	1	3	1
25	Retail trade turnover (total)	hard	1	1	3	2
26	Consumption expenditure by households	hard	1	1	3	2
27	Unemployment	hard	0	1	3	1
28	World Trade	hard	1	1	3	2
29	Imports	hard	1	1	3	2
30	Exports	hard	1	1	3	2
31	M1	financial	1	2	3	2
32	M3	financial	1	2	3	2
33	Interest rate (short-term)	financial	0	1	3	1
34	Interest rate (long-term)	financial	0	1	3	1

Continued on next page...

Table 2.5 – Continued

No.	Variable	Type	Transformation ln.	dif.	filter	Lag
35	Exchange rate, US-Dollar per Euro	financial	0	1	3	1
36	Loans to the private sector	financial	1	1	3	2
37	Loans on mortgage (nominal rate 5 to 10 years mortgage)	financial	0	1	3	2
38	Share index, AEX	financial	1	1	3	1
39	Share index, Amsterdam Midkap	financial	1	1	3	1
40	Share index, Dow Jones Euro Stoxx 50	financial	1	1	3	1
41	Share index, Dow Jones Euro Stoxx Industrials	financial	1	1	3	1
42	Consumer price index, total	price	1	2	3	2
43	Consumer price index, underlying inflation	price	1	2	3	2
44	World market commodity prices, industrial materials	price	1	2	3	2
45	Producer prices, final products, domestic market	price	1	2	3	2
46	Producer prices, investment goods, domestic market	price	1	2	3	2
47	Producer prices, intermediate goods, domestic market	price	1	2	3	2
48	Producer prices,foreign market	price	1	2	3	2
49	Producer prices, domestic market	price	1	2	3	2
50	Terms of trade	price	1	2	3	2
51	Import prices	price	1	2	3	2
52	Export prices	price	1	2	3	2
53	Constr. confidence (headline)	soft	0	1	3	1
54	Constr. confidence - building dev. (past 3 months)	soft	0	1	3	1
55	Constr. confidence - evolution overall order books	soft	0	1	3	1
56	Constr. confidence - employment expect. (next 3 months)	soft	0	1	3	1
57	Ind. confidence (headline)	soft	0	1	3	1
58	Ind. confidence - production trend observed in recent months	soft	0	1	3	1
59	Ind. confidence - assessment of order-book levels	soft	0	1	3	1
60	Ind. confidence - assessment of export order-book levels	soft	0	1	3	1
61	Ind. confidence - assessment of stocks of finished products	soft	0	1	3	1
62	Ind. confidence - production expectations coming months	soft	0	1	3	1
63	Ind. confidence - employment expectations coming months	soft	0	1	3	1
64	Cons. confidence (headline)	soft	0	1	3	1
65	Cons. confidence - financial situation (last 12 months)	soft	0	1	3	1
66	Cons. confidence - financial situation (next 12 months)	soft	0	1	3	1
67	Cons. confidence - general ec. situation (last 12 months)	soft	0	1	3	1
68	Cons. confidence - general ec. situation (next 12 months)	soft	0	1	3	1
69	Cons. confidence - unemployment expect. (next 12 months)	soft	0	1	3	1
70	Cons. confidence - major purchases at present	soft	0	1	3	1
71	Cons. confidence - major purchases (next 12 months)	soft	0	1	3	1
72	Cons. confidence - savings at present	soft	0	1	3	1
73	Cons. confidence - savings (next 12 months)	soft	0	1	3	1
74	Cons. confidence - statement on financial situation of household	soft	0	1	3	1
75	IFO-indicator, expected business-situation	soft	0	1	3	1
76	BNB-indicator, gross-index	soft	0	1	3	1
77	Ind. confidence - Spain (headline)	soft	0	1	3	1
78	Ind. confidence - France (headline)	soft	0	1	3	1
79	Ind. confidence - Italy (headline)	soft	0	1	3	1
80	Ind. confidence - United Kingdom (headline)	soft	0	1	3	1

Notes: entries denote variable number, name, category (cat.), transformation and publication lag. Type: hard= quantitative information; financial= financial and monetary variables; price= consumer and producer prices; soft= qualitative information. Ln.: 0= no logarithm; 1= logarithm; Dif.: 1= first difference; 2= second difference. Filter: 3= change against the same month of the previous month. Lag= publication lag, e.g. publication lag is one when April figure is known on the 10th of May.

The quarterly GDP series were constructed from two seasonally adjusted quarterly time series for GDP growth in the Netherlands. The first series covers the period 2001.I–2010.IV. The second –so called ESA95– series covers the period 1991.I–2004.IV. The full sample series is constructed by backdating the GDP-series starting in 2000.I over the period 1991.I–2000.IV by using the quarter-on-quarter growth rates of the ESA95-series.

Chapter 3

Forecasting and nowcasting real GDP: comparing statistical models and subjective forecasts for the euro area and its five largest countries

This chapter conducts a systematic comparison of the short-term forecasting abilities of twelve statistical models and professional analysts in a pseudo real-time setting, using a large set of monthly indicators. The analysis covers the euro area and its five largest countries over the years 1996–2011. One of the main findings is that summarizing the available monthly information in a few factors is a more promising forecasting strategy than averaging a large number of single-indicator-based forecasts. Moreover, it is important to make use of all available monthly observations. The dynamic factor model is the best model overall, in particular for nowcasting and backcasting, due to its ability to incorporate more information (factors). Judgmental forecasts by professional analysts often embody valuable information that could be used to enhance forecasts derived from purely mechanical procedures.[1]

KEYWORDS: Forecasting competitions; Nowcasting models; Professional forecasters.

[1] This chapter is co-authored by Jos Jansen and Xiaowen Jin. This chapter was published in the International Journal of Forecasting, 32, 411-436, Jansen, Jin and de Winter, Copyright Elsevier (2016). Available online at `http://dx.doi.org/10.1016/j.ijforecast.2015.05.008`. Comments and suggestions by Jakob de Haan, Jan Jacobs, Job Swank, an associate editor, three anonymous referees and seminar and conference participants at De Nederlandsche Bank, the Computational and Financial Econometrics Conference (2012, Oviedo) and the European Meeting of the Econometrics Society (2013, Gothenburg) are gratefully acknowledged. We are particularly indebted to Marta Bańbura for supplying the Matlab-code for the estimation of the Bayesian VAR model. An early version of this chapter was circulated as DNB Working Paper 365 under the title "Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts".

3.1 Introduction

Information on economic activity and its short-term prospects is of great importance to decision makers in governments, central banks, financial markets and non-financial firms. Monetary and economic policy makers and economic agents have to make decisions in real-time with incomplete and inaccurate information on current economic conditions. A key indicator of the state of the economy is the growth rate of real gross domestic product (GDP), which is available on a quarterly basis only and is also subject to substantial publication lags. In many countries an initial estimate of quarterly real GDP is published around six weeks after the end of the quarter. Moreover, real GDP data are subject to sometimes substantial revisions, as more data becomes available to statistical offices over time.

Fortunately, there is a lot of statistical information related to economic activity that is published on a more frequent and timely basis. This information includes data on industrial production, prices of goods and services, expenditures, unemployment, financial market prices, loans and consumer and business confidence. Recently, the forecasting literature has developed several statistical approaches for exploiting this potentially very large information set in order to improve the assessment of both real GDP growth in the current quarter (nowcast) and its development in the near future. Examples of such approaches include bridge models, factor models, mixed-data sampling regression models (MIDAS) and mixed-frequency vector autoregressive (MFVAR) models. These models differ in their solutions to the practical problems of dealing with large information sets and the fact that the auxiliary variables are observed at different frequencies and with different publication lags.

Practitioners now have a wealth of statistical to choose from; but which one should they use? As each model has its own strengths and weaknesses, it is difficult to make a decision on purely theoretical grounds. The ranking of the models in terms of forecasting abilities, and the extent to which this varies with the prediction horizon or the economic circumstances, has to be determined by empirical analysis. On these issues the jury is still out, however, as large-scale comparative studies are scarce. In many papers, the empirical work refers to a single country, and usually only limited numbers of models are included. Furthermore, studies differ in the size of the information set and the sample period used.[2]

This chapter is motivated by this gap in the empirical literature. It presents the outcomes of a systematic comparison of a broad range of linear statistical models –

[2] Rünstler et al. (2009) form an important exception, comparing three factor models, a bridge model and a quarterly VAR model for ten European countries; however, their study does not include the financial crisis. Kuzin et al. (2013) analyzed the relative forecasting performance of MIDAS models versus dynamic factor models, including part of the crisis years (2008–2009). Liebermann (2014) analyzed the relative forecasting performance during the years 2001–2011, but only for the United States.

twelve models in all– that have been applied in the recent literature. For the sake of comparability and robustness, the analysis includes the euro area and its five largest countries (Germany, France, Italy, Spain and the Netherlands), and utilizes an information set that is as homogeneous as possible across geographical entities. Moreover, the sample includes the volatile episode of the financial crisis of 2008 and its aftermath, which may make it easier to discriminate between the various models. The models' forecasting abilities before 2008 are contrasted with those during the crisis period. This may be of great interest to policy makers, financial analysts and economic agents alike, as information on where the economy stands and where it is headed in the immediate short run is particularly valuable in times of great uncertainty.

The provision of cross-country evidence on the relative performance of twelve different statistical forecasting models is the first contribution of this chapter to the literature. Model forecasts are the result of purely mechanical recipes, and do not incorporate subjective elements. The second contribution concerns the potential usefulness of forecasts made by professional analysts. From a practical point of view, such forecasts are very cheap and easy to use. Moreover, as an expression of the "wisdom of crowds", they may reflect much more information than the statistical information set, which is inevitably limited. A questionnaire conducted by the European Central Bank (ECB) among the participants of the ECB Survey of Professional Forecasters found that the panelists regard 40% of their short-term GDP forecasts as being judgment-based (Meyler and Rubene, 2009). It is investigated to which extent the subjective forecasts by analysts in the sample contain information beyond that generated by the best mechanical statistical models.

The remainder of this chapter is structured as follows. Section 3.2 describes the various statistical models and discusses how they deal with the challenges posed by large and irregularly shaped datasets. Section 3.3 describes the data, the pseudo real-time forecast design, and other specification issues. Section 3.4 and 3.5 present the results for the mechanical models and the professional forecasts, respectively. Section 3.6 summarizes the main findings and concludes.

3.2 Linear statistical models for short-term forecasting

3.2.1 Overview

In practice, taking advantage of auxiliary information for the forecasting of real GDP growth in the immediate short run poses several challenges in practice. The first challenge is posed by the large size of the information set. There are countless potentially useful variables for forecasting GDP, and often they are interrelated. The datasets used in the empirical literature vary greatly in size, and may include more than 300 variables. Moreover, the limited length of the time series involved makes over-parametrization a real issue. The second problem relates to the fact that indicator variables are observed more frequently (monthly, weekly, daily) than GDP. Moreover, the dating of the most recent observation may vary across indicators because of differences in publication lags. This is known as the "ragged edge" problem; see Wallis (1986).

The various statistical approaches in the literature deal with these challenges in different ways. Broadly speaking, a forecasting procedure involves two transformations of the dataset of indicators to produce a final forecast: an aggregation and the application of a forecasting tool, which links auxiliary variables to real GDP growth. The two transformations can be executed in either order, representing two fundamentally different strategies. The first strategy begins by computing an indicator-specific GDP forecast for each variable, which are then aggregated into a single final forecast in the second step. This strategy is labeled the "pooling forecasts strategy". In this approach it is necessary to specify the weighting scheme for the individual forecasts. A basic scheme is the simple average, which gives each forecast an equal weight, but weights may also be recursively depending on the indicators' (recent) forecasting performances. Examples of the pooling forecasts strategy are bridge equations and VAR models. In contrast, the "aggregating information strategy" takes the aggregation step first, by summarizing the large dataset by a small number of series. This strategy exploits the fact that the auxiliary variables are correlated. Factor analysis is used to replace a large number of correlated time series with a limited number of uncorrelated (unobserved) factors representing the common information component of the original data series. The implicit weights (factor loadings) are determined by the correlation patterns in the original dataset. The factors serve as input for the forecasting procedure in the next step. Examples of this modeling strategy are dynamic factor models and factor-augmented versions of forecasting models that pool forecasts. Finally, a recent development is estimation using Bayesian shrinkage on coefficients, which translates a large set of indicators into a single GDP forecast directly, without a clear aggregation step. This approach implicitly aggregates information by applying Bayesian shrinkage to the parameters.

The specification of the forecasting tool is the second feature that distinguishes the approaches. The traditional approaches, such as bridge models and VAR models, rely on forecasting equations that are cast solely in quarterly terms. That means that (forecasts of) monthly indicator variables first have to be aggregated to quarterly averages, before they can be used for forecasting GDP. Moreover, the available monthly observations are not fully exploited by quarterly VAR models. As this may not be an efficient use of the available information, recently developed approaches accommodate both quarterly and monthly data within the same equation or system of equations. These approaches take publication lags into account. The mixed-frequency VAR (MFVAR) model treats GDP as an unobserved monthly variable in a state space framework. Monthly GDP is related to quarterly GDP via an identity. The quarterly GDP growth rate is observed only in the third month of each quarter. The mixed-data sampling (MIDAS) design relates quarterly GDP directly to a large number of lags of monthly data series using a parsimonious specification of the lag structure.

A third, and more practical, specification issue is whether or not to include GDP's own past in the forecasting tool. In general, forecasting equations can be augmented easily with auto-regressive (AR) terms. Several authors have found that the AR versions of models tend to result in modest improvements of forecasting performance (e.g. Foroni and Marcellino, 2014). This chapter analyzes twelve statistical models. They are denoted as follows: (1) bridge model (BEQ), (2) BEQ with AR terms (BEQ-AR), (3) quarterly VAR model (QVAR), (4) factor-augmented quarterly VAR (F-VAR), (5) Bayesian quarterly VAR (BVAR), (6) dynamic factor model (DFM), (7) mixed-frequency VAR model (MFVAR), (8) factor-augmented MFVAR (F-MFVAR), (9) mixed-data sampling regression model (MIDAS), (10) MIDAS with AR terms (MIDAS-AR), (11) factor-augmented MIDAS (F-MIDAS) and (12) F-MIDAS with AR terms (F-MIDAS-AR). The next three subsections discuss the forecasting models briefly, starting with the quarterly models. To improve the flow of the discussion, the selection of the weighting scheme for indicator-based forecasts is discussed in Section 3.3.3. Moreover, some technical details have been moved to Appendix 3.B.

Below, $t = 1, \ldots, T$ stands for a quarterly time index and $m = 1, \ldots, T_m + w$ for a monthly time index. T indicates the latest available quarterly observation and T_m corresponds to the third month of quarter T; hence $T_m = 3T$. $T_m + w$ indicates the latest available monthly observation, where w is the time difference in months between the most recent observation of indicator and GDP on the monthly time scale. Quarterly averages of monthly figures are denoted by superscript ^{Q}. Quarterly GDP in quarter t, y_t^Q, is assigned to month $3t$ on the monthly time scale. Formally: $y_t^Q = \frac{1}{3}(y_{3t} + y_{3t-1} + y_{3t-2})$, where y_{3t}, y_{3t-1} and y_{3t-2} are unobserved three-month GDP growth rates, i.e. growth rates vis-à-vis the same month of the previous quarter. The matrix of monthly indicators (indexed by $i = 1, \ldots, n$) is defined as $x_m = (x_{1,m}, \ldots, x_{n,m})'$. The monthly series x_m have been transformed as three-month growth rates or differences. The matrix of monthly indicators aggregated to quarterly values is defined as $x_t^Q = (x_{1,t}^Q, \ldots, x_{n,t}^Q)'$. The quarterly GDP growth forecast for quarter $t + h$ at time t is denoted as $\hat{y}_{t+h|t}^Q$. For quarterly models, time is always measured on the quarterly time scale. In mixed-frequency approaches, the quarterly and monthly time scales intermingle.

3.2.2 Quarterly models for GDP growth

Bridge equation (BEQ)

The quarterly bridge equation is a method that has been used widely for forecasting GDP using all available observations of monthly indicators; for applications see Baffigi et al. (2004), Kitchen and Monaco (2003) and Rünstler and Sédillot (2003). Bridge equations are linear regressions that "bridge" monthly variables, such as industrial confidence and retail sales, to quarterly real GDP. Usually, the monthly indicators are not known over the entire projection horizon. Various specifications are possible within this approach. Here, a simple version of the bridge equation is proposed, proceeding in two steps. Firstly, predictions of the necessary monthly values of indicator x_i are obtained over the forecasting horizon with the help of univariate autoregressive models and aggregated to appropriate quarterly values x_i^Q. Secondly, these quarterly aggregates are used to predict GDP. The bridge model for x_i is:

$$y_t^Q = \alpha + \sum_{s=0}^{p} \beta_s x_{i,t-s}^Q + \varepsilon_{i,t}^Q, \qquad \varepsilon_{i,t}^Q \sim N(0, \sigma_{\varepsilon^Q}^2) \tag{3.1}$$

where α is a constant, p denotes the number of lags in the bridge equation and ε_i^Q is a normally distributed error-term. Eq. (3.1) is estimated for each of the n indicators. The final forecast is then calculated by weighting the n indicator-specific forecasts for each horizon. The lag parameter p is determined recursively by the Schwartz information criterion (SIC), with the maximum number of lags set at 4.

Quarterly vector autoregressive model (QVAR)

The VAR approach is very similar to the bridge equation approach. Unlike bridge equations, VAR models use the information content of GDP itself to produce forecasts of GDP (e.g. Camba-Mendez et al., 2001). Moreover, it is a system approach, attempting to exploit the interdependence of indicator and real GDP dynamics. However, misspecification anywhere in the system may affect the accuracy of the GDP predictions. More importantly, due to its quarterly time frame, the QVAR model only uses monthly observations that correspond to a full quarter. Consequently, it does not fully exploit the available monthly information. In total, n quarterly bivariate VAR models that include one of the indicators and GDP growth were estimated:

$$z_{i,t}^Q = \alpha + \sum_{s=1}^{p_i} A_s z_{i,t-s}^Q + \varepsilon_{i,t}^Q, \qquad \varepsilon_{i,t}^Q \sim N(0, \Sigma_{\varepsilon^Q}) \tag{3.2}$$

where $z_{i,t}^Q = (y_t^Q, x_{i,t}^Q)'$. From each bivariate VAR an indicator-specific GDP forecast $y_{t+h|t}^Q$ is obtained. As in the case of the bridge model, the final forecast is formed as a weighted average of the individual forecasts. The lag parameter p is determined recursively by the SIC, with a maximum of 4.

Bayesian VAR Model (BVAR)

The number of variables in multivariate VAR models is usually between three and ten, because the number of unrestricted parameters that can be estimated reliably is rather limited. Bayesian vector autoregression with shrinkage is able to handle large unrestricted VARs (e.g. Carriero et al., 2015a and Giannone et al., 2015). The quarterly Bayesian VAR model is estimated along the lines proposed by Bańbura et al. (2010). Accordingly, all variables are included in log levels, except those already expressed as rates, and derive the GDP forecast $y^Q_{t+h|t}$ from the following VAR system:

$$Z^Q_t = \alpha + \sum_{s=1}^{p} A_s Z^Q_{t-s} + \vartheta^Q_t, \qquad \vartheta^Q_t \sim WN(0, \Sigma_{\vartheta^Q}) \tag{3.3}$$

where $Z^Q_t = (y^Q_t, x^Q_{i,t} \dots x^Q_{n,t})'$. Moreover, the moments for the prior distribution of the coefficients are:

$$E[(A_k)_{ij}] = \begin{cases} \delta_i, & j = i,\ k = 1; \\ 0, & \text{otherwise} \end{cases} \qquad V[(A_k)_{ij}] = \begin{cases} \frac{\lambda^2}{k^2}, & j = i; \\ \frac{\lambda^2}{k^2}\frac{\sigma_i^2}{\sigma_j^2}, & \text{otherwise} \end{cases} \tag{3.4}$$

The coefficients $A_1, \dots, A_k$ are assumed a priori to be independent and normally distributed. The covariance matrix of the residuals is assumed to be diagonal, fixed and known: $\Psi = \Sigma$, where $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$. For non-stationary variables the random walk prior is used: $\delta_i = 1$. For stationary variables, the white noise prior is used: $\delta_i = 0$. The parameter λ governs the degree of shrinkage. If $\lambda = 0$, the posterior equals the prior and the data do not influence the estimates. At the other extreme, $\lambda = \infty$, the posterior expectations coincide with the ordinary least squares (OLS) estimates. The degree of shrinkage should be chosen so as to prevent over-fitting, while preserving the relevant sample information. Bańbura et al. (2010) argue that λ should be set relative to the size of the system. The reasoning here is that if all data series contain similar information, the relevant signal can still be extracted efficiently from a large dataset, despite the higher shrinkage that is required to filter out the systematic component. λ is determined recursively in a fashion similar to Bańbura et al. (2010).[3] The factor $\frac{1}{k^2}$ is the rate at which the prior variance decreases with the lag length and $\frac{\sigma_i^2}{\sigma_j^2}$ accounts for the different scale and variability of the data. Following Bańbura et al. (2010) the maximum number of lags was set at p at 5.

[3] Bańbura et al. (2010) begin by defining FIT, an in-sample measure of fit for three key variables (output, inflation and the short-term interest rate). They select λ such that the BVAR model and an unrestricted VAR model featuring the three key variables produce the same numerical value for FIT. Real GDP, HICP inflation and the three-month interest rate are taken as key variables. FIT and λ are calculated recursively. Like Bloor and Matheson (2011), the preliminary calculations indicated that the FIT measure thus specified gave rise to over-fitting. Therefore, the procedure was applied to FIT/2.

3.2.3 Mixed-frequency models

Interest in mixed-frequency models has increased among academics and policy makers in recent years, because of the general failure of simple quarterly models to predict or signal the sharp downturn of the economy at the onset of the financial crisis. Here, the dynamic factor model, the mixed-frequency VAR and the MIDAS approach are investigated.[4] All of these models utilize the available monthly information fully.

Dynamic factor model (DFM)

Dynamic factor models summarize the information contained in the dataset using a limited number of factors, the dynamic behavior of which are specified as a vector-autoregressive process. A key feature of this approach is the use of the Kalman filter, which allows for an efficient handling of the unbalancedness of the dataset and the different frequencies of the data. The Kalman filter replaces any missing monthly indicator observations with optimal predictions, and also generates estimates of unobserved monthly real GDP subject to a temporal aggregation constraint for the quarterly observation. Dynamic factor models have been shown to produce relatively accurate macroeconomic forecasts for many countries.[5]

In this chapter the dynamic factor model proposed by Bańbura and Rünstler (2011) is analyzed. The model is used by several central banks within the euro area. The first equation of the model is:

$$x_m = \Lambda f_m + \xi_m, \qquad \xi_m \sim N(0, \Sigma_\xi) \tag{3.5}$$

which relates the n monthly indicators x_m to r monthly static factors $f_m = (f_{1,m}, \ldots, f_{r,m})'$ via a matrix of factor loadings Λ and an idiosyncratic component $\xi_m = (\xi_{1,m}, \ldots, \xi_{n,m})'$, where $r << n$. The DFM assumes that the idiosyncratic components are a multivariate white noise process, hence the covariance matrix Σ_ξ is diagonal. Furthermore, the DFM assumes that the factors follow a vector-autoregressive process of order p:

$$f_m = \sum_{s=1}^{p} A_s f_{m-s} + \zeta_m, \qquad \zeta_m \sim (0, \Sigma_\zeta) \tag{3.6}$$

where A is a square $r \times r$ matrix. Moreover, the covariance matrix of the VAR (Σ_ζ) is driven by a q dimensional standardized white noise process η_m:

$$\zeta_m = B\eta_m, \qquad \eta_m \sim N(0, I_q) \tag{3.7}$$

---

[4] Recently, Bayesian mixed-frequency regressions and VAR models have been developed and applied to nowcasting. See e.g. Carriero et al. (2015b) and Schorfheide and Song (2015). These alternative approaches are beyond the scope of this dissertation.

[5] Examples include Giannone et al. (2008) for the United States; Bańbura et al. (2011), Camacho and Perez-Quiros (2010), Rünstler et al. (2009) and Bańbura and Modugno (2014) for the euro area; Schumacher and Breitung (2008) for Germany; Schneider and Spitzer (2004) for Austria; Cheung and Demers (2007) for Canada; Camacho and Perez-Quiros (2011) for Spain; and den Reijer (2013) for the Netherlands.

where B is a $r \times q$ matrix and $q \leq r$. The final equation is a forecasting equation linking the factors to (unobserved) mean-adjusted real GDP growth:

$$y_m = \beta' f_m + \varepsilon_m, \qquad \varepsilon_m \sim N(0, \sigma_\varepsilon^2) \tag{3.8}$$

where y_m denotes the unobserved monthly GDP growth rate. The model is estimated in four steps. In the first step, the factors loadings Λ and the estimated static factors $\hat{f}_m$ are obtained. In the second step, the coefficient matrices A_s in Eq. (3.6) and β in Eq. (3.8) are estimated by OLS using $\hat{f}_m$. In the third step, ζ_m and its covariance matrix Σ_ζ are computed, and an estimate of the matrix B is obtained by principal components analysis. In the final step, the model is cast in state space from and the Kalman filter and smoother are used to re-estimate the estimated factors ($\hat{f}_m$) and monthly GDP growth.[6] To estimate the model, the number of static and dynamic common factors needs to be specified, denoted by r and q respectively. The largest possible value of r is set at 6, based on the scree test of Cattell (1966). Moreover, $q \leq r$ by definition. In view of potential misspecification and instabilities the estimation procedure follows Kuzin et al. (2013), and refrains from choosing a particular combination of r and q, but takes the (unweighted) average of forecasts over all possible parameterizations in terms of the number of static and dynamic factors and the number of lags p in Eq. (3.6), with $p \leq 6$. Thus, the total number of model specifications is $p(r+1)r/2 = 126$.[7]

Mixed-frequency vector autoregressive model (MFVAR)

Mixed-frequency VAR models (MFVAR) are VAR models that allow for data series with different frequencies. In contrast to the quarterly VAR model, the MFVAR model fully exploits all available monthly information. It shares with the QVAR model the strengths and weaknesses of a system approach. In this case the focus is on bivariate MFVAR models featuring a monthly indicator, unobserved monthly GDP and a temporal aggregation scheme.

Let $z_{i,m} = (y_m, x_{i,m})'$ be the vector of latent monthly real GDP and indicator $x_{i,m}$. The vector follows a VAR model:

$$z_{i,m} - \mu_i = \sum_{s=1}^{p} A_s(z_{i,m-s} - \mu_i) + \varepsilon_{i,m}, \qquad \varepsilon_{i,m} \sim N(0, \Sigma_\varepsilon) \tag{3.9}$$

[6] The state space form of the dynamic factor model is outlined in Appendix 3.B.2. See Bańbura and Rünstler (2011) for a more detailed description of the dynamic factor model and the estimation procedure. See also Stock and Watson (2011). See Durbin and Koopman (2012) for a comprehensive treatment of state space models and the use of the Kalman filter and smoother.

[7] Applying a different weighting scheme leads to results that are virtually the same; see Table 3.9 in Appendix 3.C. Alternatively one could choose the number of factors r and q on the basis of in-sample criteria, as described in Bai and Ng (2002, 2007). Preliminary estimates indicate that these criteria tend to indicate a relatively large number of factors, in line with the outcome in Bańbura and Rünstler (2011), leading to volatile and less accurate forecasts.

where μ_i denotes the mean of $z_{i,t}$. As documented by Kuzin et al. (2011) the means μ_i are often quite difficult to estimate. Therefore, demeaned GDP and monthly indicators are used in the estimation procedure, adding the mean back afterwards to arrive at the final indicator-based forecast. As in the dynamic factor model, the Kalman filter and smoother fills in any missing monthly indicator observations with optimal predictions and estimates unobserved monthly real GDP subject to a temporal aggregation constraint for the quarterly observation. The state space setup of the MFVAR is outlined in Section 3.B.1. The model is estimated by the expectation-maximization algorithm, as detailed in Mariano and Murasawa (2010). As in the case of the QVAR model, the final GDP forecast is formed as a weighted average of the individual forecasts derived from the n bivariate MFVAR models. Regarding the number of lags p, it is held fixed for practical reasons ($p = 1$ for the euro area, Germany, Spain and the Netherlands and $p = 2$ for France and Italy).[8]

Mixed-data sampling regression model (MIDAS)

The mixed-data sampling regression model (MIDAS) is a single horizon-specific equation that relates quarterly GDP to (various lags of) a monthly indicator (Ghysels et al., 2007; Schumacher, 2016). It generates the GDP forecast in a direct way. The MIDAS model circumvents the ragged edge problem by including as regressors a fixed (fairly large) number of the most recent lagged values of the indicator. In applied work, the MIDAS model economizes on the number of parameters to be estimated by adopting a parsimoniously parameterized lag polynomial. The efficiency gains of such an approach come at the cost of potential efficiency losses if the implied restrictions on the lag structure between the monthly indicator and quarterly real GDP happen to be invalid. The MIDAS model is based on Kuzin et al. (2011), who work with the exponential Almon lag polynomial. The indicator-specific MIDAS model for forecasting horizon h is thus defined by the following equations:

$$y^Q_{t+h} = \beta_0 + \beta_1 B(L_M;\theta) x^{(3)}_{i,m} + \varepsilon^Q_{i,t+h} \tag{3.10}$$

$$B(L_M;\theta) = \sum_{k=0}^{K} c(k,\theta) L_M^k \tag{3.11}$$

$$c(k,\theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=0}^{K} \exp(\theta_1 k + \theta_2 k^2)} \tag{3.12}$$

[8] Achieving convergence when estimating MFVAR models with four lags was very difficult. Moreover, varying the number of lags according to the SIC does not seem to be efficient in terms of forecasting performances compared to procedures that simply keep the number of lags constant through time. Therefore, the maximum number of lags is fixed to one, two or three, based on the out-of-sample performances of the nowcasts and backcasts in the first quarter of the sample (1996.I–1999.IV).

where L_M is the monthly lag operator and the T observations of the regressor $x_{i,m}^{(3)}$ are skip sampled from $x_{i,m}$ by including every third observation, starting from the final one. Thus, $x_{i,m}^{(3)} = x_{i,m}, m = 3 + w, \ldots, 3(T-2) + w, 3(T-1) + w, 3T + w$. Eq. (3.11) describes a weighting function of lagged values, while Eq. (3.12) specifies the weight for lag k as a function of k and the two parameters governing the exponential Almon lag polynomial. K is fixed at 11. The model's parameters $(\theta_1, \theta_2, \beta_0, \beta_1)$ are estimated by Nonlinear Least Squares, subject to $\theta_1 < 5$ and $\theta_2 < 0$. The final GDP forecast is computed as a weighted average of the individual forecasts derived from the n indicator-specific MIDAS models.

3.2.4 Factor- and AR-augmented models

Factor-augmented models

This section describes versions of the QVAR, MFVAR and MIDAS models in which the independent variable is a factor rather than an observed indicator. These so-called factor-augmented versions are denoted as F-VAR, F-MFVAR and F-MIDAS, respectively. See Eqs. (3.13)–(3.15), which refer to the single-factor case:

$$z_t^Q = \alpha + \sum_{s=1}^{p} A_s z_{t-s}^Q + \varepsilon_t^Q, \qquad \varepsilon_t^Q \sim N(0, \Sigma_{\varepsilon^Q}) \tag{3.13}$$

$$z_m - \mu_z = \sum_{s=1}^{p} A_s (z_{m-s} - \mu_z) + \varepsilon_m, \qquad \varepsilon_m \sim N(0, \Sigma_\varepsilon) \tag{3.14}$$

$$y_{t+h}^Q = \beta_0 + \beta_1 B(L_M; \theta) \hat{f}_m^{(3)} + \varepsilon_{t+h}^Q \tag{3.15}$$

where $z_t^Q = (y_t^Q, f_t^Q)'$ and $z_m = (y_m, f_m)'$. The number of lags in the F-MFVAR model is the same as in the corresponding MFVAR model. The factors are obtained as follows. For F-VAR the factor is derived by applying simple principal component analysis on quarterly averages of monthly data. For the mixed-frequency models the Kalman-filtered factors are generated by the dynamic factor model, averaged over all possible parameterizations. This facilitates the comparison between the dynamic factor model and factor augmented models.

An important specification issue is the number of factors used to summarize the information set. The literature typically restricts the analysis to a single factor (e.g. Marcellino and Schumacher, 2010). This issue is investigated in three ways. First, when treating the factors as separate indicators, taking a weighted average of factor-specific forecasts typically weakens the forecasting accuracy, especially for nowcasts and backcasts (see Table 3.7 in Appendix 3.C). The only exceptions are the MIDAS models for

Spain, for which the second and third factor possess substantial predictive power. Second, the marginal forecasting power of additional factors is generally very small (often zero), again with the exception of the MIDAS models for Spain (see Table 3.8). Third, when introducing more than one factor into MIDAS equations (multi-factor MIDAS), the multi-factor models tend to produce higher RMSFEs for forecasts and nowcasts; the main exception is the two-factor model for Spain. For backcasts the picture is mixed. For some countries it results in a loss of forecasting accuracy. Moreover, the empirical relationship between the number of factors and forecasting performance is erratic and difficult to interpret. The marginal forecasting power of an additional factor can be positive or negative, and may be sensitive to the exact dating of the backcast. For example, second-month backcasts may improve, while first-month backcasts deteriorate. These findings suggest that over-fitting issues are a real concern for multi-factor MIDAS models, even when the number of factors is small. Taken together, the results discussed above hint at a weakness of factor-augmented models, namely their limited ability to incorporate extra information into forecasts. This issue is discussed in more detail in Section 3.4.3. Based on this preliminary analysis, one factor was used in each of the factor-augmented models, following other authors, except for the MIDAS models for Spain. In the latter case three factors were used, which were treated as single indicators.[9]

AR-augmented models

Finally, versions of the BEQ, MIDAS and F-MIDAS models featuring an AR(1) term are considered, as GDP's own past may contain important information. The AR-augmented models are denoted as BEQ-AR, MIDAS-AR and F-MIDAS-AR, respectively. The BEQ-AR model for x_i can be written as:

$$y_t^Q = \alpha + \varphi y_{t-1}^Q + \sum_{s=0}^{p_i} \beta_s x_{i,t-s}^Q + \varepsilon_{i,t}^Q \tag{3.16}$$

As proposed in Clements and Galvão (2008), the AR term is introduced as a common factor in the MIDAS-AR and F-MIDAS-AR models:

$$y_{t+h}^Q = \beta_0 + \varphi y_{t-1}^Q + \beta_1 B(L_M;\theta)(1-\varphi L_M^h) x_{i,m}^{(3)} + \varepsilon_{t+h}^Q \tag{3.17}$$

$$y_{t+h}^Q = \beta_0 + \varphi y_{t-1}^Q + \beta_1 B(L_M;\theta)(1-\varphi L_M^h) f_m^{(3)} + \varepsilon_{t+h}^Q \tag{3.18}$$

The parameter φ is estimated simultaneously with the other parameters.

[9] The predictive power of the second and third factors in the case of Spain is a persistent feature throughout the sample period, which agents would discover quickly. Thus, the risk of this modeling assumption introducing hindsight bias is quite small.

3.3 Data, forecast design and specification issues

This section describes the dataset, the pseudo real-time setup, the weighting scheme that was used for pooling indicator-specific forecasts in case of the QVAR, BEQ, BEQ-AR, MFVAR, MIDAS and MIDAS-AR models, and the selection of the number of lags and factors in the models.

3.3.1 Dataset

The monthly dataset consists of 72 monthly time-series variables (using harmonized definitions across the countries), which cover the broad range of information that is readily available to economic agents. To facilitate cross-country comparisons, the monthly indicators were selected based on their availability in all countries for a sufficiently long time-span.[10] The indicator variables fall into four categories. The first category is hard, quantitative information on production and expenditures, such as industrial production in various sectors and countries, car sales, world trade and unemployment. The second category refers to input and output prices, such as consumer and producer prices, and oil and commodity prices. The third category contains financial variables, both quantities (money stock and credit volume) and prices (interest rates, stock prices and exchange rates). These determine financing conditions for firms and consumers. Moreover, financial market prices partly reflect financial market expectations on output developments in the near future. The fourth category is soft, qualitative information on expectations derived from surveys among consumers, retailers and firms. Moreover, three composite leading indicators compiled by the OECD were included.

Appendix 3.A provides details on the sources, availability and the applied transformations of the data series. The available monthly data are usually already adjusted for seasonality (and calendar effects). When necessary, raw data series are seasonally adjusted using the US Census X-12 method. All monthly series are made stationary by differencing or log-differencing (in the case of trending data, such as industrial production, retail sales and monetary aggregates). Finally, each variable is standardized by subtracting the mean and dividing by the standard deviation. This normalization is standard practice in order to avoid the overweighting of series with large-variances series in the extraction of common factors. The data transformations are the same for all of the statistical models, except for the Bayesian VAR.

[10] As a consequence, country-specific indicators, such as the Ifo-indicator for Germany, were not used for forecasting. Moreover, the time series for the Purchasing Managers Index (PMI) are not long enough for all countries. Of course, as economic agents will use this country-specific information in practice, the results might underestimate the forecasting accuracy of mechanical models for certain countries or periods.

3.3.2 Pseudo real-time design

The forecast design aims to replicate the availability of the data at the time forecasts are made in order to mimic the real-time flow of information as closely as possible. To this end, a dataset that was downloaded on January 16, 2012 was combined with the typical data release calendar to reconstruct the available dataset on the 16th of each month during the period July 1995-January 2012. All monthly indicator series start in January 1985, while the quarterly GDP series start in 1985.I. Thus, a pseudo real-time design is employed, which takes data publication delays into account, but ignores the possibility of data revisions for GDP and some indicators, such as industrial production. The latter implies that the forecasting accuracy of statistical models might be overestimated. However, the effects of data revisions on the final forecast may largely cancel out (although the crisis episode may have been atypical in this regard), since statistical methods typically attempt to eliminate noise in the process by either extracting factors from a large dataset or pooling a large number of indicator-based forecasts. For example, Schumacher and Breitung (2008), using real-time data vintages for Germany, did not find a clear impact of data revisions on the forecast errors of factor models. Moreover, the effect on the *relative* performance of models, which is the main focus of this chapter, is likely to be quite small (see Bernanke and Boivin, 2003). However, abstracting from data revisions may affect the comparison of mechanical forecasts and forecasts by professional analysts to a greater extent. The expectations of analysts necessarily reflect the inaccurate initial estimates of GDP's recent past and this puts them at a disadvantage vis-à-vis mechanical models in a pseudo real-time setting as the latter can take data revisions on board.

The parameters of all models are estimated recursively, using only the information that was available at the time of the forecast. For similar approaches, see Giannone et al. (2008), Kuzin et al. (2011) and Rünstler et al. (2009), among others. For a given quarter, a sequence of eleven forecasts for GDP growth is obtained in consecutive months. Table 3.1 explains the timing of the forecasting exercise, taking the forecast for the third quarter of 2011 as an example. The first forecast is made in January 2011, which is called the two-quarter ahead forecast in month one. Subsequently, a monthly forecast is produced for the next ten months through November. The last forecast is made just before the first release of GDP in mid-November. Following the conventional terminology, *forecasts* refer to one or two-quarter ahead forecasts, *nowcasts* refer to current quarter forecasts and *backcasts* refer to forecasts for the preceding quarter, before official GDP figures become available. In case of the current example 2011.III, two-quarter ahead forecasts are produced from January to March, one-quarter ahead forecasts from April to June, nowcasts from July to September, and backcasts in October and November.

Table 3.1: Timing of forecasting exercise for third quarter GDP growth

Nr.	Forecast type	Month	Forecast made in middle of
1	Two-quarter ahead	1	January
2		2	February
3		3	March
4	One-quarter ahead	1	April
5		2	May
6		3	June
7	Nowcast	1	July
8		2	August
9		3	September
10	Backcast	1	October
11		2	November

3.3.3 Weighting scheme of indicator-based forecasts

The models BEQ, BEQ-AR, QVAR, MFVAR, MIDAS and MIDAS-AR construct a large number of different indicator-specific forecasts in the first stage, which have to be aggregated in the second stage to obtain the final forecast. Taking a weighted average of a large number of forecasts may ameliorate the effects of misspecification bias, parameter instability and measurement errors in the data, that may afflict the individual forecasts (Timmermann, 2006). Three different weighting schemes were investigated: (i) equal weights (simple mean); (ii) weights that are inversely proportional to the root mean squared forecast error (RMSFE) measured from the start of the sample period until the previous quarter (recursive RMSFE scheme); and (iii) weights that are inversely proportional to the RMSFE measured over the past four quarters (moving window RMSFE scheme). Equal weights have been proven to work quite well as pooling mechanism (e.g. Stock and Watson, 2004; Clark and McCracken, 2010). The latter two methods assign weights to the indicators based on their forecasting performance in the (recent) past.

Table 3.9 in Appendix 3.C gives an overview of the RMSFE of the three weighting schemes by horizon and country for BEQ-AR, QVAR, MFVAR and MIDAS-AR. The overall picture is that the moving window RMSFE weighting scheme, which emphasizes performance in the recent past, has the smallest RMSFE on average. However, the differences with the other schemes are very small, so the results are not sensitive to the specific weighting method. In the rest of this chapter the moving window RMSFE weighting scheme is applied for all relevant models and all countries.

3.4 Empirical results for statistical models

3.4.1 Forecasting performance

Table 3.2 presents data on the forecasting performance of the statistical models for the five countries plus the euro area for the complete sample period 1996.I–2011.III (63 quarters). The underlying empirical analysis has been carried out on a monthly basis for eleven horizons. To save space, Table 3.2 (and the other tables in this chapter as well) reports results for the two and one-quarter ahead forecasts, the nowcast and the backcast, which have been calculated as the average of the corresponding monthly data. Moreover, only he AR versions of the BEQ, MIDAS and F-MIDAS models are reported.[11] The forecasting performance is measured by RMSFE. Table 3.2 reports the results of two benchmark models that have been used in the literature: the random walk (RW) with drift and a pure univariate autoregressive (AR) model.[12] Measuring against the RW benchmark reveals the advantage of using information for forecasting, including GDP's own past. A comparison with the AR benchmark focuses on the value added of monthly auxiliary information. The first column of Table 3.2 reports the RMSFE of the random walk. For the other statistical models, including the AR model, the entries refer to their RMSFE relative to that of the RW benchmark in order to improve the comparability of the results across countries and horizons. Grey cells indicate the model with the lowest RMSFE in a row (for a particular horizon). Entries in bold indicate models that have RMSFEs that are less than 10% larger than that of the best model, and are also smaller than the RMSFE of the benchmark model.[13] The 10% threshold is meant as a rough assessment of the economic significance of differences in forecasting ability. Models that meet this condition are called "competitive models", as they do not differ "too much" from the best model in terms of forecasting performance.[14]

The outcomes in Table 3.2 point to several interesting results. First, incorporating monthly information in statistical forecasting procedures pays off in terms of forecasting accuracy, in particular for nowcasts and backcasts. The large majority of the relative RMSFEs are smaller than one and fall below those of the AR model. They also tend to decline if the horizon shortens and more monthly information has been absorbed. Moreover, models that do not fully exploit the available monthly observations (QVAR and F-VAR) generally have larger RMSFEs than models that do. Second, for many models the gain is rather limited when forecasting one and two-quarters ahead. For the two-quarter-ahead forecasts, the best models have RMSFEs that are only 5% lower than

[11] The results for non-AR versions are very close to their AR counterparts. Section 3.4.4 discusses the effect of taking an AR term on board.

[12] The drift parameter is recursively estimated. Regarding the AR model, the number of lags is determined recursively by the SIC, with a maximum of four.

[13] If the best model has an RMSFE of 0.6, the cut-off point is an RMSFE of 0.66.

[14] In addition, Diebold and Mariano (1995) tests broadly paint the same picture.

Table 3.2: Forecasting performance of statistical models, 1996.I–2011.III

Frequency	Benchmark			Pooling forecasts				Pooling information			
Model	RW	AR	BVAR	BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area											
2Q-ahead	*0.64*	1.00	**0.97**	0.97	1.01	**0.97**	**0.97**	**0.99**	**1.00**	1.01	1.01
1Q-ahead	*0.63*	1.00	0.93	**0.90**	1.00	0.93	0.93	**0.90**	0.96	0.83	**0.86**
nowcast	*0.63*	0.95	0.83	0.79	0.92	0.79	0.82	**0.69**	0.88	**0.67**	0.66
backcast	*0.63*	0.85	0.71	0.69	0.79	0.62	0.70	0.51	0.79	0.62	0.66
Germany											
2Q-ahead	*0.91*	1.05	1.02	0.98	1.00	1.02	**0.99**	**0.99**	1.00	1.01	1.01
1Q-ahead	*0.91*	1.03	**0.99**	**0.97**	**1.00**	**0.97**	**0.98**	0.92	**0.98**	**0.94**	**0.92**
nowcast	*0.91*	1.03	0.94	0.90	0.99	0.92	0.90	0.77	0.94	**0.80**	**0.77**
backcast	*0.91*	1.03	0.88	0.85	0.97	0.83	0.79	0.67	0.91	0.74	0.80
France											
2Q-ahead	*0.53*	1.02	0.94	**0.96**	1.01	**0.96**	**0.97**	**0.94**	**1.00**	**1.00**	**0.97**
1Q-ahead	*0.53*	0.99	**0.90**	**0.90**	1.01	**0.91**	**0.91**	0.84	0.98	**0.85**	**0.85**
nowcast	*0.52*	0.90	0.82	0.78	0.89	0.79	0.80	**0.64**	0.89	0.62	**0.63**
backcast	*0.52*	0.84	0.74	0.73	0.79	0.71	0.72	0.52	0.80	0.62	0.63
Italy											
2Q-ahead	*0.75*	1.04	0.95	**0.99**	1.01	**0.98**	**0.99**	**0.99**	**0.99**	1.06	**0.99**
1Q-ahead	*0.75*	0.98	**0.94**	**0.95**	0.99	**0.95**	**0.95**	**0.92**	**0.97**	**0.90**	0.89
nowcast	*0.74*	0.95	0.89	0.86	0.94	0.87	0.87	**0.74**	0.90	**0.72**	0.72
backcast	*0.74*	0.92	0.81	0.80	0.85	0.80	0.78	0.64	0.80	**0.70**	**0.67**
Spain											
2Q-ahead	*0.64*	**0.92**	**0.87**	**0.91**	**0.92**	**0.9**	**0.93**	**0.88**	**0.90**	0.86	**0.93**
1Q-ahead	*0.63*	0.83	**0.77**	0.85	0.85	**0.78**	0.89	0.75	**0.81**	0.85	0.84
nowcast	*0.63*	0.74	0.62	0.78	0.76	**0.68**	0.83	**0.64**	0.75	0.71	0.82
backcast	*0.62*	0.79	0.49	0.79	0.72	0.75	0.84	**0.57**	0.75	0.66	0.85
Netherlands											
2Q-ahead	*0.72*	**0.99**	0.92	**0.96**	**0.99**	**0.97**	**0.98**	**0.99**	**0.96**	**0.95**	1.03
1Q-ahead	*0.71*	0.99	**0.89**	**0.90**	0.98	0.92	0.96	**0.90**	0.94	0.83	**0.89**
nowcast	*0.71*	0.96	0.83	0.82	0.93	0.85	0.88	**0.76**	0.89	0.74	**0.78**
backcast	*0.71*	0.90	0.75	0.77	0.84	0.76	0.84	0.68	0.81	**0.72**	0.79

Notes: entries denote the RMSFE for a Random Walk (in italics); for all other models they denote the RMSFE relative to the RMSFE of a Random Walk. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model. RW: random walk, AR: autoregressive model, BEQ-AR: bridge equation with AR term, QVAR: quarterly vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, DFM: dynamic factor model, MFVAR: mixed-frequency vector autoregressive model, F-MFVAR: factor-augmented MFVAR model, MIDAS-AR: mixed-data sampling model with AR term, F-MIDAS-AR: factor-augmented MIDAS-AR.

the benchmark. Only for Spain does the best statistical model deliver an economically significant improvement. For the one-quarter ahead forecast the average improvement by the best models is 15% on the RW benchmark, but the other models generally post gains of less than 10%. For the nowcasts and backcasts, the average gains for the best performing models amount to roughly 30% and 40%, respectively. This pattern suggests that statistical models are of greater value when they can use information that pertains to the relevant quarter. Their relative strength is in improving the assessment of the current state of the economy. Third, the dynamic factor model displays the best performance overall. Looking across countries and horizons, it works best for backcasts and is at least competitive in all other cases. The factor augmented MIDAS-AR and MFVAR models perform relatively well for one-quarter-ahead predictions and nowcasts, but not so well for backcasts. The Bayesian VAR model works best for Spain and relatively well for the Netherlands. It also delivers the best two-quarter-ahead forecasts. Fourth, many models are competitive at the two-quarter-ahead horizon in most of the countries, but the number of competitive models falls quickly as the horizon shortens. For four countries, there are no competitive models left for the backcast. This result is another sign that the predictions from statistical models incorporate little information at the two-quarter-ahead horizon. Models that exploit all of the available monthly information fully, including the traditional bridge model, generally stay competitive up to the first quarter ahead horizon. Fifth, Spain is an exceptional case within the analyzed sample of countries, as most of the statistical models perform poorly relative to the AR model, which happens to forecast pretty well. Thus, most of the models appear to have difficulty in capitalizing on their comparative advantage, the monthly information set. In contrast, the Bayesian VAR and the dynamic factor model both perform strongly in the Spanish case.

3.4.2 The marginal value of statistical models

Ranking models according to their RMSFEs gives an initial indication of the relative usefulness of each. This subsection focuses on the marginal value of the various models by investigating whether the forecasts generated by different models differ in their information content. As the various statistical approaches follow different strategies for extracting monthly information, it is conceivable that some models may be complementary. In that case, taking a weighted average of their respective forecasts may improve the forecast accuracy. Even a model that performs badly may have a positive marginal value if it is able to pick up specific useful information. The marginal value of the models is established relative to the best statistical model (lowest RMSFE) by running an encompassing test (e.g. Rünstler et al., 2009 and Stekler, 1991). The test regression is:

$$y^Q_{t+h} = \lambda \hat{y}^Q_{\mathrm{a}(t+h|t)} + (1-\lambda)\hat{y}^Q_{\mathrm{b}(t+h|t)} + \varepsilon_t \tag{3.19}$$

where y^Q_{t+h} is GDP growth in $t+h$, $\hat{y}^Q_{\mathrm{a}(t+h|t)}$ and $\hat{y}^Q_{\mathrm{b}(t+h|t)}$ are the forecasts for quarter $t+h$ on time t of the alternative and best model respectively; λ is the weight of the alternative model; and $(1-\lambda)$ is the weight of the best model. In order to get interpretable results, the restriction that λ lies between 0 and 1 was imposed. The alternative model contains additional information compared to the best model if $\lambda > 0$. λ and its standard error are estimated on the interval [0,1] by Maximum Likelihood (ML). A one-sided (asymptotically valid) test of the hypothesis $\lambda = 0$ is performed at the 5% significance level. All of the calculations refer to the complete sample period 1996.I–2011.III (63 quarters).

Table 3.3 reports the results of the encompassing test. The entries are the RMSFEs of the forecast combination relative to that of the best model, as a measure of the potential gains from using forecast combinations. The estimated weight λ itself is not reported; entries in bold signify λ estimates that are statistically greater than zero. Dots indicate that the ML algorithm returned the corner solution $\lambda = 0$.

The main message of Table 3.3 is that, in economic terms, the gains from combining forecasts from different statistical models are very limited for all countries except for Spain. In the majority of cases, the accuracy gain is zero. Although the gains tend to increase as the horizon shortens and the models have absorbed more information, they typically do not exceed 3%, even for backcasts. Moreover, no model emerges as a clear winner, although the models with the lowest RMSFEs (DFM, F-MIDAS-AR and BVAR) also tend to be the most promising in terms of marginal value. Thus, it appears that the various approaches do not differ greatly with respect to the types of information that they extract from large-scale monthly datasets. Finally, Table 3.3 shows that statistical significance and economic importance are different concepts. Most of the non-zero entries reflect a significant test result for the encompassing test, while most of the gains in forecast accuracy are very small.

3.4.3 Splitting the sample: Great Moderation versus financial crisis

The sample includes the financial crisis, when real GDP went through a particularly volatile phase across the industrialized countries. An obvious question is whether and to what extent the performances of statistical forecasting models differ between the financial crisis period and the period before the financial crisis, which was characterized by a large degree of macroeconomic stability. The latter period has been labeled the Great Moderation. Most of the existing literature on short term forecasting is based on

Table 3.3: Marginal value of statistical models, 1996.I–2011.III)

Frequency			Pooling forecasts				Pooling information			
Model	AR	BVAR	BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area										
2Q-ahead	.	.		.	.	.	**0.99**	.	.	**0.98**
1Q-ahead	.	.	.	.	.	.	.	.		.
nowcast	.	**0.97**	**0.99**	.	**0.99**	.	.	.	.	
backcast	.	.	.	.	.	.		.	.	.
Germany										
2Q-ahead	.	.		.	.	.	.	.	.	**0.99**
1Q-ahead	.	.	.	.	.	.		.	.	0.99
nowcast	.	.	.	.	.	.		.	.	**0.99**
backcast	.	.	.	.	.	.		.	.	.
France										
2Q-ahead	.		.	.	.	.	**0.98**	.	.	**0.99**
1Q-ahead	.	**0.99**	.	.	.	.		.	**0.99**	**0.99**
nowcast	.	**0.99**	.	.	.	.	**0.99**	.		0.99
backcast	.	**0.99**	.	.	.	.		.	.	.
Italy										
2Q-ahead	.		.	.	.	.	.	.	.	.
1Q-ahead	**0.99**	**0.98**	.	**0.99**	**0.98**	.	.	**0.99**	**0.98**	
nowcast	.	.	.	.	.	.	.	.	.	
backcast	.	.	.	.	.	.		.	.	0.99
Spain										
2Q-ahead	**0.98**	**0.98**	.	0.99	**0.99**	.	.	**0.99**		.
1Q-ahead	**0.95**	**0.96**	.	**0.99**	**0.97**	.		**0.99**	.	.
nowcast	.		.	.	.	.	**0.91**	.	**0.93**	**0.96**
backcast	.		.	.	.	.	**0.91**	.	**0.97**	0.99
Netherlands										
2Q-ahead	.		.	.	.	.	.	.	.	.
1Q-ahead	.	**0.99**	.	.	.	.	.	.		.
nowcast	.	**0.99**	**0.99**	.	.	.	**0.99**	.		**0.99**
backcast	.	**0.97**	.	.	.	.		.	.	.

Notes: entries denote RMSFE(forecast combination)/RMSFE(best model). Entries in bold denote that the estimated weight in the encompassing test is statistically different from zero at the 5% significance level. Dots indicate the corner solution of zero weight (λ=0). Grey cells denote models with the lowest RMSFE. AR: autoregressive model, BEQ-AR: bridge equation with AR term, QVAR: quarterly vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, DFM: dynamic factor model, MFVAR: mixed-frequency vector autoregressive model, F-MFVAR: factor-augmented MFVAR model, MIDAS-AR: mixed-data sampling model with AR term, F-MIDAS-AR: factor-augmented MIDAS-AR.

data from the Great Moderation period. Of course, forecasting in volatile times poses greater challenges, meaning that the results of a comparative analysis will be more informative on the issue of which models are most apt at absorbing monthly information. Moreover, good forecasts and nowcasts are of greater importance to economic agents and policy makers in a volatile environment.

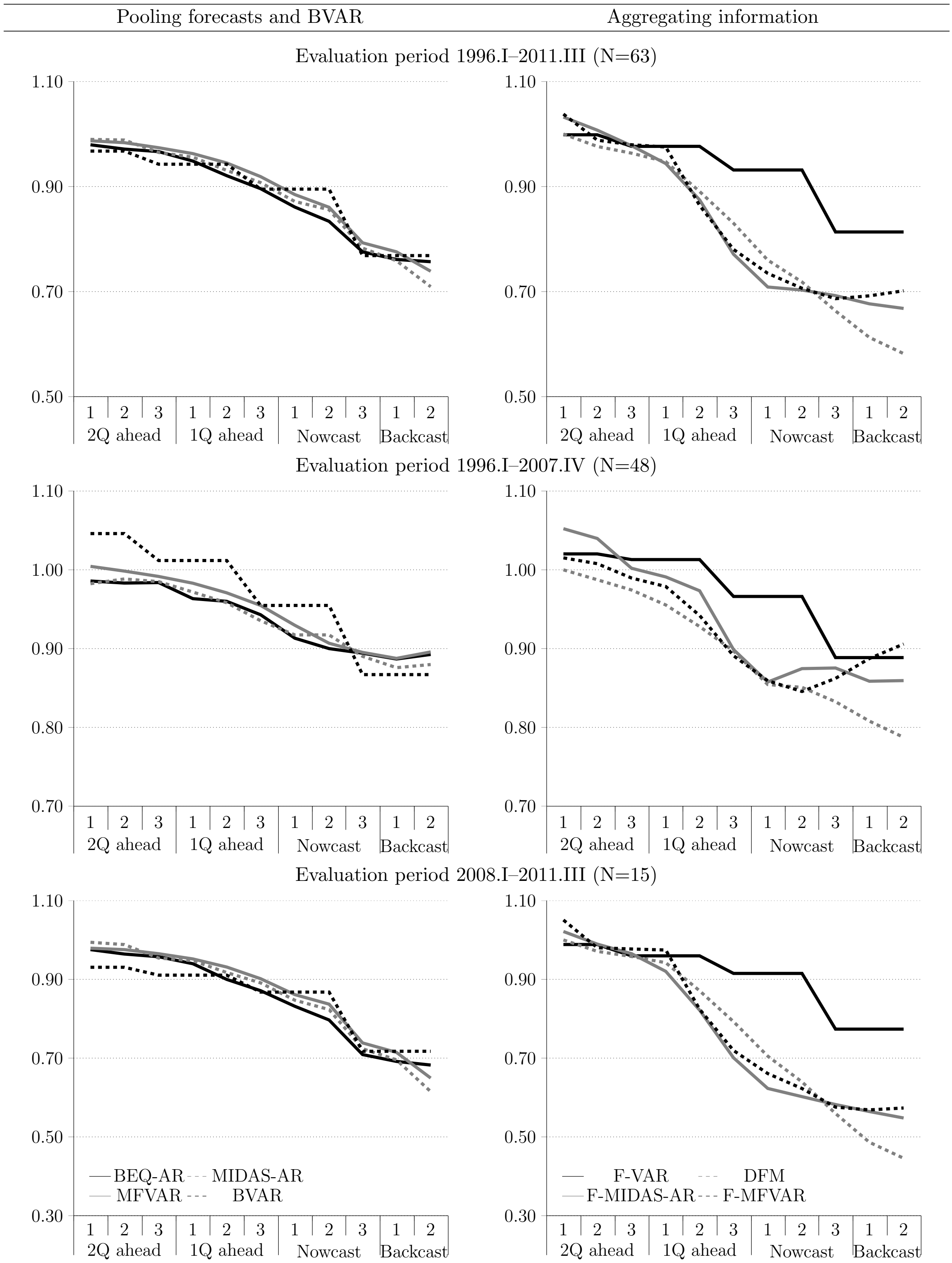
The sample period is divided into two parts: 1996.I–2007.IV (Great Moderation) and 2008.I–2011.III (financial crisis). The models' performances in both periods are discussed based on their so-called learning curves. A learning curve shows the relative decline in the RMSFE of a model as the forecasting horizon shortens, averaged over four countries plus the euro area.[15] A model's learning curve is calculated as the RMSFE standardized by the RMSFE for the first month of the two-quarter ahead DFM-forecast. Figure 3.1 shows the learning curves of selected models for the complete sample period and the two subperiods (in the rows). The graphs on the left refer to three models that aggregate indicator-specific forecasts and the Bayesian VAR, while the graphs on the right refer to the models that rely on factor analysis to summarize the indicators.

For the complete sample period, the dynamic factor model displays the steepest learning curve. Its RMSFE falls by 42% on average within 11 months. In addition, models involving factor analysis have steeper learning curves up to the nowcast, on average, than models that aggregate indicator-specific forecasts. Moreover, the learning curve is rather flat until month four for all models, reflecting the fact that the scope for predicting GDP for horizons beyond one quarter in the future is very limited. This is a stable pattern that holds during both the Great Moderation episode and the crisis episode (and also across countries). The Bayesian VAR is the fastest learning VAR model. Despite the fact that it does not use all available monthly observations, it holds up well against many of the models that do use all of the monthly information. This suggests that the models implicit aggregation by Bayesian shrinkage on coefficients, when applied in a mixed-frequency setting, may turn out to be a viable alternative approach to the information aggregation strategy of factor-based models, especially for backcasts in stable periods. For the US, Schorfheide and Song (2015) find that using within-quarter monthly information leads to drastic improvements in short-horizon forecasting performances.

Predicting GDP is much more difficult in the crisis period (see Tables 3.10–3.13 in Appendix 3.C). The RMSFE of the benchmark model is two to three times as large during the crisis period as during the Great Moderation. However, part of this deterioration can be offset, as the scope for improving forecasts through the utilization of monthly information appears to be larger in volatile times, particularly for nowcasting

[15] Spain is left out, because virtually all statistical models fail to beat the AR benchmark in the period 1996.I–2007.IV. Moreover, this avoids any possible hindsight bias related to the fact that the F-MIDAS-AR model for Spain employs a different number of factors. Country details can be found in Tables 3.10–3.13 in Appendix 3.C, which are the counterparts of Tables 3.2 and 3.3 for both subperiods.

Figure 3.1: Learning curve of statistical models, 1996.I–2011.III



Notes: the lines denote learning curves, defined as RMSFE(forecast model)/RMSFE(first month 2Q-ahead forecast DFM). All forecasts averaged across the euro area, Germany, France, Italy and the Netherlands. BEQ-AR: bridge equation with AR term, MIDAS-AR: mixed-data sampling model with AR term, MFVAR: mixed-frequency vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, DFM: dynamic factor model, F-MIDAS-AR: factor-augmented MIDAS-AR, F-MFVAR: factor-augmented MFVAR model.

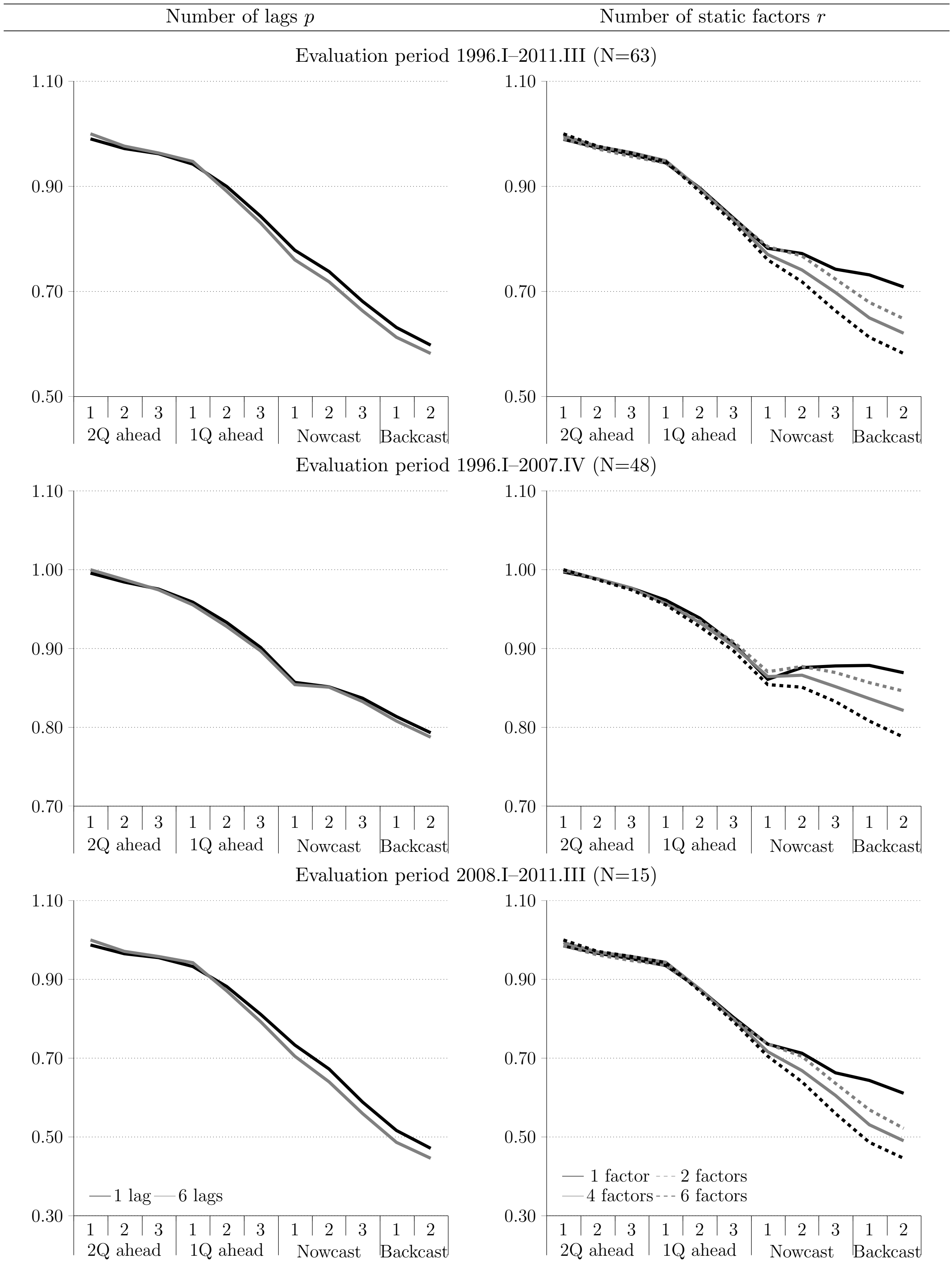
and backcasting. For example, the RMSFE of the dynamic factor model falls by 21% on average over the course of 11 months in the period before the crisis, as compared to 55% in the crisis period. The differences in forecast accuracy across models are considerably larger after the crisis than before the crisis. This also means that the number of competitive models during the Great Moderation is much larger than after the financial crisis, especially for the nowcasting and backcasting horizons. This finding is consistent with the results of D'Agostino and Giannone (2012), who show that the gain from using factor models is substantial, especially in periods of high comovement, as was the case during the financial crisis. The crisis episode poses a more demanding test to models, and consequently, fewer models manage to pass. This finding also implies that the cost of employing a suboptimal model increased after the crisis. Finally, the potential gains of combining statistical models (marginal value) tend to be markedly smaller during the financial crisis than in the preceding period.[16]

Looking at the models that rely on factor analysis, a remarkable result is the strikingly different shapes of the learning curves of factor-augmented mixed-frequency models on the one hand, and the dynamic factor model on the other hand. F-MIDAS-AR and F-MFVAR learn faster than the dynamic factor model between months four and seven, but the pace of improvement quickly levels off beyond that point. As a result, their one-quarter-ahead forecasts and early nowcasts are more accurate than their DFM counterparts on average. This good performance is attributable entirely to the financial crisis episode. During the Great Moderation, their forecasts do not show any improvement at all after month seven. In contrast, the DFM improves its forecast steadily as more and more information is absorbed, producing superior backcasts and late nowcasts in both quiet and volatile times. Moreover, the DFM delivered better or equally good forecasts over the whole horizon during the Great Moderation period.

This pattern reflects the comparative strengths and weaknesses of the two model approaches, which may play out differently at different horizons and in different circumstances. To gain additional insight into why the DFM works, Figure 3.2 shows its learning curve for different numbers of lags (p) and (static) factors (r). The learning curves indicate that the performance of the DFM for nowcasts and backcasts is strongly linked to the number of factors. However, additional factors do not offer benefits for predictions beyond the current quarter; then, one factor is then sufficient. When the DFM is restricted to only one factor, its learning curve has the same shape as that of the F-MIDAS-AR model. Moreover, the number of lags in the factor VAR process is only a minor determinant of the forecast quality at any horizon. As a consequence, the comparative strength of the DFM is its ability to include more information in the forecasting procedure. In contrast, this aspect is the weak point of factor augmented models,

[16] Moreover, the estimated weight (λ) in the encompassing is significant in only a few cases, but this can partly be attributed to the low number of observations.

Figure 3.2: Learning curve of dynamic factor model by number of lags and factors, 1996.I–2011.III



Notes: the lines denote learning curves, defined as RMSFE(forecast model)/RMSFE(first month 2Q-ahead forecast DFM with 6 lags ($p = 6$) and RMSFE(forecast model)/RMSFE(first month 2Q-ahead forecast DFM with 6 static factors ($r = 6$), respectively. All forecasts averaged across the euro area, Germany, France, Italy and the Netherlands.

which can exploit only one factor in practice (see Section 3.2.4). This hurts their performances for late nowcasts and backcasts. On the other hand, factor-augmented models (and MIDAS models in particular) may exploit their richer dynamic specification of the relationship between indicator and GDP. This may give them an edge over the DFM, in particular for one-quarter ahead predictions and early nowcasts for which the single-factor restriction is not a disadvantage. However, a flexible dynamic specification is an asset for forecasting only if it is feasible to identify stable dynamic relationships reliably. Otherwise, these models may estimate spurious dynamic relationships in-sample, which may actually reduce the accuracy of the out-of-sample predictions. The evidence in Figure 3.1 highlights this identification problem. During the Great Moderation, factor augmented models were unable to capitalize on their comparative advantage; however, that changed dramatically after the crisis hit. In volatile times, when it is easier to identify dynamic relationships, factor augmented models may deliver the best forecasts for specific horizons: (late) one-quarter ahead forecasts and early nowcasts. Dynamic flexibility tends to increase RMSFEs for two-quarter ahead forecasts in all environments, due to the very limited scope for forecasting that far into the future. Thus, these findings suggest that the question of whether or not to use factor augmented models in practice for one-quarter ahead forecasts and early nowcasts should depend on the researchers or practitioners confidence in the models abilities to uncover useful dynamic relationships.

3.4.4 Assessing model features

The fact that the conducted analysis includes many models and five countries plus euro area allows shedding some light on the issue which model features are especially valuable for forecasting and nowcasting. The following modeling choices are evaluated: (1) employing factor analysis to summarize monthly information; (2) using all available monthly information; (3) exploiting GDP's own past by adding an autoregressive term to the forecasting equation. To assess the effect of a specific model feature on the RMSFE, (pairs of) models that differ only in that aspect are compared. Moreover, the effects are averaged over four countries plus the euro area (excluding Spain again) to average out the country-specific component.

To measure the impact of utilizing factor analysis for aggregating monthly information rather than aggregating indicator-specific forecasts four pairs of models can be compared: (F-VAR, QVAR), (F-MIDAS, MIDAS), (F-MIDAS-AR, MIDAS-AR) and (F-MFVAR, MFVAR). The effect of using all available information on indicators can be measured by comparing the quarterly VAR models with their mixed-frequency counterparts: (MFVAR, QVAR) and (F-MFVAR, F-VAR). This comparison also involves the effect of making GDP a monthly latent variable in a system. For the AR effect three

Table 3.4: Effects of model features on forecasting performance, 1996.I–2011.III

Evaluation period		1996.I–2011.III (N=63)				1996.I–2007.IV (N=48)				2008.I–2011.III (N=15)			
Forecast horizon		2Q ahead	1Q ahead	nowcast	backcast	2Q ahead	1Q ahead	nowcast	backcast	2Q ahead	1Q ahead	nowcast	backcast
Aggregate information first													
F-VAR	QVAR	-0.02	-0.03	-0.04	-0.03	0.01	0.01	-0.02	-0.04	-0.03	-0.05	-0.04	-0.02
F-MFVAR	MFVAR	0.02	-0.07	-0.18	-0.08	0.01	-0.03	-0.07	0.01	0.03	-0.09	-0.27	-0.18
F-MIDAS	MIDAS	0.00	-0.07	-0.18	-0.13	0.02	-0.01	-0.06	-0.04	-0.01	-0.10	-0.28	-0.23
F-MIDAS-AR	MIDAS-AR	0.02	-0.07	-0.18	-0.09	0.04	0.00	-0.04	-0.02	0.01	-0.12	-0.28	-0.16
Average		**0.01**	**-0.06**	**-0.14**	**-0.08**	**0.02**	**-0.01**	**-0.05**	**-0.02**	**0.00**	**-0.09**	**-0.22**	**-0.15**
Full use of monthly information													
F-MFVAR	F-VAR	0.01	-0.09	-0.24	-0.15	-0.01	-0.06	-0.11	0.02	0.02	-0.11	-0.34	-0.30
MFVAR	QVAR	-0.02	-0.05	-0.09	-0.10	0.00	-0.01	-0.05	-0.04	-0.03	-0.07	-0.11	-0.14
Average		**-0.01**	**-0.07**	**-0.16**	**-0.12**	**-0.01**	**-0.04**	**-0.08**	**-0.01**	**0.00**	**-0.09**	**-0.23**	**-0.22**
Autoregressive term													
MIDAS-AR	MIDAS	0.01	-0.01	-0.03	-0.05	0.01	0.00	-0.02	-0.01	0.01	-0.01	-0.04	-0.08
F-MIDAS-AR	F-MIDAS	0.03	-0.01	-0.03	0.00	0.03	0.01	-0.01	0.00	0.04	-0.02	-0.05	-0.01
BEQ-AR	BEQ	0.00	0.00	-0.02	-0.04	0.00	0.00	-0.01	0.01	0.00	0.00	-0.03	-0.08
Average		**0.02**	**-0.01**	**-0.03**	**-0.03**	**0.01**	**0.00**	**-0.01**	**0.00**	**0.02**	**-0.01**	**-0.04**	**-0.06**

Notes: entries denote change in forecasting accuracy, calculated as (RMSFE(forecast model) - RMSFE(forecast base model))/RMSFE(forecast base model) averaged across the euro area, Germany, France, Italy and the Netherlands. BEQ: bridge equation, BEQ-AR: BEQ with AR term, QVAR: quarterly vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, MFVAR: mixed-frequency vector autoregressive model, F-MFVAR: factor-augmented MFVAR model, MIDAS: mixed-data sampling model, MIDAS-AR: MIDAS with AR term, F-MIDAS: factor-augmented MIDAS, F-MIDAS-AR: F-MIDAS with AR term.

pairs are analyzed: (BEQ-AR, BEQ), (MIDAS-AR, MIDAS) and (F-MIDAS-AR, F-MIDAS).

Table 3.4 reports the impacts of the three model features (averaged over four countries and the euro area) for the complete sample period and the two subperiods. Starting with the effect of utilizing factor analysis, it improves the forecasting accuracy substantially for all horizons, and for nowcasts in particular. The gains are much larger for mixed-frequency models than for quarterly models. For the complete sample, the average gain is 14% for nowcasts, 8% for backcasts, and 6% for one-quarter ahead forecasts. This suggests that summarizing the information from monthly data is especially helpful when the information pertains to the quarter of interest itself. When forecasting or backcasting, the inevitable loss of information due to summarizing appears to partly offset any gains that arise from the removal of noise. Moreover, there is an interesting difference between tranquil and volatile times. Using factors produces only modest gains in tranquil times, when GDP develops rather smoothly. In such periods, there is little information available in the first place, and the information losses due to summarizing may be comparatively severe relative to the gains from the removal of noise. In volatile times, when the indicators display a larger degree of comovement, the gains are much larger: up to 22% for nowcasts and 15% for backcasts. Next, the effect of using all available monthly observations is discussed. This effect is also sizable for all horizons except for the two-quarter-ahead forecast. For the full sample, the RMSFEs of nowcasts decrease by 16% and those of backcasts by 12%. Again, there is a large difference between the pre-crisis and the crisis periods. The gains from using monthly information are realized primarily in volatile episodes, as is evidenced by the 23% and 22% gains in accuracy for nowcasts and backcasts, respectively, in the crisis period. In contrast, the gains during the Great Moderation period are (very) modest, once again suggesting that the information content of the monthly dataset is low in stable environments. Finally, exploiting GDPs own past by adding an AR term has small positive effects on the forecasting accuracy of nowcasts and backcasts, but only during the crisis episode; the nowcasts improve by 3% and the backcasts by 6%.

3.5 Analysis of forecasts by professional analysts

The views of professional forecasters are an alternative and convenient source of information for policy makers and market participants. Currently, several surveys on the economic outlook are available on a regular basis. The European Central Bank undertakes a quarterly survey among professional forecasters to obtain information on inflation expectations and growth prospects for the euro area. In the US, the Federal Reserve Bank of Philadelphia runs a well-known survey. Moreover, the private sector firm Consensus Economics collects and publishes economic forecasts on a monthly basis in the publication *Consensus forecasts*. Consensus forecasts offers an overview private sector analysts' expectations for a set of key macroeconomic variables for a broad range of countries. Consensus forecasts is best known for its expectations on annual GDP growth for the current and next year. However, it also provides quarterly forecasts for GDP, which will be used in this chapter.[17] The panelists supply their forecasts for six consecutive quarters, starting from the first unpublished quarter. The numbers of respondents varies somewhat over time, but on average about nine institutions participate in the poll for the Netherlands, fifteen each for Italy and Spain, twenty for France and thirty for Germany and the euro area.

This section investigates two issues. The first issue is the quality of Consensus forecasts as a separate forecasting device, relative to the best statistical model. The second issue is the marginal value of Consensus forecasts, based on an encompassing test versus the three best models (DFM, BVAR and F-MIDAS-AR). In forming their expectations, analysts include subjective assessments of (potentially) a multitude of relevant factors, alongside presumably model-based predictions. If a mixture of model-based and (subjective) Consensus forecasts improves the accuracy of forecasts, this can be viewed as evidence that forecasts by analysts do indeed embody a different type of valuable information (subjective judgments).

The mean quarterly forecast is used as the measure of private sector expectations in the analysis. Fresh Consensus forecasts become available only once a quarter, in the second week of the last month of the quarter. For the information set, this means that Consensus forecasts are not updated in the first and second month in a quarter, while the monthly indicator series are updated every month. Moreover, at the time panelists form their expectations they have information on GDP growth in the preceding quarter. Thus, the Consensus backcast for quarter t is equal to the non-updated Consensus forecast published in the last month of quarter t.

Table 3.5 presents the results for Consensus forecasts for the complete sample period, the pre-crisis period and the crisis period.[18] For the one- and two-quarter ahead forecasts, the Consensus forecasts are better than the best statistical model in the case of Spain, and competitive for three other countries (measured over the whole sample). However, the performance relative to the best model is weak for nowcasts and particularly backcasts for all countries except for Germany and Spain. Consequently, purely mechanical models seem to be more adept at learning when monthly information regarding the quarter of interest becomes available. In the relatively stable pre-crisis period, the Consensus forecasts fare very poorly, usually ranking at the bottom of the

[17] The annual Consensus forecasts have been analyzed in several papers (e.g. Ager et al., 2009; Batchelor, 2001; Loungani and Rodriguez, 2008; Lahiri et al., 2006). The quarterly forecasts have not been used before, except for the case study for the Netherlands in Chapter 2 in this thesis.

[18] Consensus forecasts for the euro area are available from March 2002 onward only, so the results in Table 3.5 refer to the period 2003.III–2011.III for the euro area.

Table 3.5: Comparison forecasting performance of Consensus forecasts with the best statistical models, 1996.I.2011.III

Evaluation period	1996.I–2011.III (N=63)					1996.I–2007.IV (N=48)					2008.I–2011.III (N=15)				
Indicator	rRMSFE best	rank	gain BVAR	gain DFM	gain F-MIDAS-AR	rRMSFE best	rank	gain BVAR	gain DFM	gain F-MIDAS-AR	rRMSFE best	rank	gain BVAR	gain DFM	gain F-MIDAS-AR
Euro area															
2Q-ahead	0.99	1	**0.99**	**0.95**	**0.92**	1.21	10	.	.	.	0.99	1	0.99	**0.93**	**0.91**
1Q-ahead	1.07	3	**0.95**	**0.95**		1.42	10	.	.	.	1.06	2	**0.94**	**0.94**	
nowcast	1.18	4	**0.92**	**0.99**	**0.97**	1.42	10	.		**0.98**	1.17	4	**0.89**	**0.98**	**0.98**
backcast	1.37	5	**0.89**	0.99	**0.92**	1.70	10	.		**0.94**	1.32	4	**0.84**		**0.91**
Germany															
2Q-ahead	1.02	4	**0.97**	**0.99**	**0.98**	1.03	6	**0.96**	**0.98**	**0.93**	1.03	5	0.97	.	
1Q-ahead	1.02	3	**0.94**	**0.98**	**0.97**	0.98	1	**0.95**	**0.96**	**0.91**	1.06	4	0.92	.	0.99
nowcast	1.07	4	**0.87**	**0.98**	**0.95**	0.93	1	**0.90**	**0.89**	**0.84**	1.26	4	0.85	.	.
backcast	1.01	2	**0.78**	**0.93**	**0.88**	0.77	1	**0.77**	**0.76**	**0.72**	1.40	4	0.78		0.99
France															
2Q-ahead	1.05	7		**0.99**	**0.96**	1.22	10	**0.99**		**0.98**	1.06	2		**0.96**	**0.92**
1Q-ahead	1.11	8	**0.99**		**0.99**	1.27	10	**0.98**		**0.98**	1.06	5	.	**0.98**	0.99
nowcast	1.31	7	**0.96**	.	**0.98**	1.34	10	**0.95**		**0.96**	1.32	4	**0.95**	0.99	0.99
backcast	1.29	4	**0.89**	**0.98**	**0.91**	1.34	8	**0.91**	**0.99**	**0.90**	1.23	4	**0.82**	**0.98**	**0.92**
Italy															
2Q-ahead	1.13	10		.	**0.97**	1.31	10	.	.	.	1.03	2		**0.96**	**0.91**
1Q-ahead	1.15	10	.	.	.	1.38	10	.	.	.	1.01	2	**0.96**	**0.95**	0.99
nowcast	1.30	9	**0.98**	.	.	1.46	10	.	.	.	1.23	4	**0.92**	.	
backcast	1.35	10	**0.97**		**0.98**	1.40	10	**0.98**	0.99	**0.98**	1.34	4	**0.92**		0.98
Spain															
2Q-ahead	0.93	1	**0.92**	**0.89**	**0.92**	1.23	10		.	.	0.84	1	**0.82**	**0.81**	**0.84**
1Q-ahead	0.98	1	**0.93**	**0.91**	**0.82**	1.31	10		.	**0.96**	0.85	1	**0.80**	**0.82**	**0.74**
nowcast	1.03	3	**0.95**	**0.86**	**0.78**	1.44	10		**0.98**	**0.96**	0.81	1	**0.76**	**0.71**	**0.59**
backcast	1.13	2	**0.97**	**0.80**	**0.74**	1.48	7		**0.94**	**0.93**	0.69	1	**0.69**	**0.51**	**0.43**
Netherlands															
2Q-ahead	1.11	9		**0.97**	**0.98**	1.31	10	.	.		0.95	1	**0.95**	**0.85**	**0.89**
1Q-ahead	1.08	4	**0.97**	**0.95**	**0.97**	1.16	10	.	**0.99**	**0.99**	1.01	2	**0.88**	**0.88**	**0.94**
nowcast	1.17	7	**0.96**	**0.97**	**0.96**	1.25	10	0.99	**0.98**	**0.98**	1.11	4	**0.83**	**0.93**	**0.92**
backcast	1.30	10	**0.97**	**0.97**	**0.96**	1.44	10		0.99	.	1.14	2	**0.82**	**0.91**	**0.82**

Notes: entries denote rRMSFEs: RMSFE(Consensus)/RMSFE(best statistical model); ranks: ranking among 10 procedures (9 statistical models and Consensus forecasts) and gains: RMSFE(combination of Consensus and statistical model)/RMSFE(statistical model), respectively. Grey cells denote models with the lowest RMSFE. Entries in bold denote that the estimated weight in the encompassing test is statistically different from zero at the 5% significance level. Full sample and Great Moderation sample start in 2003.III for the euro area. BVAR: Bayesian quarterly VAR model, DFM: dynamic factor model, F-MIDAS-AR: factor-augmented MIDAS model with AR term.

list. However, they do very well in the case of Germany, which suggests that analysts were able to assess the economic conditions better during and after the extraordinary episode of German reunification in the early 1990s. In contrast, Consensus forecasts perform much better during the crisis period, when GDP and the monthly indicators displayed extreme fluctuations. At the one- and two-quarter ahead horizons, the Consensus forecasts belong in the top three models in most cases. For Spain and the Netherlands, the difference in forecasting precision is substantial. This suggests that analysts are able to handle extreme observations of GDP growth and auxiliary indicators better once they have occurred, while the quality of recursively estimated models in mechanical procedures is more susceptible to extreme observations in the sample, particularly when truly forecasting. The main findings support those of Lundquist and Stekler (2012), who conclude that professional forecasters are very responsive to the latest information about the state of the economy and adjust their predictions quickly. Despite this head start, private sector forecasts still fall behind the best model in most cases as the horizon becomes shorter and more timely monthly information becomes available to improve forecasts. For example, leaving aside Spain, the RMSFEs of backcasts by Consensus forecasts are between 14% and 40% larger than those associated with the best model (dynamic factor model).

Despite the fact that the Consensus forecasts are a rather poor predictor of GDP on their own, the results for the encompassing test show that they often still contain valuable extra information. The effects are generally smaller for the best statistical model, which usually differs across horizons. This implies that a combination of model-based predictions and Consensus forecasts narrows the differences between the best models. This would lower the cost of using a single model for all horizons for practitioners. In many cases, an accuracy improvement of around 10% is feasible. The effects tend to be stronger for backcasts by analysts, even though these actually reflect relatively dated information. During the crisis period, Consensus forecasts, unlike their statistical competitors, still offer great added value compared to the best statistical model for Spain, and to a lesser extent for the Netherlands. Moreover, they can be used to improve the predictions of near-best models significantly for almost all countries. The added value is smaller in the pre-crisis period, except for Germany. All in all, the outcomes of the encompassing test suggest that subjective private sector forecasts potentially contain information that sophisticated mechanical forecasting procedures are unable to pick up.[19]

[19] As fresh Consensus forecasts become available only in the third month of the quarter, the month-by-month pattern of the results is also interesting. The relative RMSFE of Consensus forecasts (versus the best statistical model) improves in third months, when they are new, and deteriorates in other months, when statistical models can take advantage of newly available monthly information. However, the results for the encompassing test on a month-by-month basis show that the value added of Consensus forecasts does not significantly decrease with their age. Even Consensus forecasts that are one or two months old often contain valuable extra information. This finding reinforces the conjecture

3.6 Conclusion

This chapter makes two contributions to the empirical literature on forecasting real GDP in the short run. The first contribution is a systematic comparison of twelve statistical linear models for five countries (Germany, France, Italy, Spain and the Netherlands) and the euro area, utilizing the same information set across countries plus the euro area. The sample period (1996.I–2011.III) allows comparison of the models' forecasting abilities in the period before the financial crisis of 2008 (Great Moderation) and the much more volatile period that followed (the financial crisis and its aftermath). The second contribution concerns the potential usefulness of (subjective) forecasts made by professional analysts. Such forecasts are very cheap and easy to use, and may incorporate valuable information that goes beyond purely statistical data.

The main findings can be summarized in five points. First, monthly indicators contain valuable information that can be extracted by mechanical statistical procedures, particularly as the horizon shortens and more monthly information is processed. The largest accuracy gains are for nowcasts and backcasts, suggesting that statistical models are especially helpful when they are able to use information that pertains to the quarter of interest. Moreover, statistical models are generally more efficient at extracting monthly information in volatile periods. Thus, their relative strength is to improve the assessment of the current state of the economy. In contrast, the predictions from statistical models generally incorporate little information at the two-quarter ahead horizon.

Second, the dynamic factor model displays the best forecasting capabilities overall, particularly for backcasts and nowcasts. Its ability to incorporate more than one factor, and thus, more information, is key to this result. Factor-augmented MFVAR and MIDAS models produce better one-quarter-ahead predictions after the financial crisis, due to their richer dynamic specifications. However, the latter feature does not appear to be an advantage in stable times. The Bayesian quarterly VAR is the best quarterly model. It performs quite well for Germany, the Netherlands and Spain in the more stable period of the Great Moderation. Remarkably, all of the other models, including the dynamic factor model, perform (very) poorly in the case of Spain during the Great Moderation. These findings suggest that Bayesian estimation is a fundamentally different way of extracting information from a large data set, which may deliver benefits, even if the model is inefficient in its use of the available monthly information.

Third, regarding crucial model features, employing factor analysis to summarize the available monthly information clearly delivers better results than the alternative of averaging single-indicator-based forecasts in the case of one-quarter ahead forecasts and nowcasts. Strategies that aggregate information work better than strategies that pool forecasts. Moreover, it is important for a model to make use of all of the available

---

that analysts' forecasts embody information that differs in nature from the information that can be filtered out of statistical data.

monthly observations. On the other hand, allowing for autoregressive terms (GDP's own past) in forecasting equations leads to only minor improvements in forecast reliability. All of these effects are more pronounced during the crisis period, implying that the cost of employing a suboptimal forecasting model is larger in periods of high volatility.

Fourth, statistical models differ significantly in the rates at which they are able to absorb monthly information as time goes by. However, the information content of the resulting forecasts appears to overlap to a large extent, and the unique model-specific component appears to be small (relative to the best model). The different models do not seem to have any comparative advantage of extracting certain types of information, offering perspectives that complement each other. The scope for improving GDP forecasts by combining the 'views' of various models is rather limited in economic terms, although there are some exceptions. This is particularly true during volatile episodes, when reliable assessments of the current situation and short run prospects are most needed, unfortunately.

Lastly, forecasts by professional analysts, which contain judgmental elements, appear to be a different category. Such forecasts are in many cases a rather poor predictor of GDP compared to the best statistical model. However, they tend to perform better during the crisis, when it really counts, and they often embody information that sophisticated mechanical forecasting procedures fail to pick up. Thus, subjective private sector analysts' forecasts seem to offer the potential of enhancing mechanical forecasts.

The results of the large-scale comparative analysis may be useful to policy makers, financial analysts and economic agents, as information on where the economy stands and where it is heading to in the short run is particularly valuable in times of great uncertainty. The dynamic factor model and factor-augmented statistical models are obvious candidate models for generating short-term forecasts in practice.

Appendix

3.A Dataset

The main source of the monthly data is the ECB statistical datawarehouse. World trade and world industrial production are from the CPB world trade monitor. Commodity prices and most financial market indicators were taken from Thomson Reuters datastream and most of the survey data from the European Commission. Table 3.6 provides an overview of all monthly variables, the applied transformations and the starting date of the monthly series at the data source. Note that only observations from January 1985 onwards have been used in estimation.

Table 3.6: Description monthly dataset

No.	Variable	Type	Transformation			Country					
			ln.	dif.	filter	EA	DE	FR	IT	ES	NL
1	World Trade (CPB)	Sales	1	1	3	’77	’77	’77	’77	’77	’77
2	World Industrial Production (CPB)	Sales	1	1	3	’91	’91	’91	’91	’91	’91
3	Ind. production United States	Sales	1	1	3	’77	’77	’77	’77	’77	’77
4	Ind. production United Kingdom	Sales	1	1	3	’77	’77	’77	’77	’77	’77
5	Ind. production (excl. construction)	Sales	1	1	3	’77	’77	’77	’77	’77	’77
6	Ind. production, cons. goods ind.	Sales	1	1	3	’80	’80	’77	’77	’77	’90
7	Ind. production, energy	Sales	1	1	3	’80	’80	’77	’80	’80	’90
8	Ind. production, interm. goods ind.	Sales	1	1	3	’77	’80	’77	’77	’77	’95
9	Ind. production, capital goods	Sales	1	1	3	’77	’80	’77	’77	’77	’77
10	Ind. production, manufacturing	Sales	1	1	3	’77	’78	’77	’77	’80	’77
11	Ind. production, construction	Sales	1	1	3	’85	’78	’85	’95	’88	’85
12	New orders manufacturing	Sales	1	1	3	’95	’91	’00	’90	’00	’95
13	New passenger cars (reg.)	Sales	1	1	3	’90	’90	’90	’90	’90	’90
14	New commercial vehicles (reg.)	Sales	1	1	3	’90	’90	’90	’90	’90	’90
15	Retail trade volume	Sales	1	1	3	’77	’77	’77	’90	’95	’77
16	Unemployment rate	Sales	0	1	3	’83	’91	’83	’83	’86	’83
17	Unemployment rate United Kingdom	Sales	0	1	3	’83	’83	’83	’83	’83	’83
18	Unemployment rate United States	Sales	0	1	3	’83	’83	’83	’83	’83	’83
19	Exports	Sales	1	1	3	’00	’89	’89	’89	’89	’89
20	Imports	Sales	1	1	3	’00	’89	’89	’89	’89	’89
21	Total HICP-index	Prices	1	2	3	’77	’77	’77	’77	’77	’77
22	Core HICP-index	Prices	1	2	3	’77	’77	’77	’77	’77	’77
23	CPI, food	Prices	1	2	3	’90	’77	’77	’77	’93	’77
24	CPI, energy	Prices	1	2	3	’90	’77	’77	’77	’77	’77
25	HICP, services	Prices	1	2	3	’90	’85	’90	’87	’92	’87
26	Producer prices (total, excl. constr.)	Prices	1	2	3	’81	’77	’77	’77	’77	’77
27	World commodity prices, total	Prices	1	2	3	’77	’77	’77	’77	’77	’77
28	World commodity prices, raw mat.	Prices	1	2	3	’77	’77	’77	’77	’77	’77
29	World commodity prices, food	Prices	1	2	3	’77	’77	’77	’77	’77	’77
30	World commodity prices, metals	Prices	1	2	3	’77	’77	’77	’77	’77	’77
31	World commodity prices, energy	Prices	1	2	3	’77	’77	’77	’77	’77	’77
32	Oil price (1 month future Brent)	Prices	1	2	3	’77	’77	’77	’77	’77	’77

Continued on next page...

Table 3.6 – Continued

No.	Variable	Type	Transformation ln.	dif.	filter	Country EA	DE	FR	IT	ES	NL
33	M1	Finan.	1	1	3	'77	'80	'77	'80	'80	'80
34	M3	Finan.	1	1	3	'77	'77	'77	'77	'77	'77
35	Interest rate on mortgage	Finan.	0	1	3	'03	'82	'80	'95	'84	'80
36	3 month interest rate euro	Finan.	0	1	3	'94	'77	'77	'77	'77	'77
37	10 year government bond yield	Finan.	0	1	3	'77	'94	'77	'77	'80	'77
40	Headline stock-index	Finan.	1	1	3	'77	'77	'77	'77	'87	'77
41	Basic Material-index	Finan.	1	1	3	'77	'77	'77	'77	'87	'77
42	Industrials stock-index	Finan.	1	1	3	'77	'77	'77	'77	'87	'77
43	Consumer goods stock-index	Finan.	1	1	3	'77	'77	'77	'77	'87	'77
44	Consumer services stock-index	Finan.	1	1	3	'77	'77	'77	'87	'77	'77
45	Financials stock-index	Finan.	1	1	3	'77	'77	'77	'77	'87	'77
46	Technology stock-index	Finan.	1	1	3	'77	'88	'77	'86	'99	'85
47	Loans to the private sector	Finan.	1	1	3	'91	'80	'80	'83	'80	'82
48	Exchange rate, US-Dollar per Euro	Finan.	1	1	3	'80	'80	'80	'80	'80	'80
49	Real effective exchange rate (CPI)	Finan.	1	1	3	'77	'77	'77	'77	'77	'77
50	Ind. conf. - headline	Survey	0	1	3	'85	'85	'85	'85	'87	'85
51	Ind. conf. - order-book expect.	Survey	0	1	3	'85	'85	'85	'85	'87	'85
52	Ind. conf. - stocks expect.	Survey	0	1	3	'85	'85	'85	'85	'87	'85
53	Ind. conf. - production expect.	Survey	0	1	3	'85	'85	'85	'85	'87	'85
54	Ind. conf. - employment expect.	Survey	0	1	3	'85	'85	'85	'85	'87	'85
55	Cons. conf. - headline	Survey	0	1	3	'85	'85	'85	'85	'86	'85
56	Cons. conf. - financial sit.	Survey	0	1	3	'85	'85	'85	'85	'86	'85
57	Cons. conf. - general ec. sit.	Survey	0	1	3	'85	'85	'85	'85	'86	'85
58	Cons. conf. - unemployment expect.	Survey	0	1	3	'85	'85	'85	'85	'86	'85
59	Cons. conf. - major purchases expect.	Survey	0	1	3	'85	'85	'85	'85	'86	'85
60	Constr. conf. - headline	Survey	0	1	3	'85	'85	'85	'85	'89	'85
61	Constr. conf. - order book (evolution)	Survey	0	1	3	'85	'85	'85	'85	'89	'85
62	Constr. conf. - employment expect.	Survey	0	1	3	'85	'85	'85	'85	'89	'85
63	Retail conf. - headline	Survey	0	1	3	'85	'85	'85	'85	'88	'86
64	Retail conf. - current Stocks (volume)	Survey	0	1	3	'85	'85	'85	'85	'88	'86
65	Retail conf. - orders expectations	Survey	0	1	3	'85	'85	'85	'85	'88	'86
66	Retail conf. - business expect.	Survey	0	1	3	'85	'85	'85	'85	'88	'86
67	Retail conf. - employment expect.	Survey	0	1	3	'86	'85	'85	'86	'88	'86
68	PMI United States	Survey	0	1	3	'77	'77	'77	'77	'77	'77
69	PMI United Kingdom	Survey	0	1	3	'92	'92	'92	'92	'92	'92
70	OECD composite leading ind. UK	Other	0	1	3	'77	'77	'77	'77	'77	'77
71	OECD composite leading ind. US	Other	0	1	3	'77	'77	'77	'77	'77	'77
72	OECD composite leading ind.	Other	0	1	3	'77	'77	'77	'77	'77	'77

Notes: entries denote variable number, name, category, transformation and starting year for each country in the dataset. Type: sales= quantitative information; prices= consumer and producer prices; finan.= financial and monetary variables; survey= qualitative information; other= other. Ln.: 0= no logarithm; 1= logarithm. Dif.: 1= first difference; 2= second difference. Filter: 3= change against the same month of the previous month. Country: EA: euro area; DE: Germany; FR: France; IT: Italy; ES: Spain; NL: the Netherlands.

Quarterly GDP data for France, Italy, the Netherlands and Spain were taken from the OECD's main economic indicators database. The source of the German GDP data is the Deutsche Bundesbank, who constructed GDP series using only GDP data for West Germany pre 1991.I and the re-unified Germany from 1991.I onwards. A synthetic

GDP series for the euro area was constructed using the database in the ECB's area wide model, that is supplemented with data from the OECD's main economic indicator database.

3.B State space representations

3.B.1 Mixed-frequency VAR

This section describes the state space representation of the mixed-frequency VAR described in Section 3.2.3. Let $p^* = \max(p, 3)$ and the transition equation of state vector is as follows:

$$\begin{bmatrix} z_{i,t+1} - \mu_{z_i} \\ z_{i,t} - \mu_{z_i} \\ \dots \\ z_{i,t-p^*+2} - \mu_{z_i} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \dots & A_p & 0_{2\times 2(3-p^*)} \\ I_{2(p^*-1)} & & & & 0_{2(p^*-1)\times 2} \end{bmatrix} \begin{bmatrix} z_{i,t} - \mu_{z_i} \\ z_{i,t-1} - \mu_{z_i} \\ \dots \\ z_{i,t-p^*+1} - \mu_{z_i} \end{bmatrix} + \begin{bmatrix} \Sigma_\varepsilon^{1/2} \\ 0_{2(p^*-1)\times 2} \end{bmatrix} v_t, \tag{3.20}$$

where $v_t \sim N(0, I_2)$. The measurement equation is:

$$z_{i,t}^Q - \mu_{z_i^Q} = \begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0_{1\times(p^*-6)} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0_{1\times(p^*-6)} \end{bmatrix} \begin{bmatrix} z_{i,t} - \mu_{z_i} \\ z_{i,t-1} - \mu_{z_i} \\ \dots \\ z_{i,t-p^*+1} - \mu_{z_i} \end{bmatrix} \tag{3.21}$$

Since y_t^Q is assigned to the third month of the quarter, the missing observations in months 1 and 2 are replaced with a random draw from the standard normal distribution $N(0, 1)$, as in Mariano and Murasawa (2010). The measurement equation of month 1 and month 2 is modified in accordance with the missing observation treatment. For months for which y_t^Q is unavailable, the upper row of the matrix on the right hand side of Eq. (3.21) is set equal to zero and white noise is added.

3.B.2 Dynamic factor model

The equations of the DFM, Eqs. (3.5)–(3.8), can be cast in state space form as illustrated below for the case of $p = 1$. The aggregation rule is implemented in a recursive way in Eq. (3.23) by introducing a latent cumulator variable Ξ for which: $\Xi_t = 0$ for t corresponding to the first month of the quarter and $\Xi_t = 1$ otherwise. The monthly state space representation is given by the following observation equation:

$$\begin{bmatrix} x_t \\ y_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \xi_t \\ \varepsilon_t^Q \end{bmatrix} \tag{3.22}$$

and the transition equation:

$$\begin{bmatrix} I_r & 0 & 0 \\ -\beta' & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} f_{t+1} \\ y_{t+1} \\ \hat{y}_{t+1}^Q \end{bmatrix} = \begin{bmatrix} A_{r1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Xi_{t+1} \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \varepsilon_t \\ 0 \end{bmatrix} \tag{3.23}$$

The application of the Kalman filter and smoother provides the minimum mean square linear estimates (MMSLE) of the state vector $\alpha_t = (f_t, y_t, \hat{y}_t^Q)$ and enables the forecasting of quarterly GDP growth y_t^Q and dealing efficiently with an unbalanced dataset of missing observations at the beginning and at the end of the series by replacing the missing data with optimal predictions. Moreover, when compared with using principal components technique alone, the two-step estimator allows for dynamics of the common factors and cross-sectional heteroskedasticity of the idiosyncratic component.

3.C Additional results

Tables 3.7 and 3.8 present a sensitivity analysis regarding the number of factors in factor-augmented versions of the MIDAS-AR, MFVAR and QVAR models. The factors are derived by applying simple principal component analysis in case of the F-VAR model. For the mixed-frequency models the Kalman-filtered factors generated by the dynamic factor model were used, averaged over all possible parameterizations. The maximum number of factors is set at 6. Table 3.7 treats each factor as a separate indicator and reports the RMSFE of a weighted average of factor-specific forecasts relative to the RMSFE of the one-factor versions reported in Table 3.2. For all countries except Spain the incorporation of additional factors tend to push up the RMSFE for all horizons. For Spain, using two extra factors improves forecasting accuracy for the F-MIDAS-AR model by 20%. For the F-MFVAR and F-VAR models the effect is fairly small. Table 3.8 looks into the marginal value of the factors when they are sequentially added to the encompassing test regression Eq. (3.19). This analysis leads to the same conclusion.

Table 3.9 presents a sensitivity analysis regarding the three weighting schemes for indicator-specific indicators used in the BEQ-AR, QVAR, MIDAS-AR and MFVAR models. In addition, the table looks into the weighting scheme for the different varieties of the dynamic factor model in terms of the number of static factors, dynamic factors and lags (126 parametrizations in total). All weighting schemes produce similar results.

Finally, Tables 3.10–3.13 are the counterparts of Tables 3.2 and 3.3 in the main text, focusing on the forecasting accuracy and marginal value of models during the Great Moderation (1996.I–2007.IV) and the financial crisis (2008.I–2011.III).

Table 3.7: Relative RMSFE multiple single-factor models versus one-factor model, 1996.I–2011.III

	F-MIDAS-AR					F-MFVAR					F-VAR				
♯ factors	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
Euro area															
2Q-ahead	1.03	1.04	1.04	1.05	1.05	0.97	0.97	0.97	0.97	0.97	1.00	0.99	1.00	1.00	1.00
1Q-ahead	1.04	1.07	1.07	1.09	1.11	0.97	0.97	0.97	0.97	0.97	1.01	1.01	1.02	1.02	1.02
Nowcast	1.04	1.06	1.05	1.07	1.08	1.00	0.99	0.99	0.98	0.99	1.02	1.03	1.04	1.05	1.05
Backcast	1.02	1.02	1.02	1.03	1.05	0.98	0.97	0.97	0.97	0.96	1.01	1.02	1.03	1.04	1.04
Germany															
2Q-ahead	1.02	1.03	1.04	1.04	1.04	0.99	0.98	0.98	0.97	0.97	0.99	0.99	0.99	0.99	0.99
1Q-ahead	1.01	1.02	1.02	1.03	1.04	0.99	0.98	0.98	0.98	0.98	1.00	1.01	1.01	1.01	1.01
Nowcast	1.02	1.05	1.06	1.08	1.10	0.98	0.98	0.98	0.98	0.99	1.02	1.04	1.04	1.04	1.04
Backcast	0.98	0.99	1.00	1.01	1.03	0.97	0.98	0.97	0.97	0.97	1.04	1.06	1.06	1.07	1.07
France															
2Q-ahead	1.00	1.01	1.01	1.02	1.03	0.99	1.00	1.00	1.00	0.99	0.99	0.98	0.99	1.00	1.00
1Q-ahead	1.03	1.05	1.06	1.07	1.08	0.98	0.99	1.00	0.99	0.99	1.00	0.98	1.00	1.00	1.00
Nowcast	1.03	1.07	1.07	1.10	1.12	0.98	0.97	0.97	0.97	0.97	1.01	0.99	1.00	1.01	1.00
Backcast	1.00	1.00	0.99	0.99	1.01	0.98	0.96	0.96	0.96	0.96	1.02	1.01	1.03	1.04	1.04
Italy															
2Q-ahead	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.01	1.01	1.01
1Q-ahead	1.02	1.05	1.05	1.06	1.07	0.99	0.99	0.99	0.99	0.98	1.00	1.01	1.01	1.01	1.02
Nowcast	1.04	1.06	1.07	1.09	1.12	0.99	0.98	0.98	0.98	0.97	1.01	1.02	1.03	1.04	1.04
Backcast	1.00	1.00	1.00	1.00	1.01	1.01	0.99	0.99	0.98	0.98	1.03	1.05	1.06	1.07	1.07
Spain															
2Q-ahead	0.93	0.91	0.91	0.92	0.92	0.98	0.97	0.97	0.97	0.97	1.00	1.02	1.04	1.04	1.04
1Q-ahead	0.92	0.87	0.87	0.86	0.84	1.01	0.98	0.99	1.00	1.00	0.99	1.04	1.06	1.08	1.08
Nowcast	0.87	0.80	0.80	0.78	0.78	1.02	0.99	1.00	1.00	1.00	0.96	1.02	1.07	1.09	1.10
Backcast	0.89	0.80	0.81	0.80	0.79	1.02	1.00	1.01	1.01	1.01	0.97	1.02	1.07	1.10	1.12
Netherlands															
2Q-ahead	1.01	1.02	1.03	1.02	1.02	0.97	0.97	0.97	0.97	0.97	1.01	1.01	1.02	1.02	1.02
1Q-ahead	1.00	1.01	1.03	1.04	1.05	0.98	0.99	0.99	0.99	0.99	1.01	1.02	1.03	1.03	1.04
Nowcast	0.99	0.99	1.01	1.02	1.03	1.00	1.00	1.01	1.00	1.00	1.01	1.03	1.04	1.05	1.05
Backcast	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.02	1.04	1.06	1.06	1.07

Notes: entries denote RMSFE(multiple single-factor models)/RMSFE(one-factor model). Weighting of factor-specific forecasts by four quarter moving window RMSFE scheme. F-VAR: factor-augmented quarterly VAR model, F-MFVAR: factor-augmented mixed-frequency VAR model, F-MIDAS-AR: factor-augmented mixed-data sampling model with AR term.

Table 3.8: Marginal value of additional factors, 1996.I–2011.III

	F-MIDAS-AR					F-MFVAR					F-VAR				
♯ factors	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
Euro Area															
2Q-ahead	.	.	.	.	.	**0.95**	.	.	.	.	.	0.99	.	.	.
1Q-ahead	.	.	.	.	.	**0.95**	.	.	.	.	.	.	.	.	.
Nowcast	.	.	.	.	.	0.99	**0.99**	.	.	.	.	.	.	.	.
Backcast	.	0.99	.	.	.	**0.97**	.	.	.	**0.98**	.	.	.	.	.
Germany															
2Q-ahead	.	.	.	.	.	**0.98**	.	**0.99**	.	.	0.99	.	.	.	.
1Q-ahead	.	.	.	.	.	**0.97**	.	.	.	.	.	.	.	.	.
Nowcast	.	.	.	.	.	**0.96**	.	.	.	.	.	.	.	.	.
Backcast	**0.98**	.	.	.	.	**0.96**	.	.	0.99	.	.	.	.	.	.
France															
2Q-ahead	.	.	.	.	.	**0.98**	.	.	.	0.99	0.99	**0.99**	.	.	.
1Q-ahead	.	.	.	.	.	**0.97**	.	.	.	.	.	**0.97**	.	.	.
Nowcast	.	.	.	.	.	**0.98**	**0.99**	.	.	.	.	**0.97**	.	.	.
Backcast	0.99	.	**0.98**	.	.	**0.97**	**0.98**	.	.	.	.	0.99	.	.	.
Italy															
2Q-ahead	.	.	.	.	.	.	.	0.99	.	0.99	.	.	.	.	.
1Q-ahead	.	.	.	.	.	**0.98**	.	.	.	0.99	.	.	.	.	.
Nowcast	.	.	.	.	.	**0.99**	**0.99**	.	0.99	.	.	.	.	.	.
Backcast	.	0.99	.	.	.	.	**0.96**	.	.	.	.	.	.	.	.
Spain															
2Q-ahead	**0.93**	**0.98**	0.99	.	.	**0.98**	**0.97**	.	.	.	0.99	.	.	.	.
1Q-ahead	**0.92**	**0.92**	.	.	**0.97**	.	**0.95**	.	.	.	**0.99**	.	.	.	.
Nowcast	**0.83**	**0.88**	.	.	.	.	**0.95**	.	.	.	**0.97**	.	.	.	.
Backcast	**0.88**	**0.81**	.	.	.	.	**0.98**	.	.	.	**0.97**	.	.	.	.
Netherlands															
2Q-ahead	.	.	.	.	.	**0.95**	.	.	.	.	.	.	.	.	.
1Q-ahead	.	.	.	.	.	**0.98**	.	.	.	.	.	.	.	.	.
Nowcast	0.99	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Backcast	.	.	0.99	.	.	**0.98**	.	.	.	.	.	.	.	.	.

Notes: entries denote RMSFE(forecast model)/RMSFE(forecast model with one factor). Entries in bold denote that the estimated weight in the encompassing test is statistically different from zero at the 5% significance level. Dots indicate the corner solution of zero weight (λ=0). F-MIDAS-AR: factor-augmented mixed-data sampling model with AR term, F-MFVAR: factor-augmented mixed-frequency VAR model, F-VAR: factor-augmented quarterly VAR model.

Table 3.9: Forecasting performance of pooling schemes, 1996.I–2011.III

	BEQ-AR			QVAR			DFM			MIDAS-AR			MFVAR		
	AV	RC	4Q	AV	RC	4Q	AV	RC	4Q	AV	RC	4Q	AV	RC	4Q
Euro Area															
2Q-ahead	0.62	0.62	0.62	0.65	0.65	0.64	0.63	0.63	0.63	0.62	0.62	0.62	0.62	0.62	0.62
1Q-ahead	0.58	0.59	0.57	0.63	0.64	0.63	0.57	0.57	0.57	0.59	0.59	0.59	0.60	0.59	0.59
Nowcast	0.51	0.52	0.50	0.58	0.58	0.58	0.44	0.44	0.43	0.51	0.53	0.50	0.53	0.50	0.52
Backcast	0.44	0.45	0.43	0.50	0.50	0.49	0.32	0.32	0.31	0.45	0.45	0.39	0.46	0.47	0.44
Germany															
2Q-ahead	0.90	0.90	0.90	0.92	0.93	0.92	0.90	0.90	0.90	0.93	0.93	0.93	0.91	0.90	0.91
1Q-ahead	0.88	0.86	0.88	0.91	0.92	0.91	0.84	0.84	0.84	0.89	0.91	0.89	0.89	0.88	0.89
Nowcast	0.84	0.82	0.82	0.89	0.90	0.90	0.70	0.70	0.70	0.85	0.86	0.84	0.85	0.81	0.82
Backcast	0.81	0.77	0.78	0.87	0.87	0.88	0.61	0.61	0.60	0.82	0.78	0.75	0.81	0.74	0.72
France															
2Q-ahead	0.51	0.51	0.51	0.53	0.53	0.54	0.50	0.50	0.50	0.51	0.51	0.51	0.52	0.51	0.51
1Q-ahead	0.48	0.46	0.48	0.51	0.51	0.53	0.44	0.44	0.44	0.49	0.49	0.48	0.48	0.48	0.48
Nowcast	0.42	0.41	0.41	0.46	0.46	0.46	0.34	0.34	0.34	0.43	0.42	0.42	0.43	0.41	0.42
Backcast	0.39	0.39	0.38	0.42	0.42	0.41	0.28	0.28	0.28	0.39	0.38	0.37	0.38	0.37	0.38
Italy															
2Q-ahead	0.73	0.73	0.74	0.75	0.75	0.76	0.74	0.74	0.74	0.74	0.73	0.73	0.74	0.74	0.74
1Q-ahead	0.71	0.70	0.71	0.73	0.73	0.74	0.68	0.68	0.68	0.72	0.72	0.71	0.71	0.71	0.71
Nowcast	0.65	0.65	0.64	0.70	0.68	0.70	0.55	0.55	0.55	0.65	0.65	0.65	0.65	0.65	0.65
Backcast	0.60	0.59	0.59	0.63	0.62	0.63	0.47	0.47	0.47	0.60	0.58	0.59	0.58	0.58	0.58
Spain															
2Q-ahead	0.59	0.57	0.58	0.59	0.58	0.59	0.56	0.56	0.56	0.59	0.59	0.57	0.61	0.59	0.59
1Q-ahead	0.56	0.54	0.53	0.55	0.54	0.54	0.47	0.47	0.46	0.51	0.49	0.49	0.58	0.56	0.56
Nowcast	0.54	0.54	0.49	0.49	0.46	0.47	0.40	0.39	0.39	0.47	0.50	0.42	0.56	0.55	0.52
Backcast	0.56	0.65	0.49	0.48	0.45	0.45	0.36	0.35	0.35	0.55	0.57	0.47	0.57	0.56	0.52
Netherlands															
2Q-ahead	0.69	0.69	0.69	0.71	0.71	0.71	0.71	0.71	0.71	0.69	0.70	0.69	0.71	0.71	0.70
1Q-ahead	0.66	0.65	0.64	0.70	0.70	0.70	0.64	0.64	0.64	0.66	0.66	0.65	0.68	0.68	0.69
Nowcast	0.61	0.60	0.58	0.66	0.66	0.66	0.54	0.54	0.53	0.62	0.60	0.60	0.64	0.63	0.62
Backcast	0.57	0.55	0.55	0.59	0.59	0.59	0.48	0.48	0.47	0.57	0.55	0.54	0.62	0.61	0.60

Notes: entries denote RMSFEs. Grey cells denote weighting scheme with lowest RMSFE. AV: equal weights (simple average), RC: weights inversely proportional to RMSFE, recursively calculated, 4Q: weights inversely proportional to RMSFE, calculated over the last four quarters.BEQ-AR: bridge equation with AR term, QVAR: quarterly VAR model, DFM: dynamic factor model, MIDAS-AR: mixed-data sampling model with AR term, MFVAR: mixed-frequency VAR model.

Table 3.10: Forecasting performance of statistical models, 1996.I–2007.IV

Frequency	Benchmark			Pooling forecasts				Pooling information			
Model	RW	AR	BVAR	BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area											
2Q-ahead	*0.34*	**1.00**	1.08	**0.96**	**0.99**	0.96	**0.98**	**0.99**	1.02	1.00	1.01
1Q-ahead	*0.34*	**0.99**	1.03	**0.92**	**0.97**	**0.91**	**0.94**	**0.92**	**1.00**	**0.91**	0.91
nowcast	*0.34*	0.96	0.93	0.87	0.93	**0.85**	0.89	**0.82**	0.93	0.86	0.78
backcast	*0.34*	0.91	0.83	0.86	0.88	0.82	0.84	0.72	0.85	0.79	0.88
Germany											
2Q-ahead	*0.62*	1.04	1.01	0.97	**0.97**	**0.98**	**0.99**	**0.99**	1.00	1.07	1.01
1Q-ahead	*0.62*	1.05	**0.97**	**0.98**	0.97	**0.99**	**0.98**	**0.97**	**1.00**	1.04	**0.99**
nowcast	*0.62*	1.05	**0.92**	**0.94**	**0.99**	**0.95**	0.91	**0.93**	**0.97**	**1.00**	**0.97**
backcast	*0.62*	1.06	0.90	**0.91**	**0.99**	**0.93**	**0.91**	**0.92**	**0.96**	**0.97**	1.06
France											
2Q-ahead	*0.35*	1.00	1.05	**0.95**	**0.98**	**0.95**	**0.96**	0.90	1.01	**0.99**	**0.96**
1Q-ahead	*0.35*	0.98	1.00	**0.90**	0.95	0.92	0.92	0.82	0.99	**0.89**	0.91
nowcast	*0.35*	0.94	0.92	0.83	0.89	0.86	0.85	0.70	0.91	**0.73**	**0.72**
backcast	*0.35*	0.93	0.85	0.86	0.88	0.84	0.84	0.65	0.89	0.76	0.78
Italy											
2Q-ahead	*0.50*	1.03	1.00	**0.99**	0.99	**0.99**	**0.99**	1.01	1.00	1.10	1.01
1Q-ahead	*0.50*	1.02	**1.00**	**0.99**	**0.98**	**0.98**	**0.97**	**0.93**	1.00	**0.98**	0.91
nowcast	*0.50*	1.00	0.96	0.91	0.96	0.92	0.92	**0.85**	0.95	**0.86**	0.80
backcast	*0.50*	0.97	0.91	0.88	0.90	0.89	0.88	**0.82**	**0.84**	**0.86**	0.79
Spain											
2Q-ahead	*0.32*	**0.96**	0.93	**0.95**	1.02	**0.99**	**0.96**	**0.96**	1.00	**0.98**	1.02
1Q-ahead	*0.31*	**0.91**	0.88	**0.95**	**0.93**	**0.92**	0.97	**0.93**	0.97	1.10	1.03
nowcast	*0.31*	**0.87**	0.82	1.00	0.95	0.92	0.99	0.96	1.02	1.02	1.06
backcast	*0.31*	0.91	0.78	1.12	1.02	1.05	1.24	1.00	1.16	1.05	1.19
Netherlands											
2Q-ahead	*0.53*	1.00	**0.93**	**0.97**	**0.99**	**0.95**	**0.98**	**0.97**	**0.96**	0.91	**0.94**
1Q-ahead	*0.53*	**1.00**	**0.89**	**0.92**	0.98	**0.91**	**0.97**	**0.94**	**0.93**	0.89	**0.90**
nowcast	*0.53*	0.99	0.85	**0.92**	0.97	**0.92**	0.94	**0.90**	**0.90**	**0.88**	0.94
backcast	*0.52*	0.97	0.82	0.91	0.95	**0.88**	0.96	**0.87**	**0.87**	**0.88**	1.00

Notes: entries denote the RMSFE for a Random Walk (in italics); for all other models they denote the RMSFE relative to the RMSFE of a Random Walk. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model. RW: random walk, AR: autoregressive model, BEQ-AR: bridge equation with AR term, QVAR: quarterly vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, DFM: dynamic factor model, MFVAR: mixed-frequency vector autoregressive model, F-MFVAR: factor-augmented MFVAR model, MIDAS-AR: mixed-data sampling model with AR term, F-MIDAS-AR: factor-augmented MIDAS-AR.

Table 3.11: Forecasting performance of statistical models, 2008.I–2011.III

Frequency	Benchmark			Pooling forecasts				Pooling information			
Model	RW	AR	BVAR	BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area											
2Q-ahead	*1.15*	1.00	0.94	**0.97**	1.02	**0.97**	**0.97**	**0.99**	**0.99**	1.02	1.01
1Q-ahead	*1.15*	1.00	0.90	0.90	1.01	0.93	0.93	0.89	0.95	0.80	**0.85**
nowcast	*1.14*	0.95	0.80	0.77	0.92	0.78	0.80	**0.65**	0.87	0.61	**0.63**
backcast	*1.14*	0.83	0.68	0.63	0.76	0.55	0.65	0.43	0.78	0.57	0.58
Germany											
2Q-ahead	*1.50*	1.05	1.03	**0.99**	1.02	1.04	**0.99**	**0.99**	1.01	0.97	1.02
1Q-ahead	*1.50*	1.03	1.01	**0.96**	1.01	0.96	0.97	**0.89**	0.97	**0.89**	0.87
nowcast	*1.50*	1.01	0.95	0.89	0.99	0.90	0.90	**0.66**	0.92	**0.67**	0.63
backcast	*1.49*	1.01	0.86	0.82	0.96	0.76	0.72	0.48	0.88	0.58	0.61
France											
2Q-ahead	*0.88*	1.03	0.87	0.97	1.03	0.97	0.97	0.96	0.99	1.00	0.98
1Q-ahead	*0.88*	0.99	**0.84**	0.91	1.03	0.91	0.91	**0.84**	0.97	**0.82**	0.81
nowcast	*0.87*	0.88	0.76	0.75	0.88	0.75	0.77	0.61	0.88	0.55	**0.58**
backcast	*0.87*	0.79	0.67	0.65	0.75	0.64	0.65	0.45	0.75	0.52	0.54
Italy											
2Q-ahead	*1.24*	1.04	0.92	**0.98**	1.02	**0.98**	**0.99**	**0.98**	**0.99**	1.05	**0.98**
1Q-ahead	*1.24*	0.95	**0.90**	**0.94**	0.99	**0.94**	**0.94**	**0.91**	0.95	0.86	**0.87**
nowcast	*1.23*	0.92	0.85	0.84	0.94	0.84	0.84	**0.68**	0.88	0.63	**0.67**
backcast	*1.23*	0.89	0.76	0.75	0.82	0.74	0.72	0.52	0.78	0.60	0.60
Spain											
2Q-ahead	*1.17*	**0.91**	**0.85**	**0.91**	**0.90**	**0.88**	0.92	**0.86**	**0.87**	0.83	**0.91**
1Q-ahead	*1.16*	0.81	**0.74**	0.82	0.83	**0.75**	0.87	0.70	0.77	0.78	0.79
nowcast	*1.15*	0.71	**0.57**	0.72	0.70	0.61	0.79	0.53	0.67	0.62	0.75
backcast	*1.15*	0.76	0.40	0.68	0.63	0.67	0.71	**0.41**	0.62	0.52	0.75
Netherlands											
2Q-ahead	*1.12*	**0.99**	0.90	**0.96**	**0.99**	**0.98**	**0.98**	1.01	**0.97**	**0.97**	1.08
1Q-ahead	*1.11*	0.99	0.89	0.88	0.98	0.92	0.96	0.88	0.94	0.78	0.88
nowcast	*1.11*	0.94	0.82	0.73	0.90	0.80	0.83	**0.64**	0.88	0.61	**0.65**
backcast	*1.10*	0.85	0.70	0.66	0.75	0.67	0.75	0.51	0.76	0.58	0.60

Notes: entries denote the RMSFE for a Random Walk (in italics); for all other models they denote the RMSFE relative to the RMSFE of a Random Walk. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model. RW: random walk, AR: autoregressive model, BEQ-AR: bridge equation with AR term, QVAR: quarterly vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, DFM: dynamic factor model, MFVAR: mixed-frequency vector autoregressive model, F-MFVAR: factor-augmented MFVAR model, MIDAS-AR: mixed-data sampling model with AR term, F-MIDAS-AR: factor-augmented MIDAS-AR.

Table 3.12: Marginal value of statistical models, 1996.I–2007.IV

Frequency			Pooling forecasts				Aggregating information			
Model	AR	BVAR	BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area										
2Q-ahead	.	.	.	.		.	.	.	.	.
1Q-ahead	**0.98**	**0.99**	**0.97**	**0.99**	**0.97**	**0.98**	0.99	**0.99**	0.99	
nowcast	.	**0.98**	.	.	0.99	.	0.99	.	.	
backcast	.	0.99	.	.	.	.		.	0.99	.
Germany										
2Q-ahead	.	.		.	**0.99**	.	.	.	.	.
1Q-ahead	.	**0.98**	.		.	.	**0.98**	.	.	**0.98**
nowcast	.	**0.97**	.	.	.		**0.99**	.	0.99	0.99
backcast	.		**0.95**	0.99	**0.97**	**0.95**	**0.96**	**0.98**	**0.97**	**0.98**
France										
2Q-ahead	.	.	.	.	.	.		.	.	.
1Q-ahead	.	.	.	.	.	.		.	.	.
nowcast	.	.	.	.	.	.		.	**0.98**	**0.98**
backcast	.	.	.	.	.	.		.	.	.
Italy										
2Q-ahead	.	.	.		.	.	.	.	.	.
1Q-ahead	.	.	.	.	.	.	.	.	.	
nowcast	.	.	.	.	.	.	.	.	.	
backcast	.	.	.	.	.	.	0.99	.	.	
Spain										
2Q-ahead	.		**0.99**	.	.	.	**0.98**	.	**0.99**	**0.98**
1Q-ahead	.		0.99	.	0.99	.	**0.97**	.	.	**0.98**
nowcast	**0.99**		.	.	.	.	**0.97**	.	**0.98**	**0.98**
backcast	.		.	.	.	.	**0.98**	.	**0.98**	0.99
Netherlands										
2Q-ahead	.	**0.98**	.	.	.	.	.	.		.
1Q-ahead	.	**0.97**	0.99	.	0.99	.	.	.		.
nowcast	.		**0.99**	.	0.99	**0.99**	**0.98**	**0.99**	**0.98**	**0.98**
backcast	.		0.99	.	0.99	0.99	**0.98**	**0.98**	0.99	0.99

Notes: entries denote RMSFE(forecast combination)/RMSFE(best model). Entries in bold denote that the estimated weight in the encompassing test is statistically different from zero at the 5% significance level. Dots indicate the corner solution of zero weight (λ=0). Grey cells denote models with the lowest RMSFE. AR: autoregressive model, BEQ-AR: bridge equation with AR term, QVAR: quarterly vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, DFM: dynamic factor model, MFVAR: mixed-frequency vector autoregressive model, F-MFVAR: factor-augmented MFVAR model, MIDAS-AR: mixed-data sampling model with AR term, F-MIDAS-AR: factor-augmented MIDAS-AR.

Table 3.13: Marginal value of statistical models, 2008.I–2011.III

Frequency			Pooling forecasts				Pooling information			
Model	AR	BVAR	BEQ-AR	QVAR	MIDAS-AR	MFVAR	DFM	F-VAR	F-MIDAS-AR	F-MFVAR
Euro area										
2Q-ahead	.		.	.	.	.	0.99	.	.	0.99
1Q-ahead	.	.	.	.	.	.	.	.		.
nowcast	.	0.98	.	.	.	.	.	.		.
backcast	.	.	.	.	.	.		.	.	.
Germany										
2Q-ahead	.	.	.	.	.	.	.	.		.
1Q-ahead	.	.	.	.	.	.	.	0.99	.	
nowcast	.	.	.	.	.	.	.	.	.	
backcast	.	.	.	.	.	.		.	.	.
France										
2Q-ahead	.		.	.	.	.	.	.	.	.
1Q-ahead	0.98	**0.95**	0.99	0.99	0.98	.	.	0.99	**0.96**	
nowcast	.	0.99	.	.	.	.	.	.		.
backcast	.	0.98	.	.	.	.		.	.	.
Italy										
2Q-ahead	.		.	.	.	.	.	.	.	.
1Q-ahead	.	.	.	.	.	.	.	.		0.98
nowcast	.	.	.	.	.	.	.	.		.
backcast	.	.	.	.	.	.		.	.	.
Spain										
2Q-ahead	0.98	.	.	.	0.99	.	.	0.99		.
1Q-ahead	**0.95**	0.98	.	.	**0.97**	.		0.98	.	.
nowcast	**0.90**	**0.92**	.	.	**0.97**	.		**0.97**	.	.
backcast	.		.	.	.	.	**0.81**	.	**0.95**	0.99
Netherlands										
2Q-ahead	.		.	.	.	.	.	.	.	.
1Q-ahead	.	.	.	.	.	.	.	.		.
nowcast	.	.	0.99	0.99	.	.	0.99	.		0.98
backcast	.	.	.	.	.	.		.	.	.

Notes: entries denote RMSFE(forecast combination)/RMSFE(best model). Entries in bold denote that the estimated weight in the encompassing test is statistically different from zero at the 5% significance level. Dots indicate the corner solution of zero weight (λ=0). Grey cells denote models with the lowest RMSFE. AR: autoregressive model, BEQ-AR: bridge equation with AR term, QVAR: quarterly vector autoregressive model, BVAR: Bayesian QVAR model, F-VAR: factor-augmented QVAR model, DFM: dynamic factor model, MFVAR: mixed-frequency vector autoregressive model, F-MFVAR: factor-augmented MFVAR model, MIDAS-AR: mixed-data sampling model with AR term, F-MIDAS-AR: factor-augmented MIDAS-AR.

Chapter 4

Improving model-based near-term GDP forecasts by subjective forecasts: a real-time exercise for the G7 countries

This chapter investigates to what extent it is feasible to improve model-based near-term GDP forecasts by combining them with judgmental (quarterly) forecasts by professional analysts (Consensus forecasts) in a real-time setting. The analysis covers the G7 countries over the years 1999–2013. The weighted average and the linear combination are considered. Incorporating subjective information delivers sizable gains in forecasting ability of statistical models for all countries except Japan in 1999–2013, even when subjective forecasts are somewhat dated. Accuracy gains are much more pronounced in the volatile period after 2008 due to a marked improvement in predictive power of Consensus forecasts. Since 2008, Consensus forecasts are superior at the moment of publication for most countries. For some countries Consensus forecasts can be enhanced by model-based forecasts in between the quarterly release dates of the Consensus survey, as the latter embody more recent monthly information.[1]

KEYWORDS: Forecast combination; Encompassing test; Nowcasting; Factor models; Judgment.

[1] This chapter is co-authored by Jos Jansen. This chapter is under review at the Oxford Bulletin of Economics and Statistics. Comments and suggestion by Marta Bańbura, Peter van Els, Jakob de Haan, Job Swank, seminar and conference participants at De Nederlandsche Bank, the European Central Bank and the International Symposium on Forecasting (2015, Riverside) are gratefully acknowledged. An early version of this chapter was circulated as DNB Working Paper 507 under the title "Improving model-based near-term GDP forecasts by subjective forecasts: A real-time exercise for the G7 countries".

4.1 Introduction

Policy makers and economic agents have to make decisions in real-time on the basis of incomplete and inaccurate information on current economic conditions. For example, data on real gross domestic product (GDP), which is the broadest measure of aggregate economic activity, are released on a quarterly basis with a substantial time lag (six weeks in many advanced countries); they are also subject to revisions. However, a wealth of statistical information that is directly and indirectly related to economic activity is nowadays available from public and private sources. Policy makers, firms and financial market participants may exploit this vast body of statistical information to form expectations on the current state of the economy and its near-term development. This requires solving the practical problem of handling a large-scale information set of potentially hundreds of time series that are observed at different frequencies and with different publication lags (the so-called ragged edge problem). The recent nowcasting literature has developed several statistical methodologies for generating near-term GDP forecasts based on large mixed-frequency datasets with ragged edges. Examples are bridge models, factor models, mixed-data sampling regression models (MIDAS), mixed-frequency vector autoregressive (MFVAR) models and Bayesian VARs.[2]

Apart from model-based predictions, policy makers and economic agents may also take advantage of published forecasts made by professional analysts. From a practical point of view, such forecasts are cheap and easy to use. Currently, several surveys on the economic outlook are available on a regular basis. The Federal Reserve Bank of Philadelphia and the European Central Bank (ECB) both maintain a regular Survey of Professional Forecasters. Moreover, the survey firm Consensus Economics publishes a well-known compilation of macroeconomic forecasts by professional forecasters for many countries. Model-based forecasts are the result of purely mechanical recipes using statistical data and do not incorporate subjective elements. By contrast, forecasts by professional analysts reflect much more information than statistical data, which are inevitably limited. For example, Meyler and Rubene (2009) report that the participants of the ECB Survey of Professional Forecasters consider forty percent of their short-term GDP forecasts to be judgment-based. Based on in-sample encompassing tests in a pseudo real-time set-up, Chapter 3 of this thesis finds that subjective predictions by private sector analysts often embody valuable information that sophisticated mechanical forecasting procedures fail to pick up. Liebermann (2014) presents a similar result for the US in the period 2000–2010 in a real-time setting.

This empirical evidence suggests that publicly available subjective forecasts offer the potential of enhancing real-time model-based GDP forecasts, and thus a better

[2] See among others Baffigi et al. (2004), Stock and Watson (2011), Kuzin et al. (2011), Ghysels et al. (2007), Foroni and Marcellino (2014), Bańbura et al. (2010), Carriero et al. (2015b) and Chapter 3 of this thesis.

assessment of the current state of the economy. The main purpose of this chapter is to investigate whether predictions by analysts are actually able to improve GDP forecasts generated by purely statistical procedures in real-time. The reverse question, whether model-based forecasts can be used to enhance subjective forecasts, is also addressed. The proposed procedure takes into account the information availability constraints facing practitioners when running forecasting models and forming expectations. Predictions produced by a dynamic factor model are used as a benchmark. The quarterly forecasts published by Consensus Economics are employed as the measure of judgmental forecasts. The analysis covers the G7 countries in the period 1999–2013, contrasting the experience in the volatile post-crisis period of 2008–2013 with that in the more tranquil period of 1999–2007.

The remainder of this chapter is structured as follows. Section 4.2 describes the real-time dataset and the Consensus forecasts, the benchmark forecasting model and the way predictions are generated in real-time. Section 4.3 presents the empirical results and Section 4.4 concludes.

4.2 Dataset, benchmark model and forecast design

4.2.1 Dataset

Data on real GDP and monthly indicators are released with different publication lags and are possibly subject to revisions at a later stage, which at least for GDP may be sizable. Many papers on nowcasting employ a pseudo real-time design, which takes publication delays into account and applies recursive estimation, but disregards data revisions of GDP and monthly indicators, such as industrial production. Such an approach is unsuitable for two reasons. First, the aim is to investigate whether it is in practice, hence in real-time, possible to enhance mechanical forecasts by judgmental forecasts. Second, as noted by Croushore (2011), data revisions could affect the results of forecast evaluations and comparisons of different (statistical) approaches. This criticism is even more relevant for the analysis in this chapter, which involves comparing model-based forecasts and forecasts by professional analysts. The expectations of analysts at a certain point in time necessarily reflect the then available information, including inaccurate initial estimates of GDP's recent past and that of key monthly indicators. This puts them at a disadvantage vis-à-vis statistical approaches in a pseudo real-time setting, where the latter can take data revisions on board for model estimation and projections. Therefore, in this chapter, real-time datasets were compiled. This means that the model-based forecasts used in the forecast evaluation exercise only incorporate information that was available at the time of forecasting.[3]

[3] Studies using real-time data include Schumacher and Breitung (2008), Camacho and Perez-Quiros (2010), Lahiri and Monokroussos (2013) and Liebermann (2014).

The real-time monthly datasets that have been compiled consist of similar variables across countries and cover the broad range of information that is readily available to economic agents. The variables were selected in the spirit of Bańbura et al. (2013), who focus on "headline" macroeconomic variables that financial market participants and the media primarily pay attention to. Bańbura et al. (2011) and Bańbura and Modugno (2014) provide evidence that the marginal impact of disaggregated data on forecasting accuracy is very small. Accordingly, the selected indicators refer to the aggregate level. Possibly available disaggregated information by sector, subcategory, region, etc. is not included. The dataset for a country consists of three parts. The first part concerns information about the domestic economy, the second part refers to variables related to global economic activity, and the third part contains key data on the two most important trading partners of the respective countries in the dataset.[4]

The indicator variables that refer to the domestic economy fall into four categories. The first category is hard, quantitative information on production and expenditures, such as industrial production, car sales, retail sales, exports, imports and unemployment. The second category refers to consumer and producer prices. The third category contains financial variables, both quantities (monetary aggregates) and prices (interest rates, stock prices and exchange rates). The latter category determines the financing conditions for firms and consumers. Moreover, financial market prices partly reflect financial market expectations on output developments in the near future. The fourth category is soft, qualitative information on expectations derived from surveys among consumers, retailers and firms. Also included is the composite leading indicator compiled by the OECD. Following Golinelli and Parigi (2014), the global variables, which are common to all countries, include oil and commodity prices, semiconductor sales, the Baltic freight index (BFI), the standard and poors exchange volatility index (VIX) and world trade. The key data on a country's two most important trading partners comprise three variables: imports, industrial production and the composite leading indicator compiled by the OECD. Finally, for all European countries four closely-watched confidence indices from Germany (Ifo), France (INSEE), Italy (ISEA) and Belgium (BNB) were also included.

The statistical monthly information set reflects real-time public knowledge in the middle of the second week of the month. The number of monthly indicators varies from around 30 for Japan and Canada to around 36 for the other countries. All monthly indicator series start in January 1985, while the quarterly GDP series start in 1985.I. Table 4.7 in Appendix 4.A provides details on the exact composition of each country's

[4] The two most important trading partners were determined on the basis of the OECD trade in value-added database, which focuses on the value added contribution of (bilateral) exports. The most important trading partners are the US and UK for Canada, the US and France for Germany, the US and Germany for France, France and Germany for Italy, the US and China for Japan, the US and Germany for the UK, and Canada and Mexico for the US.

statistical dataset and the data sources. Monthly data are seasonally (and calendar effects) adjusted at the source, except for prices and financial variables. All monthly series are made stationary by taking three-month differences, log-differences (in case of trending data, such as industrial production) or double log-differences (in case of prices). Finally, each variable is standardized by subtracting the mean and dividing by the standard deviation. This normalization is standard practice in order to avoid the overweighting of series with large-variances series in the extraction of common factors.

The subjective quarterly GDP forecasts by professional analysts were collected from paper copies of the monthly publication *Consensus forecasts*, published by the private sector firm Consensus Economics. Consensus forecasts offers a survey of private sector analysts' expectations for a set of key macroeconomic variables for a broad range of countries. Consensus forecasts is best known for its expectations on annual GDP growth for the current and next year, which have been analyzed in several papers (e.g. Ager et al., 2009; Batchelor, 2001; Loungani and Rodriguez, 2008; Lahiri et al., 2006). Once a quarter, this publication also provides averaged forecasts for quarterly GDP over a horizon of six quarters, starting with the nearest quarter for which no officially released figure is available. The number of respondents varies somewhat over time and across countries.[5] Fresh quarterly Consensus forecasts become available in the second week of the last month of the quarter. The survey date (deadline for respondents) is typically the second Monday of the third month of a quarter; publication is usually 3 days later on Thursday. The timing of the survey is therefore in line with the timing of the monthly data vintages that were collected. For the information set this means that Consensus forecasts are not updated in the first and second months in a quarter, while monthly indicators are updated every month. Moreover, at the time analysts form their expectations they have official information on GDP growth in the preceding quarter. Quarterly Consensus forecasts for the G7 countries are available from the early 1990s onwards.

4.2.2 Benchmark model: dynamic factor model

The model used to generate model-based forecasts is a dynamic factor model (DFM). Dynamic factor models summarize the information of the dataset in a limited number of factors, whose dynamic behavior is specified as a vector-autoregressive process. A key feature of this approach is the use of the Kalman filter, which allows for an efficient

[5] The average number of participants is about 15 for Canada and Italy, about 20 for France and Japan and about 28 for Germany, the UK and the US. Consensus forecasts publishes the simple average of the forecasts by all respondents, but no individual forecasts. This is not a serious limitation, as Genre et al. (2013) show that the potential gains of combining expert forecasts with alternative schemes are very limited. Moreover, a potential advantage is that analysts may be more likely to submit their true expectations, as this procedure guarantees their anonymity, reducing motives for possible strategic behaviour (Lamont, 2002; Laster et al., 1999).

handling of the unbalancedness of the dataset and the different frequencies of the data. The Kalman filter replaces any missing monthly indicator observations with optimal predictions and also generates estimates of unobserved monthly real GDP subject to a temporal aggregation constraint for the quarterly observation. One of the main conclusions of the comparative study in Chapter 3 of this thesis was that the DFM is the best statistical procedure overall, in particular for nowcasting and backcasting. Thus, it is used as the benchmark statistical model in this chapter. DFMs have been applied to many countries, generally delivering relatively accurate macroeconomic forecasts.[6]

In this chapter the DFM proposed by Bańbura and Rünstler (2011) is analyzed. The model is used by several central banks within the euro area. The first equation of the model is:

$$x_m = \Lambda f_m + \xi_m, \qquad \xi_m \sim N(0, \Sigma_\xi) \tag{4.1}$$

which relates the n monthly indicators $x_m = (x_{1,m}, \ldots, x_{n,m})'$ to r monthly static factors $f_m = (f_{1,m}, \ldots, f_{r,m})'$ via an $n \times r$ matrix of factor loadings Λ and an idiosyncratic component $\xi_m = (\xi_{1,m}, \ldots, \xi_{n,m})'$, where $r << n$. m is a monthly time index. As explained above, the monthly indicators $x_{i,m}$ are normalized three-month growth rates or differences. The DFM assumes that the idiosyncratic components are a multivariate white noise process, hence the covariance matrix Σ_ξ is diagonal. Furthermore, the DFM assumes that the factors follow a vector-autoregressive process of order p:

$$f_m = \sum_{s=1}^{p} A_s f_{m-s} + \zeta_m, \qquad \zeta_m \sim (0, \Sigma_\zeta) \tag{4.2}$$

where A is a square $r \times r$ matrix. Moreover, the covariance matrix of the VAR (Σ_ζ) is driven by a q dimensional standardized white noise process η_m:

$$\zeta_m = B\eta_m, \qquad \eta_m \sim N(0, I_q) \tag{4.3}$$

where B is a $r \times q$ matrix and $q \leq r$. The final equation is a forecasting equation linking the factors to mean-adjusted real GDP growth:

$$y_m = \beta' f_m + \varepsilon_m, \qquad \varepsilon_m \sim N(0, \sigma_\varepsilon^2) \tag{4.4}$$

where y_m denotes the (unobserved) three-month growth rate of monthly real GDP, i.e. the growth rate vis-à-vis the same month of the previous quarter. Quarterly real GDP growth in quarter t, y_t^Q, is assigned to month $3t$ on the monthly time scale.

[6] Examples are Giannone et al. (2008) and Liebermann (2014) for the United States; Bańbura et al. (2011), Camacho and Perez-Quiros (2010), Rünstler et al. (2009) and Bańbura and Modugno (2014) for the euro area; Schumacher and Breitung (2008) for Germany; Schneider and Spitzer (2004) for Austria; Cheung and Demers (2007) for Canada; Camacho and Perez-Quiros (2011) for Spain and den Reijer (2013) and de Winter (2011) for the Netherlands.

The relation between the quarterly and monthly GDP growth rates is given by $y_t^Q = \frac{1}{3}(y_{3t} + y_{3t-1} + y_{3t-2})$.

The model is estimated in four steps. In the first step, the factors loadings Λ and the estimated static factors $\hat{f}_m$ are obtained. In the second step, the coefficient matrices A_s in Eq. (4.2) and β in Eq. (4.4) are estimated by OLS using $\hat{f}_m$. In the third step, ζ_m and its covariance matrix Σ_ζ are calculated, and an estimate of the matrix B is obtained by principal components analysis. In the final step, the model is cast in state space form and uses the Kalman filter and smoother to re-estimate the estimated factors ($\hat{f}_m$) and monthly GDP growth.[7] Forecasts of quarterly GDP growth are calculated by applying Eq. (4.4) to forecasts of monthly factors generated by Eq. (4.2), and then aggregating to quarterly values.

To estimate the model, the number of static and dynamic common factors need to be specified, denoted by r and q, respectively. The largest possible value of r is set at 6, based on the scree test of Cattell (1966). Moreover, $q \leq r$ by definition. In view of potential misspecification and instabilities described in Chapter 3, the presented outcomes are based on the (unweighted) average of forecasts over all possible parameterizations in terms of the number of static and dynamic factors and the number of lags p in Eq. (4.2), with $p \leq 2$. Thus, the total number of model specifications is $p(r+1)r/2 = 42$. This strategy avoids any hindsight bias.[8]

4.2.3 Forecast design

The forecast design entails the construction of six consecutive forecasts for real GDP growth for each quarter in the period 1999.I–2013.IV. Table 4.1 explains the timing of the forecasting exercise, taking the forecast for a second quarter as an example. The first forecast is made in mid-March, just after the release of a new Consensus survey. Subsequently, a monthly forecast is produced for the next five months. The last forecast is made in mid-August. Following the conventional terminology, *forecasts* (F) refer to predictions made prior to the start of the quarter of interest, *nowcasts* (N) refer to current quarter forecasts and *backcasts* (B) refer to forecasts for the preceding quarter, as long as official (target) GDP figures are not yet available. These forecasts (or horizons) are referred to as F3, N1, N2, N3, B1 and B2, respectively. A Consensus survey consists of new F3 and N3 forecasts. There is no genuine Consensus backcast, since analysts have official information on GDP growth in the preceding quarter when

[7] The state space form of the DFM is outlined in Appendix 4.B. See Bańbura and Rünstler (2011) and Stock and Watson (2011) for a more detailed description of the DFM and the estimation procedure. See Durbin and Koopman (2012) for a comprehensive treatment of state space models and the use of the Kalman filter and smoother.

[8] A different approach is to choose the number of factors r and q on the basis of in-sample criteria, as described in Bai and Ng (2002, 2007). Chapter 3 of this thesis and Bańbura and Rünstler (2011) report that these criteria tend to indicate a relatively large number of factors, leading to volatile and less accurate forecasts.

Table 4.1: Timing of forecasting exercise for second quarter GDP growth

Nr.	Forecast type	Month	DFM	Consensus
F3	Forecast	3	March	new
N1	Nowcast	1	April	1 month old
N2		2	May	2 months old
N3		3	June	new
B1	Backcast	1	July	1 month old
B2		2	August	2 months old

they form their expectations. In the first and second months of a quarter the (non-updated) forecast dating from the third month of the previous quarter are used as Consensus forecast. For example, the Consensus backcasts made for quarter t are equal to the (non-updated) Consensus nowcast published in the last month of quarter t. Forecasts further ahead than one quarter are not analyzed, as earlier research shows that the forecastability of GDP growth is very low beyond this horizon. This holds for statistical models as well as professional forecasters; see for instance Stark (2010) and Chapter 3 of this thesis. Estimation of the DFM and the subsequent calculation of forecasts happens recursively on the basis of the most recent 15 years of data on monthly indicators and quarterly real GDP. The dataset is real-time; all data used for estimation and prediction were actually available at the time of estimation.

An important issue in a real-time forecasting exercise is how to measure the realized outcome, "actual GDP". The latest-available GDP data represent the current state-of-thinking about the "true" history of real GDP. However, these data partly reflect benchmark revisions, which both analysts and forecasting models cannot foresee. Using the latest-available data thus introduces noise in a comparison of forecasting performance between professional forecasters and mechanical statistical procedures, making this data concept unsuitable for the analysis in this chapter. Statistical agencies publish a sequence of preliminary GDP estimates, with the first release (flash estimate) receiving by far the most attention in the media. It is therefore reasonable to assume that analysts are primarily focused on predicting the flash or early estimates rather than a number that will be released far into the future. The real-time measure of actual GDP used in this chapter, is determined by the latest officially released information for the preceding quarter that analysts know at the moment when they formulate predictions for the current and subsequent quarters. For five countries (and the UK before 2008) this is the flash release, which is published six weeks after the reference quarter has ended. For the US and the UK (as from 2008) it is the first revision to the flash, as in these cases the flash is already released four weeks after the quarter has ended. This choice implies the analysis presents two backcasts for all countries.[9]

[9] Section 4.3.1 discusses results for the DFM using the most recent GDP vintage as "actual GDP"

4.3 Empirical results

4.3.1 Forecasting performance dynamic factor model and subjective forecasts

Table 4.2 presents data on the forecasting performance of the DFM for the full sample period 1999.I–2013.IV, the relatively tranquil pre-crisis period 1999.I–2007.IV and the volatile post-crisis period 2008.I–2013.IV. Throughout this chapter the empirical results will be presented in this way on account of the large difference in the level of volatility before and after 2008. Forecasting performance is measured by the root mean squared forecast error (RMSFE). As usual in the literature, the random walk (RW) with drift is used as a purely statistical benchmark model. The first column of the table reports the RMSFE of the random walk to gauge the overall level of volatility.[10] The other columns report the DFM's RMSFE relative to that of the RW benchmark in order to improve the comparability of the results across countries and horizons. Entries in bold indicate that the RMSFE differs by more than 10% from the RW's RMSFE. The 10 per cent threshold is meant as a rough informal assessment of the economic importance of the gain in forecasting accuracy from using auxiliary monthly information. In addition, Diebold and Mariano (1995) tests are presented as formal tests of statistical significance at the conventional levels (denoted by asterisks).[11] Non-starred, normal-type entries thus indicate models that are equal in terms of forecasting accuracy, both statistically and economically. In this chapter, the same two-way approach to statistical/economic significance is featured in all tables that present RMSFEs.

Table 4.2 demonstrates that incorporating monthly information pays off in terms of forecasting accuracy. In the period 1999–2013, virtually all relative RMSFEs are smaller than one and tend to decline if the horizon shortens and more monthly information has been absorbed. However, there is a marked contrast between the pre-crisis and post-crisis periods. The scope for exploiting monthly information for prediction is significantly smaller in the tranquil pre-crisis period, when real GDP growth was more predictable in general, as indicated by the much lower RMSFE of the RW benchmark compared to that of the post-crisis period. The DFM performs, on average, around 11% better than the RW for nowcasts and around 18% for backcasts before 2008; after

as a sensitivity check. Besides, the effect of conducting a pseudo real-time exercise rather than a real-time exercise is discussed.

[10] To save space RW's RMSFE is only reported for the third nowcast (N3), as it hardly varies with the horizon. The drift parameter is recursively estimated in real-time on the most recent 15 years of GDP data. Results for both the DFM and RW model were also calculated for an estimation windows of 10 years and a recursively expanding window, but the results where qualitatively the same.

[11] The Diebold-Mariano tests broadly paints the same picture as the informal 10% improvement criterion, although the two do not always match. In some cases, large differences in accuracy are not statistically significant, whilst the reverse also happens. This suggest that statistical significance and economic importance are different concepts. Moreover, the power of the Diebold-Mariano tests may be low due to the small number of observations.

Table 4.2: Forecasting performance dynamic factor model, 1999.I–2013.IV

	RW	B2	B1	N3	N2	N1	F3
	RMSFE	Relative RMSFE DFM vs. Random Walk					
		Full sample 1999.I–2013.IV					
Canada	*0.57*	**0.61****	**0.66****	**0.69***	**0.76***	**0.81***	**0.86**
France	*0.46*	**0.67***	**0.77***	**0.81***	**0.85**	**0.87**	0.93
Germany	*0.82*	**0.72***	**0.76***	**0.84**	**0.88**	**0.88***	0.92*
Italy	*0.64*	**0.62***	**0.71***	**0.76***	**0.81***	**0.80***	**0.88***
Japan	*1.08*	**0.67***	**0.72***	**0.78***	0.93	1.04	1.06
UK	*0.63*	**0.76***	**0.78***	**0.83***	0.90	0.92	0.91*
US	*0.61*	**0.57***	**0.62***	**0.70***	**0.80**	**0.86**	**0.89**
		Pre-crisis period 1999.I–2007.IV					
Canada	*0.39*	**0.69*****	**0.73*****	**0.82****	0.94	0.92**	0.94
France	*0.33*	**0.79****	**0.88***	**0.83*****	**0.80*****	**0.83*****	0.95
Germany	*0.42*	**0.83**	**0.87**	0.99	0.97	0.90	**0.87****
Italy	*0.39*	0.91*	0.95	0.97	0.97	0.92*	0.95
Japan	*0.86*	0.92	0.94	0.94	0.97	0.97	1.01
UK	*0.25*	**0.77****	**0.76****	**0.79***	**0.78***	**0.66****	**0.67*****
US	*0.49*	**0.71*****	**0.75*****	**0.82*****	0.91*	0.96	0.98
		Post-crisis period 2008.I–2013.IV					
Canada	*0.77*	**0.57***	**0.62***	**0.63***	**0.68***	**0.76**	**0.82**
France	*0.60*	**0.62***	**0.72**	**0.80**	**0.87**	**0.89**	0.93
Germany	*1.19*	**0.70***	**0.73***	**0.81**	**0.86**	**0.88**	0.92
Italy	*0.89*	**0.50***	**0.63***	**0.70***	**0.75***	**0.76***	**0.86***
Japan	*1.35*	**0.45***	**0.53***	**0.67**	0.90	1.08	1.09*
UK	*0.94*	**0.76***	**0.78***	**0.84**	0.91	0.94	0.93
US	*0.75*	**0.46**	**0.51**	**0.61**	**0.72**	**0.78**	**0.83**

Notes: entries denote the RMSFE for a Random Walk for the third nowcast (in italics); for all other cases they refer to the RMSFE relative to the RMSFE of a Random Walk. Entries in bold denote a deviation by more than ± 10% from 1. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. The alternative hypothesis of the Diebold-Mariano test is MSFE > 1 or MSFE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively. DFM: dynamic factor model.

2008 the gains are twice as large: 20% and 39%, respectively. For Italy and Japan, the accuracy gains vis-à-vis the random walk before 2008, are less than 10% for all horizons. Moreover, the DFM's predictive ability before the crisis develops unevenly over the six months of the forecasting exercise for several countries, particularly for Germany and the UK. By contrast, the DFM delivers improvements in forecasting accuracy that are large and steadily increasing with the forecast horizon for all countries in the volatile post-crisis period 2008–2013. The DFM thus helps forecasters to compensate for the generalized deterioration in predictability that characterize volatile times, in particular for nowcasting and backcasting. The evidence also suggests that the DFM's relative strength is to exploit information pertaining to the quarter under consideration and to improve the assessment of the current state of the economy. The findings in Table 4.2 are broadly consistent with the empirical literature on DFMs referred to in Section 4.2.2.

Table 4.8 in Appendix 4.C, presents the impact of two features of the forecasting design in the main text, namely (i) its real-time nature; and (ii) the use of initial estimates as the measure of target GDP. The left-hand side of the table reports the DFM's forecasting performance obtained in the corresponding pseudo real-time set-up, which also implies that prediction errors have been computed using the last vintage for GDP. A comparison of the RMSFEs of the random walk model reveals that forecasting errors tend to be larger for the pseudo real-time design on account of the extra noise generated by the data revisions. At the same time, the DFM's relative RMSFEs tend to decline more steeply in the pseudo real-time procedure. This phenomenon is especially pronounced for nowcasts in the pre-crisis period. A pseudo real-time set-up may thus overestimate the potential gains from exploiting monthly disaggregated information when nowcasting GDP, especially in stable environments. The right-hand side of Table 4.8 presents the outcomes for the real-time procedure, when RMSFEs are calculated using the final GDP vintage as measure of actual GDP. The first column of this side of Table 4.8 reports the RMSFE of the first estimate as a predictor of the final vintage. These results suggest that revisions are a substantial source of unpredictability as the RMSFE of the first estimate varies from 0.2 and 0.4 for six countries, while for Japan it is even larger. Moreover, using the final GDP vintage as the target rather than the first estimate tends to lead to a more optimistic view of the potential gains from using monthly information before the crisis, but this does not hold after the crisis.

Turning to the Consensus forecasts, panel A of Table 4.3 presents the RMSFE of the Consensus forecasts, relative to the RW benchmark model. Note, that only the entries for the third one-quarter ahead forecast and the third nowcast (columns F3 and N3) are based on a fresh Consensus forecast. The other entries refer to Consensus forecasts that are one or two months old. In the period 1999–2013, Consensus forecasts outperform the RW model for all countries and all horizons, often by large margins, with the exception of the F3 forecasts in the Japanese case. Moreover, Consensus forecast display clear learning behaviour: the relative RMSFEs of consecutive fresh quarterly Consensus forecasts (N3 versus F3) decline steeply in most cases, varying from 12% for the UK to 31% for Canada.[12] Again, the forecasting performance differs markedly between the pre-crisis and the post-crisis periods. Before 2008, the Consensus F3 forecast is mostly unable to beat the random walk. It does 10% better in the case of the UK, but does 18% and 9% worse in the cases of Canada and Japan, respectively. The Consensus third nowcast (column N3) has a better record, reducing forecasting inaccuracy markedly for most countries compared to the RW. After 2008, and similar to the experience with the DFM, Consensus forecasts perform much stronger for almost all countries and all horizons. The RMSFE of the third nowcast is 29% to 65% smaller than the RW RMSFE. The most remarkable case concerns Canada: Consensus forecasts predict badly before 2008, but very well after 2008 (and especially during the crisis episode in 2008–2009). Comparing the 1999–2007 and 2008–2013 episodes, Consensus forecasts show only modest and uneven improvements for Germany and Japan.

To gain further insights into the relative strengths of strictly model-based and judgmental predictions, panel B of Table 4.3 presents the RMSFE of the Consensus forecasts relative to that of the DFM. Newly released Consensus nowcasts (column N3) do better than the DFM for almost all countries in the evaluation period 1999–2013, and often by a significant margin. The advantage is much less pronounced for fresh one quarter ahead forecasts (column F3), which clearly beat the DFM for only the UK and the US. Looking at the results across subperiods, the relatively positive score of the Consensus F3 and N3 forecasts is largely driven by their performance after the financial crisis. In the stable pre-crisis period, Consensus forecasts basically do worse or at most marginally better than the DFM across the board; the N3 nowcast for Germany is the one clear favourable exception. However, new Consensus forecasts (N3 and F3) have the edge over the DFM after the crisis in all cases but Japan. The relative RMSFE versus the DFM generally declines between horizons F3 and N3, which suggests that the value added of subjective insights may be greater when analysts know at least some hard and soft data on the quarter of interest. In between the quarterly release dates, the performance of Consensus forecasts versus DFM forecasts deteriorates, as the latter are updated. This catching-up by the DFM is stronger after 2008. However, at the B1 horizon, a two-month old Consensus nowcast still outperforms the DFM by more than 10% for the UK over the sample period, while this holds for Canada and Germany in one of the subperiods.

[12] The generally small changes in relative RMSFEs in the months without a Consensus survey (columns B2, B1, N2 and N1) are due to revisions of GDP data, which may change the RW forecasts, and the fact that the Consensus survey was released one month earlier than normal in 2001.IV.

Table 4.3: Forecasting performance Consensus forecasts, 1999.I–2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	A. Relative RMSFE CF versus RW						B. Relative RMSFE CF versus DFM					
	Full sample 1999.I–2013.IV						Full sample 1999.I–2013.IV					
Canada	**0.56***	**0.57***	**0.57***	**0.84**	**0.82**	**0.82**	0.92	**0.86**	**0.82**	**1.11**	1.02	0.96
France	**0.68****	**0.68****	**0.68****	**0.84****	**0.86***	**0.86***	1.01*	**0.88****	**0.84*****	0.98**	0.98*	0.92*
Germany	**0.68****	**0.68****	**0.68****	**0.89***	0.90	0.90	0.94	0.90**	**0.81*****	1.01	1.02	0.98
Italy	**0.69****	**0.69****	**0.69****	**0.81****	**0.84***	**0.84***	**1.12**	0.97	0.90	1.01	1.05	0.95
Japan	**0.81***	**0.81***	**0.81***	0.98	0.98	0.98	**1.22***	**1.14**	1.04	1.06	0.94	0.93
UK	**0.57***	**0.57***	**0.57***	**0.64***	**0.65***	**0.65***	**0.75***	**0.73***	**0.69***	**0.71***	**0.71***	**0.71***
US	**0.57***	**0.57****	**0.57****	**0.82**	**0.80**	**0.80**	1.00	0.92	**0.81*****	1.03	0.94	0.90*
	Pre-crisis period 1999.I–2007.IV						Pre-crisis period 1999.I–2007.IV					
Canada	**0.89****	0.91*	0.91*	**1.23****	**1.18****	**1.18****	**1.29****	**1.24***	**1.11**	**1.31****	**1.29****	**1.26****
France	**0.87***	**0.87***	**0.87***	0.94	1.01	1.00	**1.11**	0.99*	1.05	**1.18**	**1.21**	1.06
Germany	**0.66****	**0.68****	**0.68*****	0.94	1.00	1.00	**0.80**	**0.78**	**0.69****	0.97	**1.12**	**1.15**
Italy	0.92	0.93	0.92	0.98	1.09**	1.09**	1.01	0.98	0.96	1.01	**1.19***	**1.15***
Japan	0.96	0.96	0.97	0.94	0.96	0.96	1.04	1.03	1.03	0.98	0.99	0.96
UK	**0.81***	**0.81***	**0.80***	**0.85**	**0.89**	0.90	1.05	1.06	1.01	1.09	**1.35*****	**1.35*****
US	**0.74*****	**0.73*****	**0.73*****	1.00	0.96	0.96	1.04	0.98	0.90**	1.10	1.00	0.98
	Post-crisis period 2008.I–2013.IV						Post-crisis period 2008.I–2013.IV					
Canada	**0.35***	**0.35***	**0.35***	**0.63***	**0.63***	**0.63***	**0.61*****	**0.56****	**0.55****	0.93	**0.83*****	**0.76****
France	**0.58***	**0.58***	**0.58***	**0.78****	**0.78****	**0.78****	0.94***	**0.81*****	**0.72*****	0.90***	**0.88****	**0.84*****
Germany	**0.68****	**0.68****	**0.68****	**0.88***	**0.88***	**0.88***	0.97	0.93*	**0.84*****	1.02	1.00	0.95
Italy	**0.61*****	**0.61*****	**0.61*****	**0.76*****	**0.76*****	**0.76*****	**1.21**	0.96	**0.87**	1.00	0.99	**0.88****
Japan	**0.71***	**0.71***	**0.71***	1.00	0.99	0.99	**1.59**	**1.32**	1.06	**1.11**	0.92	0.91
UK	**0.54***	**0.54***	**0.54***	**0.61****	**0.61****	**0.62****	**0.71****	**0.69****	**0.65****	**0.67****	**0.65****	**0.66****
US	**0.43**	**0.43**	**0.43**	**0.68**	**0.68**	**0.68**	0.93	**0.84***	**0.70*****	0.95	**0.88***	**0.82***

Notes: entries denote the RMSFE of the Consensus Forecasts relative to the RMSFE of a Random Walk and the dynamic factor model, respectively. Entries in bold denote a deviation by more than ± 10% from 1. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. The alternative hypothesis of the Diebold-Mariano test is MSFE $>$ 1 or MSFE $<$ 1, depending on whether the RMSFE is greater or smaller than 1, respectively. CF: Consensus forecasts, RW: Random Walk, DFM: dynamic factor model.

Whether judgmental forecasts by analysts embody valuable additional information can be more formally investigated by running encompassing tests of the Consensus forecasts versus the DFM forecasts. Even if Consensus forecasts are a poor predictor on their own, they may still possess a positive marginal value provided they are able to pick up specific useful information. In that case taking a combination of the Consensus and DFM forecasts may improve forecasting accuracy. In the empirical literature encompassing tests may take on slightly different forms. In this chapter two versions of the encompassing test are employed, namely a pure weighted average of both forecasts or an unconstrained linear combination. The respective test regressions are:

$$y^Q_{t+h} = \beta \hat{y}^Q_{\mathrm{CF}(t+h|t)} + (1-\beta)\hat{y}^Q_{\mathrm{DFM}(t+h|t)} + \varepsilon_t \tag{4.5}$$

$$y^Q_{t+h} = \alpha + \beta \hat{y}^Q_{\mathrm{CF}(t+h|t)} + \gamma \hat{y}^Q_{\mathrm{DFM}(t+h|t)} + \varepsilon_t \tag{4.6}$$

where y^Q_{t+h} is GDP growth in quarter $t+h$, $\hat{y}^Q_{\mathrm{CF}(t+h|t)}$ and $\hat{y}^Q_{\mathrm{DFM}(t+h|t)}$ are the predictions for quarter $t+h$ on time t by the Consensus survey and the DFM, respectively. These equations can be seen as extreme forms of a linear combination. Eq. (4.6) is fully unconstrained, while Eq. (4.5) is the most constrained version.[13] Eq. (4.5) goes back to Bates and Granger (1969) and has been applied by Stekler (1991) and the previous chapter, amongst others. Eq. (4.6) was proposed by Granger and Ramanathan (1984) and has been used by Fair and Shiller (1990), Liebermann (2014) and Hubert (2014). The main advantage of Eq. (4.6) is its ability to neutralize any biases in the underlying forecasts; negative estimates of β or γ are theoretically possible (Timmermann, 2006). Its disadvantage is that it may deliver imprecise estimates in small samples if the underlying forecasts are (highly) correlated. By construction, Eq. (4.6) will obtain a better fit of observed GDP than Eq. (4.5) for a given estimation sample. Consensus forecasts contain additional information in relation to the DFM if $\beta > 0$ in Eq. (4.5) or $\beta \neq 0$ in Eq. (4.6). Note that the encompassing test is an in-sample backward-looking test, which is just meant to signal the potential of benefitting from combining model-based forecasts with judgmental forecasts.

Table 4.4 reports the estimated weight of Consensus forecasts in Eq. (4.5).[14] β and its standard error are estimated on the interval [0,1] by Maximum Likelihood (ML). Moreover, Table 4.4 presents asymptotically valid tests of the hypotheses $\beta = 0$ and $\beta = 1$.

[13] Eq. (4.5) imposes the following restrictions on Eq. (4.6): $\alpha = 0$, $0 \leq \beta \leq 1$, $0 \leq \gamma \leq 1$ and $\beta + \gamma = 1$. Empirical studies have also applied intermediate forms of the encompassing test. For example, Rünstler et al. (2009) impose $\beta + \gamma = 1$, but neither $0 \leq \beta \leq 1$ nor $0 \leq \gamma \leq 1$. There is a large literature on the combination of forecasts; see Timmermann (2006) for an overview.

[14] Table 4.11 in Appendix 4.C reports the results of Eq. (4.6). These results display the same pattern as the ones for Eq. (4.5). As expected, in some cases, the estimates of β or γ are larger than one or less than zero.

Table 4.4: Encompassing test: weighted average of DFM and Consensus forecasts, 1999.I–2013.IV

	B2	B1	N3	N2	N1	F3
	Weight(β) of Consensus forecast					
	Full sample 1999.I–2013.IV					
Canada	**0.66***	**0.74***	0.84*	**0.29***	**0.45***	**0.63***
France	**0.48***	**0.64***	**0.70***	**0.52***	**0.52***	**0.65***
Germany	0.67*	0.77*	1.00*	**0.48***	**0.42***	0.59*
Italy	**0.35***	**0.56***	0.72*	**0.49***	**0.35***	0.65*
Japan	**0.18**	**0.26**	**0.39***	**0.35***	0.68*	0.81*
UK	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*
US	**0.51***	0.72*	1.00*	**0.42***	0.73*	0.89*
	Pre-crisis period 1999.I–2007.IV					
Canada	**0.00**	**0.09**	**0.28**	**0.02**	**0.06**	**0.01**
France	**0.35***	**0.52***	**0.43***	**0.26**	**0.19**	**0.35**
Germany	0.82*	0.80*	0.95*	**0.55***	**0.25**	**0.12**
Italy	**0.46***	0.56	0.63*	**0.45**	**0.00**	**0.00**
Japan	**0.41***	**0.45**	**0.43***	**0.55***	**0.52***	0.61*
UK	**0.47***	**0.46***	**0.49***	**0.44***	**0.27***	**0.23***
US	**0.32**	0.67	1.00*	**0.00**	0.52	0.72
	Post-crisis period 2008.I–2013.IV					
Canada	1.00*	1.00*	1.00*	0.67*	0.95*	1.00*
France	**0.55***	**0.70***	0.82*	**0.61***	**0.65***	0.76*
Germany	0.59	0.76	1.00*	0.44	0.49	0.78
Italy	**0.32***	**0.56***	0.76*	**0.50***	0.52	0.89*
Japan	**0.00**	**0.07**	**0.32**	**0.18**	0.82*	1.00*
UK	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*
US	**0.58***	0.73*	0.97*	0.59*	0.78*	0.92*

Notes: entries refer to the Maximum Likelihood estimate of the weight of the Consensus forecast β on the interval $[0, 1]$. The estimated weight of the forecast of the dynamic factor model is $1 - \beta$. * denotes the estimated weight in the encompassing test is statistically different from zero at the 5% significance level. Entries in bold denote that the estimated weight is statistically different from 1 at the 5% significance level.

The main message of Table 4.4 is that both Consensus forecasts and DFM forecasts contain useful information. The point estimates lie mostly between 0 and 1; corner solutions are rare, except for Canada and the UK. The relative value added of Consensus forecasts differs across time periods. In the pre-crisis period the estimated weight of Consensus forecasts is typically lower than 0.5, and often not statistically significant for F3 forecasts and early nowcasts. After the crisis, they are often greater than 0.5 for all horizons. For the UK the estimated weight of the Consensus forecasts even equals one for all horizons, suggesting that DFM forecasts do not possess any extra information compared to forecasts by professional analysts; the same holds for the backcasts for Canada. For Japan, the added value of the Consensus nowcast is low. Consensus forecasts that are one or two months old often still offer the potential of improving DFM forecasts, even though the latter incorporate more recent monthly information. This finding is yet another piece of evidence that analysts' forecasts contain information that is fundamentally different from the information that statistical models are able to pick up.

4.3.2 Enhancing model-based forecasts in real-time

The analysis in the previous section suggests that there is ample room for improving mechanical model-based nowcasts and backcasts, by combining them with judgmental predictions by professional forecasters. This section investigates what benefits can be realistically expected from such a strategy in practice, by simulating the forecasting procedure on the basis of real-time data. The procedure consists of two steps. In the first step, the DFM forecasts are obtained in real-time, the second step determines the combination formula and computes the forecast combination. Both the weighted average and the linear combination are used as combination schemes, as it is difficult to choose between them on theoretical or methodological grounds (see e.g. Clemen, 1986). Although Eq. (4.6) is able to remove the effect of possible biases in the DFM and Consensus forecasts on the combined forecast, the simpler Eq. (4.5) may still prove, despite possible bias, to be a more efficient combination rule. The combination schemes Eq. (4.5) and Eq. (4.6) are re-estimated every month according to the schedule in Table 4.1.[15] This requires a choice on the length of the estimation period. As it is difficult to motivate a specific number on a priori grounds, the prediction rules were estimated using a moving window of 4–8 years. Next, the resulting five predictions were averaged.[16]

[15] Preliminary calculations indicate that the linear combination scheme occasionally produced implausible forecasts. To deal with this problem a lower and upper bound to the forecast was set. The lower bound is the minimum of the DFM and Consensus forecasts minus 0.3 percentage points; the upper bound is the maximum of the DFM and Consensus forecasts plus 0.3 percentage points.

[16] As a robustness check, the effect on forecasting performance of employing shorter or longer estimation windows was investigated. This effect was found to be minor. Table 4.5 was also calculated

Table 4.5 reports the RMSFEs of both forecast combinations, relative to the RMSFEs of the DFM. The main conclusion of the exercise is that both combination methods are suitable to enhance DFM forecasts in economically meaningful ways, Japan being the only exception. The latter is consistent with the poor performance of the Japanese Consensus forecast vis-à-vis the DFM forecasts (see panel B of Table 4.3). The scope for improvement is in general considerably greater in the volatile post-crisis period than in the tranquil times before 2008. In the post-crisis period gains in prediction accuracy of 20% to 35% are feasible for at least some horizons for many countries. Gains are on average more limited (often less than 10%) before 2008, but still reach 33% for Germany and 16% for the UK for the third month nowcast using the weighted average scheme (30% and 9% for the linear combination scheme). After 2008, the forecasting accuracy of the combined forecasts is better than the DFM forecasts for the majority of countries and forecasting horizons. The greatest advantage of combining DFM and Consensus forecasts generally occurs in the third month of the quarter when just-released Consensus forecasts are available. Although the relative RMSFE of the combined forecast versus the DFM forecasts tends to increase in the subsequent two months, it remains below one in most cases. This is evidence that Consensus forecasts incorporate information that goes beyond the information represented by the statistical information set, and that this type of information is still valuable even if it is somewhat dated. Moreover, enhancing model-based forecasts with subjective forecasts may offer some insurance against a weak performance of mechanical models. The experience of the UK before 2008 is a case in point. As Table 4.2 shows, the DFM loses predictive power between horizons F3 and N3. This loss is roughly neutralized by utilizing additional, subjective information.

Overall, the weighted average scheme tends to work better than the linear combination scheme before 2008, although the relative difference in RMSFE typically does not exceed 10%. After 2008, the linear combination scheme appears to have a slight edge on average. It performs extremely well in case of all of Italy's predictions and Germany's backcasts, but does a generally poor job at all horizons for the US. Thus, the empirical evidence points to quite some country heterogeneity in the relative performance of both schemes, which may even differ across horizons, as the Canadian and British experience illustrates. Looking at the evidence over the whole sample period across all forecasting horizons, the weighted average scheme performs better for France, Japan and the US. The linear combination scheme is better for Germany and Italy, whilst there is no clear winner for Canada and the UK. Another pattern in Table 4.5 is that the linear combination scheme is riskier than the weighted average scheme, in the sense that combining forecasts may actually hurt forecasting accuracy appreciably; see the US results.

---

with moving windows of 3–6 years and 5–10 years as well as a recursively expanding estimation window. The results where qualitatively the same.

Table 4.5: Forecasting performance of forecast combination schemes, 1999.I–2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus DFM											
	A. Weighted average scheme						B. Linear combination scheme					
	Full sample 1999.I–2013.IV						Full sample 1999.I–2013.IV					
Canada	**0.82***	**0.78***	**0.77****	0.98	0.94*	0.90*	**0.83***	**0.80***	**0.80***	0.92	0.90*	**0.86***
France	0.91**	**0.86*****	**0.84*****	0.90***	0.91***	**0.89****	0.95**	**0.89****	**0.87*****	0.98***	0.98***	0.91*
Germany	0.96	0.91**	**0.81*****	1.01	1.00	0.98	**0.87****	**0.82*****	**0.79*****	1.00	1.02	1.01
Italy	1.00	0.96	0.97	0.98	0.98	0.96*	**0.83****	**0.83****	**0.80****	0.92	0.90*	0.94
Japan	1.01	1.00	1.00	1.02	1.03	0.94	1.04	1.05	1.05	1.07*	1.04	1.05
UK	**0.76****	**0.74***	**0.70***	**0.71***	**0.70***	**0.71***	**0.79***	**0.78***	**0.74***	**0.69***	**0.66***	**0.65***
US	0.94*	0.92*	**0.85****	0.97	0.94*	0.92*	1.01	1.01	0.95	1.08	1.07	1.05
	Pre-crisis period 1999.I–2007.IV						Pre-crisis period 1999.I–2007.IV					
Canada	1.00	1.00	0.98	1.01	1.02	1.01	0.95	0.99	1.02	0.99	1.01	0.99
France	0.98**	0.92**	0.95**	0.99**	1.01***	0.99	1.03*	0.95**	0.99**	1.05*	1.09**	1.06
Germany	**0.78****	**0.76****	**0.67****	0.96	1.05	1.06	**0.80****	**0.76****	**0.70****	0.90	0.98	1.00
Italy	0.98	0.96*	0.94**	1.00	1.02*	1.02*	0.98	0.96	0.96	0.97	1.02	1.02
Japan	0.99	0.98	0.99	0.97	0.98	0.96	1.03	1.04	1.04	1.03	1.04	1.02
UK	**0.88**	**0.85**	**0.84**	**0.85**	0.96	0.99	0.94	0.95	0.91	**0.88**	0.99	1.00
US	0.98*	0.98*	0.95*	1.03*	0.99	0.99	1.04	1.07	1.04	**1.16***	**1.14***	**1.15***
	Post-crisis period 2008.I–2013.IV						Post-crisis period 2008.I–2013.IV					
Canada	**0.70*****	**0.62****	**0.59*****	0.95*	**0.89*****	**0.85****	**0.74****	**0.67****	**0.61****	**0.85**	**0.84****	**0.78***
France	**0.86*****	**0.81*****	**0.78*****	**0.87*****	**0.87*****	**0.85*****	**0.88*****	**0.85*****	**0.81*****	0.96***	0.93***	**0.83*****
Germany	1.00	0.95	**0.84*****	1.02	1.00	0.96*	**0.88****	**0.83*****	**0.82*****	1.02	1.03*	1.01
Italy	1.01	0.96	0.98	0.97	0.96*	0.93***	**0.65*****	**0.73*****	**0.70*****	0.90	**0.85****	0.91**
Japan	1.06	1.04	1.01	1.05**	1.05**	0.93	1.06**	1.07***	1.05*	1.10*	1.04*	1.06**
UK	**0.74****	**0.73****	**0.68****	**0.69****	**0.68****	**0.69****	**0.77***	**0.76***	**0.72***	**0.68****	**0.64****	**0.63****
US	**0.88***	**0.84****	**0.72*****	0.91**	0.90*	**0.84***	0.96	0.93	**0.83*****	1.00	1.00	0.95

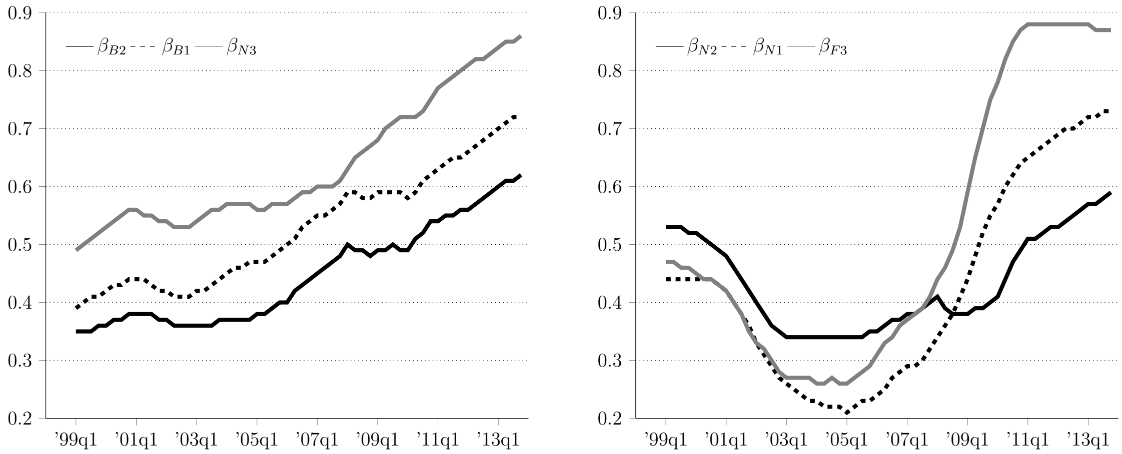
Notes: entries in bold denote a deviation by more than ± 10% from 1. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. The alternative hypothesis of the Diebold-Mariano test is MSFE > 1 or MSFE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

These findings thus suggest that the linear combination scheme seems to be less robust to overfitting in small samples, especially in stable macroeconomic environments. As combining forecasts is a symmetric operation, it is also interesting to look at the marginal value of the DFM forecasts. Table 4.9 in Appendix 4.C offers this alternative perspective, presenting the RMSFE of both forecast combinations in terms of the Consensus forecast. In the period 1999–2013 the gains in forecasting accuracy by this measure are modest in most cases. The linear combination scheme works well for Canada and Italy, but not for the US. Moreover, the marginal value of DFM forecasts is generally larger in the pre-crisis period than in the post-crisis period, which is consistent with the improvement in prediction capabilities of Consensus forecasts in the latter period. In fact, it is even significantly negative in a number of cases after 2008, reducing forecasting accuracy. An important empirical result is that it is difficult to beat a fresh Consensus forecast (columns N3 and F3) in the post-crisis period. With some exceptions, both combination rules perform worse than the Consensus forecast for both types of forecast after 2008. This finding suggests that during that period the Consensus forecast, which reflects the aggregate of both model-based predictions and subjective insights of all survey respondents, may incorporate all available information at the moment of release. Hence, the additional value of the DFM appears to be quite limited in the post-crisis period, and the optimal forecast is the Consensus forecast or something very close to it. The DFM may still serve as a means to enhance (non-updated) Consensus forecasts in the two months after the quarterly release. Looking at backcasts (B1 and B2), substantial accuracy gains are possible for Japan and Italy, and modest gains for France and Germany. There is also scope for improvement for the nowcasts N2 and N1, but it is smaller than for the corresponding backcasts.

A constant refrain in the results is that subjective Consensus forecasts have improved in predictive power relative to predictions from the DFM over time. This development is visualized in Figure 4.1, which depicts the weight of the Consensus forecast β in Eq. (4.5) over the years 1999–2013.[17] The weights of the Consensus nowcasts all show a clear upward trend (left-hand panel). Moreover, as expected, they shift downwards as they become older and the DFM exploits new information. The weight for the F3 forecasts steeply increases after the financial crisis (right-hand panel). In the last years of the sample, the weight of the Consensus forecasts is 80–90% for the N3 and F3 forecasting horizon and thus only 10–20% for the DFM forecasts. This observed trend may reflect several phenomena. First, the procedure that was used may flatten the forecasting performance of the DFM in the early part of the sample, because analysts may in reality have used statistical models that were less sophisticated than the DFM, such as pure time series models, bridge models and VAR models. Second, ana-

[17] For presentational reasons β was averaged across countries and estimation windows. The resulting time series are smoothed by applying a centered eight-quarter moving average.

Figure 4.1: Weight of Consensus forecast in weighted average combination scheme, 1999.I–2013.IV



Notes: β was averaged across countries and estimation windows. Time series are smoothed by applying a centered eight-quarter moving average.

lysts may put more effort in making their predictions in recessions and times of high volatility. This is consistent with the findings of the literature that forecasters adjust their forecast more frequently during recessions or when the information set changes a lot. The findings support the findings of Lundquist and Stekler (2012), who conclude that professional analysts are very responsive to the latest information about the state of the economy and adjust their predictions quickly. Third, the DFM forecasts are derived by a mechanical procedure that may be ill-suited to deal with large shocks, such as the financial crisis, although taking averages over all possible specifications and using a rolling estimation window offer some protection. By contrast, analysts base their subjective assessments on potentially a multitude of relevant time-varying factors and they may adjust their models, data and estimation modalities in response to large shocks. Recent work by Castle, Clements, and Hendry (2015) and Castle, Doornik, Hendry, and Pretis (2015) demonstrates the value of employing a statistical procedure that fully incorporates the possibility of location shifts. Finding out how such a forecasting procedure would alter the relative performance of statistical models and subjective forecasts is an interesting topic for future research.

Finally, given the variation in relative performance of both combination schemes across countries and forecasting horizons, the extent to which it is feasible for forecasters to identify in real-time what combination scheme they should apply at a particular moment is investigated. Two strategies are considered. The first strategy does not involve a choice, but consists of simply averaging the predictions generated by the two schemes (for all estimation windows). The empirical literature has found that such a strategy leads to less volatile predictions and possibly an improvement in accuracy (e.g. Kuzin et al., 2013; Timmermann, 2006). The second strategy tries to select the best forecasting rule among ten candidates (two combination schemes estimated using estimation windows of 4–8 years) on the basis of their recently observed (out-of-sample) forecasting ability. The latter is measured over a moving evaluation window that varies between 1–4 years. In view of the difficulty to optimally choose the length of the evaluation window on a priori grounds, again averaging is applied. The strategy first selects the pair of combination scheme and estimation window that delivers the best forecasts for each evaluation window. Next, the four resulting predictions are averaged.

The relative RMSFEs (versus the DFM) of the two strategies are presented in Table 4.6.[18] The main finding is that the differences between them are often quite small and do not show a clear pattern. This is true for both subperiods and the whole sample. In many cases, the difference in forecasting performance between the weighted average and linear combination schemes does not appear to be large and persistent enough to be exploitable by a forward-looking selection strategy in real-time. Moreover, a comparison of Tables 4.5 and 4.6 shows that averaging tends to lead to somewhat lower RMSFEs than either of the combination schemes. Thus, using a simple average of combination schemes may provide a valuable hedge against misspecification and instability of the combination schemes for practitioners. However, as the Italian case illustrates, it may pay off to monitor the prediction performance of the selection strategy at the same time, and possibly switch strategies.

4.4 Conclusion

This chapter investigates to what extent subjective information incorporated in forecasts by professional analysts may enrich mechanical forecasting procedures exploiting monthly statistical data in a truly real-time context, in which both statistical models and analysts have to deal with possibly inaccurate initial GDP estimates. Judgmental forecasts are taken from the quarterly Consensus survey (averaged over the panelists). The model-based forecasts are generated by a DFM that is estimated using real-time monthly vintages. For the sake of robustness, the analysis covers seven large countries (the so-called G7 countries) over the years 1999–2013, allowing a systematic comparison of the tranquil period 1999–2007 before the financial crisis and the volatile post-crisis period after 2008. Moreover, two different schemes to combine model-based and subjective forecasts were analyzed: (i) the weighted average and (ii) the linear combination.

The main findings can be summarized in five points. First, in keeping with other work, monthly statistical indicators seem to contain valuable information that can be extracted by the DFM in real-time, in particular as the horizon shortens and more monthly information is processed. The largest gains in predictive accuracy are for late

[18] Table 4.10 in Appendix 4.C reports the relative RMSFEs versus the Consensus forecasts.

Table 4.6: Forecasting performance of averaged or selected forecast combinations, 1999.I–2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus DFM											
	A. Average of forecast combinations						B. Selection of best forecast combination in the recent past					
	Full sample 1999.I–2013.IV						Full sample 1999.I–2013.IV					
Canada	**0.80****	**0.77***	**0.77***	0.92*	0.90**	**0.86***	**0.76****	**0.75***	**0.77***	0.92	**0.89**	**0.84***
France	0.90**	**0.86*****	**0.83*****	0.92***	0.93***	**0.89****	0.95**	**0.89****	**0.87*****	0.94***	0.95***	0.92*
Germany	**0.89*****	**0.84*****	**0.79*****	0.99	1.00	0.98	**0.89*****	0.93	**0.85*****	1.04	1.01	1.01
Italy	**0.87****	**0.85****	**0.85****	0.92*	0.91**	0.92*	**0.81****	**0.86****	**0.81****	0.91	**0.89***	0.94
Japan	1.02	1.02	1.02	1.04	1.03	0.98	1.06	1.05	1.03	1.08	1.04	1.03
UK	**0.76***	**0.74***	**0.70***	**0.68***	**0.67***	**0.67***	**0.77***	**0.74***	**0.71***	**0.69***	**0.65***	**0.66***
US	0.96	0.94	**0.88****	1.01	1.00	0.97	0.97	0.95	0.91	1.00	0.97	0.95
	Pre-crisis period 1999.I–2007.IV						Pre-crisis period 1999.I–2007.IV					
Canada	0.93	0.96	0.98	0.99	1.00	0.98	**0.89**	0.95	1.01	1.02	1.04*	1.00
France	1.00**	0.93**	0.96**	1.01**	1.04**	1.02	1.01**	0.97**	1.00**	1.03*	1.06**	1.04
Germany	**0.76****	**0.74****	**0.67****	0.90	0.97	0.97	**0.78****	**0.76****	**0.69****	**0.89**	0.96	0.96
Italy	0.96	0.94	0.93*	0.96	0.99	0.99	0.95	0.97	0.97	0.98	1.02	1.01
Japan	1.00	1.00	1.01	1.00	1.00	0.99	1.03	1.03	1.01	0.99	1.00	0.98
UK	0.90	**0.89**	**0.86**	**0.86**	0.97	0.98	0.91	**0.89**	**0.86**	**0.88**	0.99	0.99
US	0.98	0.99	0.97	1.07*	1.05	1.06	1.01	1.04	1.02	1.06	1.04	1.05
	Post-crisis period 2008.I–2013.IV						Post-crisis period 2008.I–2013.IV					
Canada	**0.71*****	**0.63****	**0.59****	**0.86****	**0.85*****	**0.80****	**0.67*****	**0.61****	**0.56****	**0.84****	**0.78****	**0.74****
France	**0.83*****	**0.81*****	**0.76*****	**0.89*****	**0.88*****	**0.83*****	0.90***	**0.84*****	**0.80*****	0.90***	0.91***	**0.87*****
Germany	0.92***	**0.87*****	**0.81*****	1.01	1.00	0.98	0.92**	0.97	**0.88*****	1.08	1.03	1.02
Italy	**0.77*****	**0.79*****	**0.80*****	0.90*	**0.87*****	0.90***	**0.65*****	**0.78*****	**0.71*****	**0.88***	**0.84*****	0.92*
Japan	1.05**	1.05**	1.03	1.06*	1.04**	0.98	**1.13***	1.10*	1.05	**1.13****	1.07**	1.05**
UK	**0.74****	**0.72****	**0.69****	**0.66****	**0.65****	**0.65****	**0.75****	**0.72***	**0.70****	**0.67****	**0.63****	**0.64****
US	0.91	**0.87***	**0.77*****	0.94**	0.94	**0.89**	0.91	**0.83****	**0.76*****	0.94	0.90*	**0.86***

Notes: entries in bold denote a deviation by more than $\pm$ 10% from 1. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. The alternative hypothesis of the Diebold-Mariano test is MSFE $>$ 1 or MSFE $<$ 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

nowcasts and backcasts, when the model is able to use statistical data that pertain to the quarter of interest. Thus, its relative strength is to improve the assessment of the current state of the economy. Moreover, the DFM is generally more efficient in extracting monthly information in volatile times.

Second, the forecasting abilities of Consensus forecasts (averaged over the panelists) remarkably improve after the financial crisis, making them a tough competitor for the mechanical DFM since 2008. In the stable pre-crisis period, the DFM tends to outperform professional analysts. But in the volatile post-crisis period, Consensus nowcasts and forecasts constitute superior predictions at the moment of their release. This pattern suggests that analysts pay more attention and devote more effort to forecasting in volatile times. This outcome is consistent with the earlier finding that analysts quickly adjust their predictions during recessions or when the information set changes a lot. It also suggests that strictly mechanical procedures may be more susceptible to extreme observations in the estimation sample, even when taking averages across specifications and using rolling windows in estimation.

Third, the difference in forecasting performance between professional analysts and the DFM tends to be greater for a fresh Consensus nowcast (N3) than for a fresh Consensus one-quarter ahead forecast (F3). This suggests that the value added of subjective insights may be greater when analysts know at least some hard and soft data relating to the quarter of interest. In between the quarterly release dates, the performance of Consensus forecasts versus DFM forecasts deteriorates, as the latter are able to benefit from newly released monthly data.

Fourth, enhancing model-based forecasts with subjective information via simple combination schemes delivers sizable gains in forecasting ability of statistical models for all countries except Japan in the years 1999–2013, even when the Consensus forecasts are somewhat dated. Accuracy gains are often modest in the rather stable pre-crisis years 1999–2007, with DFM and Consensus forecasts contributing about equally to the combined forecast. The advantage of adding judgmental information are much more pronounced in the volatile period after 2008 due to a marked improvement in predictive power of Consensus forecasts. Towards the end of the sample, Consensus forecasts are the main determinant of the forecast combination, suggesting that the marginal information content of the DFM forecasts has become rather low for many countries. Consequently, the benefits from using a combination scheme when measured against the Consensus forecasts' performance have generally diminished over time, and are mostly small or absent, except for Italy and Japan since 2008.

Fifth, both the weighted average scheme and the linear combination scheme are suitable to enhance DFM forecasts in economically meaningful ways. Albeit some heterogeneity in forecasting capabilities of the weighting schemes, both across countries and time, the results suggest that in many cases forecasters are unlikely to be able to reliably identify in real-time what combination scheme they should apply. Thus, in practical applications, using a simple average of combination schemes may offer the best hedge against misspecification and instability of combination schemes.

These outcomes may be useful to policy makers, financial analysts and economic agents, as information on where the economy stands and where it is heading to in the short run is particularly valuable, especially in times of great uncertainty. The analysis in this chapter demonstrates that judgmental forecasts by professional analysts contain valuable additional information that can be extracted in real-time. This chapter does so in a mechanical fashion, but forecasters may in practice follow more flexible approaches to take this information on board. Moreover, forecasters may use Consensus forecasts, which represent the view of their peers, as a cross-check on their own model predictions and judgmental views on the near-term prospects of the economy.

The finding that for many countries (freshly released) Consensus forecasts are hard to improve upon since 2008 does not imply that practitioners are better off without making their own model-based forecasts. Model-based forecasts are still valuable for a number of reasons. First, Consensus forecasts are released just once a quarter. Especially in volatile times, when it really counts, they run the risk of becoming outdated rather quickly. Second, an important disadvantage of forecast surveys, such as the Consensus survey, is their black-box nature; if forecasts change, it is unclear for what reasons. As argued by Bańbura et al. (2011), a crucial aim of nowcasting is the structured way of updating forecasts on the basis of the continuous flow of new statistical data. In this way, the forecasting process generates information on the relative marginal information content of the various statistical indicators. Moreover, several types of models, including the DFM, allow the decomposition of forecasts into the contributions of the various (types of) statistical variables in the dataset (e.g. Bańbura and Rünstler, 2011), which may also assist analysts in their reading of economic conditions and their near-term development. These analytical by-products of statistical procedures may also serve to detect implausible aspects of model-based forecasts, which constitute crucial ingredients for the judgmental component of the forecast. For many forecasters, the story behind the forecast is at least as important as the number itself. Third, the expectations by the Consensus survey participants reflect both model-based and subjective information. As different forecasters will employ different models, estimation procedures and datasets, the model-based component differs across analysts. The strong performance of the mean Consensus forecast is partly attributable to the fact that different statistical approaches have been used to filter the available statistical data.

Appendix

4.A Data set

Table 4.7 provides an overview of the monthly indicators that have been used for the estimation of the DFMs. As discussed in the main text, the data can be split-up into three parts: the domestic economy, global economic activity and the main trading partners; see the headings in the table. Furthermore the data can be classified into four categories: hard, quantitative information (hard), consumer and producer prices (price), financial variables (financial) and soft, qualitative information (soft). Real-time vintages are collected for all time series, starting in January 1985, if possible.[19]

The main data source for the real-time database for the United States is the ALFRED-database, the US real-time database maintained by the Federal Reserve Bank of St. Louis. The real-time data vintages were used over the period January 1992–September 2014. The variables in the ALFRED-database are updated with each subsequent release of one of the series. Based on these release dates the OECD main economic indicators original data release and revisions database (OECD RTDB) is used, since it mimics the release pattern of the main data source for the other G7-countries.

The real-time database is augmented with indicators for global economic activity, financial variables and qualitative information on expectations derived from surveys among consumers, retailers and firms. Concerning the indicators for global economic activity real-time vintages on world trade from the CPB world trade monitor are used.[20] The other indicators of global activity and the financial variables are not subject to revisions. For these indicators the latest data vintage to construct backdated vintages based on the release pattern is used. The main source for survey data is the European Commission. Moreover, country-specific business survey data for Germany, France, Italy and Belgium are collected: the Ifo business climate index, the INSEE business cycle indicator, the ISAE consumer confidence indicator and the BNB business survey, respectively.

Quarterly GDP data for the US are taken from the ALFRED-database. For the other G7 countries, the OECD RTDB is the main source. As the latter contains no German GDP data before 1999.I, German GDP data before 1999.I are taken from the Deutsche Bundesbank.

[19] There are a few exceptions: the 10-year treasury bill-rate for Japan (January 1989), the 3-month treasury bill-rate for Japan (July 1985), negotiated wages for Germany (January 1990), the Baltic freight index (May 1985), the VIX standard and poor's 500-index (January 1986) and the crude West-Texas intermediate oil price (January 1986).

[20] The world trade monitor series start in January 1991 and were backdated for the period January 1985–December 1990, using monthly series on import- and export-volumes from the IMF.

Table 4.7: Description monthly dataset

Nr.	Variable name	Type	Source	CA	DE	FR	IT	JA	UK	US
Domestic Economy Variables										
1	Industrial production index	hard	RTDB, ALFRED	X	X	X	X	X	X	X
2	Construction production index	hard	RTDB	X	X	X	.	.	.	.
3	Personal consumption expenditure	hard	ALFRED	.	.	.	.	.	.	X
4	Permits issued for dwellings	hard	ALFRED	X	.	.	.	.	.	.
5	Housing starts	hard	ALFRED	.	.	.	.	.	.	X
6	New residential sales	hard	ALFRED	.	.	.	.	.	.	X
7	Real disposable personal income	hard	ALFRED	.	.	.	.	.	.	X
8	Unemployment rate	hard	RTDB, ALFRED	X	X	X	X	X	X	X
9	Employment	hard	RTDB, ALFRED	X	.	.	.	X	.	X
10	Machinery orders, private sector	hard	COJ	.	.	.	.	X	.	.
11	Manufacturers' new orders, durable goods	hard	ALFRED	.	.	.	.	.	.	X
12	Imports	hard	RTDB, ALFRED	X	X	X	X	X	X	X
13	Exports	hard	RTDB, ALFRED	X	X	X	X	X	X	X
14	Retail and food service sales	hard	RTDB, ALFRED	X	X	X	X	X	X	X
15	Passenger car registration	hard	MEI	X	X	X	X	X	X	X
16	Producer price index	price	MEI, ALFRED	X	X	.	.	X	.	X
17	Consumer price index, all urban consumers	price	RTDB, ALFRED	X	X	X	X	X	X	X
18	Hourly earnings manufacturing	price	RTDB	X	.	.	X	X	X	.
19	Negotiated wages and salaries per hour	price	Buba	.	X	.	.	.	.	.
20	Share price index	financial	MEI, ALFRED	X	X	X	X	X	X	X
21	10-year treasury bill rate	financial	RTDB, ALFRED	X	X	X	.	X	X	X
22	3-month treasury bill rate	financial	RTDB, ALFRED	X	X	X	X	X	X	X
23	Real effective exchange rate (CPI weights)	financial	MEI	X	X	X	X	X	X	.
24	Trade weighted exchange index, maj. curr	financial	ALFRED	.	.	.	.	.	.	X
25	M1	financial	MEI	X	X	X	X	X	X	.
26	Ifo business climate index	soft	Ifo	.	X	X	X	.	X	.
27	BNB business survey	soft	BNB	.	X	X	X	.	X	.
28	INSEE business cycle indicator	soft	INSEE	.	X	X	X	.	X	.
29	ISAE consumer confidence indicator	soft	ISAE	.	X	X	X	.	X	.
30	Composite leading indicator	soft	RTDB	X	X	X	X	X	X	X
31	Economic sentiment indicator	soft	EC	.	X	X	X	.	X	.

Continued on next page...

Table 4.7 – Continued

Nr.	Variable name	Type	Source	CA	DE	FR	IT	JA	UK	US
32	Industrial confidence indicator	soft	EC	.	X	X	X	.	X	.
33	Consumer confidence indicator	soft	EC	.	X	X	X	.	X	.
34	Retail trade confidence indicator	soft	EC	.	X	X	X	.	X	.
35	Construction confidence indicator	soft	EC	.	X	X	X	.	X	.
36	Index of business conditions	soft	COJ	.	.	.	.	X	.	.
37	General business conditions US	soft	Philadelphia FED	.	.	.	.	.	.	X
38	Conf. board consumer confidence index	soft	Conference Board	.	.	.	.	.	.	X
39	Bloomberg consumer comfort index	soft	Bloomberg	.	.	.	.	.	.	X
40	Purchasing manager index, manufact.	soft	Markit	.	.	.	.	.	.	X
Global Variables										
41	World trade	hard	CPB	X	X	X	X	X	X	X
42	Worldwide semiconductor sales	hard	SIA	X	X	X	X	X	X	X
43	Crude oil, west Texas intermediate	price	ALFRED	X	X	X	X	X	X	X
44	World market ind. mat. price	price	HWWA	X	X	X	X	X	X	X
45	World market metals price, metals	price	HWWA	X	X	X	X	X	X	X
46	Baltic freight index	price	BALTEX	X	X	X	X	X	X	X
47	VIX Standard and Poor's 500 index	financial	CBOE	X	X	X	X	X	X	X
Trading Partners										
48	CA: Industrial production index	hard	RTDB	.	.	.	.	.	.	X
49	CA: Imports	hard	RTDB	.	.	.	.	.	.	X
50	CA: OECD Composite leading indicator	soft	RTDB	.	.	.	.	.	.	X
51	CH: OECD Composite leading indicator	soft	RTDB	.	.	.	.	X	.	.
52	DE: Industrial production index	hard	RTDB	.	.	X	X	.	X	.
53	DE: Imports	hard	RTDB	.	.	X	X	.	X	.
54	DE: OECD Composite leading indicator	soft	RTDB	.	.	X	X	.	X	.
55	FR: Industrial production index	hard	RTDB	.	X	.	X	.	.	.
56	FR: Imports	hard	RTDB	.	X	.	X	.	.	.
57	FR: OECD Composite leading indicator	soft	RTDB	.	X	.	X	.	.	.
58	MX: Industrial production	hard	RTDB	.	.	.	.	.	.	X
59	MX: Imports	hard	RTDB	.	.	.	.	.	.	X
60	MX: OECD Composite leading indicator	soft	RTDB	.	.	.	.	.	.	X

Continued on next page...

Table 4.7 – Continued

Nr.	Variable name	Type	Source	CA	DE	FR	IT	JA	UK	US
61	US: Industrial production index	hard	ALFRED	X	X	X	.	X	X	.
62	US: Imports	hard	ALFRED	X	X	X	.	X	X	.
63	US: General business conditions	soft	Philadelphia FED	X	X	X	.	X	X	.
64	US: Conference board cons. conf.	soft	ALFRED	X	X	X	.	X	X	.
65	UK: Industrial production	hard	RTDB	X	.	.	.	.	.	.
66	UK: Imports	hard	RTDB	X	.	.	.	.	.	.
67	UK: OECD Composite leading indicator	soft	RTDB	X	.	.	.	.	.	.
N				32	39	37	35	30	37	36

Notes: entries denote variable number, name, category, data source and availability for each country in the dataset. Type: hard= quantitative information; financial= financial and monetary variables; price= consumer and producer prices; soft= qualitative information. Source: ALFRED: Archival federal reserve economic data; BNB: Banque nationale de Belgique; Buba: Bundesbank; CBOE: Chicago board options exchange; Baltex: Baltic exchange London; CPB: Cpb economic policy analysis; COJ: Cabinet office Japan; EC: European commission; HWWA: Hamburgisches welt-wirtschafts-archiv; Ifo: Ifo institute for economic research; INSEE: National institute of statistics and economic studies; ISAE: Italian institute for studies and economic analysis; RTDB: OECD real-time data and revisions database; MEI: OECD main economic indicators; SIA: Semiconductor industry association. Country: CA: Canada; DE: Germany; FR: France; IT: Italy; JA: Japan; UK: United Kingdom; US: United States.

4.B State space representation dynamic factor model

The equations of the DFM, Eqs. (4.1)–(4.4), can be cast in state space form as illustrated below for the case of $p = 1$. The aggregation rule is implemented in a recursive way in Eq. (4.8) by introducing a latent cumulator variable Ξ for which: $\Xi_t = 0$ for t corresponding to the first month of the quarter and $\Xi_t = 1$ otherwise. The monthly state space representation is given by the following observation equation:

$$\begin{bmatrix} x_t \\ y_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \xi_t \\ \varepsilon_t^Q \end{bmatrix} \tag{4.7}$$

and the transition equation:

$$\begin{bmatrix} I_r & 0 & 0 \\ -\beta' & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} f_{t+1} \\ y_{t+1} \\ \hat{y}_{t+1}^Q \end{bmatrix} = \begin{bmatrix} A_{r1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Xi_{t+1} \end{bmatrix} \begin{bmatrix} f_t \\ y_t \\ \hat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \varepsilon_t \\ 0 \end{bmatrix} \tag{4.8}$$

The application of the Kalman filter and smoother provides the minimum mean square linear estimates (MMSLE) of the state vector $\alpha_t = (f_t, y_t, \hat{y}_t^Q)$ and enables the forecasting of quarterly GDP growth y_t^Q and dealing efficiently with an unbalanced dataset of missing observations at the beginning and at the end of the series by replacing the missing data with optimal predictions. Moreover, when compared with using the principal components technique alone, the two-step estimator allows for dynamics of the common factors and cross-sectional heteroskedasticity of the idiosyncratic component.

4.C Additional results

Table 4.8 is a counterpart of Table 4.2 in the main text. It presents relative RMSFEs versus the RW for two different set-ups. Panel B reports results for the real-time procedure when the prediction errors are computed using the last vintage for GDP as measure of actual GDP. Panel A reports results for the corresponding pseudo real-time procedure, which also implies that the last vintage for GDP serves as the measure of actual GDP. Tables 4.9 and 4.10 are the respective counterparts of Tables 4.5 and 4.6 in the main text, where the relative RMSFE is expressed in terms of the RMSFE of the Consensus forecast. Finally, Table 4.11 reports the estimated coefficients of Eq. (4.6), which display the same pattern as the ones for Eq. (4.5).

Table 4.8: Forecasting performance of dynamic factor model in alternative set-ups, 1999.I–2013.IV

	RW	B2	B1	N3	N2	N1	F3	Flash	B2	B1	N3	N2	N1	F3
	RMSFE	Relative RMSFE DFM versus Random Walk						RMSFE	Relative RMSFE DFM versus Random Walk					
		A. Pseudo real-time							B. Final vintage is target					
		Full sample 1999.I–2013.IV							Full sample 1999.I–2013.IV					
Canada	*0.66*	**0.60***	**0.64***	**0.72***	**0.76***	**0.82**	**0.88**	*0.27*	**0.63***	**0.66***	**0.72***	**0.78***	**0.82**	**0.88**
France	*0.55*	**0.57****	**0.61****	**0.67****	**0.73***	**0.77***	**0.84***	*0.22*	**0.60****	**0.66****	**0.73****	**0.77***	**0.80***	**0.86***
Germany	*0.90*	**0.63***	**0.67***	**0.75***	**0.79***	**0.84***	**0.88***	*0.39*	**0.70***	**0.72***	**0.79***	**0.84**	**0.84***	**0.89***
Italy	*0.78*	**0.67***	**0.68***	**0.77***	**0.79***	**0.74***	**0.81***	*0.33*	**0.62***	**0.66***	**0.74***	**0.77***	**0.78****	**0.86****
Japan	*1.12*	**0.69***	**0.73***	**0.80***	0.96	1.05	1.07*	*0.74*	**0.70***	**0.74***	**0.81**	0.96	1.09	1.08
UK	*0.82*	**0.74**	**0.77**	**0.80**	**0.85**	0.91	0.92	*0.43*	**0.82**	**0.83**	**0.86**	0.91	0.93	0.94
US	*0.70*	**0.63***	**0.62***	**0.69***	**0.78**	**0.85**	0.90	*0.36*	**0.67***	**0.70***	**0.75***	**0.83**	**0.88**	0.92
		Pre-crisis period 1999.I–2007.IV							Pre-crisis period 1999.I–2007.IV					
Canada	*0.45*	**0.78***	**0.81***	**0.87**	**0.89**	0.97	0.98	*0.24*	**0.81****	**0.84***	**0.86***	0.92	0.96	0.97
France	*0.36*	**0.63*****	**0.68*****	**0.66*****	**0.69*****	**0.73*****	**0.84****	*0.21*	**0.66*****	**0.70*****	**0.68*****	**0.69*****	**0.73*****	**0.86****
Germany	*0.59*	**0.79*****	**0.81*****	**0.85*****	**0.88*****	**0.87****	0.90**	*0.44*	**0.84*****	**0.85*****	0.91**	0.94	0.90**	0.93*
Italy	*0.45*	**0.86*****	**0.87****	0.90*	0.90*	**0.85*****	**0.89****	*0.32*	**0.88*****	**0.87****	**0.89****	**0.88****	**0.85*****	**0.89****
Japan	*0.68*	0.99	0.97	0.94	0.96	0.96	0.96*	*0.85*	1.01	1.00	0.96	0.95	0.94	0.96
UK	*0.46*	1.09	1.06	1.01	1.01	1.00	0.96	*0.47*	0.97	0.96	0.96	0.96	0.95	0.96
US	*0.52*	**0.79****	**0.79****	**0.83****	0.90*	0.91**	0.95	*0.35*	**0.80****	**0.81*****	**0.84****	0.92*	0.93*	0.96
		Post-crisis period 2008.I–2013.IV							Post-crisis period 2008.I–2013.IV					
Canada	*0.89*	**0.52***	**0.57***	**0.65**	**0.71**	**0.75**	**0.83**	*0.31*	**0.54**	**0.57**	**0.66**	**0.71**	**0.76**	**0.84**
France	*0.75*	**0.54***	**0.58***	**0.67***	**0.75**	**0.78**	**0.84**	*0.22*	**0.58***	**0.65***	**0.74**	**0.80**	**0.83**	**0.86**
Germany	*1.22*	**0.56***	**0.61***	**0.71***	**0.76***	**0.83***	**0.87***	*0.29*	**0.64***	**0.66***	**0.74***	**0.80***	**0.82***	**0.87***
Italy	*1.11*	**0.61***	**0.63***	**0.73***	**0.76***	**0.72***	**0.79***	*0.33*	**0.53***	**0.60***	**0.69***	**0.74***	**0.76****	**0.85****
Japan	*1.57*	**0.58****	**0.64****	**0.75***	0.96	1.07*	1.10***	*0.53*	**0.58****	**0.65***	**0.77**	0.96	**1.12***	**1.11****
UK	*1.16*	**0.63**	**0.68**	**0.74**	**0.81**	**0.89**	0.92	*0.37*	**0.77**	**0.80**	**0.84**	0.90	0.92	0.94
US	*0.91*	**0.53**	**0.52**	**0.60**	**0.72**	**0.83**	**0.87**	*0.38*	**0.60**	**0.63**	**0.70**	**0.79**	**0.85**	0.91

Notes: entries in italics denote RMSFEs; column labeled "RW" denotes a Random Walk for the third nowcast, column labeled "Flash" denotes the RMSFE of the flash as predictor of the final vintage. All other entries report the RMSFE relative to random walk's RMSFE. Entries in bold denote a deviation by more than $\pm$ 10% from 1. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. The alternative hypothesis of the Diebold-Mariano test is MSFE > 1 or MSFE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

Table 4.9: Forecasting performance of forecast combination schemes (relative to Consensus forecast), 1999.I–2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus CF											
	A. Weighted average scheme						B. Linear combination scheme					
	Full sample 1999.I–2013.IV						Full sample 1999.I–2013.IV					
Canada	0.90	0.90	0.94	**0.88***	0.92	0.94	0.90	0.92	0.98	**0.83****	**0.89***	0.90
France	0.90**	0.97	1.00	0.92	0.92*	0.97*	0.93*	1.01	1.03	1.00	0.99	0.99
Germany	1.02	1.02	1.00	1.00	0.98	1.00	0.92	0.91	0.98	0.99	1.00	1.03
Italy	**0.89**	0.99	1.07*	0.97	0.93	1.00	**0.74**	**0.85**	**0.88**	0.92	**0.86****	0.98
Japan	**0.83***	**0.88**	0.96	0.96	1.09	1.01	**0.85**	0.92	1.00	1.02	1.10	**1.13***
UK	1.01	1.01	1.02	0.99	0.99	0.99	1.05	1.06	1.08	0.98	0.93	0.92*
US	0.95	1.00	1.05**	0.95	1.01	1.02	1.01	1.10***	**1.17****	1.05*	**1.14*****	**1.17*****
	Pre-crisis period 1999.I–2007.IV						Pre-crisis period 1999.I–2007.IV					
Canada	**0.77****	**0.80***	**0.88**	**0.78****	**0.79****	**0.80****	**0.74**	**0.80**	0.92	**0.76****	**0.78****	**0.79***
France	**0.89*****	0.93	0.91**	**0.84*****	**0.83*****	0.93**	0.93*	0.96	0.95*	0.90***	0.90**	1.00
Germany	0.99	0.97	0.98	0.99	0.94	0.92	1.01	0.97	1.02	0.93	**0.88**	**0.87**
Italy	0.96	0.98	0.99	0.99	**0.85***	**0.89**	0.97	0.98	1.00	0.96	**0.86***	**0.89**
Japan	0.95*	0.96*	0.96*	0.99	0.99	1.00	0.99	1.01	1.01	1.06*	1.05	1.06*
UK	**0.84****	**0.80*****	**0.82*****	**0.78*****	**0.71*****	**0.74*****	0.90	**0.89**	0.90	**0.81*****	**0.73*****	**0.74*****
US	0.95	1.00	1.06**	0.94	0.99	1.01	1.00	1.10**	**1.16***	1.06*	**1.15****	**1.17****
	Post-crisis period 2008.I–2013.IV						Post-crisis period 2008.I–2013.IV					
Canada	**1.16**	**1.11***	1.08*	1.03	1.07*	**1.11***	**1.23***	**1.20****	**1.12****	0.92	1.01	1.03
France	0.91	1.01	1.08	0.96	0.98	1.00	0.94	1.05	**1.11*****	1.06	1.06	0.98
Germany	1.03*	1.03***	1.00	1.00	1.00	1.01	0.91	0.90	0.97	1.00	1.03	1.06*
Italy	**0.83**	1.00	**1.12*****	0.97	0.97	1.06	**0.54**	**0.76**	**0.80**	0.90	**0.86**	1.03
Japan	**0.67***	**0.78**	0.96	0.95	**1.15**	1.02	**0.67**	**0.81**	1.00	0.99	**1.14**	**1.16***
UK	1.05	1.06	1.06	1.03	1.05	1.04	1.09*	1.10*	**1.12****	1.01	0.98	0.95
US	0.95	1.00	1.03*	0.96	1.02	1.03*	1.04	**1.11****	**1.18***	1.05	**1.14*****	**1.16*****

Notes: entries in bold denote a deviation by more than $\pm$ 10% from 1. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. The alternative hypothesis of the Diebold-Mariano test is MSFE $>$ 1 or MSFE $<$ 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

Table 4.10: Forecasting performance of averaged or selected forecast combinations (relative to Consensus forecast), 1999.I–2013.IV

	B2	B1	N3	N2	N1	F3	B2	B1	N3	N2	N1	F3
	Relative RMSFE combination DFM and CF versus CF											
	A. Average of combined forecasts						B. Selection of best combined forecast in the recent past					
	Full sample 1999.I–2013.IV						Full sample 1999.I–2013.IV					
Canada	**0.87**	**0.89**	0.94	**0.83****	**0.89***	0.90	**0.82**	**0.87**	0.94	**0.83****	**0.87****	**0.87****
France	**0.89****	0.97	0.99	0.94	0.94	0.97	0.93*	1.01	1.04	0.95	0.96	1.01
Germany	0.95	0.94	0.97	0.98	0.98	1.00	0.95	1.04	1.04	1.04	0.99	1.03
Italy	**0.77**	**0.88**	0.94	0.91	**0.86****	0.97	**0.72**	**0.88**	0.90	0.91	**0.85****	0.99
Japan	**0.83***	**0.89**	0.98	0.98	1.09	1.06*	**0.87**	0.92	0.99	1.02	**1.11**	**1.11**
UK	1.01	1.01	1.02	0.95**	0.94**	0.94***	1.03	1.01	1.04	0.97	0.92*	0.92*
US	0.96	1.03	1.08**	0.98	1.06***	1.08***	0.98	1.04	**1.12****	0.98	1.03**	1.06***
	Pre-crisis period 1999.I–2007.IV						Pre-crisis period 1999.I–2007.IV					
Canada	**0.72***	**0.78**	**0.88**	**0.76****	**0.77****	**0.78****	**0.69****	**0.77**	0.91	**0.78****	**0.81****	**0.79****
France	0.90**	0.94	0.92**	**0.86*****	**0.86*****	0.96	0.91*	0.98	0.95	**0.88*****	**0.87*****	0.98
Germany	0.96	0.94	0.98	0.93	**0.87***	**0.84***	0.99	0.97	1.01	0.91	**0.86**	**0.84***
Italy	0.95**	0.96**	0.97	0.95**	**0.83****	**0.86***	0.94**	0.98	1.01	0.97	**0.86***	**0.88***
Japan	0.96	0.98	0.98	1.02	1.01	1.03	0.99	1.00	0.98	1.02	1.00	1.03
UK	**0.86****	**0.84****	**0.85****	**0.79*****	**0.71*****	**0.73*****	**0.87****	**0.84****	**0.84****	**0.81*****	**0.73*****	**0.74*****
US	0.95	1.01	1.07*	0.98	1.06**	1.08***	0.98	1.06**	**1.13****	0.96	1.04*	1.07**
	Post-crisis period 2008.I–2013.IV						Post-crisis period 2008.I–2013.IV					
Canada	**1.18**	**1.13***	1.09*	0.93	1.02	1.05*	1.10***	1.10*	1.02	0.90	0.94	0.97
France	**0.88**	1.00	1.05*	0.99	1.00	0.98	0.95	1.04	**1.11****	1.00	1.03	1.03
Germany	0.94	0.94	0.97*	0.99	1.00	1.03	0.94	1.05*	1.05	1.06*	1.02	1.07*
Italy	**0.64**	**0.82**	0.91	0.90	**0.88**	1.03	**0.54**	**0.81**	**0.81**	**0.88**	**0.84***	1.05
Japan	**0.66**	**0.79**	0.97	0.96	**1.14**	1.07	**0.71**	**0.83**	0.99	1.02	**1.16***	**1.16***
UK	1.04*	1.05*	1.06**	0.99	0.99	0.98	1.06*	1.04	1.08	1.00	0.96	0.96
US	0.98	1.05*	1.10*	0.98	1.07**	1.08**	0.98	0.99	1.09	0.99	1.02	1.05***

Notes: entries in bold denote a deviation by more than ± 10% from 1. *, ** or *** denotes the Diebold-Mariano test is significant at the 10%, 5% or 1% level, respectively. The alternative hypothesis of the Diebold-Mariano test is MSFE > 1 or MSFE < 1, depending on whether the RMSFE is greater or smaller than 1, respectively.

Table 4.11: Encompassing test: coefficients of linear combination of Consensus forecasts (β) and DFM (γ), 1999.I–2013.IV

		B2	B1	N3	N2	N1	F3
		Full sample 1999.I–2013.IV					
Canada	β	0.65***	0.74***	0.83***	0.38	0.56**	0.65**
	γ	0.35*	0.27*	0.22	0.97***	0.71***	0.77**
France	β	0.69***	0.76***	0.82***	0.66***	0.59***	0.59***
	γ	0.17	0.09	0.02	0.25**	0.32**	0.27*
Germany	β	1.15***	1.26***	1.47***	0.90*	0.92	1.01
	γ	0.49*	0.40	0.17	0.63**	0.71**	0.53*
Italy	β	0.49***	0.71***	0.71***	0.63***	0.50***	0.69***
	γ	0.77***	0.63***	0.75***	0.84***	1.05***	0.73**
Japan	β	0.32	0.33	0.52**	0.36	0.40	0.51
	γ	0.89***	0.76***	0.69**	0.67	0.24	0.04
UK	β	1.06***	1.05***	1.12***	1.37***	1.40***	1.40***
	γ	0.27	0.30	0.20	0.49**	0.42**	0.41**
US	β	0.57***	0.79***	1.10***	0.46**	0.77***	0.90**
	γ	0.49***	0.28	-0.11	0.70***	0.37	0.21
		Pre-crisis period 1999.I–2007.IV					
Canada	β	-0.25	-0.02	0.26	-0.28	-0.43	-0.51
	γ	0.94***	0.79***	0.53**	0.64**	0.69**	0.75*
France	β	0.58***	0.73***	0.77***	0.62*	0.45	0.33
	γ	0.28	0.12	0.05	0.13	0.19	0.36
Germany	β	0.72***	0.70***	0.80***	0.33	0.17	0.07
	γ	0.20	0.29	0.03	0.36	0.60**	0.78**
Italy	β	0.33	0.44*	0.50*	0.50*	0.08	0.15
	γ	0.67	0.53	0.43	0.34	0.92**	0.67
Japan	β	0.60*	0.64*	0.60*	0.79*	0.75	0.65
	γ	0.75*	0.71	0.70	0.52	0.64	-0.12
UK	β	0.41***	0.44***	0.46***	0.53***	0.40**	0.35***
	γ	0.51***	0.54***	0.54***	0.67***	0.85***	0.91***
US	β	0.35	0.63	1.09**	-0.02	0.38	0.65
	γ	0.74*	0.42	-0.19	0.82**	0.30	-0.18
		Post-crisis period 2008.I–2013.IV					
Canada	β	1.13***	1.18***	1.13***	0.85**	1.19***	1.27***
	γ	-0.07	-0.13	-0.10	0.99**	0.38	0.49
France	β	0.88*	1.00**	1.21***	0.40	0.30	0.32
	γ	0.09	0.01	-0.13	0.30*	0.38*	0.31*
Germany	β	2.19***	2.38***	2.65***	2.43*	2.56*	3.03*
	γ	0.11	-0.04	-0.28	0.45	0.40	0.08
Italy	β	0.35	0.74**	0.55*	0.24	0.31	0.57
	γ	0.79***	0.60**	0.84**	0.98***	1.08***	0.74*
Japan	β	0.05	-0.02	0.51	-0.14	0.28	0.68
	γ	1.02***	0.94**	0.65	0.73	0.14	-0.22
UK	β	1.16***	1.13***	1.24***	1.54***	1.62***	1.63***
	γ	0.16	0.20	0.06	0.44*	0.27	0.23
US	β	0.41*	0.66***	1.09***	0.54**	0.72*	0.83*
	γ	0.58***	0.38*	-0.06	0.69***	0.43	0.36

Notes: entries denote the estimated coefficients of Consensus forecasts (β) and the dynamic factor model (γ) in the (unconstrained) encompassing test in Eq. (4.6), respectively. *, ** or *** denotes that the estimated coefficient is statistically different from 0 at the 10%, 5% or 1% significance level, respectively.

Chapter 5

Forecasting and nowcasting economic growth in the euro area using factor models

Many empirical studies have provided evidence that the use of factor models, which use large datasets of economic variables, can contribute to the computation of more accurate forecasts. In this study, we examine the performances of four different factor models in a pseudo real-time forecasting competition for the euro area and five of its largest countries. The aim is to identify empirically the best factor model approach for forecasting and nowcasting of the quarterly gross domestic product growth rate. Besides, this chapter proposes some modifications of existing factor model specifications, with the aim of improving their forecasting performances empirically. The main conclusion of this chapter is that factor models consistently outperform the benchmark autoregressive model, both before and during the crisis. Moreover, the highest forecasting accuracy is generally produced by the collapsed dynamic factor model.[1]

KEYWORDS: Factor models; State space method, Forecasting competitions.

---

[1] This chapter is co-authored by Irma Hindrayanto and Siem Jan Koopman. This chapter was published in the International Journal of Forecasting, 32, 2184-1305, Hindrayanto, Koopman and de Winter, Copyright Elsevier (2016). Available online at `http://dx.doi.org/10.1016/j.ijforecast.2016.05.003`. We are grateful to Ide Kearney, Job Swank, an anonymous editor, an anonymous associate editor, three anonymous referees, seminar and conference participants at De Nederlandsche Bank, the Conference on Computational and Financial Econometrics (2012, Oviedo) and the International Symposium on Forecasting (2014, Rotterdam) for their helpful comments. An early version of this chapter was circulated as DNB Working Paper 415 and Tinbergen Institute Discussion Papers 14-113/III under the title "Nowcasting and forecasting economic growth in the euro area using principal components".

5.1 Introduction

It is acknowledged widely that the forecasting of macroeconomic time series is of critical importance, both for economic policy makers and for the general public. Reliable short-term forecasts are in particularly high demand when the economic environment is uncertain. Many different methodologies for producing such forecasts exist, ranging from basic time series models to sophisticated structural dynamic models. Over the last decade, dynamic factor models have become a popular tool for short-term forecasting amongst both practitioners and econometricians, due to their good forecasting performances in many studies; see for example, Giannone et al. (2008) and Stock and Watson (2002b) for the United States, Angelini et al. (2011) and Rünstler et al. (2009) for the euro area and Schumacher and Breitung (2008) for Germany. In all empirical studies concerning dynamic factor models, there are various decisions that need to be made before forecasting can start. We provide three examples. First, the optimal number of factors in the model needs to be determined by following procedures such as those of Ahn and Horenstein (2013), Alessi et al. (2010), Bai and Ng (2002), Hallin and Liška (2007) and Onatski (2010). Second, the selection of the database for extracting the factors, and its size, are important determinants of a successful forecasting procedure; see for example the discussions by Boivin and Ng (2005), Caggiano et al. (2011) and den Reijer (2013). Third, the number of lagged terms of the target variable in the forecasting model needs to be set. The gain in forecasting accuracy from including one or more lags of the target variable in the forecast equation has not been documented well. However, recent studies indicate that including more autoregressive terms may increase the forecasting accuracy; see for example Clements and Galvão (2008), Kuzin et al. (2011) and Chapter 3 of this thesis. It is an empirical question as to whether this finding holds for all factor model specifications. Such matters have also been discussed in related empirical studies; see for example, Bańbura et al. (2011), Jungbacker and Koopman (2015), Lahiri and Monokroussos (2013), Liebermann (2014) and Matheson (2013).

This chapter compares the short-term forecasting performances of different factor models for quarterly gross domestic product (GDP) growth in the euro area and its five largest countries, before, during and after the financial crisis. The one- to three-month ahead forecasts for the current quarter are referred to as nowcasts. The short-term forecasting of key economic variables using dynamic factor analysis has been reviewed by Bai and Ng (2008), Breitung and Eickmeier (2006), and Stock and Watson (2011), with Luciani (2014) discussing more recent contributions. Four estimation procedures for the dynamic factor model are considered: the basic principal components method of Stock and Watson (2002b), who initiated the current literature on factor models; the widely used two-step approach of Doz et al. (2011); the more elaborate quasi-maximum likelihood method of Doz et al. (2012); and the more recently proposed maximum likelihood method of Bräuning and Koopman (2014), based on a collapsed dynamic factor model. All of these estimation approaches rely to some extent on principal components that summarize the information in a large set of monthly indicators. The estimation methods proposed by Bräuning and Koopman (2014) and Doz et al. (2011) use the principal components as approximations of the dynamic factors. Doz et al. (2012) use the principal components as an initialisation of quasi-maximum likelihood estimation. The Kalman filter plays a key role in all three of these approaches.

For all dynamic factor approaches, the target variable and the common factors are analyzed simultaneously in a multivariate unobserved component time series model. All modeling frameworks allow for panels with mixed-frequencies and with the monthly time series having different publication delays and starting dates. This leads to a data matrix of monthly time series with so-called "ragged" edges at the beginning and end of the sample. The two-step approach developed by Doz et al. (2011) was applied to the euro area by Angelini et al. (2011) and Bańbura and Rünstler (2011). The first step involves the computation of the principal components and the estimation of their dynamic properties by means of a vector autoregressive model. The second step involves obtaining the factor estimates and forecasts from the Kalman filter and smoother. Doz et al. (2011) provide the asymptotic properties of the factor estimates and use the model to forecast the quarterly GDP growth using monthly variables that contain ragged edges at the beginning and end of the sample. Bańbura and Rünstler (2011) developed this approach further by including the quarterly GDP growth as a latent variable in the state vector, so that the contributions of different variables to the forecasts can be quantified using the algorithms of Koopman and Harvey (2003). The quasi-maximum likelihood approach of Doz et al. (2012) was applied to the euro area by Bańbura et al. (2011) and Bańbura and Modugno (2014). It is shown that this approach obtains consistent estimates of the factors as the size of the cross-section goes to infinity. Bańbura and Modugno (2014) extend the framework of Doz et al. (2012) by introducing modifications in relation to missing entries (at random) and the dynamic treatment of idiosyncratic effects; see also Luciani (2014).

The collapsed dynamic factor model of Bräuning and Koopman (2014) effectively adopts a low-dimensional unobserved components time series model for both the target variable and a set of principal components. This multivariate model is then used to forecast the target variable based on its past realizations and the principal components. The idiosyncratic part of the target variable is modeled explicitly and dealt with jointly with the dynamic factors. It mitigates the challenge of estimating the factors and forecasting the target variable in a joint analysis based on a large macroeconomic panel. The unknown parameters in this parsimonious model are estimated by maximum likelihood: the loglikelihood function is evaluated by the Kalman filter and maximized numerically with respect to the unknown parameters. The score function can be evaluated using a corresponding smoothing algorithm. The forecasts of the target variable are generated by the Kalman filter.

The main contributions of this chapter are twofold. First, small modifications for the different estimation approaches are proposed, with the aim of placing them on a somewhat more equal footing. For example, the model of Doz et al. (2011) is extended by including more autoregressive terms, as in Bräuning and Koopman (2014) and Stock and Watson (2002b). Besides, an alternative –more effective– way of handling the ragged edges for the collapsed dynamic factor method of Bräuning and Koopman (2014) is proposed, consisting of three steps, i.e.: (*i*) analyze each univariate time series by using an unobserved components model to extract the main signal for imputing the ragged edges; (*ii*) extract the principal components; (*iii*) estimate the parameters simultaneously. This handling of the ragged edges improves the forecasting accuracy. Second, a empirical study is conducted to compare the forecasting accuracy of the different modeling treatments for the euro area and its five largest countries. One of the main conclusions is that the investigated factor modeling approaches systematically produce more accurate forecasts than those of the benchmark autoregressive model. This good performance is not limited to the pre-financial crisis period: the factor models also outperform the benchmark model during and after the financial crisis by up to 77%, in terms of mean squared forecast errors, depending on the factor model, country and projection horizon. Overall, the collapsed dynamic factor approach is the most accurate model for forecasting and nowcasting in the empirical study.

The remainder of this chapter is organized as follows. Section 5.2 presents an overview of the four different dynamic factor model approaches considered, and discusses some possible modifications. Section 5.3 provides details of the construction of the database, the forecast setup and the model specification, together with selection details such as the number of common factors and lags. Section 5.4 discusses the empirical results. Finally, Section 5.5 summarizes the main findings.

5.2 Dynamic factor model approaches

This section considers four existing dynamic factor approaches to nowcasting and forecasting: (i) the autoregressive model for the target variable, with lagged principal components as covariates, proposed by Stock and Watson (2002b); (ii) the dynamic factor approach of Doz et al. (2011), as implemented by Bańbura and Rünstler (2011); (iii) the quasi-maximum likelihood approach for dynamic factor models of Doz et al. (2012), as implemented by Bańbura and Modugno (2014); and (iv) maximum likelihood estimation for the collapsed dynamic factor model of Bräuning and Koopman (2014).

All of these dynamic factor model approaches use principal components in their forecasting procedures, but they do so in different ways. Stock and Watson (2002b) use the principal components as covariates in the autoregressive model of the target variable. Bańbura and Modugno (2014) and Bańbura and Rünstler (2011) use principal components as proxies for the dynamic factors which facilitate the estimation of the many parameters in their large-dimensional state space model[2]. Bräuning and Koopman (2014) adopt a low-dimensional state space model which deals with the target variable and the principal components jointly. Their parsimonious representation allows standard maximum likelihood estimation.

The focus is on forecasting the quarterly GDP growth (quarter on quarter), denoted by $y^Q_{t_q}$, where $t_q = 1, \ldots, T_q$ is the quarterly time index. Following the statistical convention, the quarterly GDP growth rate at the monthly frequency y^M_t is set equal to the growth rate ($y^Q_{t_q}$) in the third month of each quarter ($t = 3t_q$) and to a missing value otherwise, where $t = 1, \ldots, T$ is the monthly time index. The time dimensional relation is $T_q = \lfloor T/3 \rfloor$. The latent monthly GDP growth rate, y_t, is the 3-month growth rate with respect to the corresponding month of the previous quarter, and y^*_t as the mean-adjusted series of y_t, that is $y^*_t = y_t - \mu$ where μ is the in-sample mean of $y^Q_{t_q}$.

The remainder of this section describes the forecasting procedures based on factor models. All procedures use a monthly time series of $r \times 1$ vectors of principal components, which are denoted F_t. The principal components are obtained from a N-dimensional standardized stationary monthly time series of candidate predictors, X_t, for $t = 1, \ldots, T$. The matrix of eigenvalues (or factor loadings) is denoted as Λ. The vector $F^Q_{t_q}$ contains the r quarterly factors, calculated by taking the three-month averages of F_t.

5.2.1 Principal components approach

Stock and Watson (2002b) designed a method for the forecasting of a single time series of length T, using a large number N of candidate predictor series, where typically $N >> T$. This high-dimensional problem is reduced to an univariate autoregressive model for the key economic time series of interest through the inclusion of a small number of principal components that are used as predictors. The autoregressive model is for the target variable with a specific forecast horizon h. The forecasts of the target variable $y^Q_{t_q+h}$ are based on the current and past values $F^Q_{t_q}, y^Q_{t_q}, F^Q_{t_q-1}, y^Q_{t_q-1}, \ldots$. The forecasting model is given by the dynamic model:

$$y^Q_{t_q+h} = \alpha_h + \sum_{j=0}^{m} \beta_{h,j} F^Q_{t_q-j} + \sum_{k=0}^{n} \gamma_{h,k} y^Q_{t_q-k} + \varepsilon^Q_{t_q+h}, \qquad t_q = 1, \ldots, T_q, \tag{5.1}$$

where α_h is the constant term, $\beta_{h,j}$ and $\gamma_{h,k}$ are regression coefficients, for $j = 0, \ldots, m$

[2] These models adopt a two-step procedure in which the estimates based on principal components are used as starting values for the quasi-maximum likelihood estimation for which the Kalman filter is used.

and $k = 0, \ldots, n$, and $\varepsilon^Q_{t_q+h}$ is the disturbance. The lag dimensions m and n are assumed to be set *a-priori*. It is common practice to set both dimensions m and n equal to two.

All parameters are subject to the forecasting horizon h. Although the model remains the same, it is assumed that the coefficients of the model may differ for each forecasting horizon h. Hence, Stock and Watson's procedure re-estimates the model coefficients for each forecast horizon h while keeping the selection of the explanatory variables fixed. The principal components F_t are obtained from a balanced sub-sample of covariates X_t, which is obtained by discarding the rows that have missing values at the end of the estimation period. Typically, this only involves removing the last few rows that are not complete due to publication delays. The missing values at the beginning of the sample are dealt with using the expectation maximization (EM) algorithm; see Appendix A of Stock and Watson (2002b).

Forecasting is then carried out following a two-step procedure: first, the factors (or principal components) are obtained from the set of candidate predictors; second, the parameters of the autoregressive model are estimated using the regression method (ordinary least squares), from which the forecasts can then be generated. This allows the easy computation of any forecast, from $y^Q_{T_q+1}$ to $y^Q_{T_q+h}$, for some $h > 1$; however, the two-step procedure must be repeated for each forecast horizon.

5.2.2 Two-step approach

The forecasting procedure of Bańbura and Rünstler (2011) is based on the dynamic factor model of Giannone et al. (2008) and the two-step estimation approach of Doz et al. (2011). The treatment is based on the state space framework, which is particularly convenient for nowcasting, as the framework makes it easy to deal with variables with missing values (at the beginning and end of the sample) and variables with different data frequencies. The first step involves carrying out principal component analysis to produce best-guess estimates of the latent factors and factor loadings. Other parameters in the model are estimated via standard regressions. The second step uses the Kalman filter and smoother to conduct the treatment of missing values, the computation of the forecasts, and the computation of the final smooth estimates of the latent factors, by casting the complete model in state space form. When the two-step method is repeated based on the smoothed estimates of the factors, the parameter estimates converge to the maximum likelihood estimates; see Doz et al. (2011). The Bańbura and Rünstler (2011) model is given by:

$$X_t = \Lambda f_t + u_t, \qquad u_t \sim \mathrm{NIID}(0, \Sigma_u), \tag{5.2}$$

$$f_t = \sum_{j=1}^{p} \Phi_j f_{t-j} + \zeta_t, \qquad \zeta_t \sim \mathrm{NIID}(0, \Sigma_\zeta), \tag{5.3}$$

for $t = 1, \ldots, T$, where Λ is the loading matrix, f_t is a $r \times 1$ vector of latent dynamic factors, u_t is a normally, identically and independently distributed (NIID) disturbance with mean zero and variance matrix Σ_u, Φ_j is the autoregressive coefficient matrix, for $j = 1, \ldots, p$, and ζ_t is a NIID disturbance with mean zero and variance matrix Σ_ζ, and the two vector disturbance series u_t and ζ_s are mutually independent of each other for all combinations of $t, s = 1, \ldots, T$. The variance matrix Σ_u is typically assumed to be diagonal. The latent dynamic stochastic process for f_t is modeled explicitly as a stationary vector autoregressive process with lag dimension p. The time index t refers to months.

The dynamic factor model is presented in state space form, which facilitates the treatment of missing values, the imputation of ragged edges and the computation of the forecasts via the Kalman filter and smoother. Bańbura and Rünstler (2011) argue that exploiting the dynamics of the estimated latent factors directly can be beneficial in improving the forecasting accuracy. However, they recommend that factors should not be very noisy. To ensure some smoothness in the factors f_t, the rank of matrix Σ_ζ can be reduced further, to $q \leq r$.

The values for the unknown parameter matrices Λ, Σ_u, $\Phi_1, \ldots, \Phi_p$, and Σ_ζ are determined from the r principal components F_t, as outlined by Giannone et al. (2008). The principal components F_t are based on the eigendecompostion of the sample variance matrix of the data $[X_1, \ldots, X_T]'$, denoted by the $N \times N$ positive definite matrix S_X. The columns of the $N \times r$ loading matrix Λ in Eq. (5.2) are set equal to the r eigenvectors associated with the r largest eigenvalues of S_X, respectively. The variance matrix Σ_u is set to be a diagonal matrix, with the ith diagonal element equal to the (i, i) element of the sample variance matrix $[(X_1 - \Lambda F_1), \ldots, (X_T - \Lambda F_T)]'$. The matrix parameters $\Phi_1, \ldots, \Phi_p$, and Σ_ζ in Eq. (5.3) are set equal to their corresponding regression estimates, obtained from the vector autoregressive model in Eq. (5.3), for which f_t is replaced by F_t, for $t = 1, \ldots, T$.

The incorporation of the quarterly target series $y^Q_{t_q}$ in the monthly state space model in Eqs. (5.2)–(5.3) is required for its forecasting. For this purpose, the procedure follows Mariano and Murasawa (2003) in the forecasting of mean-adjusted quarterly GDP growth in a mixed-frequency modeling framework. Denote the univariate mean-adjusted latent monthly variable as y^*_t and model it by:

$$y^*_t = \beta' f_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NIID}(0, \sigma^2_\varepsilon), \qquad t = 1, \ldots, T, \tag{5.4}$$

where β is an $r \times 1$ vector of coefficients and ε_t is a NIID disturbance that is mutually independent of u_t and ζ_t at all time index combinations. The link with y^*_t and the observed quarterly GDP growth rate $y^Q_{t_q}$ is established by creating a monthly time series y^M_t of missing values except at time $t = 3t_q$, where it is set equal to $y^Q_{t_q}$. A recursive latent cumulator variable y_t^{*C} is generated by:

$$y_{t+1}^{*C} = \delta_t y_t^{*C} + \frac{1}{3} y_{t+1}^{*}, \qquad \delta_t = \begin{cases} 0, & t = 3t_q, \\ 1, & \text{otherwise}, \end{cases} \tag{5.5}$$

for $t = 1, \ldots, T$ and $t_q = 1, \ldots, T_q$, with the cumulator variable being initialized as $y_1^{*C} = \frac{1}{3} y_1^*$. By construction, for $t = 3t_q$, y_t^{*C} equals the average of the latent monthly series y_t^* within quarter t_q, which again equals to the mean-adjusted quarterly growth rate of that quarter. Since y_t^M should be equal to the observed quarterly growth rate $y_{t_q}^Q$ for $t = 3t_q$, the mean of $y_{t_q}^Q$ is added back into y_t^{*C}, such that $y_t^M \equiv y_{t_q}^Q = y_t^{*C} + \mu$, where μ is defined as the in-sample mean of $y_{t_q}^Q$.

The missing values of y_t^* at $t \neq 3t_q$ can be estimated via the Kalman filter and smoother applied to the state space model in Eqs. (5.6)–(5.7) given below. Estimates for the unknown parameters β and σ_ε^2 in Eq. (5.4) are obtained by regression, applied to the model:

$$y_{t_q}^Q = \beta' F_{t_q}^Q + e_{t_q}^Q, \qquad e_{t_q}^Q \sim \text{NIID}(0, \sigma_e^2).$$

The estimate of σ_ε^2 itself is obtained from the relationship $\sigma_\varepsilon^2 = \sigma_e^2 \,/\, 3$.

The nowcasting and forecasting of quarterly GDP growth is based on the Kalman filter and smoother applied to the state space model, as given by the observation equation:

$$\begin{pmatrix} X_t \\ y_t^M \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{bmatrix} \Lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ y_t^* \\ y_t^{*C} \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \end{pmatrix}, \tag{5.6}$$

where μ is the sample average of the observed quarterly GDP growth rates $y_{t_q}^Q$, and hence of y_t^M, and the transition equation is given by:

$$\begin{bmatrix} I_r & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 \\ -\beta' & 0 & 1 & 0 \\ 0 & 0 & -1/3 & 1 \end{bmatrix} \begin{pmatrix} f_{t+1} \\ f_t \\ y_{t+1}^* \\ y_{t+1}^{*C} \end{pmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & 0 & 0 \\ I_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_t \end{bmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ y_t^* \\ y_t^{*C} \end{pmatrix} + \begin{pmatrix} \zeta_t \\ 0 \\ \varepsilon_{t+1} \\ 0 \end{pmatrix}, \tag{5.7}$$

for $t = 1, \ldots, T$. All of the variables are introduced in Eqs. (5.2)–(5.5). This state space representation is based on $q = 1$ and $p = 2$, but it is straightforward to amend it for other values of q and p. Note that the time series y_t^M contains many missing values. The ragged edges in data matrix $(X_1, \ldots, X_t)$ can also be regarded as a missing value problem. The treatment of missing values, the computation of forecasts and the estimation of f_t and y_t^* rely on the Kalman filter and smoother, which are discussed in detail in Durbin and Koopman (2012). The transition equation (Eq. (5.7)) is non-standard, given the pre-multiplication of the state vector on the left-hand-side of the equation; a minor modification provides the standard updating equation, but is somewhat less intuitive.

Modification of two-step approach

Earlier empirical studies, such as that of Chapter 3 of this thesis, have shown that adding autoregressive terms to the forecast equation can improve the forecasting accuracy significantly for GDP growth. These autoregressive terms, or lagged values of y_t, can be included in Eq. (5.4) by considering:

$$y_t^* = \rho_1 y_{t-1}^* + \rho_2 y_{t-2}^* + \beta' f_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{5.8}$$

for $t = 1, \ldots, T$, where ρ_1 and ρ_2 are the additional coefficients of the autoregressive process. Next, the state space form is adjusted accordingly. To illustrate, the observation equation for $r = 1$ and $p = 2$ is defined as:

$$\begin{pmatrix} X_t \\ y_t^M \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{bmatrix} \Lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ y_t^* \\ y_{t-1}^* \\ y_t^{*C} \end{pmatrix} + \begin{pmatrix} e_t \\ 0 \end{pmatrix}, \tag{5.9}$$

where μ is the in-sample mean of $y_{t_q}^Q$. The transition equation is given by:

$$\begin{bmatrix} I_r & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 \\ -\beta' & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1/3 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_{t+1} \\ f_t \\ y_{t+1}^* \\ y_t^* \\ y_{t+1}^{*C} \end{pmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_1 & \rho_2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_t \end{bmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ y_t^* \\ y_{t-1}^* \\ y_t^{*C} \end{pmatrix} + \begin{pmatrix} \zeta_t \\ 0 \\ \varepsilon_{t+1} \\ 0 \\ 0 \end{pmatrix}. \tag{5.10}$$

The values of he parameters ρ_1, ρ_2, β and σ_ε^2 are obtained via the regression method applied to the model:

$$\widetilde{y}_t^{*M} = \rho_1 \widetilde{y}_{t-1}^{*M} + \rho_2 \widetilde{y}_{t-2}^{*M} + \beta' F_t + e_t^M, \qquad e_t^M \sim \mathrm{NIID}(0, \sigma_\varepsilon^2),$$

for $t = 1, \ldots, T$ and where $\widetilde{y}_t^{*M}$ is the monthly time series of the linearly interpolated mean-adjusted quarterly series of $y_{t_q}^Q$, that is $\widetilde{y}_t^{*M} = y_{s_t}^{*Q}/3 + (t - 3\, s_t)(y_{s_{t+1}}^{*Q} - y_{s_t}^{*Q})/3$ for $t = 1, \ldots, T$ and $s_t = \lfloor t/3 \rfloor$, where $y_{t_q}^{*Q} = y_{t_q}^Q - \mu$ and with $y_0^{*Q} = y_1^{*Q}$.

5.2.3 Quasi-maximum likelihood approach

The dynamic factor model with its parameters estimated by the quasi-maximum likelihood approach of Doz et al. (2012), and also considered by Bańbura and Modugno

(2014), can be interpreted as a parametric alternative to the non-parametric methods based on principal components used by Bańbura and Rünstler (2011) and Giannone et al. (2008), among others. In general, the maximum likelihood approach is not considered to be a feasible approach for datasets with large cross-sections. However, Doz et al. (2012) have shown that the maximum likelihood method allows one to obtain consistent estimates of the factors even if the size of the cross-section goes to infinity.

Bańbura and Modugno (2014) propose the inclusion of serially correlated idiosyncratic components in Eq. (5.2). They also modify the EM algorithm of Shumway and Stoffer (1982) and Watson and Engle (1983) for the estimation of parameters in a dynamic factor model when a dataset contains random patterns of missing data. They show that the EM procedure for maximum likelihood estimation remains computationally feasible for large datasets. The EM method is an iterative process for finding the maximum likelihood estimates of the parameters by exploiting the cross-sectional and time series information simultaneously. The algorithm is initialised using the principal components extracted from the dataset; for applications of the methodology, see Bańbura et al. (2013) for the US and Bańbura and Modugno (2014) for the euro area.

In addition to the estimation technique, another contribution of Bańbura and Modugno (2014) is related to Eq. (5.2); the details of this modification are only summarized here. The forecasting procedure of Bańbura and Rünstler (2011) can be adopted straightforwardly as described above. The model for the stationary N-dimensional standardized vector process X_t in Eq. (5.2) is replaced by:

$$X_t = \Lambda f_t + \epsilon_t + u_t, \qquad u_t \sim \mathrm{NIID}(0, \Sigma_u), \tag{5.11}$$

$$\epsilon_t = \Theta\epsilon_{t-1} + e_t, \qquad e_t \sim \mathrm{NIID}(0, \Sigma_e), \tag{5.12}$$

where f_t is the $r \times 1$ vector of latent dynamic factors modeled as Eq. (5.3), and ϵ_t is the idiosyncratic component that is modeled as an autoregressive process with a diagonal autoregressive coefficient matrix Θ and a diagonal variance matrix Σ_e; the remaining coefficients are discussed below Eq. (5.3). The normally distributed disturbance vector e_t is uncorrelated with all other disturbances at all leads and lags.

Only small changes in the state space formulations are required in order to allow for the dynamics in the idiosyncratic component ϵ_t in the forecast procedure; see Bańbura and Modugno (2014). The dimension of the state vector needs to be increased by N elements to allow for the dynamics of ϵ_t. This will slow down the computations, but the procedure remains feasible. Reis and Watson (2010) propose that $X_t - \Theta X_{t-1}$ be considered as the observational vector instead of X_t, and include f_{t-1} in the observation equation (Eq. (5.2)). In this case, there is no need to place ϵ_t in the state vector. However, this option is only valid when no missing values occur in X_t or X_{t-1}. A more detailed discussion and a computational feasible solution to the missing value problem are presented in Jungbacker et al. (2011).

5.2.4 Collapsed dynamic factor approach

The collapsed dynamic factor model of Bräuning and Koopman (2014) is effectively a low-dimensional multivariate unobserved components time series model where the target series and a set of r principal components are treated jointly as dependent variables. Relative to the previously-discussed factor models, the collapsed dynamic factor model has a much lower dimension, such that the number of unknown parameters to be estimated is relatively small. Maximum likelihood estimation is carried out via the numerical maximization of the loglikelihood function that is evaluated using the Kalman filter.

As was summarized in Bräuning and Koopman (2014), the collapsed factor model procedure is a two-step process. The first step involves carrying out a principal component analysis for dimension reduction of the large panel of indicators. The model of Bańbura and Rünstler (2011) has the same first step, but with the aim of producing estimates of the factor loadings. In the second step, Bräuning and Koopman (2014) model these estimated principal components jointly with the target variable in a state space model that includes a small number of parameters. The unknown parameters are then estimated simultaneously using the maximum likelihood method in a standard manner. This step differs from the second step in Bańbura and Rünstler's method, where the parameters are estimated outside the state space framework. Finally, the Kalman filter and smoother method is used to obtain the in-sample estimates and out-of-sample forecasts of the target variable.

The model of Bräuning and Koopman (2014) is based on Eq. (5.2), which is extended with the target series of quarterly GDP growth in a specific way. The monthly and quarterly series are accommodated by formulating the extension of the model in terms of the unobservable series y_t^*, and obtain:

$$\begin{pmatrix} X_t \\ y_t^* \end{pmatrix} = \begin{bmatrix} \Lambda & 0 \\ \Gamma & 1 \end{bmatrix} \begin{pmatrix} f_t \\ \psi_t \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \end{pmatrix}, \tag{5.13}$$

where Γ is a loading matrix with the coefficients of the dynamic factors for the monthly unobserved series y_t^*, ψ_t is a univariate latent dynamic process for the target series, and the definitions for the other matrices and variables remain as above. The dynamic factors f_t are modeled as the vector autoregressive process in Eq. (5.3), while the unobserved component ψ_t for the target series can also be modeled as an autoregressive process, for example:

$$\psi_{t+1} = \phi_1 \psi_t + \phi_2 \psi_{t-1} + \eta_t, \qquad \eta_t \sim \mathrm{NIID}(0, \sigma_\eta^2), \tag{5.14}$$

where ϕ_1 and ϕ_2 are autoregressive coefficients and η_t is a NIID disturbance that is not related to any other disturbance in the model. The monthly series y_t^* is linked with the (partially) observed monthly series y_t^M in the relation $y_t^M = \mu + y_t^{*C}$ when $t = 3t_q$ and zero otherwise, and where y_t^{*C} is constructed as in Eq. (5.5).

The collapsed dynamic factor model is developed based on the insight that the principal component F_t is a linear combination of X_t, that is $F_t = AX_t$ for $t = 1, \dots, T$ and for a matrix A with property $A\Lambda = I$. Pre-multiplying (5.13) by matrix:

$$\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix},$$

Obtaining:

$$\begin{pmatrix} F_t \\ y_t^* \end{pmatrix} = \begin{bmatrix} I_r & 0 \\ \Gamma & 1 \end{bmatrix} \begin{pmatrix} f_t \\ \psi_t \end{pmatrix} + \begin{pmatrix} v_t \\ 0 \end{pmatrix}, \qquad v_t \sim \text{NIID}(0, \Sigma_v), \tag{5.15}$$

where $v_t = Au_t$, for $t = 1, \dots, T$. Since F_t is standardized, the variance matrix Σ_v can be restricted such that $\text{Var}(F_t) = \text{Var}(f_t) + \Sigma_v = I_r$. Hence, there is no need to estimate Σ_v because it is a function of other parameters in the model. Finally, a observation disturbance could be added to y_t^*; however, its variance has been estimated at zero in almost all cases in the current study.

The vector series of principal components F_t and the quarterly GDP growth series, transformed into the monthly series y_t^M, are treated as the observation vector. The specification of the collapsed dynamic factor model is as in Eq. (5.3) with $p = 2$, (5.14) and (5.15). The state space form for a set of r latent factors consists of the observation equation:

$$\begin{pmatrix} F_t \\ y_t^M \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{bmatrix} I_r & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ \psi_t \\ \psi_{t-1} \\ y_t^* \\ y_t^{*C} \end{pmatrix} + \begin{pmatrix} v_t \\ 0 \end{pmatrix},$$

and the transition equation:

$$\begin{bmatrix} I_r & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -\Gamma & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1/3 & 1 \end{bmatrix} \begin{pmatrix} f_{t+1} \\ f_t \\ \psi_{t+1} \\ \psi_t \\ y_{t+1}^* \\ y_{t+1}^{*C} \end{pmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta_t \end{bmatrix} \begin{pmatrix} f_t \\ f_{t-1} \\ \psi_t \\ \psi_{t-1} \\ y_t^* \\ y_t^{*C} \end{pmatrix} + \begin{pmatrix} \zeta_t \\ 0 \\ \eta_t \\ 0 \\ \varepsilon_t \\ 0 \end{pmatrix},$$

for $t = 1, \dots, T$. The unknown parameters Φ_1, Φ_2, Γ, ϕ_1, ϕ_2, Σ_ζ, σ_η^2 and σ_ε^2 are estimated by maximum likelihood. The number of unknown parameters is $4(r+1)$; that is, 8 and 12 for $r = 1$ and $r = 2$, respectively.

Modification of the collapsed dynamic factor approach

In contrast to the approach of Bańbura and Rünstler (2011), the collapsed dynamic factor approach requires a pre-analysis to treat the ragged edges in the data matrix $(X_1, \ldots, X_T)$, because the model requires a balanced dataset to compute the principal components in F_t. Adopting the EM method of Stock and Watson (2002b) for the purpose of computing the F_t's could be a feasible alternative. However, a univariate model for each variable in X_t is considered instead. The resulting models are adopted for interpolating and extrapolating the missing values using the Kalman filter and smoother. The details are as follows. For the ith time series X_{it} in X_t, the following stationary decomposition model is considered:

$$X_{it} = \theta_{it} + \kappa_{it}, \qquad \kappa_{it} \sim \text{NIID}(0, \sigma^2_{\kappa,i}),$$

for $t = 1, \ldots, T$, where θ_{it} is typically modeled as a univariate autoregressive process (for example, with two lags) and κ_{it} is a measurement error, for $i = 1, \ldots, N$. After estimating the unknown parameters, the Kalman filter and smoother replace the missing entries by their corresponding estimates, for each variable $i = 1, \ldots, N$. Thus, obtaining a balanced dataset $(X_1, \ldots, X_T)$, allowing the principal components to be constructed.

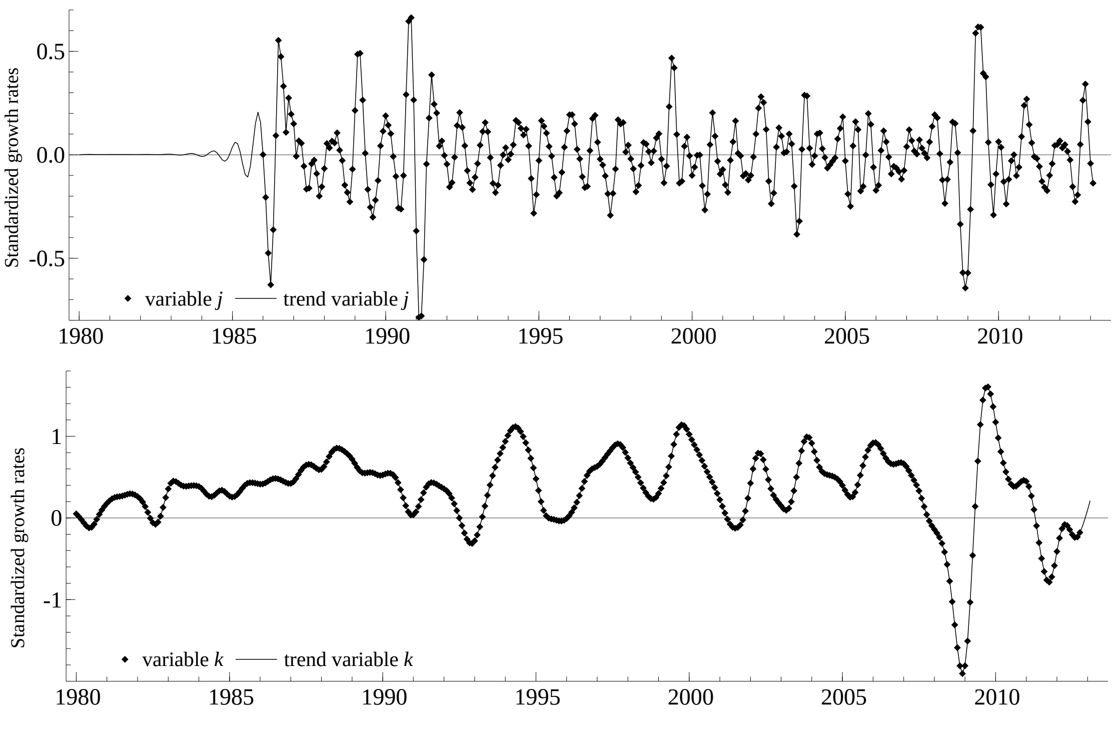


Figure 5.1: Treatment of missing values for time series j and k

The use of a stationary autoregressive process for θ_{it} ensures that the balanced variable returns to its long-term mean of zero when a long sequence of entries is missing. Figure 5.1 presents two examples in the dataset. Time-series j has missing values only at the beginning of the sample, while time-series k has missing values only at the end of the sample.

5.3 Data, forecast design and specification issues

5.3.1 Dataset

The monthly dataset of predictors consists of 52 time series variables for the euro area and its five largest countries. The variables selected are based on harmonized definitions across the euro area and its countries, and fall into four predefined categories: production & sales, prices, monetary & financial indicators, and surveys. Table 5.7 in Appendix 5.A provides descriptions of all of the variables used, together with the transformations used in the analysis and the starting date of the monthly series for each country in the sample. The monthly data are usually adjusted for seasonality (and calendar effects). When necessary, raw data series are seasonally adjusted using the US Census X-12 method. All monthly series are made stationary by differencing or log-differencing (in the case of trending data, such as industrial production, retail sales and monetary aggregates). Finally, each variable is standardized by subtracting the mean and dividing by the standard deviation. This normalization is standard practice in order to avoid the overweighting of series with large-variances series in the extraction of common factors.

5.3.2 Pseudo real-time design

The forecast design aims to replicate the availability of the data at the time forecasts are made, in order to approximate the real-time flow of information as closely as possible. For this purpose, a dataset on March 4, 2013 was collected. The typical data release calendar was used to reconstruct the available dataset on the 4th of each month over the years 1992–2012. The database was constructed such that the earliest starting date is January 1980 for the monthly series, and the first quarter of 1980 for GDP. Hence, a pseudo real-time design was employed, which takes data publication delays into account, but ignores the possibility of data revisions for GDP and some indicators, such as industrial production. This might imply that the forecasting accuracy is overrated. However, there are not yet any large real-time datasets available for the countries considered in this chapter. Moreover, the effects of data revisions on the forecasts of factors may also cancel overall; see, for example, Bernanke and Boivin (2003) for the US and Schumacher and Breitung (2008) for Germany.

For all models the parameters are estimated recursively, using only the information available at the time when the forecast would have been made, see Giannone et al.

(2008), Kuzin et al. (2011) and Rünstler et al. (2009), among others, for a similar approach. A sequence of eleven forecasts of GDP growth in a given quarter is constructed, obtained in consecutive months. Table 5.1 illustrates the timing of the forecasting exercise, taking the forecast for 2012.III as an example. The first forecast is made on January 4, 2012 and is referred to as the two-quarter ahead forecast in month one. A monthly forecast is subsequently produced for the next ten months, from February until November. The last forecast is made on November 4, 2012, approximately one and a half week before the flash release of GDP in mid-November. Following the conventional terminology, *forecasts* refer to one or two-quarter ahead forecasts, *nowcasts* refer to current quarter forecasts and *backcasts* refer to forecasts for the preceding quarter, for which official GDP figures are not yet available. In case of the current example 2012.III, two-quarter ahead forecasts are produced from January to March, one-quarter ahead forecasts from April to June, nowcasts from July to September, and backcasts in October and November.

Table 5.1: Timing of forecasting exercise for third quarter GDP growth

No.	Name	Forecast made on the 4^{th} of
1		January
2	2Q ahead	February
3		March
4		April
5	1Q ahead	May
6		June
7		July
8	Nowcast	August
9		September
10		October
11	Backcast	November

5.3.3 Model specification

The GDP forecasts generated from the Bańbura and Rünstler (2011) method can be based either on Eq. (5.4) or on adding lagged dependent variables in the forecasting equation, as in Eq. (5.8). Table 5.2 compares the forecasting accuracies of these two options, presenting the mean squared forecast errors (MSFEs) of both models. Grey cells indicate the model with the lowest MSFE averaged over all horizons. Entries in bold indicate models where the best model specification has a MSFE that is less than 10% larger than that of the other model specification. The 10% threshold is chosen as a rough indication of the economic significance of differences in forecast precision. Models that meet this condition are referred to as "competitive models"; their forecasting performance is similar to that of the best model.

Table 5.2: Sensitivity analysis for the Bańbura and Rünstler (2011) model, 1992.I–2012.IV

	EA	DE	FR	IT	ES	NL
	MSFE					
Bańbura and Rünstler (2011) model						
1 factor	**0.31**	**0.63**	0.37	**0.47**	**0.45**	**0.44**
2 factors	**0.30**	**0.64**	0.35	0.45	**0.44**	**0.41**
3 factors	**0.30**	**0.65**	0.34	0.44	0.40	0.40
4 factors	**0.30**	**0.65**	0.34	0.43	0.38	0.40
5 factors	**0.29**	**0.64**	0.33	0.43	0.37	0.40
6 factors	**0.29**	**0.64**	0.33	0.43	0.37	0.39
7 factors	**0.29**	**0.64**	0.33	0.43	0.36	0.39
average 1-7 factors	**0.30**	**0.64**	0.34	0.44	0.40	0.41
Augmented Bańbura and Rünstler (2011) model						
1 factor	0.29	0.63	0.28	0.46	0.43	0.41
2 factors	0.28	0.62	0.28	**0.45**	0.43	0.41
3 factors	0.28	0.62	0.28	**0.44**	**0.41**	**0.41**
4 factors	0.28	0.63	0.28	**0.44**	**0.39**	**0.41**
5 factors	0.28	0.63	0.28	**0.44**	**0.38**	**0.41**
6 factors	0.29	0.63	0.28	**0.45**	**0.38**	**0.41**
7 factors	0.29	0.63	0.28	**0.45**	**0.38**	**0.40**
average 1-7 factors	0.28	0.63	0.28	**0.45**	**0.40**	**0.41**

Notes: entries denote MSFEs. Grey cells denote models with the lowest MSFE. Entries in bold denote MSFEs that are up to 10% larger than that of the best model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Table 5.3: Sensitivity analysis imputation method for the Bräuning and Koopman (2014) model, 1992.I–2012.IV

	EA	DE	FR	IT	ES	NL
	MSFE					
Bräuning and Koopman (2014) with EM algorithm						
1 factor	0.30	**0.60**	**0.26**	0.46	0.41	**0.41**
2 factors	0.30	**0.63**	0.26	**0.45**	0.45	**0.42**
3 factors	0.31	**0.65**	0.26	**0.45**	0.48	0.43
4 factors	0.30	**0.68**	0.25	0.46	0.51	0.44
5 factors	0.31	**0.65**	0.25	0.47	0.53	0.44
6 factors	0.34	**0.66**	0.25	0.48	0.55	0.44
7 factors	0.37	**0.69**	0.24	0.49	0.56	0.43
average 1-7 factors	0.32	**0.65**	0.25	0.47	0.50	0.43
Bräuning and Koopman (2014) with AR(2)						
1 factor	0.27	0.57	0.24	0.42	0.36	0.38
2 factors	0.26	0.61	0.23	0.41	0.35	0.38
3 factors	0.26	0.63	0.23	0.41	0.35	0.38
4 factors	0.26	0.63	0.22	0.42	0.34	0.37
5 factors	0.26	0.64	0.22	0.42	0.34	0.37
6 factors	0.26	0.64	0.22	0.43	0.34	0.36
7 factors	0.26	0.64	0.22	0.43	0.33	0.36
average 1-7 factors	0.26	0.62	0.23	0.42	0.35	0.37

Notes: entries denote MSFEs. Grey cells denote models with the lowest MSFE. Entries in bold denote MSFEs that are up to 10% larger than that of the best model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Overall, the forecasting accuracy improves when the forecasts are based on Eq. (5.8) , though the differences are typically small. The only exception is France, for which the MSFE of the augmented Bańbura and Rünstler model is more than 10% smaller than that of the original model. Based on these results, the empirical section of this chapter will consider only the forecasts with the lagged dependent variables included in the forecast function. Moreover, Bańbura and Rünstler (2011) reduce the rank of the covariance matrix of the idiosyncratic component from r to $q \leq r$ to ensure the smoothness of the factors. The same strategy is adopted as that described above by averaging the forecasts obtained from the models with $q = 1, \ldots, r$.

The Bräuning and Koopman (2014) method also requires some way of dealing with the ragged edges of the principal components. Table 5.3 compares the forecasting accuracies of the two options considered in Section 5.2.4: the EM algorithm and interpolation via an autoregressive specification. For the euro area, France, Spain, Italy and the Netherlands, the differences in forecasting accuracy are considerable, with the autoregressive specification clearly being optimal for nearly all factor specifications. For Germany, it does not make much difference to the forecasting accuracy whether the ragged edges are dealt with by the EM algorithm or an autoregression. Thus, the empirical results section will present outcomes based on the autoregressive solution.

5.3.4 Model selection

There are many different approaches to determining r, the number of factors in F_t. One basic method is the scree test of Cattell (1966), which is based on a graph of the normalized eigenvalues calculated from the set of candidate predictors. According to this method, the point in the eigenvalue graph at which the eigenvalues begin to level off with a slow and steady decrease is the estimate of the sufficient number of factors. Figure 5.2 displays the scree plots for the euro area and its five largest countries, presenting the normalized eigenvalues of the largest thirty principal components. In general, the first principal component is able to explain 20% to 30% of the comovement in the set of candidate predictors. For most countries, the explanatory power increases only very slightly after the fifth or sixth principal component.

Figure 5.3 presents the correlations of the first four estimated principal components with the matrix of candidate predictors for the euro area. The x-axis shows the candidate variables that correspond to the numbers in Table 5.7 in Appendix 5.A, and the y-axis shows the correlations in percentages. The first principal component is correlated strongly with a broad range of variables apart from prices, which is in accordance with the high eigenvalue. This indicates that the bulk of the covariance of the candidate predictors can be explained by the first factor. The second and third principal components are correlated strongly with price variables, such as the harmonised index of consumer prices (HICP), commodity prices and the oil price.

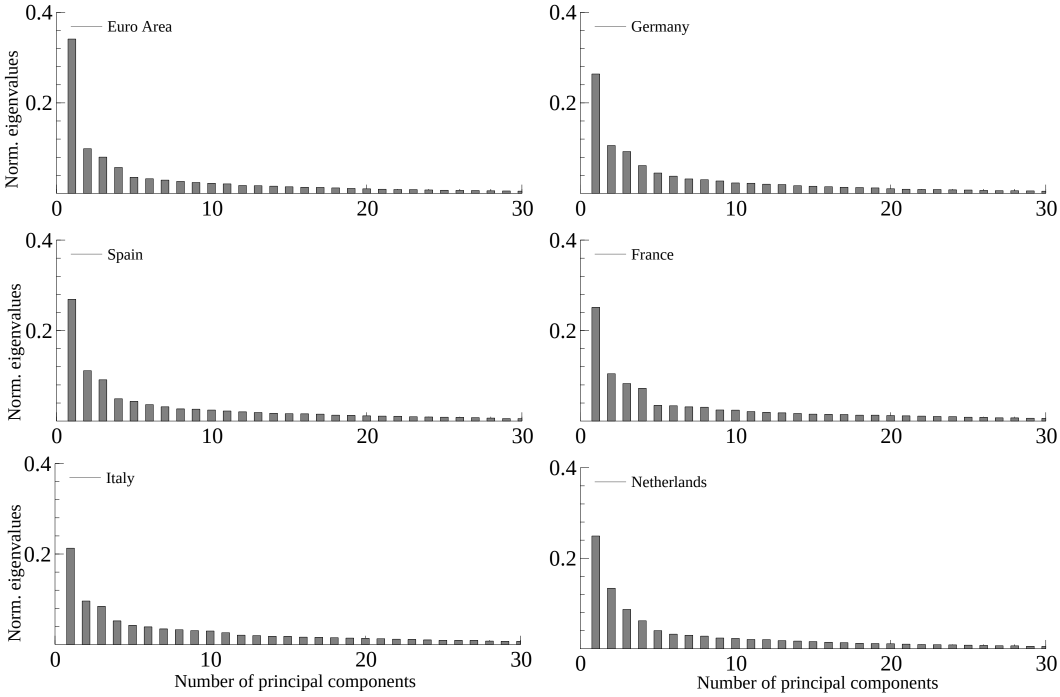


Figure 5.2: Scree plots of normalized eigenvalues from the candidate predictors, 1992.I–2012.IV

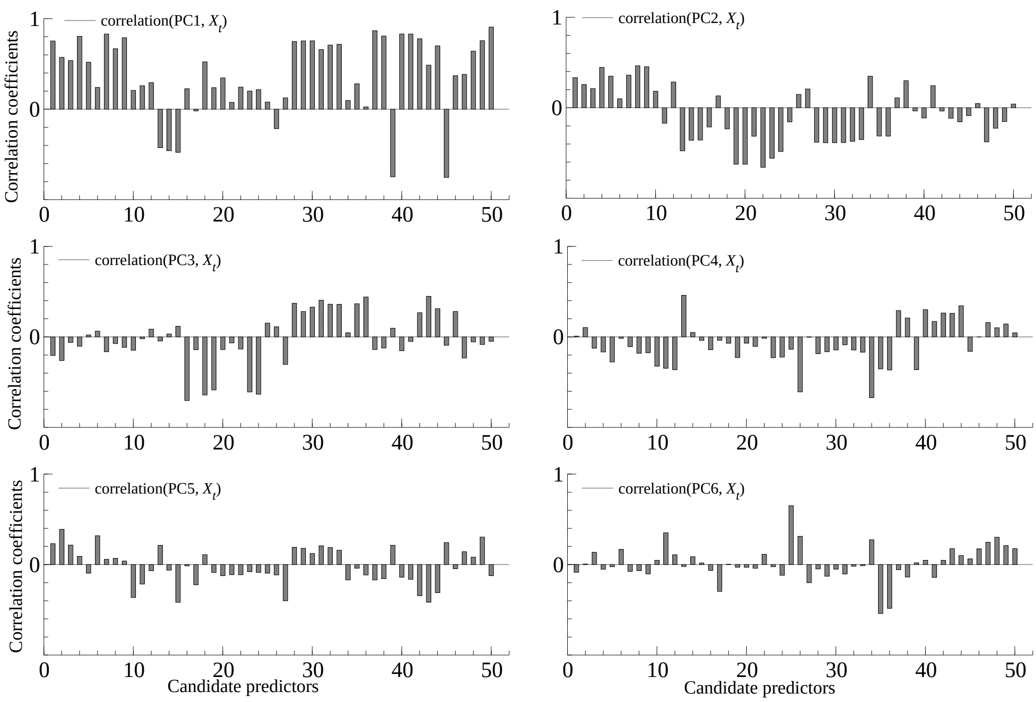


Figure 5.3: Correlation between principal components and the predictors for the euro area, 1992.I–2012.IV

The fourth principal component is correlated strongly with survey indicators. The fifth and sixth principal components are correlated strongly with the international variables, such as world trade, industrial production, OECD leading indicators in the US and the UK, and the (real effective) exchange rate of the euro.

In accordance to the intuition behind the Cattell (1966) test, Bai and Ng (2002) propose to select the number of factors by minimizing the variance of the idiosyncratic component. Their criterion has been assessed rigorously in many studies, including those of Bańbura and Rünstler (2011), Bernanke and Boivin (2003), Boivin and Ng (2005) and Giannone et al. (2005). Overall, these studies have concluded that the number of factors selected from the Bai and Ng is too large, with more parsimonious models with fewer factors being preferable for forecasting purposes. Other and modified criteria have been proposed by Ahn and Horenstein (2013), Alessi et al. (2010) and Onatski (2010). The different criteria for determining the number of factors are considered in Table 5.4.[3]. The criterion of Bai and Ng points to the use of seven static factors for the euro area and the individual countries, while the other tests mostly suggest smaller numbers of factors. Thus, since the different criteria clearly do not agree on the numbers of static factors required, the forecasts from the various models with different numbers of factors were pooled. Chapter 3 of this thesis, Kuzin et al. (2013) and others provide evidence that an unweighted average of the forecasts based on different factors leads to a high forecasting accuracy at all horizons. Therefore, the forecasting comparison was conducted using the averaged forecasts of models with one to six factors. Tables 5.8–5.13 in Appendix 5.B present the individual forecasting results for factor models with one to seven factors.

Table 5.4: Statistical tests for the number of static factors

	Bai & Ng	Onatski	Alessi et al.	Ahn & Horenstein
EA	7	4	4	7
DE	7	4	6	3
FR	7	7	5	2
IT	7	2	3	3
ES	7	2	5	3
NL	7	4	3	1

Notes: entries denote the number of factors from the Bai and Ng (2002), Onatski (2010), Alessi et al. (2010) and Ahn and Horenstein (2013) test, respectively. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

[3] The tests are conducted over the complete sample of monthly time series, without missing values at the beginning or and of the sample (for most countries the period 1986.1–2012.10)

5.4 Empirical results

5.4.1 Forecasting performance using the complete sample

This section presents and discusses the forecasting accuracy of the four factor models described in Section 5.2 relative to that of a benchmark model, which is simply an autoregressive model of order two, AR(2). The factor models are the principal component model of Stock and Watson (SW), the augmented dynamic factor model of Bańbura and Rünstler (BR), the dynamic factor model of Bańbura and Modugno (BM) and the collapsed dynamic factor model (CFM) of Bräuning and Koopman. The forecasting performance is analyzed for the euro area (EA) and its five largest countries, Germany (DE), France (FR), Italy (IT), Spain (ES) and the Netherlands (NL). The forecasting accuracy is measured by the MSFE.

Table 5.5 presents the forecasting performances of the four factor models and the benchmark model for the euro area and the five countries for the complete quarterly data sample of 1992-2012. The forecasts have been generated on a monthly basis for eleven forecast horizons. Table 5.5 reports the average forecasting accuracies for the one- and two-quarter ahead forecasts, the nowcasts and the backcasts.[4] Moreover, the MSFEs presented are averaged over model specifications with one to six factors. The rows labeled AR(2) report the MSFEs of the benchmark model, while the rows for the four factor models present the MSFEs relative to those of the benchmark model. Grey cells indicate the models with the lowest MSFEs for particular forecast horizons and countries. Entries in bold indicate models that have MSFEs that are less than 10% larger than that of the best model, and also smaller than that of the benchmark model. The 10% threshold is chosen as a rough indication of the economic significance of differences in forecast precision. Models that meet this condition are referred to as "competitive models"; their forecasting performances are similar to that of the best model. Table 5.5 also reports if the smallest MSFE is significantly smaller than those of all other models according to the Diebold and Mariano (1995) test at a 10% significance level. The results in Table 5.5 reveal various interesting insights.

First, the incorporation of monthly information into a factor model improves the forecasting accuracy, especially for nowcasts and backcasts. Averaged over all forecasting horizons and countries, the best models improve on the benchmark AR(2) model by around 15%. The results also indicate that the predictions from the factor models deteriorate for longer forecast horizons. These results confirm that factor models are particularly suitable for generating nowcasts and backcasts, but less suitable for one- and two-quarter ahead forecasting, confirming previous empirical findings in Chapter 3 and Bańbura and Rünstler (2011), Giannone et al. (2008) and Rünstler et al. (2009), among others.

[4] The forecast for the months within the quarters are available from the authors upon request.

Table 5.5: Forecasting performance dynamic factor models, 1992.I–2012.IV

	EA	DE	FR	IT	ES	NL
	MSFE					
AR(2)						
Backcast	*0.32*	*0.76*	*0.16*	*0.47*	*0.31*	*0.45*
Nowcast	*0.39*	*0.78*	*0.21*	*0.53*	*0.35*	*0.49*
1Q ahead forecast	*0.45*	*0.81*	*0.28*	*0.60*	*0.41*	*0.53*
2Q ahead forecast	*0.49*	*0.81*	*0.34*	*0.66*	***0.46***	*0.56*
All horizons	*0.42*	*0.79*	*0.26*	*0.57*	*0.39*	*0.51*
	Relative to MSFE AR(2)					
SW						
Backcast	0.69	0.85	0.90	0.83	0.82	**0.67**
Nowcast	0.83	0.97	**0.88**	0.90	**0.78**	0.71
1Q ahead forecast	0.96	1.14	**0.94**	**0.86**	**0.92**	0.95
2Q ahead forecast	0.99	1.22	1.01	**0.85**	1.07	1.03
All horizons	0.90	1.07	0.95	0.86	**0.92**	0.87
BR						
Backcast	0.49	0.64	1.01	0.63	1.04	0.64
Nowcast	**0.57**	**0.74**	1.12	0.69	1.06	0.68
1Q ahead forecast	**0.71**	0.84	1.11	**0.83**	0.89	0.81
2Q ahead forecast	**0.80**	**0.91**	1.09	**0.87**	0.96	0.97
All horizons	**0.68**	0.80	1.09	**0.78**	0.98	0.80
BM						
Backcast	0.44	0.39	1.03	0.63	0.84	**0.65**
Nowcast	0.62	0.73	1.16	**0.74**	**0.82**	**0.65**
1Q ahead forecast	0.81	0.95	1.15	**0.82**	**0.96**	**0.71**
2Q ahead forecast	0.97	1.09	1.11	**0.84**	1.10	**0.87**
All horizons	0.76	**0.83**	1.12	**0.78**	0.95	**0.73**
CFM						
Backcast	0.36*	0.57	0.77*	0.53*	0.64	**0.66**
Nowcast	0.55	0.83	0.84	**0.70**	0.75	0.61
1Q ahead forecast	0.65	**0.87**	0.86	0.79	**0.93**	0.68
2Q ahead forecast	0.76	0.88*	0.89*	0.83	**0.99**	0.85
All horizons	0.62*	**0.81**	0.86*	0.74*	0.86	0.71

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts averaged across model specifications with **one–six** factors. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). * denotes the Diebold-Mariano test is significant at the 10% level. AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Second, the collapsed dynamic factor model of Bräuning and Koopman (2014) displays the highest forecasting accuracy for most countries and horizons. Averaged over all horizons, the collapsed dynamic factor model is the best model in all countries except for Germany, for which the model of Bańbura and Rünstler (2011) is best. However, the collapsed dynamic factor model is still a competitive model. This is also true for the other cases in which the collapsed dynamic factor model does not have the lowest MSFE, with the exception of the nowcasts and backcasts for Germany. The collapsed dynamic factor model posts the highest gains in forecasting accuracy over the benchmark model for the euro area, with average improvements ranging from 24% for the two-quarter ahead forecasts to 64% for the backcasts.

Third, the collapsed dynamic factor model is the only model that consistently outperforms the benchmark model in terms of forecasting accuracy; the forecasting performances of the other three factor models are rather less favorable. For example, the augmented Bańbura and Modugno (2014) and Bańbura and Rünstler (2011) models fail to produce lower MSFEs than the benchmark model for France for all forecast horizons; the model of Stock and Watson (2002b) is unable to outperform the benchmark model with regard to the one- and two-quarter ahead forecasts for Germany, France, Spain and the Netherlands; the forecasts from Bańbura and Modugno (2014) are less accurate than those from an AR(2) model for one- and two-quarter ahead forecasts, for Germany and Spain; and the backcasts and nowcasts of Bańbura and Rünstler (2011) are less accurate than those of the AR(2) model for Spain.

These outcomes provide empirical evidence that the predictions from dynamic factor models are especially suitable for nowcasting and backcasting. The results also suggest that the collapsed dynamic factor model of Bräuning and Koopman (2014) displays a significantly greater ability to absorb monthly information than the other three dynamic factor models that are considered in this study. It is also the only model that is able to deliver significantly smaller MSFEs than any of the other competing models according to the Diebold and Mariano (1995) test. The good forecasting performance of the collapsed dynamic factor model is relatively robust to the model specification, as is shown in Tables 5.8–5.13 in Appendix 5.B. The tables show the forecasting accuracies for factor model specifications with one to seven factors, respectively, for all factor models.

5.4.2 Forecasting performance during the Great Moderation and the financial crisis

The sample also includes the period of the financial crisis, during which there was a sharp downturn in a broad range of indicators, including manufacturing, confidence indicators and exports. As a consequence, real GDP growth dropped sharply across all industrialized countries. An interesting question is whether, and to what extent, the performances of the factor models differ between the volatile financial crisis and the previous years, which can be characterized as a relatively stable period. Forecasting in times of crisis clearly poses greater challenges. Hence, a comparative analysis that focuses on these periods may be informative as to the question of which factor model is most suitable for forecasting GDP growth. In order to determine the influence of the financial crisis on the forecasting accuracy of the factor models, the sample is divided into two periods: the "Great Moderation" (1992–2007), and the financial crisis and its aftermath (2008–2012). Table 5.6 presents the outcomes of the forecasting performances of the four factor models and the benchmark model for the euro area and the five countries, during both periods. A comparison of these two distinct periods points to some interesting results.

First, the prediction of GDP growth is more difficult during and after the financial crisis than during the Great Moderation. The MSFE of the benchmark model during and after the financial crisis is two to six times larger than that during the Great Moderation, depending on the country. However, the relative improvements in forecasting accuracy from the factor models are larger during the financial crisis, especially for nowcasting and backcasting. For example, the relative MSFE of the collapsed dynamic factor model in the euro area improves by 69% during the financial crisis, compared to 33% during the Great Moderation. This finding is consistent with the results obtained in other studies; see for example Chapter 3 of this thesis, and D'Agostino and Giannone (2012). Both studies show that the gain in forecasting accuracy is especially sizeable in periods of large swings in GDP and a high degree of comovement in the monthly predictors, as was the case during the financial crisis.

Second, when averaged over all horizons, the collapsed dynamic factor model of Bräuning and Koopman (2014) is highly competitive during the Great Moderation. In various cases, it is able to deliver significantly smaller MSFEs than any of the competing models according to the Diebold and Mariano (1995) test. This indicates that the collapsed dynamic factor model is quite suitable for processing monthly information in more quiet times. This conclusion also holds for most countries when analyzing the forecasting performances for each forecast horizon separately. The greatest gain in forecasting accuracy against the benchmark model was 47%, recorded for the backcasts in the euro area. However, Spain is the exception, as the collapsed dynamic factor model is not competitive for either nowcasting or one-quarter ahead forecasting.

Third, the collapsed factor model is still a competitive model for most countries during and after the financial crisis, but the overall results are mixed. On average across all forecast horizons, the collapsed dynamic factor model is a competitive model for the euro area, and for three of the five individual countries, i.e. France, Italy and the Netherlands. Again, Spain is an exception, as it is not very competitive. For Spain and

Table 5.6: Forecasting performance dynamic factor models during the Great Moderation and the financial crisis, 1992.I–2012.IV

	EA	DE	FR	IT	ES	NL	EA	DE	FR	IT	ES	NL
	Great Moderation (1992.I–2007.IV)						financial crisis (2008.I–2012.IV)					
	MSFE											
AR(2)												
Backcast	*0.18*	*0.43*	***0.12***	*0.25*	*0.24*	*0.28*	*0.80*	*1.78*	*0.31*	*1.17*	*0.55*	*0.99*
Nowcast	*0.19*	*0.44*	***0.13***	*0.27*	*0.25*	*0.30*	*1.03*	*1.89*	*0.47*	*1.39*	*0.67*	*1.09*
1Q ahead forecast	*0.21*	*0.44*	***0.15***	*0.29*	*0.28*	*0.32*	*1.23*	*1.98*	*0.68*	*1.60*	*0.84*	*1.18*
2Q ahead forecast	*0.23*	***0.44***	*0.18*	*0.31*	***0.30***	*0.35*	*1.32*	*2.01*	*0.83*	*1.77*	*0.96*	*1.23*
All horizons	*0.20*	*0.44*	***0.15***	*0.28*	***0.27***	*0.32*	*1.12*	*1.93*	*0.60*	*1.51*	*0.77*	*1.13*
	Relative to MSFE AR(2)											
SW												
Backcast	0.92	1.15	1.06	1.10	1.07	0.97	0.54	0.62	0.70	0.64	0.49	**0.38**
Nowcast	0.94	1.10	1.00	1.02	1.06	**0.90**	0.77	0.87	0.78	0.82	0.44	0.55
1Q ahead forecast	1.16	1.25	1.04	0.91	1.17	0.96	0.85	1.06	**0.87**	0.84	0.65	0.94
2Q ahead forecast	1.21	1.34	1.09	**0.93**	1.25	1.02	0.87	1.13	**0.95**	0.80	**0.90**	1.05
All horizons	1.08	1.21	1.05	0.98	1.15	0.96	0.79	0.96	**0.86**	0.79	0.66	0.78
BR												
Backcast	0.65	0.87	1.18	0.76	1.03	**0.92**	0.38	0.47	0.81	0.54	1.06	0.37
Nowcast	0.70	**0.93**	1.24	**0.75**	1.05	**0.92**	0.49	0.59	1.01	0.66	1.06	0.47
1Q ahead forecast	**0.74**	0.91	1.15	**0.81**	0.90	0.92	**0.70**	**0.78**	1.08	0.85	0.89	0.72
2Q ahead forecast	0.79	0.97	1.05	**0.85**	**0.97**	0.95	**0.81**	**0.88**	1.12	0.88	0.95	0.99
All horizons	**0.73**	**0.92**	1.14	**0.80**	**0.98**	**0.93**	**0.64**	0.71	1.05	0.77	0.97	0.68
BM												
Backcast	0.74	**0.76**	0.96	0.84	0.96	0.88	0.22	0.11	1.11	**0.48**	0.68	0.44
Nowcast	0.91	1.06	1.11	0.85	0.98	**0.90**	0.45	0.48	1.21	**0.67**	0.63	0.43
1Q ahead forecast	1.10	1.21	1.15	0.97	1.20	0.89	**0.65**	0.77	1.15	0.73	0.72	0.56
2Q ahead forecast	1.23	1.24	1.07	1.04	1.36	**0.89**	0.83	0.99	1.14	0.72	0.84	0.85
All horizons	1.03	1.10	1.09	0.94	1.15	**0.89**	**0.60**	0.64	1.15	0.68	0.73	**0.59**
CFM												
Backcast	0.53	0.74	**0.96**	0.64*	0.80	**0.95**	**0.23**	0.44	0.53	0.45	0.43	**0.38**
Nowcast	0.60	0.85*	0.98*	0.73	0.87	0.88	0.52	0.82	0.70	**0.69**	0.60	0.37
1Q ahead forecast	0.67*	**0.93**	0.91*	0.79	0.99	0.79	0.64	**0.83**	0.83	**0.80**	0.85	**0.59**
2Q ahead forecast	**0.80**	**0.97**	0.89*	0.84	0.93	0.81*	0.74	0.82*	0.90	0.82	1.05	**0.88**
All horizons	0.67*	0.88*	0.93*	0.77*	0.91*	0.85	0.59	0.76	0.80	**0.73**	0.81	0.58

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts averaged across model specifications with **one–six** factors. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). * denotes the Diebold-Mariano test is significant at the 10% level. AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Germany, the forecasts from the models of Bańbura and Modugno (2014) and Stock and Watson (2002b) outperform those from the other models. Finally, the forecasting accuracy of the Stock and Watson (2002b) model is low during the Great Moderation, even relative to the benchmark model.

5.5 Conclusion

This chapter has studied the use of factor models for forecasting the growth rate of real GDP for the euro area and its five largest countries, i.e. Germany, France, Italy, Spain and the Netherlands. Four model approaches were considered, all relying on the use of dynamic factors, namely those of Bräuning and Koopman (2014), Doz et al. (2011, 2012), and Stock and Watson (2002b). The common factors for the euro area and the five countries are computed from a harmonized monthly dataset of 52 variables. The full quarterly sample ranges from the first quarter of 1992 to the fourth quarter of 2012. This sample length allows making a distinction between the factor models' performances during the volatile financial crisis and the more tranquil years before. The model of Doz et al. (2011), as implemented by Bańbura and Rünstler (2011), was modified by introducing more autoregressive terms into the model for the target variable (the GDP growth rate). Furthermore, the collapsed dynamic factor model of Bräuning and Koopman (2014) was modified by considering the use of an autoregressive model of order two, AR(2), for dealing with the ragged edges at the beginning and the end of the estimation period.

The empirical findings can be summarized in the following four points. (i) The monthly factors extract information that is valuable for the short-term forecasting of real GDP growth. The largest forecasting accuracy gains are obtained for nowcasting and backcasting. Overall, the monthly factors are especially useful for forecasting the corresponding quarter. (ii) The gains in forecasting accuracy in the period during and after the financial crisis are larger than those during the Great Moderation period. This finding underscores the usefulness of factor models for forecasting the growth rate of real GDP during volatile periods. (iii) The collapsed dynamic factor approach of Bräuning and Koopman (2014) has been shown to produce the highest forecasting accuracy overall for the euro area and its five largest countries. However, there is a marked contrast between the period of the Great Moderation and the period during and after the financial crisis. During the Great Moderation, the model of Bräuning and Koopman (2014) has the highest forecasting accuracy for most forecasting horizons and most countries considered. During and after the financial crisis, the relative forecasting accuracy among the factor models considered is much more diversified, although that of Bräuning and Koopman (2014) is still the most competitive one. (iv) Interpolating missing values by using an AR(2) model in the Bräuning and Koopman (2014) model has been shown to improve the forecasting accuracy for most countries considered. The inclusion of an autoregressive term of the target variable GDP in the Bańbura and Rünstler (2011) model improves its forecasting accuracy slightly, although the gains are generally small.

Appendix

5.A Dataset

The main data source is the ECB statistical datawarehouse. The world trade series are taken from the world trade monitor of the Netherlands Bureau of Policy Analysis (CPB). Since these series start in 1991, the series have been backdated using the world trade data from the International Monetary Fund (IMF). Time series on industrial production for the US are obtained from the board of governors of the Federal Reserve System. The commodity prices and most of the financial market indicators are taken from Thomson Reuters datastream. The survey data are provided by the European Commission (EC) while the purchasing managers indices for US and the UK are from Markit services.

The quarterly GDP series for Germany, France, Italy, Spain and the Netherlands are taken from the ECB statistical datawarehouse. A synthetic GDP series was constructed for the euro area using the database in the ECB's area wide model supplemented with data from the ECB statistical datawarehouse.

Table 5.7: Description monthly dataset

Nr.	Variable	Transformation			Starting year					
		sa	ln.	dif.	EA	DE	FR	IT	ES	NL
I. Production & sales (N=*15*)										
1	World Trade	1	1	1	'77	'77	'77	'77	'77	'77
2	Ind. prod. US	1	1	1	'60	'60	'60	'60	'60	'60
3	Ind. prod. UK	1	1	1	'68	'68	'68	'68	'68	'68
4	Ind. prod. (excl. constr.)	1	1	1	'60	'60	'60	'60	'61	'62
5	Ind. prod., consumer goods	2	1	1	'80	'80	'63	'60	'65	'90
6	Ind. prod., energy	2	1	1	'80	'91	'63	'80	'80	'00
7	Ind. prod., interm. goods	1	1	1	'60	'80	'63	'77	'65	'00
8	Ind. prod., capital goods	1	1	1	'60	'80	'63	'77	'65	'70
9	Ind. prod., manufacturing	2	1	1	'60	'78	'60	'71	'80	'70
10	Ind. prod., construction	2	1	1	'85	'78	'85	'95	'88	'85
11	Passenger car registration	1	1	1	'77	'77	'77	'77	'77	'79
12	Retail trade volume	2	1	1	'70	'68	'70	'90	'95	'60
13	Unemployment rate	1	0	1	'83	'62	'83	'83	'86	'83
14	Unemployment rate UK	1	0	1	'83	'83	'83	'83	'83	'83
15	Unemployment rate US	1	0	1	'83	'83	'83	'83	'83	'83
II. Prices (N=*9*)										
16	Total HICP-index	2	1	2	'60	'60	'60	'60	'60	'60
17	Core HICP-index	2	1	2	'62	'62	'60	'60	'76	'61
18	Producer prices	2	1	2	'81	'60	'62	'70	'60	'60
19	Commod. prices, tot.	2	1	2	'60	'60	'60	'60	'60	'60
20	Commod. prices, ind. mat.	2	1	2	'60	'60	'60	'60	'60	'60
21	Commod. prices, food-bev.	2	1	2	'60	'60	'60	'60	'60	'60

Continued on next page...

Table 5.7 – Continued from previous page...

Nr.	Variable	Transformation			Starting year					
		sa	ln.	dif.	EA	DE	FR	IT	ES	NL
22	Commod. prices, metals	2	1	2	‘60	‘60	‘60	‘60	‘60	‘60
23	Commod. prices, energy	2	1	2	‘60	‘60	‘60	‘60	‘60	‘60
24	Oil price	2	1	2	‘85	‘85	‘85	‘85	‘85	‘85
III. Monetary & financial indicators (N=*14*)										
25	M1	2	1	1	‘70	‘80	‘80	‘80	‘80	‘80
26	M3	2	1	1	‘70	‘70	‘70	‘70	‘70	‘70
27	Int. rate mortgage	2	0	1	‘03	‘82	‘80	‘95	‘84	‘80
28	3 month interest rate	2	0	1	‘94	‘60	‘64	‘60	‘60	‘60
29	10 year gov. bond yield	2	0	1	‘70	‘60	‘70	‘60	‘80	‘60
30	Headline stock-index	2	1	1	‘73	‘73	‘73	‘73	‘87	‘73
31	Basic material-index	2	1	1	‘73	‘73	‘73	‘73	‘87	‘73
32	Industrials stock-index	2	1	1	‘73	‘73	‘73	‘73	‘87	‘73
33	Cons. goods stock-index	2	1	1	‘73	‘73	‘73	‘73	‘87	‘73
34	Cons. service stock-index	2	1	1	‘73	‘73	‘73	‘73	‘87	‘73
35	Financials stock-index	2	1	1	‘73	‘73	‘73	‘73	‘87	‘73
36	Loans to the private sector	2	1	1	‘80	‘80	‘80	‘83	‘80	‘82
37	Exchange rate, $ per EUR	2	1	1	‘74	‘74	‘74	‘74	‘74	‘74
38	Real eff. exchange rate	2	1	1	‘70	‘70	‘70	‘70	‘70	‘70
IV. Surveys (N=*14*)										
39	Ind. conf. - headline	1	0	1	‘85	‘85	‘85	‘85	‘87	‘85
40	Ind. conf. - orders	1	0	1	‘85	‘85	‘85	‘85	‘87	‘85
41	Ind. conf. - stocks	1	0	1	‘85	‘85	‘85	‘85	‘87	‘85
42	Ind. conf. - prod. expect.	1	0	1	‘85	‘85	‘85	‘85	‘87	‘85
43	Ind. conf. - empl. expect.	1	0	1	‘85	‘85	‘85	‘85	‘87	‘85
44	Cons. conf. - headline	1	0	1	‘85	‘85	‘85	‘85	‘86	‘85
45	Cons. conf. - exp. fin. sit.	1	0	1	‘85	‘85	‘85	‘85	‘86	‘85
46	Cons. conf. - exp. ec. sit.	1	0	1	‘85	‘85	‘85	‘85	‘86	‘85
47	Cons. conf. - exp. unemp.	1	0	1	‘85	‘85	‘85	‘85	‘86	‘85
48	Cons. conf. - exp. maj. pur.	1	0	1	‘85	‘85	‘85	‘85	‘86	‘85
49	PMI United States	1	0	1	‘60	‘60	‘60	‘60	‘60	‘60
50	OECD leading ind. UK	1	1	1	‘60	‘60	‘60	‘60	‘60	‘60
51	OECD leading ind. US	1	1	1	‘60	‘60	‘60	‘60	‘60	‘60
52	OECD comp. leading ind.	1	1	1	‘70	‘61	‘70	‘62	‘76	‘61

Notes: entries denote variable number, name, category, transformation and starting year for each country in the dataset. Sa: 1= seasonal adjustment at the source; 2= seasonal adjustment by US Census X-12 method. Ln.: 0= no logarithm; 1= logarithm. Dif.: 1= first difference; 2= second difference. Country: EA: Euro area; DE: Germany; FR: France; IT: Italy; ES: Spain; NL: Netherlands.

5.B Number of factors in dynamic factor models

Tables 5.8–5.13 show the forecasting performance for the factor models considered in the main text of Chapter 5 with one to five and seven factors, respectively.

Table 5.8: Forecasting performance dynamic factor models, 1992.I–2012.IV, one factor

	EA	DE	FR	IT	ES	NL
	MSFE					
AR(2)						
Backcast	*0.32*	*0.76*	*0.16*	*0.47*	*0.31*	*0.45*
Nowcast	*0.39*	*0.78*	*0.21*	*0.53*	*0.35*	*0.49*
1Q ahead forecast	*0.45*	*0.81*	*0.28*	*0.60*	*0.41*	*0.53*
2Q ahead forecast	*0.49*	*0.81*	***0.34***	*0.66*	***0.46***	*0.56*
All horizons	*0.42*	*0.79*	*0.26*	*0.57*	*0.39*	*0.51*
	Relative to MSFE AR(2)					
SW						
Backcast	0.64	0.79	0.77	0.72	0.83	0.74
Nowcast	0.78	0.92	0.78	0.87	0.73	0.79
1Q ahead forecast	0.88	1.08	0.88	0.90	0.87	0.90
2Q ahead forecast	0.89	1.12	**0.98**	**0.88**	**0.98**	0.96
All horizons	0.83	1.00	0.88	0.86	0.86	0.86
BR						
Backcast	**0.49**	**0.63**	1.00	0.64	1.11	0.62
Nowcast	**0.56**	**0.69**	1.13	0.73	1.28	**0.69**
1Q ahead forecast	0.74	0.85	1.15	0.86	1.05	0.84
2Q ahead forecast	**0.82**	**0.92**	1.11	**0.88**	**1.03**	0.97
All horizons	**0.69**	**0.79**	1.12	**0.80**	1.11	**0.81**
BM						
Backcast	0.74	0.86	1.29	0.86	1.26	0.77
Nowcast	0.75	0.96	1.33	0.87	1.11	0.79
1Q ahead forecast	0.89	1.06	1.22	0.86	1.20	0.81
2Q ahead forecast	1.01	1.17	1.10	0.83	1.24	0.85
All horizons	0.87	1.03	1.21	0.85	1.20	**0.81**
CFM						
Backcast	0.49	0.57	0.97	0.58	**0.87**	0.69
Nowcast	0.52	0.63	0.98	0.64	0.88	0.67
1Q ahead forecast	0.65	0.76	**0.92**	0.77	**0.92**	0.74
2Q ahead forecast	0.79	0.87	0.95	0.83	0.95	**0.86**
All horizons	0.64	0.73	**0.95**	0.73	**0.91**	0.75

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts with model specification with **one** factor. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Table 5.9: Forecasting performance dynamic factor models, 1992.I–2012.IV, two factors

	EA	DE	FR	IT	ES	NL
	MSFE					
AR(2)						
Backcast	*0.32*	*0.76*	*0.16*	*0.47*	*0.31*	*0.45*
Nowcast	*0.39*	*0.78*	*0.21*	*0.53*	*0.35*	*0.49*
1Q ahead forecast	*0.45*	*0.81*	*0.28*	*0.60*	*0.41*	*0.53*
2Q ahead forecast	*0.49*	*0.81*	***0.34***	*0.66*	***0.46***	*0.56*
All horizons	*0.42*	*0.79*	*0.26*	*0.57*	*0.39*	*0.51*
	Relative to MSFE AR(2)					
SW						
Backcast	0.67	0.81	0.81	0.75	**0.82**	**0.69**
Nowcast	0.80	0.93	0.79	0.88	0.74	0.76
1Q ahead forecast	0.89	1.10	0.89	0.90	0.84	0.94
2Q ahead forecast	0.89	1.13	**0.97**	**0.88**	**0.98**	0.99
All horizons	0.84	1.01	0.89	0.87	0.86	0.87
BR						
Backcast	**0.49**	**0.62**	1.01	**0.61**	1.11	0.64
Nowcast	**0.55**	0.72	1.13	0.69	1.26	**0.68**
1Q ahead forecast	0.70	**0.83**	1.14	0.84	1.03	0.81
2Q ahead forecast	**0.80**	**0.91**	1.11	**0.88**	**1.02**	0.96
All horizons	**0.66**	**0.79**	1.11	**0.78**	1.09	**0.79**
BM						
Backcast	0.83	0.82	1.34	0.87	1.19	0.73
Nowcast	0.91	0.95	1.40	0.87	1.06	0.75
1Q ahead forecast	1.06	1.12	1.31	0.88	1.18	0.83
2Q ahead forecast	1.18	1.25	1.18	**0.86**	1.26	**0.91**
All horizons	1.03	1.06	1.29	0.87	1.18	**0.82**
CFM						
Backcast	0.46	0.60	**0.87**	0.57	0.77	0.70
Nowcast	0.51	**0.73**	0.91	0.62	0.86	0.68
1Q ahead forecast	0.63	0.81	0.89	0.76	0.95	0.72
2Q ahead forecast	0.78	0.88	0.92	0.83	0.97	0.85
All horizons	0.62	0.77	**0.91**	0.72	**0.91**	0.75

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts averaged across model specifications with **one–two** factors. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). * denotes the Diebold-Mariano test is significant at the 10% level. AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Table 5.10: Forecasting performance dynamic factor models, 1992.I–2012.IV, three factors

	EA	DE	FR	IT	ES	NL
	MSFE					
AR(2)						
Backcast	*0.32*	*0.76*	*0.16*	*0.47*	*0.31*	*0.45*
Nowcast	*0.39*	*0.78*	*0.21*	*0.53*	*0.35*	*0.49*
1Q ahead forecast	*0.45*	*0.81*	*0.28*	*0.60*	*0.41*	*0.53*
2Q ahead forecast	*0.49*	*0.81*	***0.34***	*0.66*	***0.46***	*0.56*
All horizons	*0.42*	*0.79*	*0.26*	*0.57*	*0.39*	*0.51*
	Relative to MSFE AR(2)					
SW						
Backcast	0.67	0.82	0.82	0.76	**0.80**	**0.71**
Nowcast	0.79	0.94	0.80	0.88	0.77	0.78
1Q ahead forecast	0.89	1.12	**0.89**	0.89	0.86	0.94
2Q ahead forecast	0.90	1.16	**0.96**	**0.87**	1.01	1.00
All horizons	0.84	1.03	0.89	0.86	0.87	0.88
BR						
Backcast	0.49	**0.63**	1.00	0.61	1.08	0.65
Nowcast	**0.56**	0.72	1.12	**0.68**	1.17	**0.70**
1Q ahead forecast	**0.70**	0.83	1.13	**0.83**	0.97	0.82
2Q ahead forecast	**0.79**	**0.91**	1.10	**0.87**	**0.99**	0.95
All horizons	**0.66**	0.79	1.10	**0.77**	1.04	**0.80**
BM						
Backcast	0.70	**0.63**	1.31	0.83	1.09	0.72
Nowcast	0.84	0.84	1.44	0.86	1.00	0.75
1Q ahead forecast	0.99	1.05	1.35	0.90	1.14	0.85
2Q ahead forecast	1.09	1.17	1.22	**0.86**	1.24	0.94
All horizons	0.94	0.95	1.32	0.87	1.13	0.83
CFM						
Backcast	0.44	0.60	**0.88**	0.54	0.75	**0.70**
Nowcast	0.52	**0.77**	0.91	0.65	0.85	0.66
1Q ahead forecast	0.64	**0.84**	0.89	0.77	0.97	0.71
2Q ahead forecast	0.77	0.88	0.92	0.82	0.98	0.85
All horizons	0.62	0.79	**0.90**	0.72	**0.91**	0.74

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts averaged across model specifications with **one–three** factors. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). * denotes the Diebold-Mariano test is significant at the 10% level. AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Table 5.11: Forecasting performance dynamic factor models, 1992.I–2012.IV, four factors

	EA	DE	FR	IT	ES	NL
			MSFE			
AR(2)						
Backcast	*0.32*	*0.76*	*0.16*	*0.47*	*0.31*	*0.45*
Nowcast	*0.39*	*0.78*	*0.21*	*0.53*	*0.35*	*0.49*
1Q ahead forecast	*0.45*	*0.81*	*0.28*	*0.60*	*0.41*	*0.53*
2Q ahead forecast	*0.49*	*0.81*	*0.34*	*0.66*	***0.46***	*0.56*
All horizons	*0.42*	*0.79*	*0.26*	*0.57*	*0.39*	*0.51*
			Relative to MSFE AR(2)			
SW						
Backcast	0.68	0.85	**0.87**	0.77	**0.77**	**0.70**
Nowcast	0.82	0.95	0.85	0.87	0.76	0.78
1Q ahead forecast	0.92	1.12	**0.93**	0.88	0.85	0.94
2Q ahead forecast	0.94	1.18	**0.99**	**0.85**	1.03	1.00
All horizons	0.87	1.04	**0.93**	0.85	0.87	0.88
BR						
Backcast	0.49	0.64	1.00	0.62	1.05	0.65
Nowcast	**0.56**	0.73	1.12	**0.68**	1.11	**0.69**
1Q ahead forecast	**0.70**	0.83	1.12	**0.83**	0.94	0.81
2Q ahead forecast	**0.79**	**0.91**	1.10	**0.87**	0.98	0.96
All horizons	**0.66**	0.79	1.10	**0.77**	1.01	**0.80**
BM						
Backcast	0.58	0.49	1.23	0.75	1.02	**0.69**
Nowcast	0.74	**0.77**	1.39	0.82	0.96	**0.70**
1Q ahead forecast	0.92	1.03	1.32	0.89	1.13	0.79
2Q ahead forecast	1.05	1.15	1.20	**0.86**	1.21	**0.92**
All horizons	0.87	0.90	1.28	0.84	1.10	**0.79**
CFM						
Backcast	0.39	0.59	0.83	0.53	0.70	**0.70**
Nowcast	0.52	**0.79**	**0.87**	0.67	**0.79**	0.65
1Q ahead forecast	0.64	**0.86**	0.87	0.78	0.95	0.70
2Q ahead forecast	0.76	0.88	0.90	0.82	**0.99**	0.85
All horizons	0.62	**0.80**	0.87	0.73	**0.89**	0.73

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts averaged across model specifications with **one–four** factors. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). * denotes the Diebold-Mariano test is significant at the 10% level. AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Table 5.12: Forecasting performance dynamic factor models, 1992.I–2012.IV, five factors

	EA	DE	FR	IT	ES	NL
			MSFE			
AR(2)						
Backcast	*0.32*	*0.76*	*0.16*	*0.47*	*0.31*	*0.45*
Nowcast	*0.39*	*0.78*	*0.21*	*0.53*	*0.35*	*0.49*
1Q ahead forecast	*0.45*	*0.81*	*0.28*	*0.60*	*0.41*	*0.53*
2Q ahead forecast	*0.49*	*0.81*	*0.34*	*0.66*	***0.46***	*0.56*
All horizons	*0.42*	*0.79*	*0.26*	*0.57*	*0.39*	*0.51*
			Relative to MSFE AR(2)			
SW						
Backcast	0.69	0.84	0.90	0.79	0.75	**0.69**
Nowcast	0.84	0.96	**0.88**	0.87	0.76	0.75
1Q ahead forecast	0.94	1.12	0.95	**0.86**	0.89	0.95
2Q ahead forecast	0.98	1.19	1.00	**0.84**	1.07	1.02
All horizons	0.89	1.05	**0.95**	0.85	**0.89**	0.88
BR						
Backcast	0.49	0.64	1.01	0.62	1.04	0.64
Nowcast	**0.56**	0.73	1.12	0.68	1.07	**0.69**
1Q ahead forecast	**0.70**	0.83	1.11	**0.83**	**0.91**	0.81
2Q ahead forecast	**0.80**	**0.91**	1.09	**0.87**	0.97	0.96
All horizons	**0.67**	0.79	1.09	**0.77**	0.99	0.80
BM						
Backcast	0.50	0.42	1.19	0.68	0.91	**0.65**
Nowcast	0.69	**0.75**	1.35	0.79	0.87	**0.66**
1Q ahead forecast	0.88	0.98	1.29	0.87	1.02	**0.73**
2Q ahead forecast	1.03	1.12	1.20	**0.85**	1.14	**0.89**
All horizons	0.83	**0.86**	1.26	0.82	1.00	**0.75**
CFM						
Backcast	0.37	0.58	0.79	0.53	0.68	**0.68**
Nowcast	0.54	0.82	0.85	**0.69**	**0.78**	0.64
1Q ahead forecast	0.65	**0.87**	0.86	0.79	**0.94**	0.69
2Q ahead forecast	0.76	0.88	0.89	0.83	**0.99**	0.85
All horizons	0.62	**0.81**	0.86	0.74	0.88	0.72

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts averaged across model specifications with **one–five** factors. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). * denotes the Diebold-Mariano test is significant at the 10% level. AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Table 5.13: Forecasting performance dynamic factor models, 1992.I–2012.IV, seven factors

	EA	DE	FR	IT	ES	NL
	MSFE					
AR(2)						
Backcast	*0.32*	*0.76*	*0.16*	*0.47*	*0.31*	*0.45*
Nowcast	*0.39*	*0.78*	*0.21*	*0.53*	*0.35*	*0.49*
1Q ahead forecast	*0.45*	*0.81*	*0.28*	*0.60*	*0.41*	*0.53*
2Q ahead forecast	*0.49*	*0.81*	*0.34*	*0.66*	***0.46***	*0.56*
All horizons	*0.42*	*0.79*	*0.26*	*0.57*	*0.39*	*0.51*
	Relative to MSFE AR(2)					
SW						
Backcast	0.70	0.86	0.90	0.86	0.88	**0.68**
Nowcast	0.83	0.99	**0.89**	0.90	**0.77**	0.71
1Q ahead forecast	0.95	1.17	**0.94**	**0.86**	**0.91**	0.95
2Q ahead forecast	0.99	1.25	1.02	**0.85**	1.08	1.07
All horizons	0.90	1.09	0.95	0.87	**0.93**	0.88
BR						
Backcast	0.49	0.65	1.01	0.63	1.04	0.63
Nowcast	**0.57**	**0.74**	1.12	**0.70**	1.04	0.67
1Q ahead forecast	**0.72**	0.84	1.11	**0.83**	0.89	0.81
2Q ahead forecast	**0.81**	**0.92**	1.09	**0.87**	0.96	0.97
All horizons	**0.68**	0.80	1.09	**0.78**	0.97	0.79
BM						
Backcast	0.39	0.39	0.99	0.59	0.79	**0.65**
Nowcast	**0.58**	0.72	1.13	0.70	**0.78**	**0.65**
1Q ahead forecast	0.77	0.94	1.13	0.78	**0.91**	**0.70**
2Q ahead forecast	0.94	1.08	1.12	0.83	1.07	**0.86**
All horizons	0.72	**0.82**	1.11	**0.75**	**0.91**	**0.73**
CFM						
Backcast	0.34	0.56	0.75	0.53	0.63	**0.64**
Nowcast	0.56	0.84	0.83	**0.71**	0.73	0.60
1Q ahead forecast	0.66	**0.88**	0.87	**0.80**	**0.91**	0.67
2Q ahead forecast	0.76	0.88	0.90	0.83	**0.99**	0.85
All horizons	0.62	**0.81**	0.86	0.75	0.85	0.70

Notes: entries denote the RMSFE for an AR(2) (in italics); for all other models they denote the RMSFE relative to the RMSFE of an AR(2). All forecasts averaged across model specifications with **one–seven** factors. Grey cells denote models with the lowest RMSFE. Entries in bold denote models whose RMSFE is at most 10% larger than the RMSFE of the best model and smaller than the AR(2). * denotes the Diebold-Mariano test is significant at the 10% level. AR(2): autoregression of order 2, SW: Stock and Watson (2002b) model, BR: augmented Bańbura and Rünstler (2011) model, BM: Bańbura and Modugno (2014) model, CFM: Bräuning and Koopman (2014) model. EA: Euro area, DE: Germany, FR: France, IT: Italy, ES: Spain, NL: the Netherlands.

Chapter 6

Summary and conclusion

It is well documented that forecasters around the world failed to forecast the depth and duration of the financial crisis of 2008–2009. Indeed, even ascertaining the current state of the economy is a challenging task. A key indicator of the state of the economy is the growth rate of real gross domestic product (GDP), which is available on a quarterly basis only and subject to a substantial publication delay. Most countries publish an initial estimate of quarterly GDP around six weeks after the end of a quarter. The initial GDP estimate can be subject to substantial revisions, as more information becomes available to statistical offices over time. Fortunately, there is a lot of statistical information related to economic activity that is published on a more frequent and timely basis. This information includes data on industrial production, unemployment, consumer confidence, stock markets and prices of goods and services. The forecasting literature has recently developed several statistical approaches to exploit this potentially very large information set in order to improve the assessment of real GDP growth in the adjacent quarters. Examples are bridge equations (BEQ), factor models, mixed-data sampling regression models (MIDAS) and mixed-frequency vector autoregressive (MFVAR) models. These so-called "nowcasting" models differ in their approach to the practical problems of how to handle a large-scale information set and the fact that the auxiliary variables are observed at different frequencies and with different publication lags.

Apart from model-based forecasts, practitioners can also take advantage of published forecasts made by professional analysts. From a practical point of view, such forecasts are cheap and easy to use. Moreover, as an expression of the "wisdom of crowds", they may reflect much more information than the statistical information set, which is inevitably limited. So which model should practitioners use, and how should they incorporate the forecasts of professional analysts? This is not a trivial question. An accurate assessment of GDP growth in the adjacent quarters is essential as a starting point for the medium-term forecasts of macro-economic models. The ranking of the different approaches to forecasting GDP growth in the adjacent quarters and the extent to which this ranking varies with the prediction horizon or economic circumstances has to be determined by empirical analysis. On these issues the jury is still out, as large-scale comparative studies are scarce.

This dissertation is motivated by this gap in the existing literature and has made the following contributions to the literature: Firstly, it provides new insights from a large-scale comparative study of the current generation short-term forecasting models. Secondly, new evidence is presented on the usefulness of the quarterly forecasts of professional analysts by constructing a new database of these forecasts for the seven most important industrialized countries (G7), the euro area, Spain and the Netherlands. Lastly, it presents several modifications to the current generation short-term forecasting models. The remainder of this chapter summarizes the main findings of this dissertation and discusses possible avenues for future research.

Chapter 2 provides new evidence on the forecasting performance of statistical linear models and professional analysts for the Netherlands over the period 1995–2010. Furthermore, it is examined whether the forecasts of professionals could have enhanced the forecasts of the mechanical models. Chapter 3 enriches the analysis of Chapter 2 in two directions. Firstly, the number of countries is extended to the euro area and its five largest countries (Germany, France, Italy, Spain and the Netherlands). Secondly, the number of short-term forecasting models is increased from four to twelve. These extensions add robustness to the outcomes of the forecasting model “horse race” in Chapter 2. In addition, the proposed extensions can shed new light on the issue as to which model features are especially valuable for short-term forecasting. Chapter 4 addresses a new question within the empirical literature on nowcasting, i.e. whether predictions by analysts are able to improve GDP forecasts generated by statistical procedures in a *truly* real-time context. Based on the relatively good forecasting performance of the dynamic factor model in the previous chapters and the widespread use of this nowcasting model amongst practitioners, Chapter 5 presents an analysis on the most appropriate specification of the dynamic factor model.

One of the main conclusions of this dissertation is that employing factor analysis to summarize the available monthly information clearly delivers better results than pooling single-indicator models. The dynamic factor model (DFM) displays the best forecasting capabilities overall. Its ability to incorporate more than one factor, and thus more information, is key to this result. Factor-augmented MFVAR and MIDAS models produce more accurate one-quarter ahead forecasts due to their richer dynamic specification, but this advantage is limited to the period following the financial crisis. Moreover, it is important that a model uses all available monthly information and allows for autoregressive terms in the forecasting equation. All of these effects are more pronounced during the crisis period, implying that the cost of employing a suboptimal forecasting model is larger in periods of high volatility. The BVAR is the best quarterly model. It performs quite well for Germany, the Netherlands and Spain in the more stable period of the Great Moderation. This finding suggest that Bayesian estimation is a fundamentally different way of extracting information from a large dataset, which may deliver benefits, even if the model makes inefficient use of the available monthly information.

The scope for improving GDP forecasts by combining the "views" of various statistical models is rather limited in economic terms. This is particularly true during a volatile period when a reliable assessment of the current economic situation and the short-term prospect is most needed, unfortunately. The forecasts of professional analysts, which contain judgmental elements, appear to be a different category. The forecasting ability of analysts remarkably improves after the financial crisis, making them a tough competitor for the mechanical models since 2008. In the stable pre-crisis period, the DFM tends to outperform professional analysts. But in the volatile post-crisis period, newly released forecasts of professional analysts for the current and next quarter are superior to the DFM. This new insight is in line with research that suggests that analysts pay more attention and devote more effort to forecasting in volatile times. The results also suggests that the value of subjective insights of professional analysts is greater when there are at least some data available on the pertaining quarter. Another finding is that the relative forecasting advantage of professional analysts declines as their forecasts age. This is related to the fact that the DFM is able to fully exploit all newly released monthly data. However, combining the forecasts of professional analysts and the DFM delivers sizable gains in forecasting ability of statistical models for most countries, even when the forecasts of analysts are somewhat dated. A final insight from this thesis is that it is infeasible to determine the optimal rule for combining the forecasts of professional analysts and the DFM in real-time. Overall, using a simple average of the various combination rules provides the best hedge against misspecification and instability. The main conclusion from studying different dynamic factor model specifications is that the recently proposed collapsed dynamic factor model has the edge over the other factor models considered. This conclusion holds both before and during and after the financial crisis, for most countries and for most forecasting horizons. The findings in this thesis may be useful to policy makers, financial analysts and economic agents, as information on where the economy stands and is heading in the short run is of crucial importance.

Based on the results in this dissertation, several promising avenues for future research can be distinguished. Firstly, considering the relatively high forecasting accuracy of the quarterly Bayesian VAR model in Chapter 3, more research is needed on mixed-frequency Bayesian VAR models, as these could develop into viable and practical alternatives to factor models. Secondly, a relevant issue is finding ways to incorporate more than one factor into the MIDAS model to improve its capabilities for nowcasting and backcasting. Thirdly, more research is needed on making statistical procedures robust to extreme observations and structural breaks in the data generating process. Fourthly, unveiling the black-box nature of the forecasting process of professional analysts could provide important insights into which type of models they use, what indicators they use and how and when they add judgment to their forecasts. A final promising topic for future research is to investigate the usability of “big-data” sources for nowcasting GDP growth. Current research indicates that information on search queries and on-line user activity can potentially be helpful, but only for a limited number of indicators, such as: unemployment, housing prices and household consumption. New big-data sources such as scanner data, data on traffic intensities and payment card data have the potential to help econometricians model and forecast a broader range of indicators.

The new insights in this dissertation and the possible avenues for future research outlined above can hopefully contribute to keeping short-term forecasting the exciting research area it currently is.

Samenvatting (Summary in Dutch)

Je kunt de dingen pas voorspellen als ze hebben plaatsgevonden. Deze gevleugelde uitspraak van de Franse schrijver Ionesco is ook van toepassing op de financiële crisis van 2008–2009. Voor de meeste economen kwam de abrupte economische krimp bij aanvang van de crisis als een volslagen verrassing.

Ook het bepalen van de huidige stand van de economie is voor beleidsmakers een uitdagende taak. Dit komt doordat het bruto binnenlands product (BBP) –dat de bestedingen van alle economische actoren bij binnenlandse ondernemingen meet– alleen op kwartaalbasis en met een behoorlijke publicatievertraging beschikbaar komt. Zo werd de sterke krimp van de Nederlandse economie in het eerste kwartaal van 2009 pas halverwege mei –zes weken na afloop van het kwartaal– door het Centraal Bureau voor de Statistiek (CBS) gepubliceerd. Beleidsmakers beschikken wel over een grote hoeveelheid aan tijdiger en frequenter gepubliceerde indicatoren die een deel van de economie beschrijven, zoals de industriële productie, het werkloosheidspercentage, de inflatie en het sentiment onder consumenten en bedrijven. Het is echter geen sinecure deze maandelijkse stroom aan informatie te vertalen naar een inschatting van de BBP-groei. De eerste modelmatige uitdaging is dat de publicatiefrequentie van het BBP en de indicatoren verschilt. Het BBP is uitsluitend op kwartaalbasis beschikbaar, terwijl de meeste indicatoren een maandfrequentie hebben. Een tweede uitdaging is dat de maandindicatoren verschillende publicatievertragingen hebben. Zo is de informatie van financiële markten na afloop van een maand direct beschikbaar, maar wordt de groei van de industriële productie pas zes weken na afloop van een maand gepubliceerd. Gegeven deze uitdagingen is het de vraag hoe een beleidsmaker optimaal gebruik kan maken van de continue stroom aan indicatoren om de BBP-groei te voorspellen. Een benadering is gebruik te maken van een statistisch model. Hierin valt veel te kiezen. De laatste jaren is namelijk een groot aantal econometrische zogenoemde 'nowcasting' modellen ontwikkeld die –elk op hun eigen wijze– de stroom aan maandelijkse informatie vertalen naar een inschatting van de BBP-groei in de nabije kwartalen. Een andere benadering is gebruik te maken van de gepubliceerde voorspellingen van professionele analisten. Deze voorspellingen vormen voor de beleidsmaker een eenvoudig en goedkoop alternatief voor het ontwikkelen en onderhouden van een (ingewikkeld) statistisch model. Voordeel is ook dat analisten gebruik kunnen maken van meer informatie dan de indicatoren die zijn opgenomen in een mechanisch statistisch model. Welke benadering leidt tot de beste inschatting van de BBP-groei is een empirische kwestie, waarover tot op heden geen uitsluitsel bestaat. Dit komt doordat de bestaande studies meestal een klein aantal landen of modellen betreft. Bovendien hebben de meeste studies betrekking op de periode vóór de recente financiële crisis. Dit is een gemis omdat het voor beleidsmakers juist dan cruciaal is te beschikken over een accurate inschatting van de BBP-groei. Tot slot is weinig bekend over de wijze waarop beleidsmakers de voorspelling van modellen en professionele analisten in de praktijk zouden kunnen combineren om de BBP-groei nauwkeuriger te voorspellen.

Dit proefschrift heeft tot doel deze lacunes op te vullen en draagt op verschillende vlakken bij aan de wetenschappelijke literatuur over nowcasting. Uit een grootschalige vergelijking van de huidige generatie nowcasting-modellen voor verschillende landen volgen nieuwe inzichten over welk modeltype het best voorspelt, wat daarbij de cruciale modelkenmerken zijn, of het combineren van verschillende modeltypes de voorspelfout verkleint en of de rangorde van de modellen in termen van voorspelkracht wijzigt in periodes van grote volatiliteit. Daarnaast draagt het proefschrift bij aan de kennisvorming over de voorspelkracht van professionele analisten. Daartoe wordt voor elk land een database samengesteld die de kwartaalvoorspellingen van analisten voor de BBP-groei bevat. Hierdoor kan worden vastgesteld onder welke omstandigheden analisten beter of slechter voorspellen dan statistische modellen en of het combineren van de voorspelling van analisten en modellen zinvol is. Tot slot worden in dit proefschrift rekenregels ontwikkeld om de voorspellingen van modellen en professionele analisten te combineren. Onderstaand wordt een korte beschrijving gegeven van de inhoud van elk hoofdstuk.

Hoofdstuk 2 analyseert welk statistisch model het meest geschikt is om de Nederlandse BBP-groei in de nabije kwartalen te voorspellen. Daarnaast wordt onderzocht hoe de voorspelkracht van de modellen zich verhoudt tot de voorspelkracht van professionele analisten. De analyse is uitgevoerd over de periode 1995–2010. Omdat in deze periode zowel de recessie volgend op het knappen van de internetzeepbel (2001–2002) als de financiële crisis (2008–2009) vallen, kan een onderscheid gemaakt worden tussen de voorspelkracht voor en na (diepe)recessies. Het hoofdstuk beschouwt twee strategieën om de maandelijkse stroom aan indicatoren te vertalen naar een inschatting van de BBP-groei. De eerste strategie is het combineren van modelvoorspellingen die elk zijn afgeleid van één indicator. Door die voorspellingen te middelen kan een voorspelling voor de BBP-groei worden berekend. In dit hoofdstuk worden twee indicator-modellen beschouwd: het brugmodel (BEQ) en het vector autoregressieve model (VAR). In de tweede strategie wordt de gemeenschappelijke informatie in de maandindicatoren samengevat in één of enkele factoren. De factoren worden vervolgens gebruikt om de BBP-groei te voorspellen in een zogenoemd dynamisch factormodel (DFM). De voorspelkwaliteit van beide strategieën wordt vergeleken met de voorspelkwaliteit van een zogenoemd 'naïef' voorspelmodel. In zo'n model wordt geen gebruik gemaakt van de stroom aan maandindicatoren, maar wordt uitsluitend gebruik gemaakt van het (recente) beloop van het BBP. De voorspelkwaliteit van beide strategieën wordt ook afgezet tegen de voorspelkwaliteit van professionele analisten. De kwartaalvoorspellingen van professionele analisten zijn afkomstig van papieren edities van de *Consensus forecasts*, verzameld door het bedrijf Consensus Economics. In de analyse wordt de informatie die de professionele analisten voorhanden hadden ten tijde van het maken van hun kwartaalvoorspelling zo goed mogelijk nagebootst. Daarbij wordt rekening gehouden met de publicatiekalender van de indicatoren, maar *niet* met de (soms aanzienlijke) revisies in het BBP en de indicatoren (*pseudo* real-time). De empirische resultaten wijzen uit dat het loont om bij het voorspellen van de Nederlandse BBP-groei een statistisch model te gebruiken dat de beschikbare maandinformatie benut: de voorspelfout is over het algemeen kleiner dan van een naïef voorspelmodel. Het DFM heeft de hoogste voorspelprecisie, vooral bij het voorspellen van de BBP-groei in het lopende en voorgaande kwartaal. Dit resultaat wordt in belangrijke mate gedreven door de goede voorspelkwaliteit tijdens de crisisjaren. De voorspelfout van professionele analisten is groter dan de voorspelfout van het DFM, met uitzondering van de financiële crisis. Deze uitkomst suggereert dat analisten vooral in periodes van grote volatiliteit gebruik maken van meer informatie dan de maandindicatoren die zijn opgenomen in het DFM.

Hoofdstuk 3 breidt de analyse uit hoofdstuk 2 in drie richtingen uit. Ten eerste wordt het aantal geanalyseerde statische modellen uitgebreid. Ten tweede wordt getracht meer inzicht te krijgen in de 'succesfactoren' van een voorspelmodel. Zo wordt onder meer onderzocht of de voorspelkracht van indicatormodellen toeneemt indien factoren –zoals in het DFM– of individuele indicatoren –zoals in de standaard VAR– worden opgenomen. Ten derde wordt het aantal landen, waarvoor een 'voorspelcompetitie' is uitgevoerd, vergroot. Naast Nederland wordt de voorspelkracht van de modellen ook onderzocht in de eurozone en haar vier grootste lidstaten Duitsland, Frankrijk, Italië en Spanje. Dit heeft tot doel vast te stellen of de voorspelkwaliteit van statische modellen en professionele analisten verschilt tussen landen. Een van de belangrijkste bevindingen is dat het DFM voor de meeste landen en de meeste voorspelhorizonten de kleinste voorspelfout heeft. Dit komt vooral door de relatief hoge voorspelkwaliteit van het DFM tijdens de crisisjaren. Dit resultaat bevestigt de uitkomst van hoofdstuk 2 voor een groter aantal landen. De belangrijkste 'succesfactor' van het DFM lijkt de flexibiliteit om meer dan één factor op te kunnen nemen in het model. Daarnaast maakt het model relatief goed gebruik van de voorhanden zijnde maandelijkse informatie. Het combineren van verschillende modeltypes –bijvoorbeeld een DFM en een VAR– blijkt de voorspelkwaliteit niet tot nauwelijks te verbeteren. Dit ligt anders voor de voorspelling van professionele analisten. De voorspelling van analisten blijkt andere

–aanvullende– informatie te bevatten die mechanische modellen niet of onvoldoende (kunnen) meenemen, die de voorspelkwaliteit verbetert.

Hoofdstuk 4 beantwoordt een nieuwe nieuwe vraag in de literatuur: hoe en wanneer moet een beleidsmaker de voorspelling van statistische modellen en professionele analisten combineren om de BBP-groei zo accuraat mogelijk te ramen? De analyse wordt uitgevoerd in een *volledige* real-time opzet. Verschil met de *pseudo* real-time analyse in de vorige hoofdstukken is dat nu rekening wordt gehouden met het feit dat de gepubliceerde waarde van maandindicatoren en het BBP in de loop der tijd (fors) kunnen worden gereviseerd. De analyse is uitgevoerd voor de periode 1999 tot en met 2013 voor de zogenoemde G7-landen (VS, VK, Canada, Duitsland, Frankrijk en Italië). Voor elk van deze landen is een database geconstrueerd die voor elke maand een momentopname ('foto's') bevat van de beschikbare maandindicatoren. De momentopnames zijn zodanig gemaakt dat ze nagenoeg overeenstemmen met de datum waarop professionele analisten hun voorspelling maakten. In de analyse worden twee manieren beschouwd om de BBP-voorspelling van modellen en professionele analisten te combineren: een gerestricteerd gewogen gemiddelde en een ongerestricteerde lineaire combinatie. In dit hoofdstuk wordt enkel het DFM geanalyseerd, vanwege de aangetoonde goede voorspelkwaliteit in de eerdere hoofdstukken van dit proefschrift. Een van de belangrijkste inzichten uit hoofdstuk 4 is dat professionele analisten een stevige concurrent zijn voor het DFM. Tijdens en na de financiële crisis is de 'verse' –net gepubliceerde– voorspelling van professionele analisten voor het lopende en eerstvolgende kwartaal vrijwel niet te verslaan door het DFM. In de periode voor de financiële crisis verdwijnt dit voordeel en is de voorspelfout van analisten groter dan van het DFM. Dit patroon suggereert dat professionele analisten meer aandacht besteden aan het opstellen van hun BBP-voorspelling in volatiele tijden. De voorspelfout van professionele analisten is het kleinst voor 'verse' voorspellingen van de BBP-groei. De relatieve voorspelfout ten opzichte van het DFM neemt toe naarmate de voorspelling van analisten veroudert. Dit komt omdat het DFM blijft 'leren' door gebruik te maken van nieuwe maandinformatie. Een andere uitkomst is dat de voorspelfout van het DFM kan worden verkleind door deze te combineren met de voorspellingen van professionele analisten, ook wanneer laatstgenoemde enigszins gedateerd zijn. Er lijkt geen 'gouden' combinatieregel te bestaan voor het combineren van de voorspellingen van het DFM en professionele analisten. Een ongewogen gemiddelde van de onderzochte combinatieregels blijkt te leiden tot de kleinste voorspelfout.

Hoofdstuk 5 analyseert wat de meest geschikte modelspecificatie van het DFM is voor het voorspellen van de BBP-groei in de nabije kwartalen. Een grondige analyse van de specificatie van het DFM is van belang in het licht van het wijdverbreide gebruik van dit modeltype en de vastgestelde relatief goede voorspelkracht in de eerdere hoofdstukken van dit proefschrift. Het hoofdstuk analyseert de voorspelprestatie van vier veelgebruikte specificaties. In het eerste deel van het hoofdstuk worden de technische details van de modellen beschreven en worden, gebaseerd op de inzichten uit de eerdere hoofdstukken van dit proefschrift, enkele technische modificaties doorgevoerd. In het tweede deel van het hoofdstuk wordt de voorspelkracht van de (gemodificeerde) factormodellen getoetst voor de eurozone en haar vijf grootste lidstaten in de periode 1992–2012. Net als in de voorgaande hoofdstukken wordt onderscheid gemaakt tussen de voorspelkracht vóór, tijdens en na de financiële crisis. Belangrijkste conclusie is dat het recent door Bräuning and Koopman (2014) voorgestelde factormodel voor de meeste landen, de meeste voorspelhorizonten, en zowel vóór als tijdens en na de financiële crisis een hogere voorspelprecisie heeft dan de andere factormodellen. Daarnaast blijken de voorgestelde modelmodificaties te leiden tot een kleinere voorspelfout.

Hoofdstuk zes vat de belangrijkste uitkomsten samen en schetst een aantal mogelijkheden voor toekomstig onderzoek gebaseerd op de uitkomsten van dit proefschrift.

Bibliography

Ager, P., M. Kappler, and S. Osterloh (2009). The accuracy and efficiency of the Consensus forecasts: a further application and extension of the pooled approach. *International Journal of Forecasting 25*(1), 167–181.

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters 80*(23-24), 1806–1813.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of euro area GDP growth. *Econometrics Journal 14*(1), 25–44.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting 20*(3), 447–460.

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

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

Bai, J. and S. Ng (2008). Large dimensional factor analyis. *Foundations and Trends in Econometrics 3*(2), 89–163.

Bańbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-casting and the real-time data flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, pp. 195 – 237. Elsevier.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Bańbura, M., D. Giannone, and L. Reichlin (2011). Nowcasting. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*, pp. 63–90. Oxford University Press.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Bańbura, M. and G. Rünstler (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting 27*(2), 333–346.

Barhoumi, K., O. Darn, and L. Ferrara (2010). Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting 29*(1-2), 132–144.

Batchelor, R. (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the Consensus. *Applied Economics 33*(2), 225–235.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*(4), 451–468.

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

Bloor, C. and T. Matheson (2011). Real-time conditional forecasts with Bayesian VARs: an application to New Zealand. *The North American Journal of Economics and Finance 22*(1), 26–42.

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

Bräuning, F. and S. J. Koopman (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting 30*, 572–584.

Breitung, J. and S. Eickmeier (2006). Dynamic factor models. *AStA Advances in Statistical Analysis 90*(1), 27–42.

Caggiano, G., G. Kapetanios, and V. Labhard (2011). Are more data always better for factor analysis? Results for the euro area, the six largest euro area countries and the UK. *Journal of Forecasting 30*(8), 736–752.

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

Camacho, M. and G. Perez-Quiros (2011). Spain-sting: Spain short-term indicator of growth. *The Manchester School 79*(1), 594–616.

Camba-Mendez, G., G. Kapetanios, R. J. Smith, and M. R. Weale (2001). An automatic leading indicator of economic activity: forecasting GDP growth for European countries. *Econometrics Journal 4*(1), 37.

Carriero, A., T. E. Clark, and M. Marcellino (2015a). Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics 30*(1), 46–73.

Carriero, A., T. E. Clark, and M. Marcellino (2015b). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(4), 837–862.

Castle, J. L., M. P. Clements, and D. F. Hendry (2015). Robust approaches to forecasting. *International Journal of Forecasting 31*(1), 99–112.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics 3*(2), 240–264.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*(2), 245–276.

Cheung, C. and F. Demers (2007). Evaluating forecasts from factor models for Canadian GDP growth and core inflation. Working Papers 07-8, Bank of Canada.

Clark, T. E. and M. W. McCracken (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics 25*(1), 5–29.

Clemen, R. T. (1986). Linear constraints and the efficiency of combined forecasts. *Journal of Forecasting 5*(1), 31–38.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic forecasting with mixed-frequency data: forecasting output growth in the United States. *Journal of Business & Economic Statistics 26*(4), 546–554.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature 49*(1), 72–100.

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

de Winter, J. M. (2011). Forecasting GDP growth in times of crisis: private sector forecasts versus statistical models. DNB Working Papers 320, De Nederlandsche Bank.

den Reijer, A. (2013). Forecasting Dutch GDP and inflation using alternative factor model specifications based on large and small datasets. *Empirical Economics 44*(2), 435–453.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics 164*(1), 188–205.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi-maximum likelihood approach for large approximate dynamic factor models. *The Review of Economics and Statistics 94*(4), 1014–1024.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Oxford University Press.

Fair, R. C. and R. J. Shiller (1990). Comparing information in forecasts from econometric models. *American Economic Review 80*(3), 375–389.

Foroni, C. and M. Marcellino (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. *International Journal of Forecasting 30*(3), 554–568.

Garratt, A., K. Lee, and S. Vahey (2008). Real-time probability forecasts of UK macroeconomic events. *National Institute Economic Review 203*(1), 78–90.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013). Combining expert forecasts: can anything beat the simple average? *International Journal of Forecasting 29*(1), 108–121.

Ghysels, E., A. Sinko, and R. Valkanov (2007). MIDAS regressions: further results and new directions. *Econometric Reviews 26*(1), 53–90.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics 97*(2), 436–451.

Giannone, D., L. Reichlin, and L. Sala (2005). Monetary policy in real time. In *NBER Macroeconomics Annual 2004*, Volume 19, pp. 161–224. National Bureau of Economic Research.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Golinelli, R. and G. Parigi (2014). Tracking world trade and GDP in real time. *International Journal of Forecasting 30*(4), 847–862.

Granger, C. W. and R. Ramanathan (1984). Improved methods of combining forecasts. *Journal of Forecasting 3*(2), 197–204.

Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor models. *Journal of the American Statistical Association 102*(478), 603–617.

Hindrayanto, I., S. J. Koopman, and J. M. de Winter (2016). Forecasting and nowcasting economic growth in the euro area using factor models. *International Journal of Forecasting* (32), 2184–1305.

Hubert, P. (2014). Revisiting the greenbook's relative forecasting performance. Revue de l'OFCE-Débats et politiques 137, Observatoire français des conjonctures économiques.

Isiklar, G., K. Lahiri, and P. Loungani (2006). How quickly do forecasters incorporate news? Evidence from cross-country surveys. *Journal of Applied Econometrics 21*(6), 703–725.

Jansen, W. J. and J. M. de Winter (2016). Improving model-based near-term GDP forecasts by subjective forecasts: a real-time exercise for the G7 countries. DNB Working Papers 507, De Nederlandsche Bank.

Jansen, W. J., X. Jin, and J. M. de Winter (2016). Forecasting and nowcasting real GDP: comparing statistical models and subjective forecasts. *International Journal of Forecasting 32*(2), 411 – 436.

Jungbacker, B. and S. J. Koopman (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *Econometrics Journal 18*, C1–C21.

Jungbacker, B., S. J. Koopman, and M. van der Wel (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control 35*, 1358–1368.

Kitchen, J. and R. Monaco (2003). Real-time forecasting in practice: the US treasury staff's real-time GDP forecast system. *Business Economics 38*(4), 10–28.

Koopman, S. J. and A. Harvey (2003). Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control 27*(7), 1317–1333.

Kuzin, V., M. Marcellino, and C. Schumacher (2011). MIDAS vs. mixed-frequency VAR: nowcasting GDP in the euro area. *International Journal of Forecasting 27*(2), 529–542.

Kuzin, V., M. Marcellino, and C. Schumacher (2013). Pooling versus model selection for nowcasting GDP with many predictors: empirical evidence for six industrialized countries. *Journal of Applied Econometrics 28*(3), 392–411.

Lahiri, K., G. Isiklar, and P. Loungani (2006). How quickly do forecasters incorporate news? Evidence from cross-country surveys. *Journal of Applied Econometrics 21*(6), 703–725.

Lahiri, K. and G. Monokroussos (2013). Nowcasting US GDP: the role of ISM business surveys. *International Journal of Forecasting 29*(4), 644 – 658.

Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior & Organization 48*(3), 265–280.

Laster, D., P. Bennett, and I. S. Geoum (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics 114*(1), 293–318.

Liebermann, J. (2014). Real-time nowcasting of GDP: a factor model vs. professional forecasters. *Oxford Bulletin of Economics and Statistics 76*(6), 783–811.

Loungani, P. and J. Rodriguez (2008). Economic forecasts. *World Economics 9*(2), 1–12.

Loungani, P., H. Stekler, and N. Tamirisa (2013). Information rigidity in growth forecasts: some cross-country evidence. *International Journal of Forecasting 29*(4), 605–621.

Luciani, M. (2014). Forecasting with approximate dynamic factor models: the role of non-pervasive shocks. *International Journal of Forecasting 30*(1), 20–29.

Lundquist, K. and H. O. Stekler (2012). Interpreting the performance of business economists during the great recession. *Business Economics 47*(2), 148–154.

Marcellino, M. and C. Schumacher (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: a model comparison for German GDP. *Oxford Bulletin of Economics and Statistics 72*(4), 518–550.

Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics 18*(4), 427–443.

Mariano, R. S. and Y. Murasawa (2010). A coincident index, common factors, and monthly real GDP. *Oxford Bulletin of Economics and Statistics 72*(1), 27–46.

Matheson, T. D. (2013). New indicators for tracking growth in real time. *OECD Journal: Journal of Business Cycle Measurement and Analysis 2013*(2), 51–71.

Meyler, A. and I. Rubene (2009). Results of a special questionnaire for participants in the ECB survey of professional forecasters (SPF). MPRA Paper 20751, University Library of Munich, Germany.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*(4), 1004–1016.

Reinhart, C. M. and K. S. Rogoff (2008). Is the 2007 US sub-prime financial crisis so different? An international historical comparison. *American Economic Review 98*(2), 339–44.

Reis, R. and M. W. Watson (2010). Relative goods' prices, pure inflation, and the Phillips correlation. *American Economic Journal: Macroeconomics 2*(3), 128–157.

Roodenburg, O. and A. H. den Reijer (2006). Dutch GDP data revisions: are they predictable and where do they come from? *Applied Economics Quarterly (formerly: Konjunkturpolitik) 52*(4), 337–356.

Roscam Abbing, M., J. de Jong, and J. Verbruggen (2010). Forecasting in times of crisis: CPB-forecasts during the great recession. CPB Document 207, CPB Economic Policy Analysis.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. V. Nieuwenhuyze (2009). Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of Forecasting 28*(7), 595–611.

Rünstler, G. and F. Sédillot (2003). Short-term estimates of euro area real GDP by means of monthly data. Working Paper Series 0276, European Central Bank.

Schneider, M. and M. Spitzer (2004). Forecasting Austrian GDP using the generalized dynamic factor model. Working Papers 89, Oesterreichische Nationalbank.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics 33*(3), 366–380.

Schumacher, C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting 32*(2), 257 – 270.

Schumacher, C. and J. Breitung (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting 24*(3), 386–398.

Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis 3*(4), 253–264.

Stark, T. (2010). Realistic evaluation of real-time forecasts in the survey of professional forecasters. Research report, Federal Reserve Bank of Philadelphia.

Stekler, H. O. (1991). Macroeconomic forecast evaluation techniques. *International Journal of Forecasting 7*(3), 375–384.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–162.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

Stock, J. H. and M. W. Watson (2011). Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook of Economic Forecasting*, Oxford Handbook of Economic Forecasting, pp. 35–59. Oxford University Press Oxford.

Timmermann, A. (2006). Forecast combinations. In C. G. G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 135 – 196. Elsevier.

Wallis, K. F. (1986). Forecasting with an econometric model: the ragged edge problem. *Journal of Forecasting 5*(1), 1–13.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *Journal of Econometrics 23*, 385–400.

This thesis contains four chapters that cast new light on the ability of professional analysts and statistical models to assess economic growth in the current quarter (nowcast) and its development in the near future. This is not a trivial issue. An accurate assessment of the current state of the economy is important as a starting point for medium-term forecasts, especially during times of heightened volatility, such as the recent financial crisis.

Nowadays, practitioners have a wealth of statistical models to choose from; but which one should they use? Can statistical models be modified to improve their forecasting accuracy? What are the gains from combining the forecasts of different statistical models? Did the financial crisis change the forecasting performance of statistical models relative to professional analysts? Can practitioners use the near-term outlook of professional analysts to improve the forecasting accuracy of statistical models? This thesis gives answers to these questions, providing new insights of interest to both academics and practitioners. Central to this research is the construction of a new dataset, comprised of the near-term economic growth forecasts of professional analysts, and the monthly indicators available when analysts made their forecasts.

Jasper de Winter (1972) studied Economics at the University of Amsterdam. From 1997 to 2003 he worked at SEO Economics where he conducted applied research for clients in the (non) profit sector. Since 2003, he has been working at De Nederlandsche Bank as a (senior) economist. He is currently working in the Econometrics and Models Department, and is a member of the Working Group on Forecasting of the European System of Central Banks. His main research interests are short-term forecasting and modeling.