CPB Netherlands Bureau for Economic Policy Analysis

Nowcasting GDP growth

The CPB will use a new time series model to support the forecast of GDP growth.
In this publication, we study a dynamic factor model which is capable of predicting based on the most up-to-date monthly and quarterly data.
This data is published earlier and more frequently than GDP growth.

The forecasts from the dynamic factor model for GDP growth for the previous quarter (backcast) and the current quarter (nowcast) outperform those from the current supporting BVAR model.

CPB - March 2024

Rob Luginbuhl, Loes Verstegen

1 Introduction

The growth rate of Gross Domestic Product (GDP) is an important indicator that we forecast at the CPB, but is also one of many key macroeconomic statistics that are published with a delay. The CBS publishes their initial estimate of quarterly GDP growth 45 days after the end of each quarter (the "flash" estimate).[1] As a result, we begin each of our bi-annual forecasting rounds without knowing value of GDP growth from the previous quarter.

In this research we develop a dynamic factor model (DFM) to exploit the information that is published earlier and possibly at higher frequency than GDP in order to obtain a more accurate estimate of output growth before the GDP figure is released. There is a substantial number of monthly data series available that are leading indicators for GDP with a shorter publication lag. These leading series include 'hard' information, for example on industrial production or household consumption expenditures, as well as 'soft' information such as surveys on consumer and producer confidence. These series can be included in a DFM to obtain better forecasts of quarterly GDP.

Our primary goal is to specify a mixed frequency DFM that combines the information from leading indicators with different publication delays and frequencies to obtain a GDP forecast for the current quarter. This forecast is commonly referred to as a nowcast. The DFM also enables us to predict the value of the previous quarter's GDP before it is released. Such a prediction is referred to as a backcast. We can use the DFM to update both the nowcast and backcast each time a new observation for a variable in the model is released.

A DFM has a number of advantages. To begin with, a DFM is able to model a large number of variables, as many as several hundred, in what is nonetheless a parsimonious model. It is able to achieve this by reducing the time series dynamics in the model down to a function of a small number of factors. In this way the model remains tractable and is able to avoid the problem of overfitting suffered by many large time series models. A DFM also allows us to combine the mixed frequency information contained in monthly and quarterly macroeconomic series in a single time series model. Another main advantage is that we can use the only partially available data at the end of the sample period (such datasets are said to have a ragged edge) to estimate the model's parameters and then obtain forecasts (including backcasts and nowcasts). This is because the estimation method for the DFM is able to handle datasets with missing observations. This contrasts favorably with the quarterly Bayesian vector autoregressive model (BVAR) in use at the CPB, which we currently use to complement our GDP growth forecasts from our principal structural macro model, Saffier[2]. This BVAR currently in use is not able to handle mixed frequency data, nor the ragged edge at the end of sample period, when some of the variables have missing values.[3]

Results from an out-of-sample forecasting competition indicate that the DFM is able to produce more accurate nowcasts and backcasts. The competition compares the forecasting performance of our DFM with the existing quarterly CPB BVAR, as well as with a naïve random walk specification for GDP growth. Our main focus is on stable economic times, but the DFM is also better at signaling turning points. The success of the DFM corroborates research4 that has demonstrated that mixed frequency DFMs are useful tools for nowcasting.

[1] The "regular" estimate of GDP growth is published 90 days after the end of each quarter, which is based on more detailed information than the flash estimate. After 2,5 years, the GDP figures are final except for revisions that take place every 5 to 10 years.
[2] See CPB (2021) for the extensive documentation of the macromodel Saffier 3.0.
[3] Although mixed-frequency BVARs do exist in the literature, and there are also techniques that allow them to be estimated with datasets with a ragged edge, neither of these are currently implemented at the CPB.

2 Motivation

The purpose of developing a DFM is to provide the most accurate forecasts of the quarterly growth rate of GDP possible for use as an additional input for the bi-annual CEP and MEV publications on the state of the Dutch economy. The primary instrument used at the CPB to produce forecasts for the CEP and MEV publications is the large structural macro model Saffier. This model is designed with a number of criteria in mind. It attempts to provide a high degree of detail on many aspects of the Dutch economy, but it also ensures the internal consistency of the forecasts it produces. The model is also developed based on underlying macroeconomic theory, thereby aiding in the interpretation of the model's outcomes. Lastly, but not least, the model is partially estimated and partially calibrated not only with the aim of obtaining accurate forecasts, but also with an eye on the economic interpretation of the parameters. This means that Saffier is designed for multiple purposes, namely for forecasting as well as scenario and policy analysis. Past research at the CPB has demonstrated that time series models specifically selected for their forecasting accuracy could outperform Saffier for the short forecasting horizon (Elbourne et al., 2008).

The macromodel Saffier is unable to make use of any data not included in the model. It is possible to include a partial update of the variables included in Saffier to obtain updated model-consistent forecasts based on the most recent, but still incomplete, data. In this sense, Saffier is able to accommodate the ragged edge in the most recent data needed to produce forecasts. However, there is a lot of data outside of the information set of Saffier which can be helpful for the forecasts, especially for a short horizon. A DFM can include a large number of disparate macroeconomic variables to obtain a forecast of output growth given the latest information available on these variables. The CPB expert can use that information as a basis for expert judgement to adjust the Saffier outcome.

Currently the CPB already has an operational time series model, a BVAR, designed to obtain forecasts of output growth based on a substantial number of macroeconomic time series, but it has some shortcomings compared to a DFM. For one, the BVAR does not allow for the use of mixed-frequency data and therefore only allows for the inclusion of quarterly data, whereas many important leading indicators are monthly series. The current operational version also does not permit the use of datasets which end with a ragged edge. The DFM is able to accommodate both mixed-frequency datasets with monthly and quarterly data, as well as data sets with a ragged edge. The main advantage of being able to use a dataset with a ragged edge is that every single new data release can be used in the DFM. With the CPB BVAR it is necessary to wait until the variable with the largest publication delay is released.

We note that the literature has demonstrated the feasibility of specifying a mixed-frequency BVAR.[5] Estimation techniques based on the Kalman Filter also make it possible to use datasets which contain missing values, and thereby also with a ragged edge. Owing to the restricted focus of this current research we compare the DFM forecasts with those from the current CPB quarterly version of the BVAR. An interesting extension of this research is the development of mixed-frequency BVARs which allow for their estimation based on datasets with a ragged edge, which was outside the scope of this research. Since BVAR models are already in use at the CPB, it was decided to investigate the usefulness of a dynamic factor model for forecasting.

[4] See Banbura et al. (2013) for an overview of the literature.
[5] One of the leading papers on this topic is by Brave et al. (2019).

An important aspect of a DFM and a BVAR is that they both permit the formulation of a time series model that includes a large number of variables without resulting in a model that suffers from the problem of overfitting. This problem occurs in time series models which include many variables and thereby also include a large number of parameters to be estimated. This is the case for example with large vector autoregressive (VAR) models. Without going into any detail here, the BVAR effectively achieves parameter shrinkage to avoid the problem of overfitting through the use of priors designed for this purpose. The DFM also is able to avoid the pitfall of overfitting, which we discuss below.

3 The Model

3.1 Why a dynamic factor model?[6]

Policy advisors and professional forecasters employ many disparate sources of information to obtain as accurate an estimate of GDP growth as possible. If this estimate is based on a single model which includes many data series, then it is generally better to use a more parsimonious model to avoid the problem of overfitting. A dynamic factor model is useful in this respect, as it is assumed that the series in the model are driven by only a handful of shared unobserved factors and an idiosyncratic residual. The unobserved factors capture the joint dynamics of the data. Since many economic series display a high degree of co-movement, the unobserved factors can usually explain a large part of the dynamics of each series. In the literature, there are many papers providing empirical evidence that this is the case for large panels of macroeconomic variables.[7]

A dynamic factor model is a multivariate times series model that can be written in the state space form. In the literature, many papers based on VAR (vector autoregressive) and BVAR models employ the state space representation of their models to facilitate their estimation.[8] The state space form is a system of two equations, which we will show mathematically in section 3.2. The first equation is called the *measurement equation* that links the observed series to a latent state process, which is associated with the unobserved state of the economy. The second equation is called the *transition equation*. It describes how the latent state evolves over time. One can think of our dynamic factor model as a state space model where GDP growth is linked to latent state factors through the measurement equation. How these latent factors evolve is then specified in the transition equation.

The information set of forecasters contains information on a monthly basis, or perhaps at an even higher frequency. A dynamic factor model is able to deal with these mixed frequencies by treating the low frequency GDP series as high frequency data with periodically missing observations. A dynamic factor model can easily solve a missing data problem caused by this mixed frequency data set. Another example of a missing data problem is the *ragged* edge of the data set. That is, the time of the last available observation differs from series to series due to the different publication delays for each series. For example, stock prices are available almost immediately, whereas the publication of the monthly consumption series for the Netherlands is published with a delay of approximately 40 days. A dynamic factor model allows for all available information to be used, whereas many other methods require the panel of data to have realizations for all series.

[6] In this section we highlight a number of models commonly used for nowcasting. However, our aim was not to provide a comprehensive literature review. There is a large literature on nowcasting approaches and the performance of different models. See for example Banbura et al. (2013).
[7] For an overview of empirical evidence, see Stock and Watson (2011).
[8] More generally there is a large body of literature about state space time series models, see Durbin and Koopman (2001).

The state space representation of the dynamic factor model makes the use of the Kalman filter possible. The Kalman filter makes projections for both observed and state variables. Therefore, it can easily cope with the missing data problems of a mixed frequency data set with a ragged edge.[9] An additional advantage of the Kalman filter for this state space model is that the role of data releases in signaling changes in economic activity can be evaluated. The model produces forecasts for all variables and thus allows for the extraction of the unexpected component from data releases. This tells us to what extent the data release is a surprise (or *news*) compared to the projection. When the GDP nowcast is revised, the model can help to link the revision to the *news* factor from each variable in the model. As such, the dynamic factor model can also help to understand the changes in the nowcast for GDP growth over time, as well as to evaluate the significance of each data release. In our research, we do not use this approach but an alternative measure to measure the role of data releases in forecasting economic growth. This will be explained in more detail in Section 7.

[9] The R-package nowcasting that we use actually makes use of an alternative routine to the Kalman filter provided by Giannone et al. (2008).

Mixed-frequency BVAR and other approaches for nowcasting

Next to a DFM, another multivariate dynamic model that can be written in the state space form is a mixed frequency Bayesian VAR (MF-BVAR). Such a model could include time series, for example GDP, that are only measured at quarterly frequencies. The corresponding monthly values for GDP are then treated as unobserved. The resulting MF-BVAR can be represented as a state space model, and the Kalman filter is used to perform the model estimation and obtain forecasts. Like the DFM, a MF-BVAR also meets the requirements that we set out in the beginning of this section. Given that any state space model can handle missing observations through the use of the Kalman filter, a MF-BVAR can handle both mixed frequency data and data sets with ragged edges. Additionally, the news component from released data can also be extracted when using an MF-BVAR.

Nowcasting is common at different central banks and other institutions, where both approaches, DFM and MF-BVAR, are being used. A leading example is the New York Fed Staff nowcast (Bok et al., 2018). The weekly updated nowcast was published since 2016. In September 2021 The New York Fed suspended the publication of the nowcast due to the uncertainty around the pandemic and the volatility of the data, but the nowcast releases were reintroduced in September 2023. The Dutch Central Bank (DNB) also uses a dynamic factor model named DFROG. Jansen et al. (2016) compare the performance of 12 statistical models, such as bridge equation, quarterly VAR models (Bayesian and factor augmented), MIDAS models, an DFM and an MF-BVAR. The authors find that for most countries, the DFM performs best, especially for the backcast and the nowcast of GDP growth. A recent study by the IMF (Dauphin et al., 2022) describes the use of DFMs and machine learning algorithms to nowcast GDP growth across European countries. The DFMs tend to perform strongly during normal times, whereas machine learning methods are better in identifying turning points. The Chicago Fed, for example, uses a dynamic factor approach for the Chicago Fed National Activity Index (CFNAI). The European Central Bank in fact made a toolbox available on their website, which can be used to forecast with Bayesian VARs (Dieppe et al., 2016). This BEAR toolbox is also capable of estimating mixed frequency BVARs.

Other approaches for nowcasting are based on single equation rather than multivariate dynamic models, or use machine learning methods. For example, a *Bridge equation* handles the mixed frequency problem by temporally aggregating the data to a lower frequency. Moreover, ragged edges can be handled using auxiliary ARMA or VAR models. In a *MIDAS (mixed data sampling) model*, the high-frequency data is incorporated into the model equation using a lag distribution. These models typically suffer from the curse of dimensionality and can quickly become too large to estimate. Another disadvantage of this type of models is that the impact of the news component on the update of the nowcast cannot be computed. A third option for nowcasting with mixed frequency data is *blocking*. The model is specified at the lowest frequency. For any variable at a higher frequency, for example a monthly variable in a quarterly model, there will be three different time series, one for each month of the quarter. Then, by blocking one could turn a mixed frequency data set into for example a regular quarterly BVAR model for nowcasting. Next to this type of models, there is a growing literature on the usefulness of machine learning models for nowcasting, also in crisis times, e.g. Richardson et al. (2021) and Coulombe et al. (2022). At the CPB, a machine learning method is used to support the unemployment forecasts (Scheer, 2019).

3.2 What exactly is a dynamic factor model?

A dynamic factor model can be written in the state space form as follows.

1) $x_t = \mu + \Lambda f_t + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.\ N(0, \Sigma_\varepsilon)$
2) $f_t = \sum_{i=1}^{p} A_i f_{t-i} + B\, u_t, \qquad u_t \sim i.i.d.\ N(0, I_q)$

Equation (1) is the measurement equation that links the vector of observed variables x_t to an intercept μ and a vector of r unobserved common factors f_t. Here, the vector $x_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})'$ is a vector with N time series, which can initially be of different frequency. The matrix Λ specifies the combination of the unobserved factors f_t that make up each observed variable. The vector of disturbances ε_t represent the measurement errors with the diagonal covariance matrix Σ_ε. Equation (2) is the transition equation that poses a VAR structure on the factors f_t, with the number of lags given by p. The matrix A_i contains the autoregressive parameters of the factors at lag i. The matrix B determines the size of the variance of each factor's residual, and also potentially allows for the elements of the state vectors to be correlated. Both ε_t and u_t are normally distributed with covariance matrices Σ_ε and I_q. The latter identity matrix has rank q, which can be smaller than the number of factors r. In addition, it is also assumed that the idiosyncratic component ε_t is not related to u_t for all lags, i.e. $E[\varepsilon_t u_{t-k}] = 0$ for any k.

In the vector of observed variables, there may be missing observations due to the mixed frequency of the time series. Our variable of interest is GDP growth, which is a quarterly variable, but the frequency of x_t is monthly. Hence, we construct a partially observed monthly counterpart of GDP growth, where quarterly GDP is observable in the third quarter of the month and there is a missing value for the remaining two months in the quarter.

We estimate the dynamic factor model by maximum likelihood, using the Expectation-Maximization (EM) algorithm, as in Doz et al. (2012). This algorithm is popular when it comes to estimating parameters for models with unobserved components, as is the case in this state space model. There are two steps in the algorithm that are iterated until convergence is reached;

1) In the Expectation-step of the algorithm, given the parameter estimates from the previous iteration (or the initialization) the conditional expectation of the unobserved components (the latent state) is calculated using Kalman smoothing.
2) In the Maximization-step of the algorithm, the maximization of the likelihood function given the values of the unobserved components from the previous step leads to new parameter estimates.

Before the first iteration, principal components are computed and treated as if they were the true common factors. Then the model parameters are estimated by OLS regression, and these estimates are used to initialize the parameters before the first iteration.

4 Data

In our dynamic factor model, we include a data set with 20 variables related to consumption, trade, production, financial markets and confidence indicators. Initially, our DFM had 27 variables, but we excluded 7 of these variables due to their availability or forecasting performance. We elaborate further on this in section 6.1.1. In Table 4.1, an overview of the total data set is presented, with a short description for all the variables in the dataset, the source of the data, their respective frequencies and publication delays. There is a substantial number of series related to the Dutch economy, but we also include variables related to other economies, such as the US or Europe.[10] Most variables are available on a monthly basis, except for GDP growth, employee wages and the capacity utilization rate. The source of most of the data is Statistics Netherlands, whereas some indicators are taken from the Dutch Central Bank, Datastream or the OECD. Our data set consists of 115 quarterly observations, from the first quarter of 1995 until the third quarter of 2023.

Table 4.1 Overview of the data set

Variable	*Description*	*Frequency*	*Publication delay (in days)*	*Source*
GDP growth	Quarter-on-quarter growth rate of real GDP	quarterly	45	Statistics Netherlands
Producer confidence	Composite indicator for producers in Dutch industry	monthly	0	Statistics Netherlands
Consumer confidence	Composite indicator of survey among households	monthly	0	Statistics Netherlands
Consumption	Domestic consumption of households	monthly	45	Statistics Netherlands
Export	Total export of goods (incl. re-exports)	monthly	45	Statistics Netherlands
Import	Total import of goods (incl. imports for the purpose of re-exports)	monthly	45	Statistics Netherlands
US Policy Uncertainty	Three component index	monthly	0	Economic Policy Uncertainty
EU Policy Uncertainty	European News Index	monthly	0	Economic Policy Uncertainty
Volatility US	Volatility in the US measured by S&P	monthly[11]	0	Datastream
AEX stock index	Ultimo stock index of the AEX	monthly	0	Dutch Central Bank
Dollar-euro exchange rate	Monthly average	monthly	0	Datastream
Long interest rate	Monthly average	monthly	0	Dutch Central Bank
Short interest rate	Monthly average	monthly	0	Dutch Central Bank
Consumer price index	Month-on-month change	monthly	0	Statistics Netherlands

[10] If not explicitly indicated otherwise, the variable in Table 4.1 refers to the Dutch economy.
[11] The volatility index based on the US S&P has a daily frequency, which we convert into a monthly series.

Energy prices	Energy component in derived CPI	monthly	0	Statistics Netherlands
House prices	Month-on-month change	monthly	25	Statistics Netherlands
Leading indicator EU4	Composite leading indicator for 4 largest European economies	monthly	20	OECD
Leading indicator US	Composite leading indicator	monthly	20	OECD
Bankruptcies	Total number of bankruptcies per month	monthly	42	Statistics Netherlands
Industrial production	Average day production of Dutch industry	monthly	40	Statistics Netherlands
Utilization rate industry	Survey result on experienced use of capacity in the industry	quarterly (first month of quarter)	0	Statistics Netherlands
World trade monitor	Quarter-on-quarter growth trade volume	monthly	53	CPB
Checkable deposits	Deposits that can be converted into cash or can be transferred (overnight deposits)	monthly	30	Dutch Central Bank
Long-term deposits	Deposits with agreed maturity	monthly	30	Dutch Central Bank
Deposits with notice	Deposits that are redeemable at notice	monthly	30	Dutch Central Bank
Money supply M3	Excluding currency in circulation	monthly	30	Dutch Central Bank
Employee wages	Wages of employees in the market sector	quarterly	45	Statistics Netherlands

5 Forecasting competition

The forecasts accuracy of the dynamic factor model is assessed via a forecasting competition, and compared to two benchmark models: the CPB-BVAR and a random walk.[12] In this forecasting competition, for each new observation of a variable in the DFM, we re-estimate the model to obtain a new *out-of-sample* forecast of GDP growth for the as-yet unpublished previous quarter (this is the backcast), as well as for the current quarter (the nowcast) and the four quarters which follow. Note that the backcast is only produced until the GDP growth rate for the preceding quarter is published. This is a *recursive window approach*, as the first period of the data set remains 1995Q1 throughout the forecasting competition. Hence, the window that is used to estimate the model increases each time a variable is updated.[13]

[12] The forecasting competition is almost continuous, as for every new observation arriving an out-of-sample forecast is produced. Hence, it is not possible to compare the DFM to the published outcomes of Saffier, since those are only produced twice a year.
[13] We have also experimented with a rolling window approach, where the length of the sample period (window) used for estimation is fixed. This resulted in slightly less accurate forecasts for GDP growth for all forecasting horizons.

For each data update of the variables in the DFM between January 2008 and September 2023, a *quasi real-time vintage* is constructed to re-estimate the DFM.[14] *Real-time data* is data that was available in the past at the time of its initial publication and that has not been updated to reflect subsequent revisions. This data is the most appropriate data to use in forecasting competitions. However, we do not have real-time data available. Instead, we use the latest available data set which does include revisions, in order to reconstruct vintages in the past while taking into account the publication delay. As the DFM, BVAR and random walk all are based on these same quasi real-time vintages, the resulting model comparisons are fair.

The forecasts of the DFM are evaluated and compared to the benchmark models for two time periods, with and without the Great Recession and Covid period. In general, we would like the model to be set up such that it works well in relatively stable times. Including the crisis periods in the forecasting competition could have a large influence on the choices for the optimal setting of the model. For this reason, the forecasting competition first focuses on the period from the beginning of January 2010 until the end of December 2019. We refer to this period as the competition period. [15] This period was a reasonably stable one given that it excludes both the earlier Great Recession and later Covid period. However, a DFM can still be useful in economic crises, as it can help identify economic turning points earlier. This is possible thanks to the timely use of frequent data updates in the DFM to obtain new forecasts. To demonstrate this, we also have forecasting results for the DFM based on the longer sample period from 2008Q1 to 2023Q3, which we refer to as the full sample period. This is the maximum sample period we can use to evaluate the DFM's forecasting performance given the availability of the data in our model.[16] In section 6.3, we show how the DFM would have forecasted during the Great Recession and Covid period.

Our forecasting evaluation consists of forecasting GDP up to four quarters ahead. Each time a new variable in the DFM becomes available we obtain a new forecast from the DFM. These updates then continue up until the moment that the actual GDP figure is published. This occurs during the backcast quarter (the quarter after the actual date of the GDP value). As a result, we obtain forecasts for four, three, two and one quarters ahead, a quarter of nowcasts (estimated GDP values during the quarter in which the GDP value falls), and finally a (partial) quarter of backcasts. During each quarter we are able to produce multiple forecast updates because the DFM is a mixed-frequency model primarily of monthly data. Therefore, most variables are updated three times each quarter, which allows us to produce up to three updates for each variable per quarter. In practice, the number of updates is fewer due to multiple variables being published on the same day.

The nowcasting R-package produces a variable update schedule based on the calendar from each year of the sample period. As the calendar shifts each year, the result is a varying variable update schedule. The fact that the order in which variable updates become available changes from year to year makes it difficult to properly combine the forecasting results over time to obtain root mean squared forecast errors. Therefore, we make use of a representative update schedule for the variables in the DFM, which closely follows the actual publication dates in each year, if not exactly. For variables which share the same publication day, we estimate the DFM based on the updates of the combined set of variables. This results in fewer forecasting updates per quarter than would be the case if each variable update resulted in a new DFM estimation and forecasting step. This procedure then allows us to combine the forecast errors over time to obtain root mean squared forecast errors (RMSFEs).

[14] The *nowcasting* R-package by De Valk et al. (2019) is used to estimate and forecast with the dynamic factor model.
[15] Note that this period closely matches the zero lower bound episode, which might reflect a structural break in the time series. In operational use, the parameters of all specifications in the mixture model will be re-estimated, which allows for the model to incorporate changes in the relationship between the variables.
[16] We are actually able to obtain forecasts for the DFM starting in 2006Q1. To obtain forecasts up to four quarters ahead, we must begin our forecasting evaluation of the DFM therefore in 2007Q1. However, we also compare the forecasting performance of the DFM with the BVAR and a naïve random walk. Given the limitations of the BVAR, we cannot begin a fair forecasting competition before 2008Q1.

In our forecasting competition, we loop over the number of factors, *r*, in the DFM and the number of lags, *p*, in the AR process for the factors. There are a number of options for determining values for *r* and *p*. One could leave it to the judgement of the researcher, there are test criteria that could be used, and it is possible to loop over both parameters to determine which values produce the best out-of-sample forecasts.[17] We have opted for the third option and compare the entire set of DFM models based on different values of *r* and *p* to our benchmark models. We also evaluate a mixture model based on the forecasts from the DFM models with different number of factors, *r*, and lags, *p*. In addition to the mixture DFM, we also evaluate the average forecasts from the combination of the mixture DFM and BVAR, as well as the combination of the mixture DFM, BVAR and random walk.[18]

The benchmark for the nowcasting and forecasting performance of the DFM is our regular BVAR model and a naïve random walk forecast. The regular CPB-BVAR model for GDP growth has as the disadvantage of being a quarterly model which requires a balanced data set without ragged edges for forecasting. As such, for most periods within our forecasting competition, it is not possible to include the latest information in this model. Note that the BVAR includes the same (quarterly equivalent) variables as those that we use in the DFM to ensure a fair model comparison. A second benchmark is the naïve forecast, where GDP growth is assumed to follow a random walk. This implies that at all forecast horizons the forecast is equal to the last observed value, or in other words a horizontal line. The main criterion that we use to compare the forecasts from the DFM to those of the benchmarks is the *root mean squared forecast error* (RMSFE). Additionally, we also make use of the *mean absolute forecast errors* to construct the average improvement values for each variable. These improvement values are discussed in more detail below in Section 6.1.

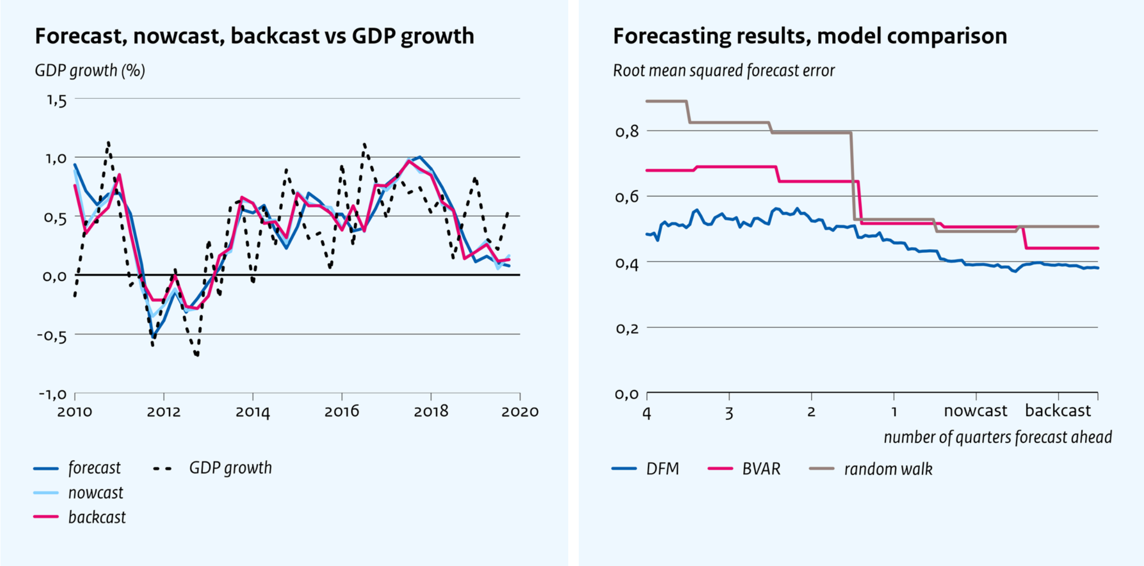
6 Results

The preferred model specification, that will be used in the future CPB forecasts for CEP and MEV, is the mixture model of the 12 DFMs with $r = 2,\ldots,5$ factors and $p = 1,\ldots3$ lags in the VAR equation. In this section, we will show that this model specification has a good forecasting performance and produces the most stable outcomes. Although forecasting accuracy is very important for this nowcasting instrument, for practical use it is also desirable that the model produces reasonable estimates instead of outliers. In Figure 6.1 (left) we show the one quarter-ahead forecast, nowcast and backcast over time from the quasi real-time nowcasting exercise for the mixture DFM with $r = 2,\ldots,5$ factors and $p = 1,\ldots3$. The forecasts, nowcasts and backcasts are given by the last available update in each of these forecast horizons. The estimates shown in this section are obtained from the competition period covering the period from 2010Q1 until 2019Q4.

[17] If the goal would be to optimize over the different model specifications for the number of factors *r* and number of lags *p*, the out-of-sample forecasts could be made on a different time period (test period) than the optimization over *r* and *p* itself (cross-validation period). See for example Scheer (2019). However, the goal of this forecasting exercise is not to optimize over the different models, but to reflect the (what turns out to be small) differences in forecasting accuracy. Because of the limited amount of time periods in our forecasting competition and the final choice for the mixture model, we believe the choice not to incorporate an extra validation period is justified.

[18] It should be noted that each time the DFM is re-estimated based on a variable update, the nowcasting R-package function takes roughly twenty minutes to convergence and produce new a forecast. As a result, producing a complete set of forecasts covering the competition period (for a given number of factors *r* and lags *p*) requires roughly two full days of computing based on parallel calculations with 15 cores. More variables, more factors, and more lags all slow the convergence of the DFM estimation even further. This limits our ability to perform an extensive analysis of the optimal model specification, both in terms of the variables to include in the model as well as in terms of the number of factors and lags in the DFM. For this reason, we also did not experiment with the option to estimate factors for different subgroups of variables, even though this is possible in the nowcasting R-package.

Figure 6.1 Predictions of DFM model compared to observed GDP growth (left) and RMSFE of DFM compared to benchmark models (right) for period between 2010Q1 and 2019Q4



When comparing the mixture DFM to the two benchmark models, we find that the DFM performs better for all forecasting horizons. Based on estimates for the competition period from the mixture DFM for $r = 2,\dots,5$ and $p = 1,\dots,3$, we obtain the RMSFEs shown in Figure 6.1 (right). Note that there are many DFM forecasting updates in each quarterly forecasting horizon, since we produce a DFM forecast for every variable update. The BVAR and random walk are quarterly models. In the case of the BVAR this means that we are only able to obtain new forecasts once a complete quarter of observations is available. For the random walk we can obtain new forecasts as soon as the flash estimate for GDP is released, which occurs 45 days after the beginning of a quarter. This explains why these models have only six updates over the six forecasting horizons, and why the timing of their updates is slightly different. The figure shows that the mixture DFM performs better than the BVAR and naïve random walk over all horizons, going from the 4 quarter ahead forecast to the backcast. For all 3 models, the RMSFE tends to decrease when the horizon becomes shorter. For a long horizon, the random walk performs much worse, but for the shorter horizons the forecasting accuracy of the three models becomes more similar.

Because the DFM is updated multiple times each quarter, we also report the average RSMFE per quarterly forecast horizon in Table 6.1. Here, the same conclusion arises. The mixture DFM outperforms the BVAR and random walk. Besides, for all models, the RMSFE tends to decrease as more information arrives and the forecasting horizon becomes shorter. In Table 6.1 we report the RMSFE for the DFM in two different ways. The first row contains the RMSFEs averaged over the updates for each forecast horizon, and then averaged over time (2010Q1-2019Q4). The second row of the table refers to the RMSFE related to the final forecast from the last update in that quarter (rather than the average over all updates). Generally, these final values (row 2) are only slightly lower than the average RMSFE over all updates in a quarter (row 1). It is noteworthy that the additional benefit of the information contained in the final update is modest compared to the RMSFE we are already able to obtain in the rest of the quarter. This is also reflected in the relatively modest decrease in the RMSFE between horizons in the first row of the table. To some extent the information contained in some of the variable updates during the quarter will be more relevant for forecasting GDP growth than the last variable update of the quarter. To test for the statistical significance of the observed differences between the RMSFEs we obtain for the mixture DFM and for the baseline models, we perform the Diebold Mariano test (DM) on the forecast errors for each update of the DFM. Table 6.1 indicates when the mixture DFM perform statistically better than the BVAR or random walk at the 10%, 5% and 1% significance level. In the case of the row of averaged RMSFEs, we conservatively report on the value of the largest p-value from the relevant forecast horizon.

Table 6.1 Average root mean squared forecast errors per forecast horizon, over competition period (2010Q1-2019Q4)

Model	Forecast horizon (quarters)					
	4	3	2	1	nowcast	Backcast
mixture DFM	$0.52^{\dagger}$	$0.54^{\dagger}$	0.50	0.42	$0.39^{*,\dagger}$	$0.39^{*,\dagger\dagger\dagger}$
mixture DFM (last update)	$0.53^{\dagger}$	$0.54^{\dagger\dagger}$	0.47	$0.39^{**,\dagger}$	$0.39^{*,\dagger\dagger\dagger}$	$0.38^{**,\dagger\dagger\dagger}$
BVAR	0.68	0.67	0.60	0.51	0.48	0.44
Random walk	0.86	0.81	0.68	0.51	0.50	0.51

Note: *, ** or *** denotes the Diebold-Mariano test for comparison with the BVAR is significant at the 10%, 5% or 1% level, respectively. †, †† or ††† denotes the Diebold-Mariano test for comparison with the random walk is significant at the 10%, 5% or 1% level, respectively.

6.1 Marginal improvements

The results of the forecasting exercise can be used to derive a measure of how much each variable in the DFM contributes to the successful prediction of GDP. These improvement values can then help us with model selection, because they help us to determine which variables can be best dropped from the model. They also provide us with a set of important contributors to the forecast of GDP growth. The forecast improvement of a variable is defined as the change in the absolute value of the forecast error relative to the previous absolute forecast error obtained before the update of the variable. If the absolute forecast error decreases, this results in a positive improvement.[19]

For both the competition period (2010Q1-2019Q4) and the entire sample period (2008Q1-2023Q3), we calculate the average improvement value for each variable of the mixture DFM. These are shown in Table 6.2. The values represent the average size of the decreases in the absolute forecast error due to the addition of a new observation of the respective variable. A strongly positive average improvement generally indicates that the variable is a leading indicator and that it helps to predict the future value of GDP. A negative average value however can be caused by a number of factors. If the variable is published with a long lag, then it will tend to be less informative. Quarterly values, such a GDP growth itself, have particularly long publication lags. Of course, some variables tend to be lagging indicators and are therefore generally not as useful in predicting GDP. Prices such as the CPI tend to be lagging indicators. We see that both CPI and GDP have large negative improvements. Bankruptcies and wages are also generally regarded as lagging indicators, while the World Trade Monitor (WTM) has a publication lag of around 50 days. These variables also have negative or small average improvement values. Negative values could point to the possibility for further improvement of the DFM. In order to find out whether excluding the variables with an average negative contribution would lead to better forecasts, the forecasting competition could be repeated looping over the variable set. Due to the limited time and scope of our analysis, this extension is left for future research.

[19] When multiple variables have updated values published on the same day, we obtain a marginal improvement by removing a variable from the set of conditioning variables to measure the resulting increase in the absolute forecast error caused by omitting the variable. For these "marginal" values we subtract the absolute value of the forecast error when the variable is used to produce the forecast from the absolute value of the forecast error without the variable. These "marginal" estimates are only used to calculate the improvements, and do not play a role in any of the other model statistics or figures.

Capacity utilization, the house price index, imports, exports and consumption generally have strongly positive average improvements.[20] This holds for both the competition period and the full sample period which includes the Great Recession and Covid period. In fact, many of the variables have a positive impact when we look at the competition period. Once we include the quarters of the Great Recession and Covid, we find that the variation among the average improvements becomes large. From Table 6.2 it seems that producer confidence, for example, has a much more valuable contribution to the forecast, whereas the leading indicator for the US would have had a negative contribution to the GDP forecast. In both periods, the leading indicator for the 4 European countries leads to a negative average improvement, which is a remarkable outcome in our view. To some extent, this could be related to the inclusion of the UK in this composite leading indicator. Compared to the average leading indicator of 19 other EU countries, this EU4 leading indicator contributes less to the GDP forecasts. Additionally we note that the average improvement value for GDP suggests that its inclusion in the DFM is more harmful than helpful in producing its own forecast. However, we need to include it in the model in order to be able to obtain GDP forecasts from the DFM.[21]

Table 6.2 Average improvements for mixture DFM model

Variable	*Competition period, 2010Q1-2019Q4*	*Full sample period, 2008Q1-2023Q3*
Utilization rate industry	0.45	0.32
House prices	0.31	0.26
Leading indicator US	0.16	-1.04
Consumer confidence	0.15	-0.18
Export	0.15	-0.05
Import	0.13	0.27
Dollar-euro exchange rate	0.10	0.19
Consumption	0.09	0.32
Employee wages	0.08	0.04
Three Component Index	0.04	-0.24
Short interest rate	0.02	0.48
Industrial production	0.02	-0.46
AEX stock index	0.01	-0.02
Producer confidence	0.01	1.28
Bankruptcies	-0.03	0.23
World trade monitor	-0.05	-0.01
Energy prices	-0.10	-0.50
Consumer price index	-0.17	-0.44
Leading indicator EU4 (FR, GE, IT, UK)	-0.40	-1.13
GDP growth	-0.52	-1.50

[20] For comparison, Bok et al. (2017) find that for the DFM used at the NY Fed, survey, consumption, manufacturing and housing data are the main contributors to changes in the nowcast.
[21] We do not explore the use of specifying subgroups of factors in the DFM. Potentially, we could explore the possibility of excluding GDP from the factors, so that GDP would only appear in the measurement equation. We leave this topic for future research.

6.1.1 Variable selection using improvements

Based on the average improvements, we were able to drop 7 variables from the initial dataset for the DFM. Initially, we started with a dataset of 27 variables. By calculating the average improvements for these 27 variables over the competition period, we were able to identify poorly performing variables to eliminate these from the model. From a preliminary search over the number of factors r and lags p, we found that the values r = 3 and p = 1 produced the smallest average RMSFEs over the competition period. For this model, we computed the average improvements per variable based on the larger dataset with 27 variables, which are shown in Table 6.3.

Table 6.3 Average improvements for DFM with factors r = 3 and lags p = 1 based on larger dataset

Variable	*Competition period, 2010Q1-2019Q4*	*Full sample period, 2008Q1-2023Q3*
Import	0.94	0.69
Export	0.74	0.50
Leading indicator EU	0.23	-5.10
House prices	0.20	0.03
World trade monitor	0.13	-0.09
Bankruptcies	0.07	0.10
Producer confidence	0.05	-0.16
Short interest rate	0.05	-0.53
Consumption	0.05	-0.02
Dollar-euro exchange rate	0.03	-0.56
Employee wages	0.01	-0.04
Checkable deposits	0.00	0.23
Consumer price index	-0.01	-0.89
Deposits with notice	-0.01	-0.39
Industrial production	-0.01	0.57
Utilization rate industry	-0.06	-1.47
Consumer confidence	-0.08	-0.27
Energy prices	-0.09	-0.84
Long-term deposits	-0.10	0.04
Leading indicator US	-0.10	-1.60
Money supply M3	-0.12	0.40
Three Component Index	-0.15	-1.01
AEX stock index	-0.16	-0.33
Long interest rate	-0.17	-0.69
Volatility US	-0.24	-0.52
EU Policy Uncertainty	-0.35	-1.00
GDP growth	-1.05	-1.08

Judging by their large negative average improvement value, a number of variables perform poorly in this DFM model. This led us to drop the long-run interest rate, the EU policy uncertainty variable, and US volatility from the DFM. We also dropped the money supply variable M3 as well as its three components, checkable deposits, deposits with notice and long-term deposits. These variables generally have low average improvement values, but more importantly these monthly series contain several breaks and cannot be automatically updated. This makes it difficult to include these variables in any operational DFM at the CPB.

The average improvement values are greatly inflated by occasional implausible forecast values which some DFM specifications produce around the Great Recession and Covid period. This problem with extreme forecasts can clearly be seen in Table 6.2 and Table 6.3 in the columns reporting on the average improvement values based on the full sample period. The extreme forecasts introduce extremely large absolute forecast errors around the time of extreme events such as the Great Recession and Covid period. This problem fortunately does not occur when we only forecast over the more stable competition period. DFM specifications with more factors ($r = 4$ or 5) and longer lag lengths ($p = 2$ or 3) are more susceptible to this problem.[22]

6.2 Model selection for the number of factors and lags

Although our final model selection is a mixture DFM, we have also analyzed individual DFMs over various numbers of factor and lags. Since the aim of this model is to help Saffier with additional information mainly for the short term, we first focus on a shorter forecasting horizon for the DFM of only one quarter ahead, the nowcast and backcast. In Table 6.4 we list the average RMSFEs over this shorter forecasting horizon for DFMs with the number of factors, r, ranging from two to five and the number of lags, p, ranging from one to three. The table also shows how the RMSFEs are affected by the selection of a different forecasting competition period.

Table 6.4 Average RMSFE for different model specifications, for a short forecasting horizon (r = # factors, p = # lags)

	Competition period				Full sample period			
	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
$p = 1$	0.47	0.47	0.42	0.38	1.70	1.87	1.59	1.55
$p = 2$	0.44	0.43	0.39	0.38	1.90	1.99	1.77	1.80
$p = 3$	0.43	0.42	0.38	0.39	2.77	2.34	1.89	2.21

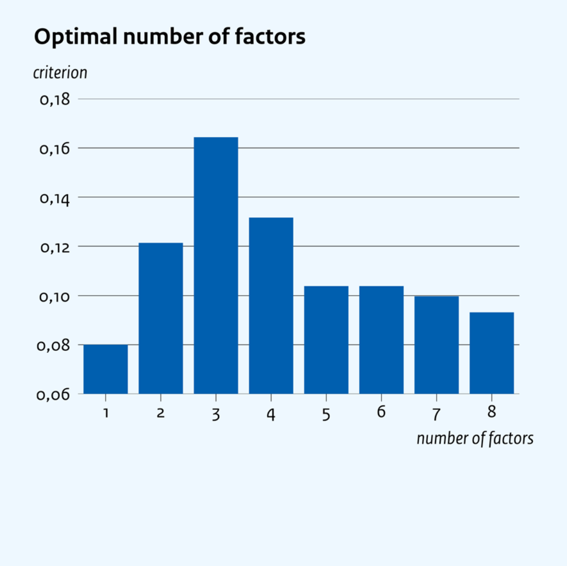
In general, we find that the choice of model specification only has a small impact on the forecast accuracy, especially when we consider the competition period. The average RMSFEs reported in Table 6.4 do not differ much over the number of factors and of AR lags in the model. The best performing DFMs have either four or five factors, regardless of the forecasting competition period used. For the more stable competition period from the start of 2010 until the end of 2019, the optimal DFM has $p = 3$, or three lags. The average RMSFEs for this optimal DFM during the competition period, with $r = 4$ and $p = 3$, are shown below in Figure 6.3 (right). For the full sample period, the optimal number of lags in the DFM is equal to 1.

We can compare the results of our forecasting competition on the optimal DFM, with the optimal values for r and p that we obtain from the model selection routines in the nowcasting package. The function ICfactors produces a criterion for the selection of the optimal number of factors, r, and ICshocks returns the optimal value for the number of lags, p. Figure 6.2 shows the negative values of the criterion for the optimal number of factors on the left-hand side. In this case we obtain an optimal value of $r = 3$. The optimal number

[22] Occasionally, the estimation function "nowcast" in the R-package, which we use for the estimation of our DFM, fails to converge to a solution, mostly around the time of the Great Recession and Covid period. More specifically, we take the forecast for the relevant quarter from the previous forecast based on the same observations only excluding for the most recent data update (which caused the estimation to fail). This allows us to generate a forecast and forecast error, but the improvement for the newly updated variable cannot reasonably be defined.

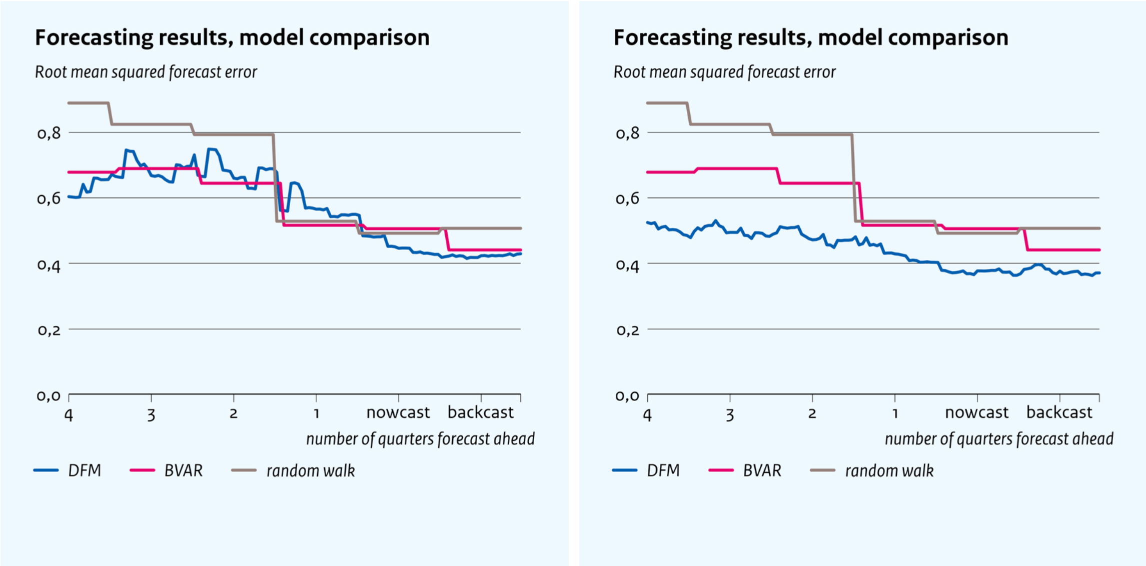
of lags returned by ICshocks is $p = 1$. These optimal values for r and p produced by the nowcasting package are not particularly robust. Depending on the exact method we specify for the criterion selection, we obtain an optimal number of factors from 1 to 16 (the maximum number we allowed for). The optimal number of lags tends to range from 1 to 3. For this reason, we have opted not to rely on these results for our model selection, and instead based it on the smallest RMSFEs from our forecasting competitions.

Figure 6.2 DFM model selection, optimal number of factors *r*



Nonetheless we reproduce the average RMSFEs for the optimal DFM according to the nowcasting package's selection criterion. In Figure 6.3, we compare the RMSFE of the optimal DFM using the selection criterion with $r = 3$ and $p = 1$ in the left figure, to the DFM with the smallest RMSFE according to Table 6.2, with $r = 4$ and $p = 3$, in the right figure. The plots show the average RMSFE for the competition period, over the different forecasting horizons. Note that both graphs also show the same results for the two benchmark models, the BVAR and random walk model. The DFM with $r = 3$ and $p = 1$ performs similarly to the BVAR, while the DFM with $r = 4$ and $p = 3$ performs better than the BVAR at all forecasting horizons. Generally, the random walk performs more poorly than the DFM and BVAR.

Figure 6.3 RMFEs for DFM with $r = 3$, $p = 1$ (left) and DFM with $r = 4$, $p = 3$ (right) for the competition period (2010Q1-2019Q4)



6.2.1 Mixture models

There is a substantial body of literature on the benefits of forecasting based on a mixture of models.[23] We find that a mixture of DFM over the number of factors and lags produces robust and accurate forecasts which are generally better than the other mixture models we explore. Our baseline mixture DFM averages the forecasts from the 12 DFMs with factors $r = 2,\ldots,5$ and lags $p = 1,\ldots,3$. This is the mixture model presented in Section 6. Due to the existence of some variability in the performance of the DFM over the specified values of r and p, we have opted to base our baseline mixture DFM on the forecasting average from the DFMs for $r = 2,\ldots,5$ and $p = 1,\ldots,3$, where all forecasts receive equal weight.[24] In addition, we also produce an alternative mixture forecast based on the DFMs with $r = 4,5$ and $p = 1,\ldots,3$. This latter mixture reflects the fact that DFMs with $r = 4$ or 5 tend to perform better than those with $r = 2$ or 3. The average RMSFEs for these two mixture models are shown in the first and fourth column of Table 6.5 for the shorter forecast horizon of one quarter ahead, the nowcast and backcast. Similarly, Table 6.6 reports the same average RMSFEs, but then for all forecasting horizons up to four quarters ahead.

There is an additional step we take when calculating the mixture of forecasts of the DFMs. Given that these models occasionally generate extreme forecasts we correct for these forecasts when generating the mixture. In those cases where a DFM generates a forecast which is greater (smaller) than the maximum (minimum) observed value of GDP growth, we drop the DFM forecast and replace it with the maximum (minimum) observed value. We feel justified in removing these extreme values because a researcher at CPB would consider these forecasts implausible and would never use these in practice.[25]

Using our baseline mixture DFM, we also obtain mixtures of forecasts based on this DFM with the BVAR and with the BVAR and random walk. There are two variants of these two mixtures. In one case we weigh the baseline mixture DFM with the total number of models in this mixture (12), while the other models are simply given a weight of one. We refer to these mixture forecasts weighted in this manner as weighted mixtures. The

[23] A description of the use of model averaging in economics can be found in Steel (2020).
[24] Most literature on mixtures of forecast models indicates that the exact weighting scheme has only a small impact on the quality of the forecasts obtained (Timmermann, 2006).
[25] However, we do not remove these extreme forecasts when analyzing the individual DFMs so as to accurately demonstrate this problem with the DFM estimates produced by the nowcasting R-package.

other weighing option we include is based on the baseline mixture DFM also being given a weight of one. We refer to these mixtures as “unweighted” ones. The average RMSFEs for these mixture models are also shown in Table 6.5 and Table 6.6.

The results in both Tables 6.5 and 6.6 generally show that the baseline mixture DFM performs as well or better than the other mixture models. There are two exceptions which are worth noting. Firstly, the mixture DFM with r = 4,5 and p = 1,...,3 (second row) performs best over the shorter forecast horizon during the more stable competition period. However, over all forecasting horizons this is no longer the case. Secondly, over all forecasting horizons, the baseline mixture DFM and BVAR with equal weights (“unweighted”) performs best during the competition period. This is possibly the result of the fact that the BVAR tends to forecast relatively well at longer forecasting horizons.

Table 6.5 Average RMSFEs for mixture DFM, short forecasting horizon (1 quarter ahead to backcast)

	Competition period			Full sample period		
Mixture type	DFM	DFM & BVAR	DFM, BVAR, RW	DFM	DFM & BVAR	DFM, BVAR, RW
Unweighted	0.40	0.41	0.41	1.51	1.61	1.74
r = 4,5 and p = 1,...,3	0.38			1.54		
Weighted		0.40	0.40		1.51	1.53

Table 6.6 Average RMSFEs for mixture DFM, all forecasting horizons (4 quarters ahead to backcast)

	Competition period			Full sample period		
Mixture type	DFM	DFM & BVAR	DFM, BVAR, RW	DFM	DFM & BVAR	DFM, BVAR, RW
Unweighted	0.47	0.42	0.45	1.64	1.64	1.75
r = 4,5 and p = 1,...,3	0.46			1.68		
Weighted		0.45	0.43		1.63	1.63

6.3 Results for different subperiods

In this section we explore how well the DFM is able to identify or even forecast economic crises, such as the Great Recession and the Covid period. Most forecasting models tend to perform poorly in times of economic crisis, and it is not our primary aim to produce a forecasting model only for such periods. The DFM is mostly designed to help predict GDP during more stable periods of economic growth. Nonetheless there are some indications that our DFM is able to detect economic turning points somewhat better than the benchmark BVAR and random walk models.

Relative to more stable economic periods, it is more difficult for the mixture DFM to forecast in times of crisis. This can be seen in Figure 6.4 where we plot the average RMSFEs we obtain based on the full sample period. The forecast errors during the crisis periods are much larger than those we obtain during the more stable competition period. However, the DFM still outperforms the BVAR and random walk, mainly for the nowcast and the backcast. In Table 6.7 we list the average RMSFEs taken over each forecasting horizon for the full sample period, and compare these to the average RMSFEs over the competition period. At every forecast horizon, the average RMSFE is substantially larger when including the quarters from the Great Recession and Covid period.

Figure 6.4 Root mean squared forecast errors, 2008Q1-2023Q3

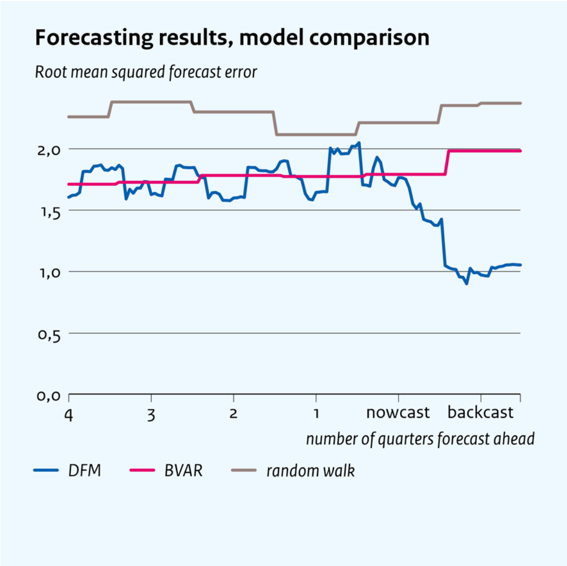


Table 6.7 Average root mean squared forecast errors per forecast horizon, for full sample period and competition period

Model	Forecast horizon (quarters)					
	4	3	2	1	nowcast	Backcast
	Full sample period (2008Q1-2023Q3)					
mixture DFM	1.75	1.72	1.76	1.84	$1.33^{\dagger}$	$1.03^{*,\dagger}$
mixture DFM (last update)	$1.73^{\dagger}$	1.58^{*}	$1.58^{*,\dagger}$	$1.70^{\dagger}$	$0.99^{*,\dagger}$	$1.05^{*,\dagger}$
BVAR	1.72	1.75	1.78	1.78	1.87	1.98
Random walk	2.32	2.34	2.21	2.16	2.28	2.37
	Competition period (20010Q1-2019Q4)					
mixture DFM	$0.52^{\dagger}$	$0.54^{\dagger}$	0.50	0.42	$0.39^{*,\dagger}$	$0.39^{*,\dagger\dagger\dagger}$
mixture DFM (last update)	$0.53^{\dagger}$	$0.54^{\dagger\dagger}$	0.47	$0.39^{**,\dagger}$	$0.39^{*,\dagger\dagger\dagger}$	$0.38^{**,\dagger\dagger\dagger}$
BVAR	0.68	0.67	0.60	0.51	0.48	0.44
Random walk	0.86	0.81	0.68	0.51	0.50	0.51

Note: the bottom half of this table contains the same results as Table 6.1, but is added for the sake of comparison. *, ** or *** denotes the Diebold-Mariano test for comparison with the BVAR is significant at the 10%, 5% or 1% level, respectively. †, †† or ††† denotes the Diebold-Mariano test for comparison with the random walk is significant at the 10%, 5% or 1% level, respectively.

One of the main motivations for a mixed-frequency model such as the DFM is that the more frequent forecasting updates could help to identify a turning point before the quarterly BVAR or random walk can. The mixed frequency DFM benefits from the monthly update of most of its variables. This should help to produce better forecasts and this in turn should help to identify turning points more quickly. Figure 6.5 shows the mixture DFM backcast during the Great Recession (left) and Covid-19 period (right) together with the backcasts from the BVAR and random walk.[26] In the case of the backcast during the Covid period (right-hand side), we can see that the DFM is able to follow the pattern of GDP growth quite well. The forecasts by the BVAR

[26] Note that in Figure 6.5 to Figure 6.7, we show the DFM forecast we obtain from the last update in the forecast horizon.

and random walk are too late in picking up the larger drops and hikes in GDP. In the case of the Great Recession (left-hand side) this difference is less pronounced.

Also for the nowcast, the DFM performs better during the turning points of the cycle than the benchmark models do. Figure 6.6 shows that again this seems to be more the case for the Covid-19 period than for the Great Recession. However, the BVAR and random walk do slightly worse here than in the backcast. The forecasting results for the one quarter ahead forecasts of the Great Recession and Covid period shown in Figure 6.7 indicate that the DFM performs somewhat better than the BVAR and random walk models, but none of these models do particularly well at this forecast horizon.

Figure 6.5 Turning points detection via the backcast, the Great Recession (left) and Covid-19 (right)

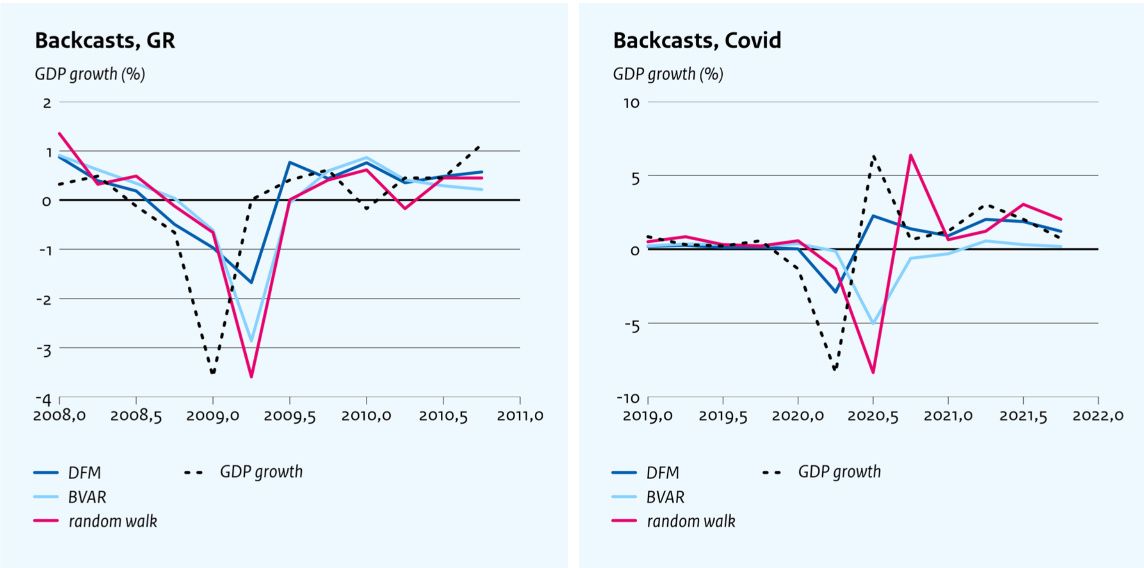


Figure 6.6 Turning points detection via the nowcast, the Great Recession (left) and Covid-19 (right)

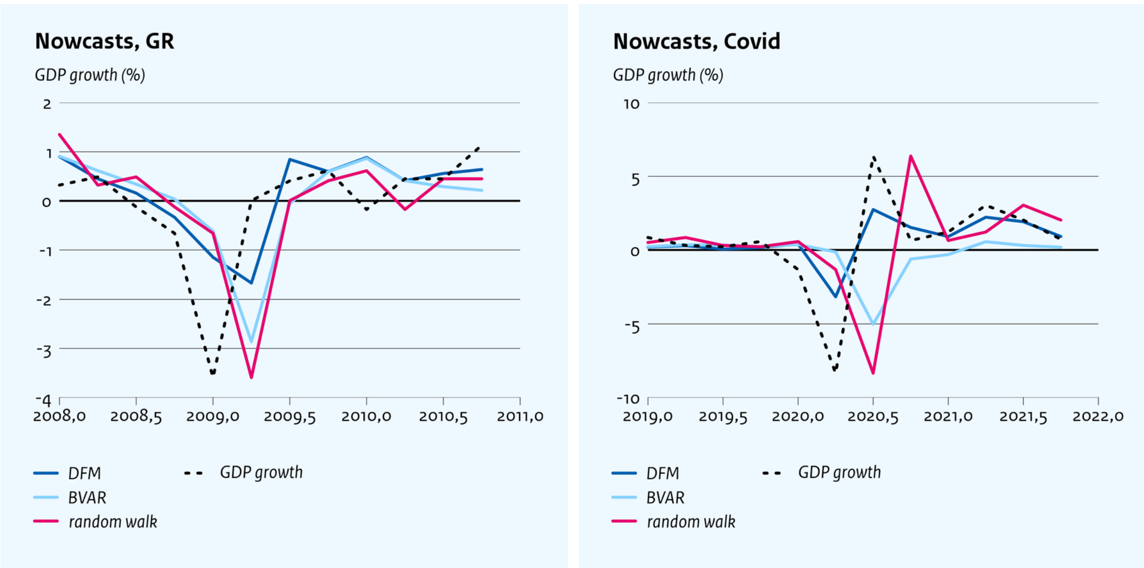
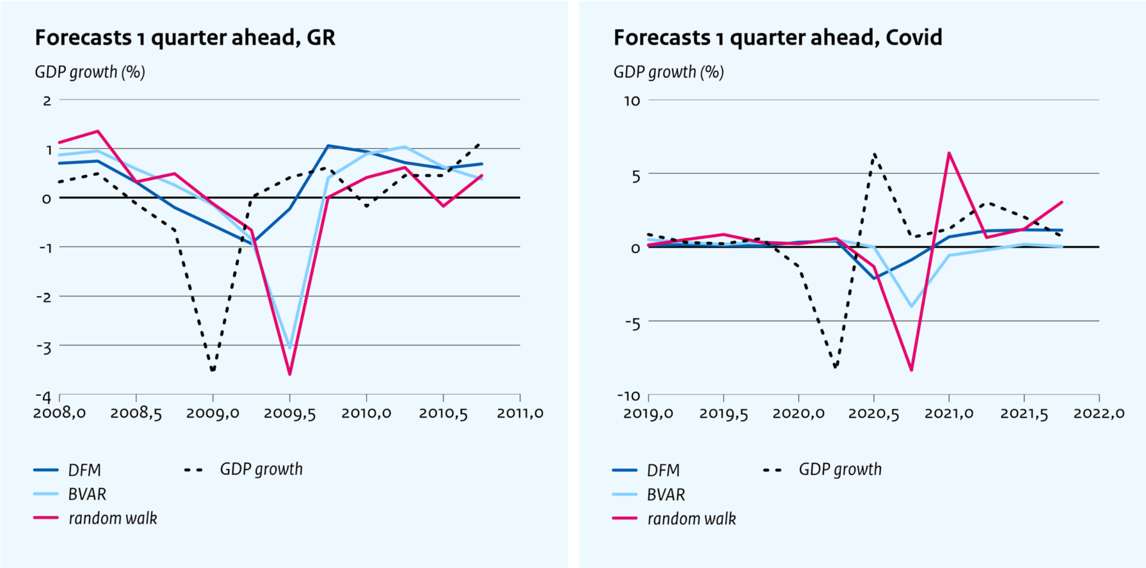


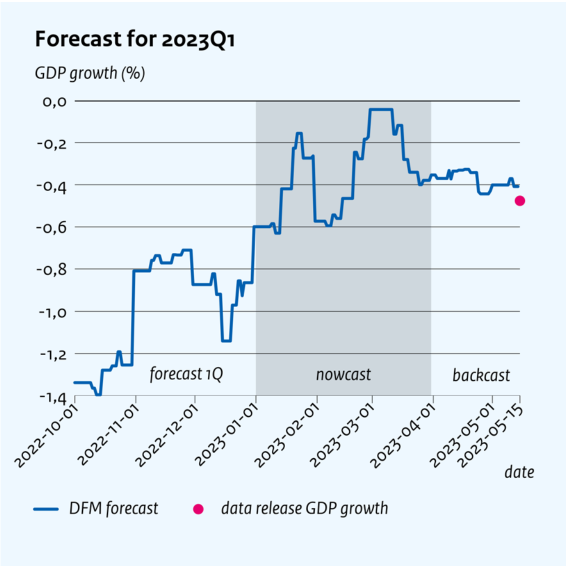
Figure 6.7 Turning points detection via the 1 quarter head forecast, the Great Recession (left) and Covid-19 (right)



7 How does a DFM work in CPB practice?

To show how the release of new data can help the forecaster in practice, we replicate a nowcast for a specific quarter in the past, namely for the first quarter of 2023. During the weeks when the CPB is working to produce a new forecast, this nowcasting model can be used several times. Each week there is new information released which affects the nowcast (and if relevant the backcast) for GDP growth. It is therefore interesting to see how the nowcast changes as new information is published. Beginning with the first day of the previous quarter until the first official release of Statistics Netherlands (45 days after the end of the quarter), we produced an updated forecast from the DFM for each day that a new figure is released. In this manner the newest data releases can be reflected in a new forecast each time they arrive. In Figure 7.1 we show the progression of the nowcast of GDP growth during of the first quarter of 2023.

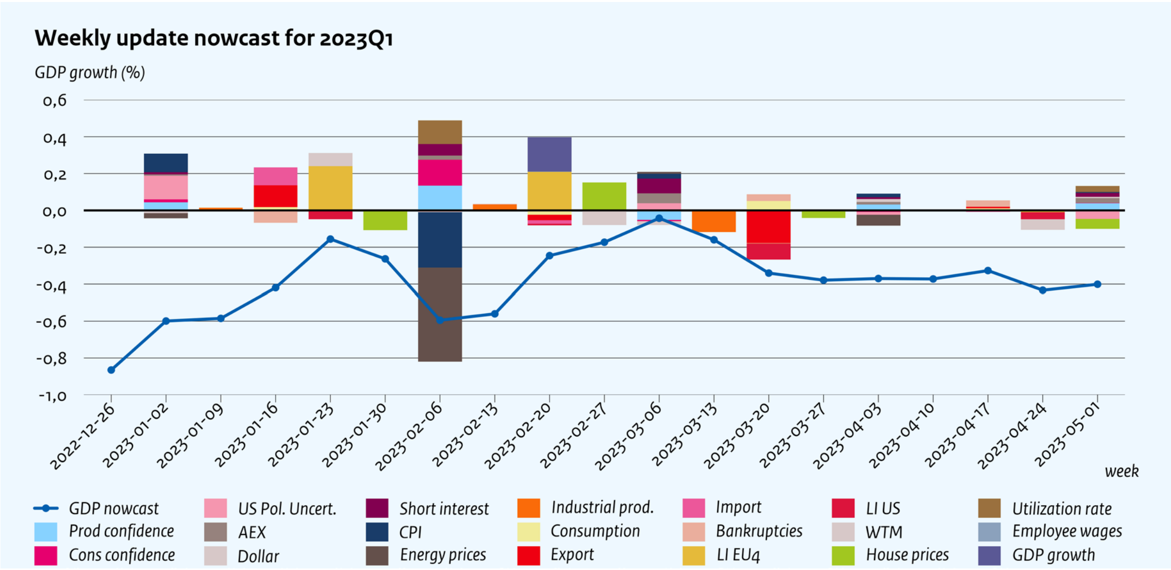
Figure 7.1 Forecast for 2023Q1 over time; from October 1st 2022 until May 15th 2023



As more information arrives, the dynamic factor model produces a better forecast. In 2023Q1, GDP suddenly declined after two very positive years. Figure 7.1 shows that the dynamic factor model was rather pessimistic about the forecast for GDP growth especially in the fall of that year. In the beginning of 2023, the nowcast became somewhat less negative, and closer to actual GDP growth. On the first of the month, new data on producer and consumer confidence, the dollar exchange rate, stock prices, interest rates and the utilization rate were released. In November, this led to a quite large upward adjustment of the nowcast for 2023Q1. Additionally, new data releases in January caused the model to adjust the forecast for 2023Q1 upward. Towards the end of the first quarter of 2023, new information such as a declining industrial production and export growth and an increasing monthly figure for imports led to a drop in the nowcast. The end result in May, just before the flash estimate of the CBS for 2023Q1 was released, was a backcast nearly equal to the realized level.

The revision of the nowcast of GDP growth can be expressed as a sum of the effects of the news contained in particular releases. For illustration purposes, we show in Figure 7.2 the weekly update of the nowcast. In addition we calculate the marginal contribution to the update from all data releases in a week. This is done by sequentially adding each data release to the vintage of the previous week and for each update re-estimating the DFM to update the forecast for 2023Q1. The impact of the data releases in each week are indicated by the colored bar for that week, where each updated variable is given a different color. Here, we find that in the beginning of the quarter, the confidence indicators and the indicator for US policy uncertainty have shifted the nowcast upwards. The same is true for the monthly data on exports and imports (from the previous quarter) as well as the leading indicator for the EU. In the beginning of February, the data releases at that time had opposing effects on the nowcast for GDP growth. Information on producer and consumer confidence as well as the utilization rate had an upward effect on the nowcast, whereas the (energy) price inflation of January shifted the nowcast down. Figure 7.2 shows that in this quarter the leading indicators, consumer confidence and exports are important in explaining the eventual GDP nowcast for 2023Q1.

Figure 7.2 Weekly forecast for 2023Q1 with individual variable contributions



8 Conclusion

The primary aim of this research was to specify a mixed-frequency DFM that provides an accurate forecast for GDP growth, using information from leading indicators with different publication delays and frequencies. We test this model against the benchmark CPB-BVAR and random walk model, particularly during more stable periods when these time series models tend to perform best. Our results indicate that the DFM produces good forecasts, and slightly better than the benchmark models, both over the shorter forecasting horizons, as well over the entire forecasting horizon of up to four quarters ahead. For all three models, the RMSFE tends to decrease when the forecasting horizon becomes shorter, while the forecasting accuracy of the models also becomes more similar. In times of crisis, the DFM has much larger forecast errors than in stable periods but still outperforms the BVAR and random walk, particularly in the case of the nowcast and the backcast.

Several model specifications and model combinations were tested. Our preferred specification is a mixture model, a combination of 12 individual DFMs with number of factors ranging from 2 to 5, and number of AR lags ranging from 1 to 3. The choice of model specification has a small impact on the forecast accuracy, especially when we consider the more stable period (2010Q1-2019Q4). The forecast errors for the individual DFMs do not differ substantially, and also the forecast error for the mixture DFM lies in the same range. The advantage of the mixture DFM is that it is less susceptible to implausible outcomes. The DFM is also combined with the benchmark models, the BVAR and random walk. We find that the baseline mixture DFM performs as well or better than these other mixture models, for the shorter forecasting horizons. When we consider all forecasting horizons, the baseline mixture DFM and BVAR with equal weights performs best. This is possibly the result of the fact that the BVAR tends to forecast relatively well at longer forecasting horizons.

For the operational use of the DFM, it is helpful to derive a measure of the extent to which each data release contributes to the forecast of GDP growth. For the past quarters, for which GDP growth is observable, this boils down to the change in the absolute forecast error from one update to the next. This tells us whether the variable update produces an improvement in the model forecast. The average improvement we obtain for each variable provides us with an indication of how informative each variable is in forecasting GDP with the DFM. For future quarters, and hence for the use at CPB, it is insightful how each data release impacts the forecast of GDP, since that can help to understand and explain the changes in the forecast over time.

References

Banbura M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-Casting and the Real-Time Data Flow. In G. Elliott, C. Granger & A. Timmermann (ed.), *Handbook of Economic Forecasting* (ed. 1, vol. 2) (pp. 195-237). Elsevier.

Bettendorf, L., Boeters, S., Kranendonk, H. & Verstegen, L. (2021). *Het macromodel van het CPB: Saffier 3.0* (CPB Notitie). Centraal Planbureau.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A., & Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics, 10*(1), pp. 615-643. https://doi.org/10.1146/annurev-economics-080217-053214

Brave, S., Butter, A., & Justiniano, A. (2019). Forecasting Economic Activity with Mixed Frequency Bayesian VARs. *International Journal of Forecasting, 34*(4), pp. 1692-1707. https://doi.org/10.1016/j.ijforecast.2019.02.010

Coulombe, P., Leroux, M., Stevanovic, D. & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics, 37*(5), pp. 920-964. https://doi.org/10.1002/jae.2910

Dauphin, J.F., Dybczak, K., Maneely, M., Sanjani, M. Suphaphiphat, N., Wang, Y., & Zhang, H. (2022). Nowcasting GDP: A scalable approach using DFM, machine learning and novel data, applied to European economies. (IMF Working Paper No. 2022/052). International Monetary Fund.

De Valk, S., de Mattos, D., & Ferreira, P. (2019). Nowcasting: An R Package for Predicting Economic Variables Using Dynamic Factor Models. *The R Journal, 11*(1), pp. 230-244.

Dieppe, A., Legrand, R., & van Roye, B. (2016). The BEAR toolbox. (ECB Working Paper No. 1934). European Central Bank.

Doz, C., Giannone, D., & Reichlin, L. (2012). A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models. *The Review of Economics and Statistics, 94*(4), pp. 1014-1024.

Durbin J., & Koopman, S.J. (2001). *Time Series Analysis by State Space Methods.* (ed. 2). Oxford University Press.

Elbourne, A., Kranendonk, H., Luginbuhl, R., Smid, B., & Vromans, M. (2008). Evaluating CPB's published GDP growth forecasts; a comparison with individual and pooled VAR based forecasts. (CPB Document 172). Centraal Planbureau.

Giannone, D., Reichlin, L, & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics, 55*(4), pp. 665-676. https://doi.org/10.1016/j.jmoneco.2008.05.010

Jansen, W., Jin, X., & de Winter, J. (2016). Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts. *International Journal of Forecasting, 32*, pp. 411-436. http://dx.doi.org/10.1016/j.ijforecast.2015.05.008

Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting, 37*(2), pp. 941-948. https://doi.org/10.1016/j.ijforecast.2020.10.005

Scheer, B. (2019). Werkloosheidsramingen met machine learning: kan het nog beter? (CPB Achtergronddocument). Centraal Planbureau.

Steel, M. (2020). Model Averaging and Its Use in Economics. *Journal of Economic Literature*, *58*(3), pp. 644-719. https://doi.org/10.1257/jel.20191385

Stock, J. & Watson, M. (2011). Dynamic Factor Models. In M. Clements & D. Hendry (ed.), *Oxford Handbook on Economic Forecasting* (pp. 35-60). Oxford University Press. https://doi.org/10.1016/j.ijforecast.2019.02.010

Timmermann, A. (2006). Forecast Combinations. In G. Elliott, C. Granger & A. Timmermann (ed.), *Handbook of Economic Forecasting* (ed. 1, vol. 1) (pp. 135-196). Elsevier.