Discussion Paper

Nowcasting GDP growth rate: a potential substitute for the current flash estimate

M. Kuiper
F.P. Pijpers

May 2020

Contents

1 Introduction 4
1.1 Nowcasting 5

2 Theoretical Background 6
2.1 Dynamic Factor Model 6
2.2 Principal Component Analysis 7
2.3 Kalman filter 7
2.4 From factors to GDP 9
2.5 Nowcast uncertainty 10

3 Data and nowcast design 11
3.1 Data selection 11
3.2 The set of predictors 11
3.3 Preprocessing data 13
3.4 Simulating real-time information flow 14
3.5 Data revisions 14

4 Preliminary data analysis 14

5 Principal Component Analysis 16
5.1 Selecting number of factors 16
5.2 Interpretation of factors: quarterly models 18
5.3 Interpretation of factors: OA approach 19

6 Results 22
6.1 Econometric Approach 23
6.2 Operational Approach 25
6.3 Model Confidence Set 26
6.4 Model Averaging 30
6.5 News Impact and Uncertainty 34
6.6 Testing some model assumptions 35

7 Year on Year growth rates 36

8 Conclusion 38

I Overview models 43

II Complete data set 43

III Preliminary data analysis 45

IV PCA Analysis Monthly Series 47

V PCA Analysis Quarterly Series. 48

VI PCA Analysis Operational Approach 50

VII model quality measures 52

VIII Model Confidence Set 57

IX Year on Year Growth 59

Abstract

Currently Statistics Netherlands provides a first estimate of the GDP growth rate 45 days after the end of the reference quarter. Since a lot of information on both the consumption and production side is still missing by that time, this first estimate often deviates substantially from the final growth rate. Therefore, this paper investigates whether nowcasting, as introduced by Giannone et al. (2008), could provide a good substitute for the current estimation method. Their approach to nowcasting is to use a large number of time series that are released more timely and/or at a higher frequency than the target variable and that are related to the target variable. This paper provides a thorough investigation in the various aspects of the nowcast model in order to obtain the best model. For that purpose, the real-time information flow of all the indicator time series is simulated over the period 2008-2017, using a dataset covering 2005-2017. This allows for proper backtesting of the models since the nowcast for a certain past quarter can be compared with the final growth rate for that quarter. We found that the best performing nowcasting model has a margin of uncertainty very close to the current 'flash' estimates. Importantly, it is demonstrated to suffer from a significantly smaller bias than the current method does.

1 Introduction

The publication of quarterly economic growth figures is of major importance for Statistics Netherlands (SN). Apart from its use as input for economic planning and government policies, the Gross Domestic Product (GDP) of each member state partly determines the level of its tax contribution to the EU. The GDP growth rate publications also receive significant attention in the media. People, and consequently companies, attach considerable value to the releases of the GDP growth rates. The way people respond in turn has an effect on the economy.
SN provides a first estimate of the quarter on quarter GDP growth rate 45 days after the end of the reference quarter. This is the so-called 'flash estimate'. The next updated growth rate is released after 90 days, while it takes more than 32 months before the final revised GDP growth rate is determined. The last couple of years, the flash estimate turned out to deviate significantly from the final growth rate. Revisions of 40 basis points have not been exceptional. In particular during the recent financial crisis, deviations of this order could turn a first positive flash announcement into a negative growth rate once definitive determinations become available. It is a matter of high priority for SN to establish where these large deviations are coming from and if SN can find a way to mitigate them.

There is clearly a trade-off between accuracy and speed. The quicker one wants to provide an estimate of the state of the economy, the less information is available, inducing inaccuracy. Basic economics tells us that GDP can be determined via three approaches, which should evidently all sum up to the same result: the expenditure, the production and the income approach. The expenditure and the production approaches are used in the construction of the flash estimate, with the latter being leading simply because there is generally more information available on the production side. Since both the consumption side and the production side should add up to the same number, data on both sides are used to construct (an estimate of) the GDP. However, the problem is that data on the components of the GDP are far from complete at the time the flash estimate is constructed, resulting in 'blind spots' in the GDP composition which require input from models and human interpretation.

1.1 Nowcasting

Producers generally do not report individual production or sales within 45 days after the end of a quarter. Consequently, it will be hard to fully eliminate the blind spots. Fortunately, there are other indicators that provide information on the current state of the economy. This is the fundamental idea of a fairly new principle called 'nowcasting', introduced by Giannone et al. (2008). Nowcasts are predictions of the recent past and current state of the economy, exploiting timely information. The underlying idea of nowcasting is to combine many high(er) frequency time series that are published in an earlier stage to nowcast the target variable with a lower frequency. In the literature, nowcasting turns out to be of added value, particularly when used at very short time horizons (e.g. Chernis and Sekkel (2017) and Rünstler et al. (2009)). In terms of the problem set out above, the GDP growth rate of the previous quarter might be nowcasted by exploiting various indicators that are already available at the time the flash estimate is being constructed. As a consequence, they have different publication moments and lags. In order to cope with the large data set a Dynamic Factor Model (DFM) will be used with r common factors. A Kalman filter will be applied to deal with non-synchronous data releases.

This paper reports on an investigation whether the nowcasting method could be an appropriate method for determining the flash estimate of the Dutch GDP growth rate. This builds upon the model introduced by Giannone et al. (2008). This model is extended by comparing different methods to get from factor nowcasts to nowcasts of the GDP growth rate. Moreover, the effect of transforming the data set to stationary time series before it enters the model is investigated. Thirdly, we will consider adding autoregressive (AR) terms of GDP growth itself to the regression of GDP on the factors. According to Jansen et al. (2012) this should improve the nowcasting performance. In addition, the effect of changing other parameters of the models on the nowcast performance will be investigated.

The remainder of this paper is set up as follows. First, a brief theoretical background is provided, supporting the econometric computation described hereafter in Section 3. It provides details on the way the data set is constructed as well as the techniques used for data processing. Moreover, it describes the econometric design used for backtesting the different nowcasting models. In order to give an idea of the data, Section 4 provides a brief preliminary analysis on the historical GDP data. Thereafter, the number of factors included in the various models is determined in Section 5. The empirical results of the nowcast models are given in Section 6. This section investigates the effect of the various parameters of the models on the nowcasting performance to find to best performing models. Hereafter, the Model Confidence Set (MCS) theorem introduced by Hansen et al. (2011) and some model averaging techniques are used to select the best models and to try to improve the nowcasts. Section 7 concisely explores whether the model is appropriate for nowcasting year on year growth rates.

2 Theoretical Background

2.1 Dynamic Factor Model

The fundamental underlying principle of nowcasting is using a large set of frequently released indicators to predict the target variable, such as the GDP growth rate. More time series will provide more information. Each additional time series can potentially provide information beyond that of the standard macroeconomic variables. Apart from additional information, forecasting using a higher variety of predictors is presumably less vulnerable to the stochastic component in individual time series. Therefore, one would like to add many time series that have some predictive power on the GDP, resulting in a high-dimensional vector of predictors $P_t = (p_{1t}, ..., p_{Nt})$. The (past) GDP itself can in principle be used as one of the predictors; in this paper variants both with and without are presented.
One way to exploit the rich set of information while keeping the estimation noise reduced is to exploit the common variation among the predictors. Especially when using a large set of (macro)economic time series, there will be correlations among (some of) the predictors. Therefore, the DFM will be appropriate here (see for example Giannone et al. (2004) and Watson (2004)).
Let Y_t be the N-dimensional vector of predictors or indicators. In what follows, most of the indicators will be monthly time series, therefore in this model the time index t refers to the month. It is assumed that Y_t follows a DFM with r latent factors, which are not observed directly but determined using (repeated) principal component analysis (PCA). F_t,

$$\begin{aligned} P_t &= \Lambda F_t + \epsilon_t, \\ \epsilon_t &\sim N(0, \Sigma_t), \end{aligned} \tag{1}$$

with P_t and ϵ_t being N-dimensional vectors and F_t an r-dimensional vector of factors with $r << N$. The $N \times r$ matrix of factor loadings Λ is assumed to be time-invariant. Its elements $\lambda_{i,j}$ are the correlations between the i-th indicator and the j-th factor. The first term on the right of Equation (1) is the common component of Y_t, which captures a large part of the comovements in the data, the joint dynamics. The second term is the idiosyncratic component, which represents the individual specific (stochastic) shock corresponding to each of the N predictors. The common component and the idiosyncratic component are two orthogonal processes.
To capture the lead and lag relationships among the predictor variables it is reasonable to assume that the factors follow an AR process

$$\begin{aligned} F_t &= AF_{t-1} + u_t, \\ u_t &\sim N(0, Q_t), \end{aligned} \tag{2}$$

in which for now the order is set to one. This will be relaxed later. The coefficient matrix A is assumed to be time invariant. The error term u_{it} is a white noise process independent of the idiosyncratic shocks ϵ_{it} at all lead and lags, such that $E[\epsilon_t u'_{t-j}] = 0$ for all j, where the prime on the vector u indicates taking the transpose. Furthermore, the covariance matrix of the idiosyncratic disturbances Σ_t is assumed to be diagonal. That means that the errors are independent and uncorrelated (over time). Both Σ_t and Q_t are considered time-invariant within one loop of the Kalman filter (see sect. 2.3). Equations (1) and (2) form the state space representation, which will later turn out to be of great use when applying the Kalman filter to obtain estimates for the predictors and the target variable. The final step is to transform the nowcasts of the factors into nowcasts of the GDP growth rate via

$$\widehat{Y}_t = \widehat{\beta} F_t + \epsilon_t, \tag{3}$$

with $\widehat{Y}_t$ the target variable, namely the (quarterly) GDP growth rate. The estimation of β is discussed in section 2.4

2.1.1 Four-step estimation

The state space model (1)-(3) will be the starting point of the nowcast model. The quarterly GDP estimation procedure essentially boils down to four steps. First, Principal Component Analysis (PCA) will be applied to estimate the common factors. To reduce the dimension of the problem, a number of r factors will be used to model the common component of the predictor series. The second step entails estimating the coefficient matrices A and Λ, by applying a Vector Autoregressive (VAR) model and Ordinary Least Squares (OLS) respectively. The third step is the Kalman filter, which is initialised by the PCA estimates of the factors. Note that the PCA provides the factor loadings, i.e. the linear combinations which are used to transform the input data to the factors. For the overall PCA analysis presented in this paper in section 5, all data from 2005 to 2017 are used. However, for testing the quality of the nowcasting, the PCA is carried out from scratch at every month, with only the data that are available up to that point. This means that also the loading factors must be treated as estimates. One should therefore expect that for a number of quarters at the beginning of the series, the loading factors have quite large uncertainties.
The last step involves transforming the nowcasted monthly factors to a nowcast of the GDP growth rate.

2.2 Principal Component Analysis

PCA provides consistent estimates of the factors (see for example Forni et al. (2000) and Bai (2003)). PCA requires a matrix without missing observations or jagged edges. Therefore, PCA is applied on a balanced data set constructed by using only the months up to which all indicators are known. The last rows with missing values due to publication lags are discarded.
The factors are given by the standardised principal components of the sample correlation matrix. Consequently, the estimated factors and loading matrix are those $\widehat{F}$ and $\widehat{\Lambda}$ that minimise the (nonlinear) sum of squared idiosyncratic shocks,

$$(\widehat{F}, \widehat{\Lambda}) = \underset{F_t, \Lambda}{\text{argmin}} \sum_{t=1}^{T} \sum_{i=1}^{N} (p_{it} - \lambda_i F_t)^2. \tag{4}$$

In order to reduce the dimensionality of the problem, only the first r principal components are used. More on factor estimation can be found in for example Stock and Watson (2011) or Tsay (2010).

2.3 Kalman filter

The third step in the estimation procedure entails the Kalman filter. Doz et al. (2006, 2011) proved the consistency of the common factors estimated by the Kalman filter. An extensive treatment of the Kalman filter can be found in for example Tsay (2010) or Durbin and Koopman (2012).

One can apply the Kalman filter to make nowcasts of the factors, which will be used via Equation (3) to provide estimates of the current state of the economy. The Kalman filter re-estimates the factors $F_{t|t}$ recursively, starting with F_0. The latter will be set equal to the first row of the initial factor estimates $\widehat{F}_t$ found by PCA.

The core of the nowcasting method is that new released data will be compared with the corresponding prediction made by the Kalman filter. If the model's prediction deviates from the actual release this is regarded as 'news', which will in turn affect the estimate of the factors. This news will then be used to update the provisional estimate $F_{t|t-1}$ to $F_{t|t}$ according to

$$F_{t|t} = F_{t|t-1} + K_t \cdot \text{news}_t, \tag{5}$$

with

$$\text{news}_t = Y_{t|t-1} - Y_t^{real}. \tag{6}$$

The matrix K_t is referred to as the Kalman gain and determines the weight that is given to the news.
Under the model assumptions we have $F_t|\Omega_s \sim N\left(F_{t|s}, S_{t|s}\right)$. In every step, the covariance matrix of the prediction, $S_{t|t-1} = \text{Var}(AF_t + u_t|\Omega_{t-1})$ is also computed. Here Ω_t represents all the information that is available at a certain point in time t. This requires taking into account publication lags. From this the forecast error and its covariance matrix G_t can be constructed. The latter two together determine the Kalman gain. In summary, we have

$$\begin{aligned} S_{t|t-1} &= AS_{t-1|t-1}A' + Q \\ G_t &= \Lambda S_{t|t-1}\Lambda' + \Sigma, \\ K_t &= S_{t|t-1}\Lambda' G_t^{-1}. \end{aligned} \tag{7}$$

where primes indicate taking the transpose of matrices. The starting value $S_{1|0}$ will be set equal to the unconditional variance of the estimated factors $\widehat{F_t}$. Finally, we obtain the updated estimates for the factors and the corresponding covariance matrix for the current period, given by

$$\begin{aligned} F_{t|t} &= F_{t|t-1} + K_t \cdot \text{news}_t, \\ S_{t|t} &= S_{t|t-1} + K_t \Lambda S_{t|t-1}. \end{aligned} \tag{8}$$

2.3.1 Handling publication lags

The Kalman filter is used to provide estimates of the current state of the economy. If we are for example in March, we would like to estimate the GDP up to the end of February. Since the indicators are suffering from publication lags there are missing observations at the end of the information matrix. The Kalman filter provides estimates for missing observations. The maximum lag in the current dataset is three months. Consequently, the filter will start making nowcast from $t-2$ up to now (t). For $t-3$ all 88 indicators have up to date information, while at $t-2$, $t-1$ and t only 62, 51 and 26 respectively. For those indicators with missing observations at the end, there will be no value to compare the prediction with. Ergo, there is no news and hence no Kalman gain to be made. Therefore, the elements of the covariance matrix Σ corresponding to the missing data points were replaced by a very large number. This implies large uncertainties, which is perfectly true if there is no observation. Since the Kalman Gain is constructed by taking the inverse of G_t this results in gains close to zero for the missing observations. As a result, the absence of observations implies no news and hence no update (see also Giannone et al. (2008)).

2.4 From factors to GDP

The fourth and final step is to transform the Kalman nowcasts of the r factors into a nowcast of the GDP growth rate via the regression (3). Since the state space model (1)-(3) is a monthly model, the Kalman filter will provide monthly factors. However, for GDP growth rate and GDP value we only have quarterly numbers at our disposal. Hence, we also need quarterly factors in order to run the regression and obtain an estimate for β.
The model described above relies on the assumption of having stationary time series. Seasonal modulation of the time series might also have some detrimental effect, but correcting for that is left for further research. Therefore the entire nowcasting procedure will be applied to stationary time series. In addition, the entire procedure will be applied to nonstationary data to investigate whether this has an effect on the nowcasting performance. The first approach will be referred to as the Econometric Approach (EA) and the latter as the Operational Approach (OA). For the EA, regression (3) will be run using the historical GDP growth rates, while for the OA the historical GDP values will be used. As will be described in the next section, the indicator series will be transformed and made to be stationary by taking (log) differences. We consider both taking monthly differences and quarterly differences. Consequently, within the EA those two different methods are used resulting in different estimates of the GDP growth rate.

For the first approach, all time series were transformed into monthly differences or monthly growth rates. Accordingly, the estimated factors will also be 'monthly' factors F_t^m. These will be aggregated to quarterly factors by using two different methods. A simple manner could be to simply take the average of the latent monthly factors to construct quarterly factors. However, to allow for different weights for the different months within a quarter, the following regression will provide the first way (M1) to estimate the quarterly GDP growth rate

$$\widehat{Y}_t^Q = \beta_0 F_t^m + \beta_1 F_{t-1}^m + \beta_2 F_{t-2}^m + \epsilon_t. \tag{9}$$

For the second aggregation approach, we use that the quarterly value of GDP, z_t, is the sum of the GDP of the corresponding months within that quarter

$$z_t^Q = z_t^m + z_{t-1}^m + z_{t-2}^m. \tag{10}$$

Further, using the approximation from Mariano and Murasawa (2003), i.e. approximating the sum of the three monthly values by their geometric mean plus a constant, we obtain the following expression for the GDP growth rate

$$\begin{aligned}\widehat{Y}_t^Q \quad &= \ln(z_t^Q) - \ln(z_{t-1}^Q),\\ &\approx \frac{1}{3}[\ln(z_t^m) + \ln(z_{t-1}^m) + \ln(z_{t-2}^m)\\ &\quad - \ln(z_{t-3}^m) - \ln(z_{t-4}^m) - \ln(z_{t-5}^m)]\\ &= \frac{1}{3}\left[\widehat{Y}_t^m + 2\widehat{Y}_{t-1}^m + 3\widehat{Y}_{t-2}^m + 2\widehat{Y}_{t-3}^m + \widehat{Y}_{t-4}^m\right], \end{aligned} \tag{11}$$

where $\widehat{Y}_t^m = \ln(z_t^m) - \ln(z_{t-1}^m)$ denotes the unobserved monthly GDP growth rate. This derivation motivates the second aggregation method (M2) to obtain quarterly factors from the monthly factors, namely

$$F_t^Q = \frac{1}{3}\left[F_t^m + 2F_{t-1}^m + 3F_{t-2}^m + 2F_{t-3}^m + F_{t-4}^m\right]. \tag{12}$$

The second set of estimates takes three-month differences of the indicator series Y_t, as to transform all monthly indicators to quarterly equivalents. The resulting factors are therefore also quarterly factors and we can regress them on the historical GDP data set. Every month quarterly estimates of the factors are constructed. Following Giannone et al. (2008), the third method (Q1) takes every third element of the estimated factors. The fourth method (Q2) takes the three-month average of the factors. For each month, the process described above is repeated, leading to a sequence of nowcasts. The sequence F_t^Q will contain missing observations for the first two months of a quarter and the quarterly growth rate on the third month of a quarter.

As a result, we have four different ways to estimate the GDP growth rate. Additionally, two methods based on the same procedure but then applied to nonstationary data will be considered. That is, every third element of the estimated factors, as well as the three-month average will be used to compute F_t^Q. In Appendix I a graphical overview of the models is provided. These six methods will be compared with one another and ultimately with the historical flash estimates of SN. For each method the Mean Squared Forecast Error (MSFE) is computed, which will provide a measure for comparing and evaluating the performance of the several methods.

2.5 Nowcast uncertainty

Apart from providing an accurate nowcast of the current state of the economy, we would like to know the (in-sample) uncertainty belonging to the estimate. Under the assumption that the common component and the idiosyncratic component are uncorrelated, we have the following expression for the corresponding uncertainty,

$$\begin{aligned} \text{Var}(\widehat{Y}_t) &= \text{Var}(\beta F_t + \epsilon_t) \\ &= \beta' \text{Var}(F_t) \beta + \Sigma_t \\ &= \beta' S_{t|t} \beta + \Sigma_t, \end{aligned} \tag{13}$$

where Σ is estimated by

$$\widehat{\Sigma} = \frac{1}{T-1} \sum_{t=0}^{T} \hat{\epsilon}_t^2 = \frac{1}{T-1} \sum_{t=0}^{T} (X_t^Q - \widehat{\beta} F_t^Q)^2. \tag{14}$$

The source of uncertainty is twofold. On the one hand uncertainty is coming from the nowcasts of the latent factors by the Kalman filter, given by $S_{t|t}$. On the other hand we have the estimation error from regression (3). One must keep in mind that the estimates rely on the assumption of Gaussian errors and a proper specification of the model.
For some aggregation methods described in the previous section, an expression for the variance of the sum of factors is needed. This requires taking into account the autocorrelation between the estimated factors. As follows, an expression for $C_{s|t} = \text{E}[(F_t - \widehat{F}_t)(F_{t-s} - \widehat{F}_{t-s})]$ is needed. It turned out that the factors only showed significant autocorrelation at lag 1. For higher lags the autocorrelation was negligible. Therefore, in addition to the variance matrix (8), the Kalman filter will also provide an estimate for $\text{Cov}(F_t, F_{t-1}|\Omega_t)$, namely

$$C_{1|t} = A S_{t-1|t-1} - K_t \Lambda A S_{t-1|t-1} = (I_r - K_t \Lambda) A S_{t-1|t-1}. \tag{15}$$

The total covariance matrix at time t will be

$$\Psi_{t|t} = \begin{bmatrix} S_{t|t} & C_{1|t} \\ C_{1|t} & S_{t|t} \end{bmatrix}. \tag{16}$$

The expression $K_t \Lambda$ in (15) reflects the autocorrelation. In the absence of autocorrelation, the factors of the previous period will not provide any new information. They will not be correlated with the news of the current period, which will cause the expression $K_t \Lambda$ to be zero. The updated covariance matrix then just equals that of the previous period multiplied by the system matrix A.

3 Data and nowcast design

The first two subsections are devoted to the selection procedure of the indicators. Subsection 3.3 deals with the initial processing of these time series. Further, Subsection 3.4 describes the way the real-time information flow is simulated and the final subsection covers some remarks concerning data revisions.

3.1 Data selection

The success of the nowcast depends heavily on the indicator time series being used (e.g. Bernanke and Doivin (2003) or Banbura and Rünstler (2011)). The indicators for the GDP growth rate need to meet two criteria. First of all, as with regular forecasting, the time series should be correlated with GDP. They should have some predictive power on the GDP growth rate. A first selection was made based on economic theory and logical reasoning. The number of mortgages granted towards households for purchasing houses might reveal some information about the state of the economy, while this might be questionable for the number of cows in The Netherlands. The former might reflect the confidence consumers have in the economy, while the cattle industry is of less importance in The Netherlands than for example (commercial) services.

Secondly, nowcasting works better with indicators that are released more frequently than the GDP time series and are timely. Therefore, time series that are only published yearly were excluded. The prior focus was set on time series with monthly cadence of availability/ publication. Some quarterly time series were added if it was expected that they would be of considerable importance in predicting GDP. For the PCA, which requires a full matrix without missing values, these time series were converted into monthly time series by simple linear interpolation between the quarterly data points. This was deemed preferable to performing a PCA on quarterly data, given the loss of information that that would imply. Higher frequency data, such as daily closing prices of the AEX index were aggregated to monthly data by taking the monthly average, although this results in some information loss. The reason is that not many daily time series are available that provide useful information on GDP. The main purpose is to design a nowcasting model in which all time series are considered on a monthly basis in order to make proper predictions of quarterly growth rates.

An additional point of consideration when selecting time series was that they should have enough data points. The time series should at least have started in 2005 in order to allow for proper backtesting. A final 'criterion' was the reliability of continuation of the time series. For using the nowcasting model in practice, one does not want to rely on time series of private parties that might stop publishing these time series sometime in the future.

3.2 The set of predictors

The total set of indicators comprised 88 economic and financial indicators. A complete overview of the time series can be found in Table II.1 in Appendix II. The main source for the time series is SN itself. Some additional time series were obtained from the Dutch Central Bank, the European Central Bank and the OECD website.

First of all, historical data on the Dutch GDP are used as indicators since these naturally contain a lot of information on the current GDP. The problem is that those time series are, as explained in the introduction, published with a significant lag. According to the expenditure approach, GDP is the total sum of final uses of goods and services. Consequently, it is the sum of household consumption (C), investments made by companies (I), government spending (G) and net export ($\Delta E \equiv Export - Import$),

$$GDP = C + I + G + \Delta E,$$

where only expenditures on final goods and services count, not on intermediate goods. Based on this approach, time series that provide information on each of these four components were added. For consumption, which in general accounts for most of the GDP, time series of income and expenditures of households were added. Since prices influence expenditures, monthly time series of price indices of both (commercial) services and several kinds of goods were also used. They are especially interesting in the context of nowcasting since they are released monthly and quite quickly after the ending of the month. Data on the investment activity of companies are contained in for example the number of loans granted towards corporations, the number of bankruptcies and average daily production of certain sectors. For energy, mining, industry, minerals and water monthly production levels are available with a forty day lag. One has to keep in mind that those numbers are provisional and can be revised quite significantly. For other sectors, only data of the previous quarter are available. Information on the government spending component is provided by adding time series on the central government's balance sheet, its main income components and its main spending components. The last component, trade, is of great importance for a small open economy like The Netherlands. This justifies the inclusion of time series of trade, such as export and import volume per month.

A second group of time series tells something about the state of the economy, without being directly related to the build-up of GDP. Predictors one can think of are time series related to the labour market, such as unemployment rates, the number of outstanding vacancies, the number of self-employed or numbers on welfare benefits. Another important indicator of the performance of the economy is the housing market, as we have learned in the previous decade. Time series such as the number of existing houses sold or an index number measuring added value of construction can be informative for the healthiness of the economy, just like the number of mortgages taken out.

The last group of indicators belongs to the set of soft information. Soft information refers to opinions, ideas and expectations about the past, current and future performance of the economy. These factors are referred to as business cycle indicators. An important indicator of potential trend change in household consumption is the consumer confidence. Monthly, SN investigates the confidence by surveying various points. Relevant for the GDP estimate are presumably the economical situation and the financial situation in both the past twelve months and the coming twelve months, as well as the willingness to buy. These average scores reflect pessimistic or optimistic feelings of society towards the state of the economy, which in turn will affect consumption.
The business counterpart is the business cycle survey carried out among companies. Some industries are surveyed on a monthly basis, whereas others only four times a year. Results of the business cycle survey for retail and commercial services were used since these are released monthly. Examples of such sentiments are expectations towards prices, revenues, sales and the economic climate in general.

3.3 Preprocessing data

The model described in the previous section requires stationary time series $y_1, ..., y_N$. Basically all the time series used are nonstationary, which is a common feature of macroeconomic and financial time series. However, the transformation needed to make the time series stationary depends on the source of the nonstationarity.

If a time series exhibits a clear trend, then the appropriate way of removing the trend should be applied. Three common (nonstationary) processes are the random walk, the random walk with drift and the time series with a deterministic trend $f(t)$, which consecutively can be written by

$$y_t = \gamma y_{t-1} + \epsilon_t, \quad \text{with} \quad \gamma = 1 \tag{17}$$

$$y_t = \beta + \gamma y_{t-1} + \epsilon_t \quad \text{with} \quad 0 < \gamma \leq 1, \tag{18}$$

$$y_t = \alpha + f(t) + \epsilon_t. \tag{19}$$

In general two main types of trends are distinguished. The first two processes are examples of processes with a stochastic trend. The random walk has a constant mean (equal to its starting value y_0), but its variance changes over time as $t\sigma^2$, while the random walk with drift (18) also has a time-varying mean. The latter results in an up- or downward linear trend, depending on the sign of β. The second type of trend is the so-called deterministic trend. It contains a time dependent component, see Equation (19). The process has constant variance, but time-varying mean. If $f(t)$ is linear $= \beta t$ then E$[y_t] = \alpha + \beta t$. Correct transformation depends on the nature of the trend. Using the wrong transformation can lead to very different results (see for example Diebold and Senhadji (1996) or Dagum and Giannerini (2006)). Processes with a deterministic (linear) trend are trend-stationary and can be made stationary by removing the time trend $\alpha + \beta t$, which can be found by regressing y_t on t. The variance of a deterministic trend is constant. Time series with a stochastic trend can be transformed and be made stationary by taking differences.

From a plot one can often not easily distinguish stochastic from deterministic trends. However, one can at least rule out the random walk process. In order to determine whether a trend is stochastic or deterministic one can apply the Augmented Dickey-Fuller (ADF) unit roottest, with the appropriate null hypotheses. From Equations (18) and (19) one can see that the variables with a deterministic trend have no unit root. They are $I(0)$ around a deterministic trend, while the time series with a stochastic trend are $I(1)$ processes with drift. If one considers the first difference $\Delta y_y \equiv y_t - y_{t-1}$ and does a regression of the type:

$$\Delta y_t = \delta y_{t-1} + \beta + \beta_1 t + \epsilon_t, \tag{20}$$

Introducing the possibility of a deterministic trend $\beta_1 t$ for the first difference Δy_t implies that the trend $f(t)$ in Eq. (19) is assumed to have terms of higher order than just a linear βt term. In a more general setting more lags $\Delta_k y_t \equiv y_t - y_{t-k}$ could also be included on the right hand side, but that is less appropriate here.

We now have two competing hypotheses: H_0) $\beta_1 = 0, \delta = 0$ and H_1) $\beta_1 \neq 0, \delta < 0$. Whenever the null hypothesis is rejected, Eq. (17) can be excluded as an appropriate model for the time series. The trend component is most likely deterministic, but as an extra test we also tested whether (at least one of) the other ADF-tests rejected the single null hypothesis of having a unit root $\delta = 0$, confirming the existence of a deterministic trend. The existence of both a unit root and a deterministic trend at the same time (i.e. $\beta_1 \neq 0, \delta = 0$) is not likely to occur (Elder and Kennedy, 2001; Perron, 1988).

3.4 Simulating real-time information flow

In order to backtest the nowcasting models, the information that was available at a certain point in time needs to be simulated. To achieve this, the flow of data releases for the various predictors needs to be simulated. The flash estimate is always made 45 days after the end of the reference quarter. Therefore, the artificial date on which nowcasting is performed is set on the fifteenth of a month. That means that if for example 'today' is the 15th of May 2009, we are nowcasting the GDP of April 2009. Whenever for a certain predictor the release date for April is before May 15, its corresponding publication lag is set to zero. This is the case for most survey indicators. When on May 15 the observations for both April and March are missing for a particular time series, the corresponding lag is two. In this way, for each predictor time series its corresponding lag is determined by looking at the publication date for a particular month. The complete data set was downloaded on May 15, 2018. It is assumed that the publication calendar remains the same throughout the backtesting period. The approach set out above, results in jagged edges at the end of the information matrix at time t. As explained, the Kalman filter can easily handle these jagged edges and exploit the information from timely released indicators.

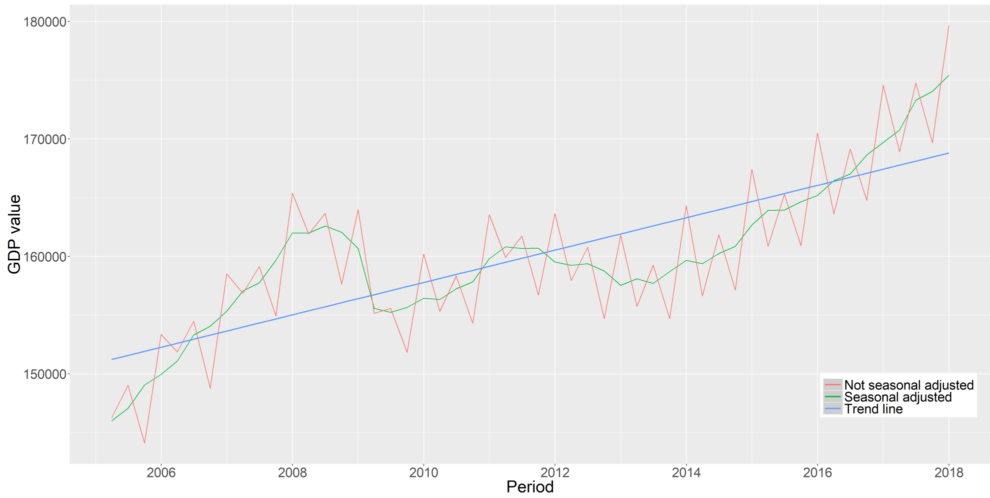
3.5 Data revisions

A final remark concerning the data is that some time series are subject to revisions. In particular the production series and the historical GDP series used for backtesting are revised several times. These time series contain the revised data up to and including 2014. For the quarters of 2015 till now provisional data are available. As a consequence, the simulated information set at a certain point in time t will not exactly replicate the actual information set as it would have been at that time. At a particular moment t people did not yet have access to the revised GDP growth rates of the first few preceding quarters, while now the simulated information set for time t contains the revised and hence the final GDP growth rates. This might result in smaller deviations of the nowcast and hence a better performance. However, Schumacher and Breitung (2008) explored the effect of data revisions on the performance of the models for the German GDP. They concluded that using either a real-time data set or a data set containing final GDP data did not lead to significantly different results. A similar conclusion was drawn by Bernanke and Doivin (2003) for the USA. As a matter of fact, for most time series the original numbers are not available, solely the final numbers. Therefore, the revised indicator time series are used for nowcasting.

4 Preliminary data analysis

First, some analysis on the historical GDP time series is conducted to give an impression of the data. The figures below provide a graphical representation of the GDP evolution. Figure 4.1 plots both the seasonal adjusted and the non-seasonal adjusted time series of GDP over the period 2005-2017. Apart from the two drops caused by the financial and subsequently the Eurocrisis, the real GDP exhibits a clear trend upwards. Denote the log-series by $X_t = \log(z_t)$, with z_t being the GDP value at time t. Conducting a unit root test as described in Section 3.3 suggests that GDP has a deterministic trend.

Figure 4.1 GDP value 2005-2017



The lag for testing was set on $p = 1$ since the partial autocorrelation function (PACF) of X_t only showed a significant lag at $p = 1$. The AIC criterion also suggested using $p = 1$. However, for higher lags the null was not rejected, indicating the existence of a stochastic trend. The autocorrelation function (ACF) of the log-series, showed exponential decay, supporting the existence of trend-stationarity of the time series (Caiado and Crato, 2005). When taking the shorter period 2005-2017, no lag resulted in rejecting the null hypothesis of having a stochastic trend. In particular for the evaluation period we can conclude that GDP has a stochastic trend. A reason for this could maybe be found in the financial unstable period during the first decade of 2000.

Figure 4.2 Quarter on quarter growth rate GDP 2005-2017

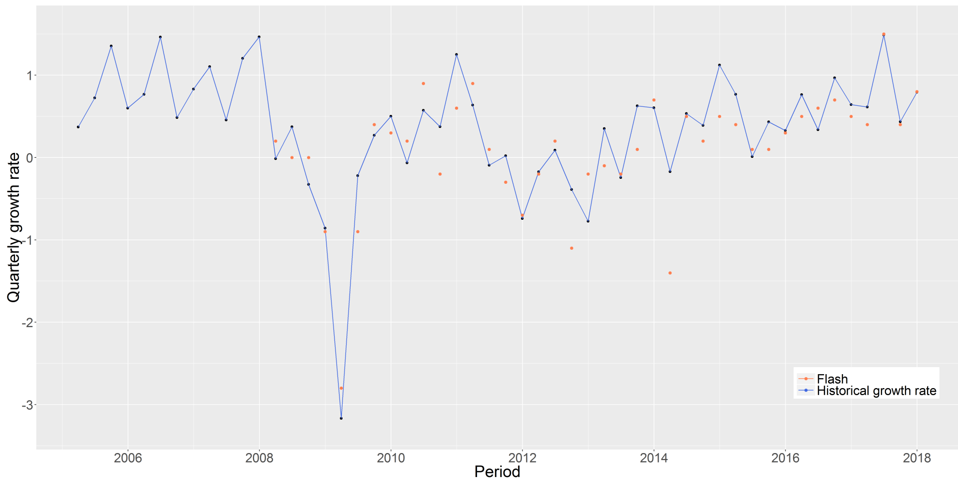


Figure 4.2 plots the quarter on quarter GDP growth rate over the period 2005-2017. SN has started publishing quarter on quarter growth rates in 2008. Therefore, the orange dots, representing the flash estimates, are only shown over the period 2008-2017. Some substantial deviations from the final growth rates can be observed. From the plot it seems that the flash often underestimates the growth rate. The historical bias of the flash is indeed -0.1246. It on average underestimates the GDP growth.

In order to examine the data more closely, a linear filter was applied. The objective of a linear filter is to decompose a time series in a level, trend, seasonal and noise component. The true

underlying evolution of the time series can then be revealed as well as the patterns of the various components. The former three can be modelled, while the latter is the random part. All time series have a level (could be zero) and noise component. The most basic filter is a moving average with equal weights. Widely used are the Henderson filters (Henderson, 1916). To the GDP series, unequal weights as proposed by Perrucci and Pijpers (2017) are applied, which are better capable of identifying seasonal patterns with not perfectly constant frequencies. It removes signals within a certain band of frequencies, for example around a year. The filter is applied to the original raw time series of GDP. The plots of the various components can be found in Appendix III. The seasonal pattern observed in Figure 4.1 is nicely filtered, with some irregularities around the financial crisis (see left panel of Figure III.1). However, the plot of the noise component (right panel of Figure III.1) suggest that there is another pattern with a frequency different than a year that is not filtered. The filter only filters the yearly seasonality. Applying the filter again results in the second set of figures. From this second step it becomes clear that there is also a periodic pattern in the GDP data with period larger than one year. The resulting noise component now looks much different (see right panel of Figure III.3). One can now distinguish the peaks around the two crises. As expected, the trend component is evident here and follows the GDP line very closely, see Figure III.4.
A Fourier transform (Figure III.3) of the time series of the GDP growth rate of Figure 4.2 has distinct peaks both at a period of 16 months and at 38 months. The origin of such a non-seasonal periodicity or cycle with a period in the range of larger than one year, up to about three years, is unclear at this stage. Repeating the filtering process a second time on the GDP series can help to suppress the 16-month periodicity, although imperfectly since that is not what it is designed to do. Nevertheless the remaining noise term (right panel of Figure III.2) is much reduced, and the signatures of the two crises become visible. The process by which the final revised GDP growth rate is determined more than 32 months after the first estimates needs to be investigated more closely, to determine whether it could have generated such an effect, but that is outside of the scope of this research.

5 Principal Component Analysis

The PCA results into N principal components. To reduce the dimension of the problem, the first r provide the estimates for the common factors $\widehat{F}$ in Equation (1). After PCA is conducted a crucial decision has to be made, namely selecting the number of principal components. Taking too few components can result in information loss and consequently less reliable estimates or predictions. On the other hand, too many factors might lead to factors with insignificant loadings and can reduce the efficiency of the model (Zwick and Velicer, 1986). In the first subsection we will determine the number of factors needed for both the EA and OA models. For the quarterly differenced models and the OA some interpretation of the factors is provided in the last subsections.

5.1 Selecting number of factors

For selecting the number of factors a couple of tests were conducted. All the corresponding tables and figures can be found in Appendices IV-VI. For the monthly models, Figure IV.2 in the Appendix provides the first fifteen eigenvalues of the correlation matrix ρ with the corresponding proportion of the total variance explained, as well as the cumulative proportions. These proportions are also presented in the scree plot in IV.1, as originally proposed by Cattell (1966). A relatively informal test is to look for the elbow in the scree plot.

Four additional tests based on the scree plot were conducted. The Acceleration Analysis, which is based on the curvature of the scree plot, essentially looks for the elbow in the scree plot by looking for the maximum acceleration factor AF_i of all eigenvalues, i.e. the second order derivative. The first eigenvalue exceeding the coordinate corresponding to the maximum acceleration factor determines r. A more solid test is the Parallel Analysis (Horn, 1965). It bootstraps numerous correlation matrices from the original data set Y and for each replication it determines the eigenvalues to obtain a (bootstrap) distribution for the eigenvalues. The number of bootstraps was set on 1000. The Parallel Analysis then compares the eigenvalues from the actual data set Y with the 95^{th} quantile of the empirical distribution. The first eigenvalue (when ordered in decreasing order) that is lower than its quantile estimation determines r (Buja and Eyuboglu, 1992). The third test is the classical Kaiser Rule, which simply compares the eigenvalue with the mean eigenvalue and only eigenvalues larger than $\bar{\lambda}$ are retained. Since the correlation matrix is used, $\bar{\lambda}$ equals one (Kaiser, 1960). A fourth test is the so called Optimal Coordinates method. It basically extrapolates the previous eigenvalue and compares the next eigenvalue with the 'optimal coordinates' obtained from the extrapolation. The Optimal Coordinates and the Acceleration Analysis methods were proposed by Raîche et al. (2013). The four tests with their respective r are provided in the table below.

Table 5.1 Optimal number of factors according to the four tests. The optimal number of factors according to the four tests are provided for the monthly ($\mathbf{r}_m$), quarterly ($\mathbf{r}_q$) and OA ($\mathbf{r}_o$) models. The second column provides the mathematical representation of the decision rule corresponding to each of the four tests.

Method	Decision rule	$\mathbf{r}_m$	$\mathbf{r}_q$	$\mathbf{r}_o$
Acceleration	$\lambda_i > \lambda_{i,\alpha}^{boot}$ & $\max(AF_i)$	3	1	2
Kaiser Rule	$\sum_i(\lambda_i > 1)$	21	18	11
Optimal Coordinates	$\sum_i [\ \lambda_i > \lambda_{i,\alpha}^{boot}$ & $\lambda_i > \lambda_i^{extra}\ ]$	9	10	7
Parallel Analysis	$\sum_i(\lambda_i > \lambda_{i,\alpha}^{boot})$	9	10	7

A more formal approach is using the Bai and Ng (2002) information criterion. Remember that PCA is finding $\hat{F}$ and $\hat{\Lambda}$ that minimise (4). An information criterion proposed by Bai and Ng (2002) is the sum of (4), denoted by $V(F,\Lambda)$, plus a penalty term. The optimal number of factors is given by the value of r that minimises the criterion

$$IC_i = \log(V(F,\Lambda)) + g_i(N,T), \tag{21}$$

where the penalty should depend on both N and T. They retain the best results by using

$$g_1(N,T) = \frac{(N+T)}{NT}\log(\min(N,T)). \tag{22}$$

We also considered a second penalty function, namely

$$g_2(N,T) = \frac{(N+T)}{NT}\log(\frac{NT}{N+T}). \tag{23}$$

Since the tests do not unequivocally point to the same best values for r any choice is slightly subjective. Based on the overall impression from the scree plot, the four tests and the information criteria the optimal number of factors for the monthly and quarterly models are close to eight and ten respectively. Together these factors are able to capture 56.24% and 71.01% of the total variance.

All the results of the PCA of the OA can be found in Appendix VI. At first glance a striking difference with the EA can be observed. The first factor is able to catch almost 30% of the total variance, as opposed to approximately 10% for the stationary situation.
Based on the tests between five and seven factors seems to be optimal. The latter two factors only had a few substantial factors loadings. Therefore, five seems to be the appropriate number of components, as is affirmed by the scree plot and the information criteria.[1] These factors together capture 79.27% of the total variance.

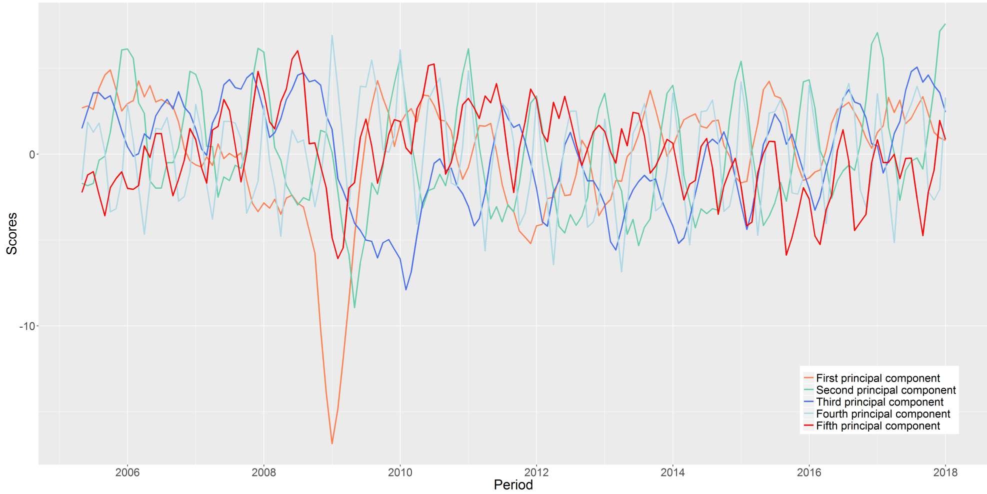
5.2 Interpretation of factors: quarterly models

As can be observed in Figure 5.1, the factors seem to be stationary. When applying unit root tests on each of the factors, they all rejected the null of having a unit root up till (at least) 8 lags (2 years).

- The first factor captures the general trend. It exhibits the clear drop during the financial crisis and subsequently the (less severe) drop around the Eurocrisis. It is highly correlated with producer sentiments and vacancy indicators, CLI and the AEX index.
- The second factor is a fairly stable factor. It shows a clear seasonal pattern. Indicators that load heavily on this factor are indicators regarding the production, revenues and prices of raw materials, minerals and energy. In addition, some CPI and governmental series are related to this factor.
- The third factor loads quite heavily on indicators that respond slowly to economic decline, such as unemployment rates, numbers on social benefits and consumption of durable goods. The drop of the crisis also shows a little delay compared to the other factors. In addition, series on the housing market are related with this factor.
- The fourth factor is highly correlated with time series of sentiment indicators. It also contains quite some time series on consumption and is highly correlated with the number of mortgages taken out. It might reflect the consumer and producer sentiment.
- The final factors do not exhibit a clear correlation with a certain group of indicators. They contain for example prices and revenues of industry and raw materials. However, also governmental time series and series on trade are captured in these factors.

1) The value of r that minimizes the criteria is four for both (see Figure VI.3). Here a modified criterion was used. Since using the Criterion (IV.3) resulted in ever decreasing values of the criterion, we adopted the criterion by replacing the term $\log(V(F,\Lambda))$ by $V(F,\Lambda)$. The value of the criterion for $r = 5$ is close to that of $r = 4$. The step towards six is bigger.

Figure 5.1 First five principal components. In order to obtain a clearer overview, only the first five components are plotted.



From Table 5.2, it becomes clear that in particular the first three factors are crucial in describing the evolution of the GDP. After the third factor the correlations between GDP and the factors decrease substantially.

Table 5.2 Correlation GDP and factors of quarterly models

Factor	1	2	3	4	5	6	7	8	9	10
Correlation	0.572	0.424	0.324	0.003	0.050	0.055	0.101	0.059	0.188	0.056

5.3 Interpretation of factors: OA approach

5.3.1 Relation to GDP

For this part, GDP itself was added to the set of predictors Y_t. This did not alter the results above significantly. For example, when GDP was added the total variance explained by the first component was 32.87% and for the first five 79.44%, compared to 32.61% and 79.27% without GDP.

As the variable of interest is GDP, one would like to have factors that, at least together, capture a large part of the variance of that variable. The common component of GDP is $C_t = \lambda_{gdp} F_t$, in which λ_{gdp} is just the row-vector of loading matrix Λ corresponding to GDP.
The sum of the squared row elements of the loading matrix, $\sum_{r=1}^{5} \lambda_{r,j}^2$, gives the communality of the j-th variable in Y_t. This is a measure of the proportion of variance in that particular variable explained by the factors together. Communalities close to one indicate that the extracted components are able to explain most of the variance. They therefore provide an idea of the performance of the model with respect to individual indicators. The communality for GDP is 0.937. Since the data were standardised, the GDP specific variance is 1-0.937 $= 0.063$, confirming that five principal components are sufficient to roughly capture the dynamics of the Dutch GDP. Adding a sixth factor led to a communality of 0.953, so not much gain is attained. The most explanatory power can be attributed to the first component, as can be seen in Table 5.3. It shows the correlation between GDP and the five factors. Here one should keep in mind that the PCA was applied to nonstationary data. The first component captures the common trend.

Figure 5.2 GDP and its common component. Top Panel: Monthly GDP and its common component (standardised). Bottom panel: Quarterly GDP and its common component

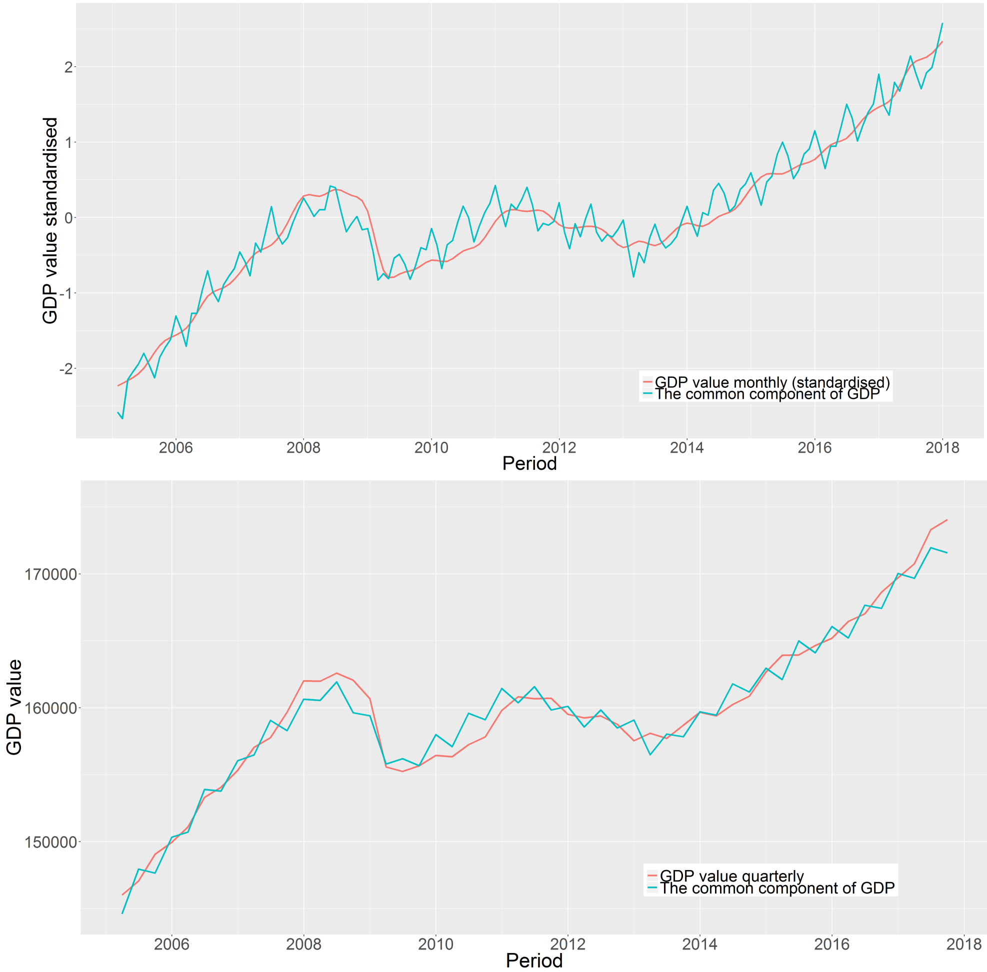


Table 5.3 Correlation GDP and factors of OA

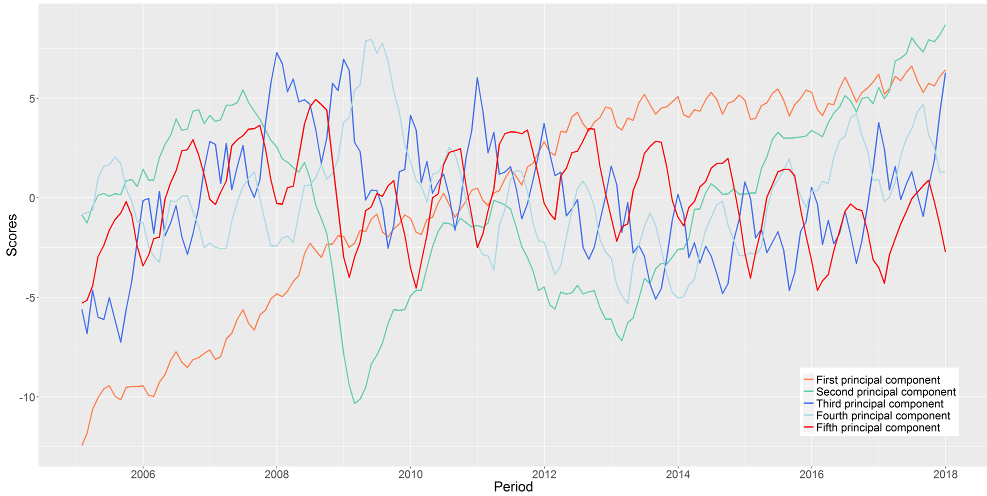
Factor	1	2	3	4	5
Correlation	0.761	0.465	0.340	0.158	0.027

In order to visualise the explanatory power of the factors with respect to GDP, the common component of GDP is plotted together with the monthly (interpolated) historical GDP. They follow each other closely. Figure 5.2 is obtained when applying aggregation in order to get quarterly data and after destandardisation. Both figures illustrate that the dynamics of the GDP is explained quite well by the common factors and that the effect of the individual specific variation is limited.

5.3.2 Interpretation of factors

This section attempts to give an economical interpretation to the factors found for the OA. One should keep in mind that the factors found by PCA are not unique. Different representations of the factors, called rotations, result in different factors. The so-called varimax rotation preserves the orthogonality, but often leads to a rotation with nice interpretations since it puts heavier weights on some variables and almost none on others. It maximises the sum of the variances of the squared loadings. Applying this rotation resulted in the loading matrix provided in Figure VI.4 in the Appendix. Figure 5.3 plots the first five principal components, where the first, second and third column were multiplied by minus one to obtain nice representations and facilitate interpretation. This also made sure that the correlations between GDP and each component were positive.

Figure 5.3 The first five principal components after varimax rotation



As expected, the resulting five factors are not stationary now, since we applied PCA to a nonstationary data set. The first factor clearly picks up the general trend in the economy. When transforming the original data set Y_t we observed that many time series exhibited a trend. The PCA applied on the nonstationary time series causes the PCA to capture this common growth. This also explains why the number of factors needed to explain a certain part of the covariance is significantly lower compared to the situation where PCA is applied to a stationary data set. The first factor can explain up to 30% of the variance, compared to around 10% in the stationary case. The scree plot therefore also has a more clear cutoff point. Nevertheless, we can find nice interpretations of the factors.

- The first component seems to represent the general underlying trend of the economy. It shows a trend that overall moves upwards, with a small drop around the recent crisis. One can think of it as representing innovations.
- The indicators of the business cycle survey conducted under Dutch companies all load heavily on the second factor. Likewise, the Composite Leading Indicator (CLI), another sentiment, and vacancy indicators (of the industry, construction sector and commercial services) appear to be dominant in this component. Therefore one could regard this factor as representing the sentiment element. These sentiments are often early indicators of changes in the economy. The light green line in Figure 5.3 exhibits a substantial and steep decline right at the beginning of the crisis. Sentiments are indicators that react quickly to changes in the economic outlook.
- The third component contains quite some prices and consumption indices, as well as some social benefit indicators. Just before the financial crisis, prices rose tremendously, which is reflected in the blue line in Figure 5.3. People thought that the economy would keep on growing forever and expenditures were high. The number of citizens receiving social benefits reached its lowest point, just like the number of bankruptcies. In this factor one can also see the two economic drops.
- In contrast to the second component, the fourth component seems to respond slowly, reaching lower levels in the period 2012-2014. Taking a closer look at the factor loadings

reveals that indeed indicators that in general respond slower to economic decline are represented in this factor. The indicators that load heavily on the fourth component can be placed in three categories: employment (unemployment rates, number of vacancies, number of unemployment benefits and total social benefits), consumer confidence (pricing index sold houses, total consumption of durable goods, total domestic consumption, consumer sentiment 'time to do big purchases' and savings of households) and indicators for the business sector, such as production of the construction sector and the number of pronounced bankruptcies in court. Observing these time series individually over time reveal that they all have one thing in common: they tend to respond to the crisis with some delay. The drop of the fourth component shows the effect of the Eurocrisis, the crisis that followed after the financial crisis in 2008. During the Eurocrisis people really started to feel the consequences of the economic depression. Consumer confidence reached its lowest level ever, the housing market collapsed in The Netherlands (partly due to some peculiarities in the Dutch housing market) and even initially healthy companies experienced troubles. Companies will in general have some reserves and will not go bankrupt within a couple of months. Therefore, there was a peak in the number of bankruptcies in 2012-2013. The same holds for the number of vacancies. When economic welfare deteriorates companies will not (be legally able to) fire employees right away. Moreover, when the economic perspectives improve, companies are in general holding back with hiring new employees. This causes the number of vacancies to increase not before 2015, resulting in high unemployment rates in the preceding years. The consumer confidence indicators also reflect this conservative behaviour. Household savings is also highly correlated with the fourth component. Households postpone big purchases and save money since they still remember the crisis. This in turn affects the housing price index and less new houses are being built. Both consumers and companies need time to recover from the impact of the crisis and to restore their confidence in the economy. In 2015, consumer confidence was equal to its pre-crises level again.
– The fifth component can be interpreted as the steady factor. Production indicators of raw materials, electricity, gas and petroleum are well represented in this factor. The red line in Figure 5.3 reveals the seasonal pattern of these indicators. During winters, prices will be higher. These indicators are hardly affected by a crisis since consumers and companies have to use gas and electricity anyhow. The index for revenues of the catering industry is also heavily (and solely) correlated with this factor. Moreover, another important indicator in this component is the consumption of 'other goods', which makes sense since it entails for example consumption of gas. Finally, some governmental indicators are represented in this last factor. They are all indicators on the income side of the balance sheet of the central government, which are alike the other indicators in general fairly steady elements. They do not fluctuate enormously over time. The income of the Dutch government is partly dependent on the income of gas, causing it to exhibit a seasonal pattern as well.

In summary, the five factors can be referred to as the 'general economic trend'-, 'sentiment'-, the 'consumption and prices'-, the 'housing market'- and the 'governmental and commodity'-factor respectively.

6 Results

The four methods of the EA set out in Section 2.4 and the OA are put to work in this section. A simple AR model of order two is used as a benchmark. Since the test is to simulate every nowcast based *only* on the data available *at that point in time* the data does not allow making nowcasts starting in 2005 when the first data are available. No attempt is made to nowcast the first few years of quarterly data The nowcasts simulated the period 2008-2017, resulting in 40 quarters to be nowcasted. The nowcast accuracy is measured by the MSFE. Since the Kalman filter needs some time to start up, even starting the simulations only in 2008 the nowcasts of the first few quarters deviate a lot and would dominate the overall MSFE. In order to make a sensible comparison between the methods, the MSFE calculation is restricted to the period 2011-2017: i.e. 28 quarters.

Before turning to comparing the six methods, the effect of the number of factors (r), the number of lags p in the VAR-regression (2) and the effect of adding lags (ly) of GDP itself to regression (3) are investigated in order to select the optimal models. Since the effects are different for the EA and the OA these effects are treated in different subsections. The MSFE as well as the average standard error (see Equation (13)) for various combinations of these parameters are provided in Appendix VII for the EA and Appendix VII.8 of the OA. In each table the lowest MSFE or lowest standard error for a certain method is underlined.

The hitting times are also included, which represent the number of times that the true historical value of GDP (growth rate) fell into the 95%-confidence interval of the particular estimate. For the OA confidence level 0.98 was used because of the larger numbers (GDP value instead of GDP growth rates). In total there are 28 observations. The number of hits h can be regarded as a binomial distributed random variable, $\sum_{i=1}^{28} \mathbf{1}_{\{x_i \in CI\}} \sim \text{bin}(28, 0.95)$. The theoretical probability that a certain point falls into the confidence interval is $p = 0.95$. We performed a binomial test with confidence level 0.95 to test whether the probability of success is 0.95. For $h \leq 24$ the null hypothesis of $p_0 = 0.95$ was rejected.

In Subsection 6.3 the Model Confidence Set (MCS) procedure is applied to reduce the set of models after which model averaging is applied to these models in Subsection 6.4. Subsection 6.5 investigates the effect of the indicators on the nowcast as well as their uncertainty. In the last subsection some of the model assumptions are tested.

6.1 Econometric Approach

6.1.1 The effect of number of factors

For the monthly differenced time series, the MSFE of the second model (M2) (see overview in appendix I) is lower than that of the first one for $r \leq 7$, see Table 6.1. Including more than seven factors does not lead to decreasing MSFE for both model M1 and M2. This is not what we would expect based on the PCA tests. However, more factors does lead to some improvement in the average standard error for model M2 (see Table VII.1), but not for model M1. This effect stops after $r = 12$. For example, for $r = 15$, $\text{MSFE}_{\text{M2}} = 0.7507$ and the average standard error is 0.5029. The uncertainty of the models can explain the coverage ratios found (Table VII.2). The first model M1 entails quite some uncertainty which results in coverage ratios of 1 for almost every r. Here a side note needs to be made. The way the standard error of M1 is calculated does not take into account the covariance between F_t and F_{t-1}. Therefore this error most likely overestimates the uncertainty.

Based on the MSFE the optimal r would be four for both models. Based on the PCA and a balanced choice between uncertainty and MSFE, we also consider taking $r = 8$. Adding more factors increasingly adds noise instead of useful information.

For the quarterly differenced time series the second method (Q2) clearly outperforms the first method of Giannone et al. (2008). Taking the sum of the monthly factors instead of 'simply' taking the last element decreases the MSFE for all r. More factors increased the MSFE (for example 0.4586 for $r = 15$), while it did not reduce the average standard error. Therefore we will try both $r = 4$ (lowest MSFE) and $r = 10$. The latter is based on the PCA analysis and a relatively low average standard error.

For the same period the MSFE for the current flash estimate is 0.2328, but this may be underestimated somewhat since the final quarters of the time series are not yet definitive which means that the current estimates are probably closer to the flash than the definitive values will turn out to be, thereby lowering its MSFE compared to other models.

Table 6.1 Mean Squared Forecast Error (MSFE). Lags in VAR-regression: 1 ($p = 1$). No lagged GDP information ($ly = 0$). For comparison $\text{MSFE}_{flash} = 0.2328$ for this period.

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.4280	0.4887	0.5403	0.4509	0.4976	0.5389	0.6274	0.6940	0.6054
M2	0.2318	0.2719	0.3968	0.3890	0.6015	0.7704	0.6355	0.6339	0.6436
Q1	0.3043	0.3182	0.3466	0.4019	0.4533	0.4643	0.4258	0.4548	0.5589
Q2	0.2415	0.2569	0.2633	0.3171	0.3467	0.3477	0.3341	0.3194	0.3462

6.1.2 The effect of order VAR regression

First of all, method M1 is not appropriate when using $p \geq 2$. Due to the fact that it already regresses on lag terms of the factors, adding lags to the VAR-regression of the factors introduces collinearity and hence no appropriate estimate of β is possible. Increasing the number of lags of the VAR regression did not improve the performance of model M2. In particular for higher r the MSFE increased significantly, see Table VII.3. The uncertainty of the model also increased when including more lags. Method M2 already includes information from previous factors. A higher order VAR model for the factors therefore does not seem to add any relevant information, merely noise. The same holds for the two methods Q1 and Q2. Only for small r the results are similar or a little worse than that of $p = 1$, but for $r \geq 6$ the MSFE and average variation increase dramatically. Another undesired consequence is that all the coverage ratios were 1. When plotting the figures, one could observe that the higher number of lags in the VAR regression caused in particular problems in the beginning of the simulation period (see for example Figure VII.1 in the Appendix). Henceforth, the VAR p parameter is fixed at $p = 1$.

6.1.3 The effect of autoregressive terms GDP

It is expected that the growth rates of previous quarters will contain valuable information for the current growth rate. In Figure 4.2 one can observe that growth rates tend to cluster somewhat. It does not jump up and down from positive to negative.

Overall, including lags of GDP led to some improvement in the MSFE of both monthly models (see Tables VII.6, VII.9 and VII.12). For M1 this effect is seen up to $ly = 1$, while for M2 the lowest MSFE is obtained when setting $ly = 2$. Interesting to see is that while the first model's uncertainty is significantly lower when ly is increased, the uncertainty of M2 is barely affected by including lags of GDP. The standard error of M1 decreases by roughly 40%. This could be due to the fact that M2 already takes into account more historical information by considering more lags of the factors. For both models, the coverage ratios are good. An observation worth mentioning is that when including lags of GDP growth rate, method M1 has the tendency to 'overshoot': it overestimates the growth rate when this is positive and underestimates when it is negative. One could argue that it puts too much weight on the previous growth rates.

Surprisingly, the quarterly models did not perform better when information on the previous growth rate was added, contradicting Jansen et al. (2012). Although the difference is not big, both models have a higher MSFE when the number of AR terms is increased. The average standard deviation is barely affected. Only the standard deviations of Q2 become a little bit smaller. However, the gains are so small compared to the loss in MSFE that we will use $ly = 0$.

In summary, for M1 the best results (in terms of MSFE) are obtained using the combination $r = 4$, $p = 1$, $ly = 2$, for M2 we have the same but $ly = 1$. For both models we also try $r = 8$, based on the PCA. For the quarterly models we have $r = 4$, $p = 1$ and $ly = 0$. Additionally, $r = 10$ is considered. The models with $r = 4$ are modest improvements on the simple AR model in terms of MSFE. The difference is not much since the MSFE of the AR model is 0.2464. The improvements in terms of uncertainty are larger. The AR model has an average standard error of 0.7764 which is substantially more than all of these best model results.

6.2 Operational Approach

For the OA two variants were compared, first taking just the third month of each quarter (Giannone et al., 2008) and secondly taking the three-month average of the factors. These two variants will be referred to as O1 and O2 respectively. The findings are summarised in Appendix VII.8. Most emphasis will be put on the MSFE since the estimation of the errors is less reliable since we are using nonstationary time series.

6.2.1 The effect of number of factors

Overall, when looking at Table VII.18, the lowest MSFE are found for $r = 5$, which is also the optimal number of factors in the PCA. In addition, the column with $r = 5$ contains a lot of good coverage ratios (see Table VII.20). A second point is that for all r method O2 outperforms the first. Thirdly, as one would expect, increasing the number of factors decreases the MSFE for both approaches. However, where this effect is more or less indisputable for O2, this effect is less unambiguous for the first method. There this effect seems to stop after $r = 6$ (which is close to the value of five found in PCA). This effect becomes smaller for larger r. After $r = 12$ this effect is no longer substantial. The same actually holds regarding the average standard error. For method O2 the uncertainty decreases when more factors are used, while for O1 this effect stops after nine factors.

6.2.2 The effect of order VAR regression

Based on MSFE, an increase of p improves the nowcast performance of both models. A side effect is that the standard errors increase a lot. For higher r in combination with more lags of the VAR model, the nowcasts for the first years deviate a lot from the realisations. The number of parameters to estimate is too large compared to the number of observations. It even results in negative GDP. For $p = 2$, some gains can be observed in terms of forecast performance when r is small. If the evaluation period is taken from 2013-2017 it became clear that more lags did improve the models, (see for example Figure VII.2 in Appendix). Especially the second approach worked extremely well when setting $p = 3$ or $p = 4$. A consequence is however that the uncertainty increases dramatically. Or one should actually say that the uncertainty can barely be estimated because of the nonstationarity of the time series and the large number of parameters. However, to make comparison of the various models possible a larger evaluation period is desirable. Therefore we set $p = 1$. The gains made in term of MSFE are also not large compared to the extra number of parameters needed to be estimated. The matrix A is for example an $(r \cdot p \times r \cdot p)$ matrix and the Kalman filter has to update the vector F_t which has length $r \cdot p$.

6.2.3 The effect of autoregressive terms GDP

The biggest gains are attained when adding lags of the target variable itself. Again these effects are the greatest (at least for the evaluation period 2011-2017) for small r. Just like with the monthly models, for the first method more than one lag does not improve the model. For O2 up to $ly = 2$ improves the nowcasting performance. Interesting to see is that for O2 increasing the number of factors above $r = 7$ no longer decreases the MSFE when lags of GDP are included in the regression. The historical GDP values already contain a lot of information, such that more factors do not contribute to the performance.

In summary, for the first approach O1 the best results in terms of MSFE are attained when using $r = 5, p = 1$ and $ly = 1, 3$. For the second model O2 the combination $r = 5, p = 1$ and $ly = 2$ provides the best estimates. Both beat the simple AR model which has a MSFE of 3.489. One should keep in mind that this model is not a proper benchmark in the nonstationary case. We also consider taking $r = 7$ to explore the effect of the number of factors.

6.3 Model Confidence Set

It is of interest to assess the relative performance of the various models chosen in the previous section. Due to taking differences and the 'starting' problems of the Kalman filter the very first quarter of 2008 is also excluded from the evaluation period. The evaluation period for the MCS is set from the second quarter of 2008 up and until 2017, resulting in only 39 quarters that can be tested. As before, the MSFE is calculated for the 28 quarters from 2011-2017. While it is acknowledged that this is a limited amount of observations, it is still possible to reduce the number of acceptable models. To this end the MCS procedure is applied, introduced by Hansen et al. (2011). The MCS will contain the best model for a chosen confidence level $1 - \alpha$. The idea is to start with a wide range of models M_0 and through an iterative process the most poorly performing models are excluded. The result is a subset $\widehat{M}_{1-\alpha} \subseteq M_0$, that contains the best model with a given level of confidence, comparable to the notion of regular confidence intervals.

In the first step the null hypothesis is that $\widehat{M} = M_0$, meaning that all models are performing equally well. If this null is rejected, the poorest model is eliminated and the procedure will be repeated until the null is no longer rejected. The surviving models will together form the final MCS. The performance of the models is assessed by means of a loss function. We will consider both the squared $L^{(1)}$ and the absolute $L^{(2)}$ loss function. The losses of the i-th model for the t-th quarter are then denoted by

$$L_{i,t}^{(1)} = \left(\widehat{Y_{i,t}} - Y_t^{real}\right)^2, \tag{24}$$

$$L_{i,t}^{(2)} = \left|\widehat{Y}_{i,t} - Y_t^{real}\right|. \tag{25}$$

Hansen et al. (2011) introduced two test-statistics. The first one is based on the relative sample loss between the models, $d_{ij,t} = L_{i,t} - L_{j,t}$ for $i, j \in M_0$. Following this notation, $\overline{d}_{ij}$ is the time-average of the relative sample loss between model i and model j. The second statistic is a measure with respect to the average loss over all the models and not pairwise, that is $d_i = \sum_j \overline{d}_{ij}$. The t-statistics and corresponding test statistics used will read

$$t_{i,j} = \frac{\overline{d}_{i,j}}{\sqrt{\widehat{\text{Var}}(\overline{d}_{i,j})}}, \quad T_1 = \max_{i,j \in M_0} |t_{i,j}|, \tag{26}$$

$$t_i = \frac{\overline{d}_i}{\sqrt{\widehat{\text{Var}}(\overline{d}_i)}}, \quad T_2 = \max_{i \in M_0} t_i. \tag{27}$$

The variance is determined by bootstrapping. Accordingly, each model is assigned a p-value $\widehat{p}_i$, which is constructed such that model $i \in \widehat{M}_{1-\alpha}$ if and only if $\widehat{p}_i \geq \alpha$ for any $i \in M_0$. As a consequence, the model with the lowest p-value will be eliminated first. The specific p-values can be found in Table 1 of Hansen et al. (2011). One has to choose a confidence level $1 - \alpha$. The lower α, the more models will end up in the MCS. The reason for this is that for low α, relatively few models will have p-values smaller than α. The p-value p_m, with $m = \{1, 2\}$, of the test is determined as the proportion of the bootstrapped test statistics, T_m^b with $b = 1, \ldots, 5000$, that is larger than the sample T-statistic. If $p_m \leq \alpha$ the null is rejected and the worst performing model is eliminated, which is the model with the highest t-statistic. For more details, see Hansen et al. (2011) or Bernardi and Catania (2018).

The MCS procedure can be applied to nonstationary data as long as the relative losses $\{d_{ij,t}\}$ are stationary. This implies that the losses L_i can be nonstationary, but the models should be affected in the same way, such that the relative losses are stationary. All time series were transformed stationary before entering the model. However, the OA nowcasts are obtained using nonstationary data. We tested the losses and observed some significant autocorrelation at lags 1, 2, 3 or 4. When inspecting the loss matrix $\{d_{ij,t}\}$, the same was observed. A unit root test was applied on the columns of the loss matrix using four lags. For all the columns, the null of having a unit root was rejected. Nevertheless, as explained above the MCS procedure builds on the bootstrap distribution of the test statistics. Since we are dealing with time series and we observed some autocorrelation within one year we set the minimal bootstrap block length k_{min} to four (quarters). The MCS procedure will determine the block length by fitting an AR model on $d_{ij,t}$. It suggested using $k = 4$ or $k = 5$. Choosing higher block lengths could lead to insignificant results in our short time frame.

Since the MCS procedure can easily handle many models, we will consider the following models: for M1 and M2 both $r = 4$ (based on MSFE) and $r = 8$ (based on PCA analysis), for Q1 and Q2, $r = 4$ (based on MSFE) and $r = 10$ (PCA analysis) and for the OA $r = 5$ (PCA) and $r = 7$ and $ly = 0, 1, 2$ and the flash estimate. In total M_0 consists of 15 models. The notation is as follows: the numbers in the subscript refer to r, p and ly respectively. The corresponding output figures of the MCS procedure can be found in Appendix VIII.

When applying the MCS procedure something remarkable occurs. Using loss function $L^{(1)}$, the average loss of method $O2_{712}$ is 29.69, while the losses of the other models are between 0.474 and 3.354. This high loss is caused by a huge deviation in the first observation (second quarter 2008). Despite its evidently higher loss, using test statistic T_1 and all models, the MCS procedure does not eliminate model $O2_{712}$ for various α (see output Figure VIII.1 in Appendix). The reason is found in the fact that the t-statistics (26) and (27) are constructed by dividing by the variance of $\overline{d}_{ij}$ and $\overline{d}_i$ respectively. The relative losses including model $O2_{712}$ are around 100 times higher than the remaining variances. Therefore, the resulting t-statistics are comparable to that of the other models. As a result, in the ranking from small to large, this method does not end up as

performing the worst and therefore it is not eliminated. We repeat to process but delete the first observation to get a more reliable MCS. The total number of quarters is now $T = 38$.
Another interesting observation is that test statistic T_2 is very sensitive to these kinds of outliers. When using all the 14 models and $L^{(1)}$, we find $T_2 = 0.9935$ as opposed to $T_2 = 1.954$ when model $O2_{712}$ is excluded and 1.7117 when only the first observation is excluded. The corresponding p-values are 1, 0.1916 and 0.3276. In the first situation, for no value of α the procedure will eliminate a model. Since the matrix of nowcasts contains quite some outliers we decide to use only the first test-statistic T_1. For the same reason the second loss function $L^{(2)}$ is more appropriate here since $L^{(1)}$ is very much influenced by large losses. As an illustration, using the entire set the p-value is 0.157 when using T_1 and the squared loss function, in contrast to 0.0302 when using the absolute loss function. The latter rises to 0.036 when dropping the first observation.

The MCS procedure is first applied to the set of stationary methods (see output Figure VIII.2 in the Appendix). For $\alpha = 0.1$, the MCS procedure only eliminates $M1_{810}$. For $\alpha = 0.2$ both $M1_{810}$ and $M2_{810}$ are trimmed. This affirms our analysis that using less factors improves the nowcasting performance. Apparently, the models with the number of factors corresponding to the PCA analysis do not perform better. It takes up to $\alpha = 0.4$ to eliminate a third model, which is $Q1_{1010}$. This is again a model with more factors.

Now turn to the OA methods (see output Figure VIII.3 in the Appendix). For $\alpha = 0.1$, none of the models is eliminated, hence $\widehat{M}_{0.9} = M_0$. For $\alpha = 0.2$ only $O1_{513}$ is eliminated and until $\alpha = 0.7$ no other models are excluded. The losses are very close to one another here. Since the squared loss function puts less weight on small deviations it is expected that it can differentiate a little bit more between the models. However, using $L^{(1)}$ also leads to elimination only of $O1_{513}$ at $\alpha = 0.2$. Only for large α, for example $\alpha = 0.7$, the MCS contains only one model, which is $O2_{511}$.

Since we are interested in assessing the performance of the various models, the MCS is now applied to the complete set of models. However, for the moment the flash estimate is still excluded. The reason for this is that we want to compare the models constructed in this paper. Including the flash could lead to different MCS since the t-statistic measures the relative performance of the models.

Table 6.2 MCS based on complete set of models, without the flash estimate (2008-2017)

α	**Model Confidence Set**	**MSFE$_{av}$**
0.1	$\{Q1_{410}, Q1_{1010}, Q2_{410}, Q2_{1010}, M1_{411}, M2_{411}, M2_{412}, M2_{810}, O1_{511}, O2_{511}, O2_{512}, O2_{712}\}$	0.5035
0.25	$\{Q1_{410}, Q1_{1010}, Q2_{410}, Q2_{1010}, M1_{411}, M2_{411}, M2_{412}, O1_{511}, O2_{511}, O2_{512}\}$	0.4376
0.4	$\{Q1_{410}, Q1_{1010}, Q2_{410}, Q2_{1010}, M1_{411}, M2_{411}, M2_{412}, O1_{511}, O2_{511}\}$	0.4288
0.5	$\{Q1_{410}, Q2_{410}, Q2_{1010}, M1_{411}, M2_{411}, M2_{412}, O1_{511}, O2_{511}\}$	0.4381

The first models that are eliminated are model M1 using eight factors and the OA model using $ly = 3$. After that, albeit at a high α, the OA method with seven factors and the quarterly methods with ten factors are trimmed. We do not want to derive strong conclusions for low confidence levels, but it again appears that the models with more factors perform worse, contradicting the PCA analysis. The last column of Table 6.2 contains the MSFE of the nowcast when averaging the nowcasts of the models included in a particular MCS (using equal weights). As expected, this MSFE decreases when the MCS contains fewer elements. The MSFE when taking the equal average over all methods, excluding the flash, is 0.5165 (over the period 2008-2017).

According to Hansen et al. (2011), informative data will result in a MCS that contains less models compared to a MCS based on less informative data. We therefore also applied the MCS procedure over the period 2011-2017 (see output Figure VIII.4 in the Appendix), leading to 28 observations. This period does not contain the big drop in the first quarter of 2009 (see Figure 4.2). Hansen et al. (2011) found, when forecasting inflation, that using a more volatile period yields a smaller MCS. The MCS is then better in distinguishing the performance of the models. Surprisingly, we found that the MCS over the period 2011-2017 contains less elements than over the period 2008-2017, despite the fact that this period is less volatile (see Table 6.3). A possible reason could be that none of the models was able to approach the growth rate of the first quarter of 2009 even closely. The corresponding error potentially dominates the overall MSFE. As observed before, the MSFE when taking the average nowcasts over all the models in a particular MCS decreases when the MCS becomes smaller. The MSFE when taking the average of all nowcasts is 0.2542 for the period 2011-2017. For the latter three MCS, the MSFE is lower than the lowest MSFE of the individual models, being 0.2179 for $M2_{411}$.

Table 6.3 MCS based on complete set of models, without the flash estimate (2011-2017)

α	Model Confidence Set	$MSFE_{av}$
0.1	{$Q1_{410}$,$Q1_{1010}$, $Q2_{410}$, $Q2_{1010}$, $M1_{411}$, $M2_{411}$, $M2_{412}$, $M2_{810}$, $O2_{512}$ }	0.2350
0.2	{$Q1_{410}$,$Q1_{1010}$, $Q2_{410}$, , $M1_{411}$, $M2_{411}$, $M2_{412}$ }	0.2161
0.3	{$Q1_{1010}$, $Q2_{410}$, $M1_{411}$, $M2_{411}$, $M2_{412}$ }	0.2084
0.4	{$Q2_{410}$, $M2_{411}$, $M2_{412}$ }	0.1956

Since the shorter time frame seems to be better at distinguishing the methods, the time span 2011-2017 also is applied to the complete set of models, including the flash estimate (see output Figure VIII.5).

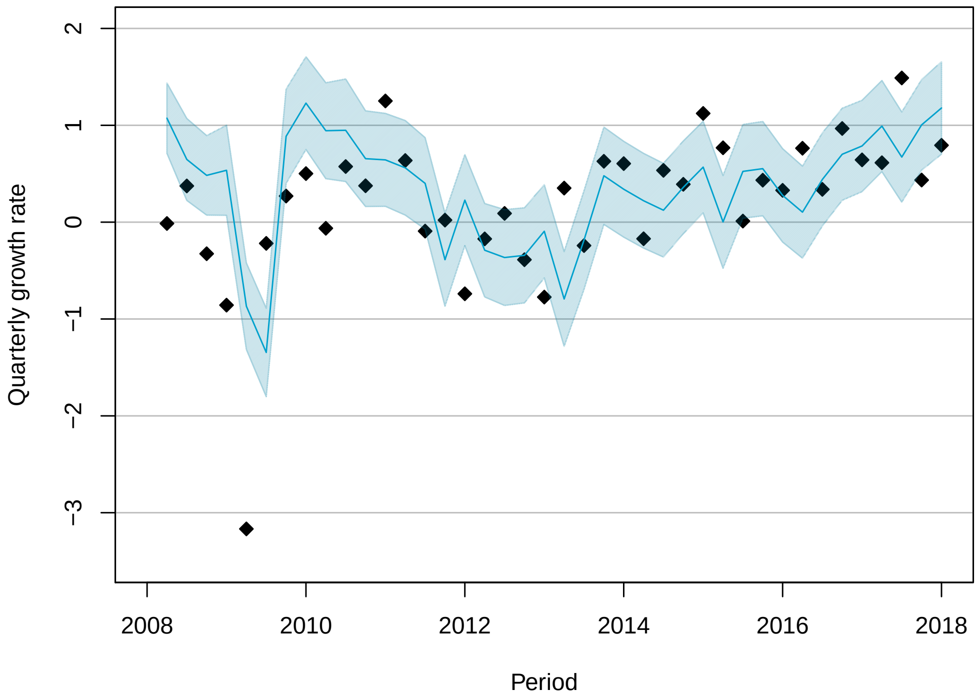
Table 6.4 MCS based on complete set of models (2011-2017)

α	Model Confidence Set	$MSFE_{av}$
0.05	{$Q1_{410}$, $Q1_{1010}$, $Q2_{410}$, $Q2_{1010}$, $M2_{411}$, $M2_{412}$, flash }	0.1872
0.1	{$Q1_{1010}$, $Q2_{410}$, $M2_{411}$, $M2_{412}$, flash}	0.1583
0.12	{$Q2_{410}$, $M2_{411}$, $M2_{412}$, flash}	0.1341
0.15	{flash}	0.1436

Under slightly different α, the same MCS were obtained when using the squared loss function. At confidence level 0.88 we can say that $MCS_{0.88}$ ={$Q2_{410}$, $M2_{411}$, $M2_{412}$, flash} contains the best model. On top of that, from Table 6.4 we can deduce that with 85% the MCS, which only contains the flash estimate, will contain the best model. Overall, it seems that the best performing models are $Q2_{410}$, $M2_{411}$ and $M2_{412}$, although none of them is capable of improving on the current flash estimate. From those three, $M2_{411}$ has the lowest MSFE, though the difference is small. However, $Q2_{410}$ has the smallest uncertainty.
The historical bias over the period 2011-2018 of the models are smaller than 0.0668 (for model $Q2_{410}$), which is substantially better than the historical downward bias of 0.1802 of the current flash estimate for this period. A two-sided t-test applied to the four methods did not lead to rejecting the null of zero mean for the nowcasting models, while it did for the flash.

In addition, averaging over the models in $\text{MCS}_{0.88}$ yields a lower MSFE than that of the flash. The MCS procedure does confirm our initial conclusion that for the quarterly, monthly and OA, the second model outperforms the first. It also substantiates the hypothesis that more factors does not improve the nowcast accuracy. As an illustration, Figure 6.1 shows the nowcast of the quarterly growth rate using model Q2_{410}.

Figure 6.1 Nowcast Q2 ($r = 4, p = 1, ly = 0$). The black dots are the realised historical GDP growth rates. The blue line reflects the nowcast. The shaded area represents the corresponding uncertainty, taking one standard error.



6.4 Model Averaging

The MCS procedure did not lead to the selection of one model that irrefutably outperforms the others. Instead of selecting one model we could exploit the results of the MCS procedure to improve the nowcast performance by combining the nowcasts of some of the better performing models. Combining forecasts in general leads to smaller MSFEs (Clemen, 1989). Since the EA models indisputably outperform the OA models, the MCS procedure was applied on the set of EA models using the period 2011-2017. Both the squared and the absolute loss function were used. At significance level $\alpha = 0.1$ and using $L^{(1)}$ model M1_{810} was eliminated. Subsequently, for $\alpha = 0.17$ also M2_{810} was trimmed. Increasing α led to the consecutive elimination of models Q1_{1010}, Q1_{410}, Q2_{1010}, M1_{411} and Q2_{410}. This leads to a sequence of decreasing model sets with best performing models. A particular subset of models performing well will be referred to by S_i in which i is the number of models contained in the subset, for example S_5 ={Q2_{410}, Q2_{1010}, M1_{411}, M2_{411}, M2_{412}}. More or less the same subsets were obtained when using loss function $L^{(2)}$, although at different significance levels. To each of these subsets six different model averaging techniques were applied in order to investigate whether combining nowcast methods leads to better nowcasting performances. The nowcast of a particular averaging method is denoted by

$$X_t = \sum_{j=1}^{i} \gamma_j Z_{j,t}, \tag{28}$$

with $Z_{j,t}$ being the nowcast for the t-th quarter of model $j \in S_i$ and γ_j the weight given to that particular model.

The first averaging method simply takes the average with equal weights over the models in S_i. A closely related approach is the trimmed mean. The two models with the highest and lowest nowcast for a certain quarter are given zero weight, and the remaining all equal weights $\gamma_j = \frac{1}{i-2}$. Another simple and closely related averaging method is taking the median. The fourth averaging method (B-G) is proposed by Bates and Granger (1969) and allows for unequal weights. The weight γ_j given to a certain model depends on its historical nowcast variance, see (29). It minimises the variance of the error of the average nowcast. By the fifth method, proposed by Granger and Ramanathan (1984), the weights are estimated by OLS. This method takes into account the covariance between the models. The weights of the first four methods sum up to unity, while the latter does not impose that restriction. It also allows for negative weights. The last averaging method (SW) we will consider is introduced by Stock and Watson (2004) and depends inversely on the historical MSFE of the models in the model set (see also Diebold and Pauly (1987)). The weights used by the latter three averaging models are given by

$$\gamma_{\text{B-G},j} = \frac{1/\hat{\sigma}_j^2}{\sum_{k=1}^{i} 1/\hat{\sigma}_k^2}, \quad \text{where} \quad \hat{\sigma}_j^2 = \frac{1}{t-1}\sum_{l=1}^{t} (z_{j,l} - \overline{z_j})^2 \tag{29}$$

$$\gamma_{\text{OLS}} = (Z'Z)^{-1}Z'W, \tag{30}$$

$$\gamma_{\text{SW},j} = \frac{m_{j,t}^{-1}}{\sum_{k=1}^{i} m_{k,t}^{-1}} \quad \text{where} \quad m_{j,t} = \sum_{l=1}^{t} \delta^{t-l}(w_l - z_{j,l})^2, \tag{31}$$

with $t = 1, \ldots, T$. Here T denotes the total number of quarters, which equals 39 since the first quarter of 2008 is lost because of taking differences. We obtained the best results when using the values 0.95 and 1 for discount factor δ. The $(t-1) \times i$ matrix Z and the $(t-1)$-dimensional vector W in (30) are the matrix of nowcasts and the vector of historical GDP growth rates respectively, both only up to time $t-1$.
The usual approach is to split the data set into two subsets: a set which is used for estimating the weights and an evaluation period (see for example Diks and Vrugt (2010) or Stock and Watson (2004)). However, the total period entails only 10 years. The time frame is too short to split the sample into a calibration and an evaluation period. In order to assess the performance of the averaging methods more thoroughly, the time-varying weights were re-estimated for every quarter t using only the information that would have been available at that point in time. That is, the nowcasts of the several models from the second quarter of 2008 up to t and the realisations of the GDP growth rate up to the preceding quarter $(t-1)$ are known, comparable to the nowcast design described in Section 3.4. The MSFE of the six averaging methods for the evaluation period 2011-2017 are provided in Table 6.5.

Overall, when model averaging is applied to a (sub)set of stationary models and for all averaging methods except OLS, the average nowcast outperforms the AR benchmark ($\text{MSFE}_{AR} = 0.2464$). In addition, when six or less models are used the MSFE is essentially always lower than the lowest MSFE of all the individual models, which is 0.2179 for M2_{411}. When the model set contains only two models, most methods are the same. The optimal set of models seems to be S_4 for the first four methods and for the latter three set S_3. This observation indicates that additional gains can be attained by first applying the MCS procedure to select a set of better performing models before using combination techniques. The gains in terms of MSFE are substantial when one compares the MSFEs of the average nowcast when using all the EA models or only particular subsets determined by the MCS procedure.

Table 6.5 MSFE of averaging methods. MSFE of averaging methods using time-varying weights and simulating real-time information flow over the period 2008-2017. The MSFE is calculated over the period 2011-2017. An underlined entry signifies the averaging method that provides the lowest MSFE for a given set of models. A star (*) indicates the model set with the lowest MSFE for a certain averaging method.

	Mean	Trimmed mean	Median	B-G	OLS	SW $\delta = 1$	SW $\delta = 0.95$
Set of models							
All models	0.2542	0.2534	0.2417	<u>0.2140</u>	1.2609	0.2433	0.2443
All OA	0.4065	0.3915	0.4054	0.3884	<u>0.3337</u>	0.3901	0.3904
All EA	0.2219	0.2244	0.2536	<u>0.2048</u>	0.3720	0.2227	0.2239
S_8	0.2307	0.2312	0.2509	<u>0.2119</u>	0.3207	0.2291	0.2304
S_7	0.2258	0.2272	0.2384	<u>0.2113</u>	0.3133	0.2289	0.2301
S_6	0.2093	0.2148	0.2318	<u>0.2058</u>	0.3060	0.2152	0.2162
S_5	0.2019	0.2044	0.2324	<u>0.1928</u>	0.2910	0.2053	0.2065
S_4	0.1911*	0.2020*	0.2020*	<u>0.1906</u>*	0.2436	0.1960	0.1967
S_3	<u>0.1934</u>	0.2153	0.2153	0.1944	0.2261*	<u>0.1934</u>*	0.1941*
S_2	0.2185	0.3844	0.2185	0.2187	0.3680	0.2185	0.2185

Figure 6.2 MSFE of the average nowcast for increasing number of EA models in the set of models

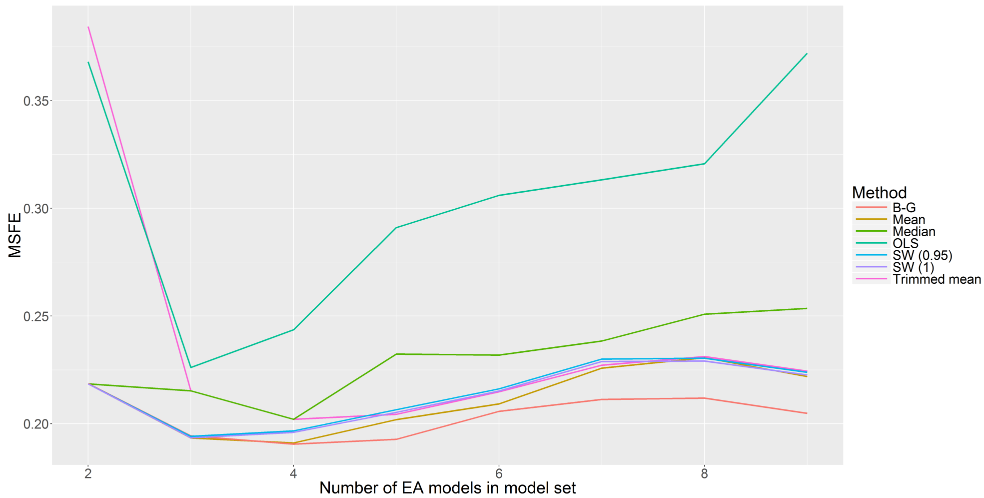


Figure 6.2 displays the behaviour of the different models for increasing model set size. The lowest MSFEs are obtained for method B-G, followed by the simple mean. The most simple method, the mean, outperforms the more complicated averaging methods. The SW methods outperform the median, but they do not perform better than the mean or trimmed mean. Only for small and larger model sets they seem to perform slightly better than the trimmed mean. In case of uncorrelated nowcasts, the SW method with $\delta = 1$ is equivalent to the B-G method. As expected, we can conclude that the nowcasting models are correlated. The OLS averaging method is the worst performing method, over all model sets. The OLS averaging method is also the only one that results in an average nowcast with a significant bias (at significance level 0.05). This result is found for both model set S_3 and S_4. Model set S_4 contained one model with a bias, namely model $M1_{411}$ with bias -0.1890.

The OLS combination method is not optimal if the nowcast sample size is relatively small compared to the number of models, due to high variability (see Wei (2009)). This is in particular a problem when the individual nowcast models are highly dependent (Wei and Yang, 2012), which is the case here. It causes difficulties in estimating the optimal weights. Although the set of models used for averaging is not as large as used in the before mentioned papers, the same conclusion can be drawn from Table 6.5. For their application, Diks and Vrugt (2010) found evidence that the OLS averaging method performs better than the simpler methods. However, their (calibration) data set contained many observations. They also found that the performance of the OLS estimator deteriorates if the calibration sample size decreases, corresponding to our results. Stock and Watson (2004) also found that simple averaging methods perform better than more complicated methods when forecasting output growth. The same conclusion is reached by Rapach and Strauss (2005, 2008) in two separate studies to combine forecasting models for employment growth. They concluded that the OLS averaging method performs worse than the simple methods and the SW method. It could be that the bad performance of the OLS method is related to the type of data set used. Apparently, for this data set the sample size is too small compared to the number of models.

The main aim is to find an appropriate substitute for the current flash estimate of the GDP growth rate. From that point of view, one would not prefer adding the flash estimate itself to the set of models to be averaged. Nonetheless, the flash estimate uses information that is available at the time the GDP growth rate is nowcasted. Therefore, it could also be argued that the flash estimate is part of the information set at time t and hence it can be used in the averaging method. When applying the MCS procedure we already observed that the MSFE of the average nowcast of the models in the MCS $M_{0.88}$ was lower than that of the flash estimate (see Table 6.4). Table 6.6 provides the MSFEs of the average nowcast when the flash is added to the best performing subsets S_3 and S_4.

Table 6.6 MSFE averaging methods including flash estimate. MSFE of averaging methods using time-varying weights and simulating real-time information flow. The MSFE is calculated over the period 2011-2017. Underlined are the entries that have a lower MSFE than the flash estimate. The lowest MSFE is marked with (*).

Method	Mean	Trimmed mean	Median	B-G	OLS	SW $\delta = 1$	SW $\delta = 0.95$
Set of models							
S_3 with flash	<u>0.1341</u>	0.1704	0.1704	0.1513	0.1584	<u>0.1109*</u>	<u>0.1137</u>
S_4 with flash	0.1447	0.1685	0.1681	0.1565	0.1602	<u>0.1149</u>	<u>0.1195</u>

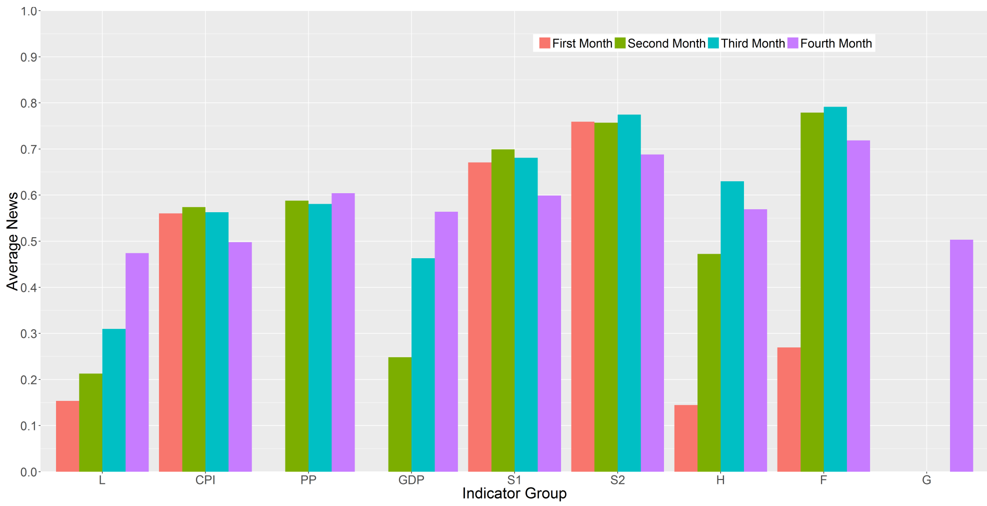
As expected, the MSFEs decrease substantially when the flash estimate is added to the set of models. The underlined entries reflect combining methods that outperform the flash estimate, based on MSFE. The best performing model uses the model set {flash, $Q2_{410}$, $M2_{411}$, $M2_{412}$} and the SW averaging method with $\delta = 1$. However, it inherits the bias from the flash estimate. The nowcast average is biased (-0.1024) while this bias was not observed (i.e. much smaller) when using model set S_3 only. The null of zero mean was not rejected, but the p-value was very close to 0.05. A bias of ten basis points is undesirable. The weights of the SW methods (31) sum up to unity, thereby relying on the assumption of unbiased individual nowcasts. The bias of the OLS method disappeared when a constant was added. The constant term adjusts for the bias, however the MSFE did increase. An appropriate and easy alternative is using the simple mean. This combining method is better than the current flash estimate and has no bias (p-value 0.594).

We can conclude that model averaging lowers the MSFE by 6% if only model averaging is applied (thus to all EA models), by 12.5% if first a few models are selected by using the MCS procedure and by 38.46% if we allow the flash to be part of the model set and use the simple mean.

6.5 News Impact and Uncertainty

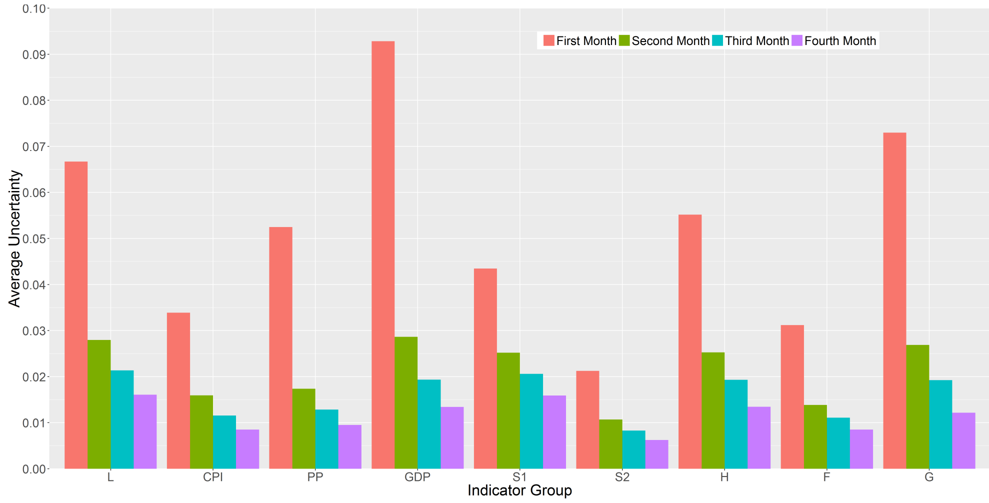
Another route towards improvements is to consider what variables affect the nowcast the most and are therefore worth tracking more closely, or spending more effort in improving. This information could also be used to improve the set of indicators being used for the nowcast. Following Bok et al. (2017), the indicators were classified to nine groups: GDP&Income (GDP), Government (G), Producer Prices (PP), Consumer Price Indices (CPI), Producer sentiments (S1), Consumer sentiments (S2), Housing (H), Labour (L) and finally a group mixed with among other things financial indicators (F). In order to measure the relevance of certain groups of time series the average news impact of the groups is taken for each of the three months of a quarter and for the one month backcast. The average news impact of the groups is provided in Figure 6.3. Since the Government group only contains time series with a three-month publication lag, this group only has an impact on the nowcasts in the fourth month. During the first month the sentiments and the consumer prices have the largest impact. These time series do remain important in the other months as well. In the second and third month the mixed group has a lot of impact on the adjustment of the nowcast. The impact of the Labour group increases throughout the quarter. Surprisingly, the nowcast is more affected by sentiments than by numbers on labour or GDP. One could conclude that soft information does indeed matter.

Figure 6.3 Average newsimpact per indicator group



The same is done for the average uncertainty of the indicators. Figure 6.4 shows the uncertainty of the common component, $\widehat{\Lambda}\text{Var}(\widehat{F}_t)\widehat{\Lambda}$. The average was taken over all the nowcasts by taking only the diagonal elements. When looking at the matrices the off-diagonal elements were indeed close to zero, justifying this choice. For all groups, the first months entails the most uncertainty. The uncertainty is the largest for the Labour, GDP and Government groups. These groups contain quite some time series with a three-month publication lag. On the other hand, sentiments are published without lag and therefore the corresponding uncertainty is lower. For all time series the uncertainty decreases throughout the quarter, when more information becomes available.

Figure 6.4 Average uncertainty (common component) per indicator group



6.6 Testing some model assumptions

In this final empirical subsection some of the assumptions that were made in the theoretical part of this paper are briefly checked. For all assumptions model $Q2_{410}$ is used and always the last nowcast, the nowcast for the fourth quarter of 2017.
First of all, the DFM (1) - (3) assumes that the covariance matrix Σ_t of the idiosyncratic errors is a diagonal matrix. Applying the test statistic constructed by Lan et al. (2015) on the covariance matrix Σ, resulted in unequivocally rejecting the null hypothesis of having a diagonal matrix. The residuals of the nowcast of the last quarter of 2017 were used. The statistic was 95.57 when $r = 5$. It decreases when r becomes larger: 88.21, 65.73 and 59.63 for $r = 8, 10$ and 15 respectively. These statistics are all far larger than 1.64. So although the diagonal elements are large compared to the off-diagonal elements and we manually set Σ to a diagonal matrix, we do reject the null of having a diagonal matrix. A possible reason could be that Y contains some disaggregated time series instead of aggregated variables only. For example, the predictor set contains production time series for various sectors instead of only the total production of the Dutch economy.

The second assumption to be verified is the stationarity of the factors. This is an important assumption. Again, the factors constructed by the Kalman filter for the last observation were used. This results in five time series of length 153 (one for each month). We conducted unit root tests with several lags. All factor series convincingly rejected the null of having a unit root for up to twelve lags (3 years). This justifies the use of a VAR model. When checking the eigenvalues of the system matrix A, all its eigenvalues were indeed smaller than one in modulus (for $r = 5$ they were: 0.902, $0.836 \pm 0.143i$, $0.287 \pm 0.494i$). The residuals $\hat{u}_t$ of the VAR regression all showed a significant lag at 12. The Ljung-Box test applied to the residual series failed to reject the null of having no serial correlation up to l lags with $l \leq 2$. However, for higher lags the null was rejected. It suggests that it could be worth adding F_{t-12} to Equation (2). However, this complicates the Kalman filter and is therefore left for later research.

Finally, the adequateness of the model is tested by applying the Ljung-Box test to the residuals ϵ_t in regression (3). Only for one lag ($l = 1$) the null of having no serial correlation is rejected. For all other lags the null is not rejected, indicating that the model is adequate.

Table 6.7 Autocorrelation function residuals ϵ_t. The autocorrelation function of the residuals of the regression of GDP growth rate on the factors. The p-values refer to the Ljung-Box test. For $p > 0.05$ we do not reject the null of no autocorrelation up to order l.

Lags(l)	1	2	3	4	8
Acf	-0.297	-0.001	0.040	-0.059	-0.221
P-value	0.031	0.097	0.190	0.291	0.264

7 Year on Year growth rates

In this section it is briefly checked whether the model would be appropriate to estimate year on year growth rates of the GDP. All the figures can be found in Appendix IX. Therefore, the time series are made stationary by taking yearly differences or transform them into yearly growth rates. The historical year on year growth rates with the corresponding flash estimates are provided in Figure IX.1 in the Appendix. The flash estimate of the year on year growth also suffers from a bias, as can be clearly seen in the graph. Over the period 2001-2017 the historical bias was -0.4456.

The PCA analysis indicated using around 7 factors. The corresponding correlations with GDP are given in Table 7.1. Again the first factors seem to be able to explain the behaviour of GDP. Conducting unit root tests revealed that all seven factors are stationary. The seven factors are plotted in Figure IX.2 in the Appendix. The first factor loads on almost exactly the same indicators as the first factor of the quarterly differenced data set, see Figure 5.1. The second factor is highly correlated with slowly responding indicators. It is comparable to the third factor in the quarterly models. This can also be seen when comparing the Figures IX.2 and IX.2. The third factor loads on indicators regarding raw materials, industry and trade, comparable to the second factor of the quarterly model. The fourth is related to CPI. The fifth factor loads heavily on consumer confidence indicators. The red line in Figure IX.2 indeed exhibits a steep decline right at the beginning of the crisis in 2008. The sixth factor is again the stable (and seasonal) factor, related to governmental, energy and catering industry series. The seventh factor does not clearly reflect a particular part of the economy. The correlations of the indicators with each of the factors is provided in Figure IX.3 in the Appendix.

Table 7.1 Correlation GDP and factors yearly model

Factor	1	2	3	4	5	6	7
Correlation	0.842	0.462	0.009	0.041	0.017	0.060	0.030

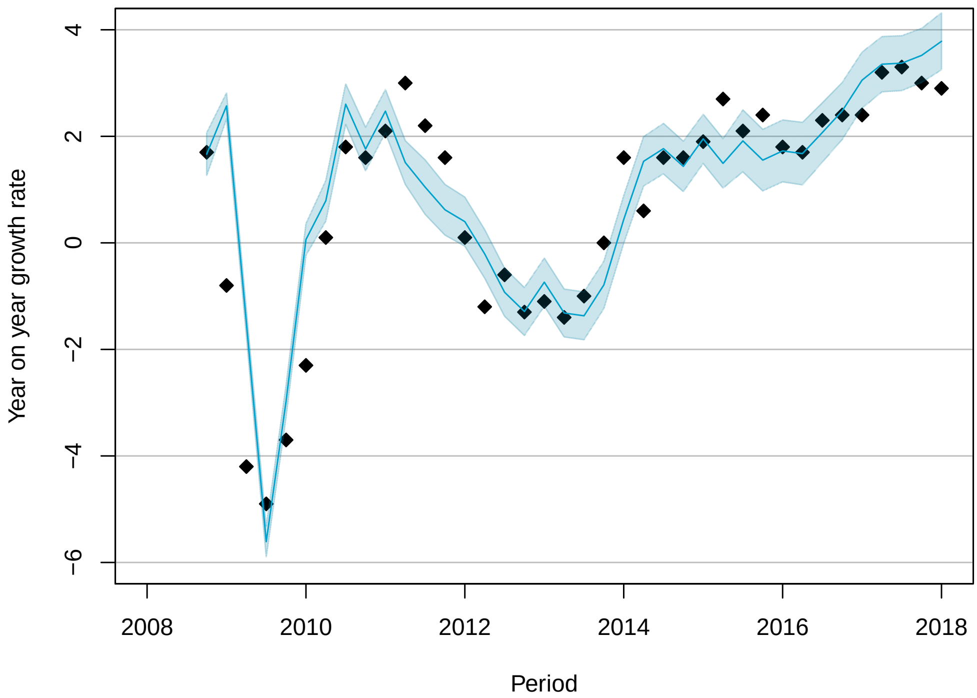
The same techniques to get quarterly factors from the monthly factor of the Kalman filter will be used as we did for the quarterly and OA models. That is, taking every third element (Y1) and the three-month average (Y2). The resulting MSFE are provided in the table 7.2. The simple AR model yields a MSFE of 0.6130.

Table 7.2 MSFE ($p = 1, ly = 0$). Over the period 2011-2017.

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
Y1	0.8158	0.7218	<u>0.7082</u>	0.7368	0.8131	0.8380	0.8302	0.7224	0.7574
Y2	0.5656	0.5042	0.4654	<u>0.4582</u>	0.5007	0.5084	0.5807	0.9305	0.7277

The lowest MSFEs are again obtained when using the three-month average of the factors instead of simply taking every third element. This model (Y2) is also able to beat the AR benchmark. Taking seven factors results in the lowest MSFE. When adding one autoregressive term of GDP the MSFE decreased to 0.4351 when using $r = 6$ and adding two increased the MSFEs again. Increasing ly more did not lead to a better nowcasting performance. The standard error did also decrease when $ly = 1$ and increase when $ly \geq 2$, see Figure IX.1. Finally, increasing the number of lags of the var regression p did not improve the nowcasting performance. The MSFEs increased for most number of factors. Therefore, the best model for nowcasting year on year growth rates is model $Y2$ with $p = 1$, $ly = 1$ and $r = 6$. The corresponding graphical representation is provided in Figure 7.1. This model is not able reduce the MSFE to a value below the flash which has a MSFE of 0.2328, which increases to 0.3545 when 2016 and 2017 are excluded.
However, the nowcast does not suffer from a bias (p-value 0.6525), while the flash has a bias of -0.2656 over the period 2010-2017.

Figure 7.1 Nowcast Y2 ($r = 6, p = 1, ly = 1$). The black dots are the realised historical GDP values. The blue line reflects the nowcast. The shaded area represents the corresponding uncertainty, taking one standard error.



8 Conclusion

This paper explores whether nowcasting is a method that could be a good substitute for the current flash estimate of the GDP growth rate of the preceding quarter. Overall, the nowcasting method as introducted by Giannone et al. (2008) seems to be a promising method for nowcasting the GDP growth rate.

In summary, the models in which the time series are preprocessed to be stationary outperform the models using the nonstationary time series. We can therefore conclude that not only theoretically but also practically transforming the set of predictors to stationary time series does indeed provide better estimates. Additionally, the nowcasting methods Q2 and M2 are able to beat a simple AR model. According to our results, exploiting the information from multiple monthly/quarterly factors improves the nowcasting performance. Those models also end up in the MCS. Based on the corresponding uncertainty, model Q2 with four quarterly factors and zero lags of GDP is preferred. It provides a good estimate of the current state of the economy, with an MSFE of 0.2415.

In terms of the mean squared forecast error it does not outperform the current flash estimate, although the difference is small. One side note that needs to be made here, is that it can happen that in the process of carrying out consistency checks leading up to determining the final GDP growth rate, that some minor adjustments are made which in the present context would favour the current flash estimate when comparing it with other methods. Therefore, a full fair comparison is not possible. In addition, the flash estimate equals the historical GDP growth rate for the last four quarters (see Figure 4.2) since those estimates are not yet revised. This results in a lower MSFE for the flash and hence an underestimate. If the MSFE is calculated over the period 2011-2016 it increases to 0.1765.

Combining nowcasts results in lower MSFEs and when the flash is included in the model set, the average nowcast beats the flash. What this means is that the expert knowledge that is used when determining the current flash estimates, incorporates some information which is not completely captured in all of the factors that are currently taken along in the PCA nowcast. On the other hand, the current flash estimate also does not contain all of the information either and in addition causes a bias which is absent from the method described here.

Advantages of the nowcasting method are that it is a more objective and much faster method than the flash estimate. In addition, the nowcasting method is reproducible and the uncertainty corresponding to the estimate can be estimated. While the choice of models and factors is of course relevant for the current nowcast method, it is applied in a more consistent way so that there is much less scope for humen interpretation and bias. A major advantage is that it this method is demonstrated not to suffer from a bias, which the current flash estimate does.

The literature on nowcasting is in general enthusiastic about its performance. Therefore, more research in the various aspects of the model is needed. From our experiment we can draw the following subconclusions.

First of all, we found that in general increasing the number of factors beyond 4 did not contribute to the nowcasting performance of the stationary models, thereby contradicting the results of the PCA analysis. The performance of the nonstationary models did improve when the number of factors increased. However, this effect only held up to a certain extent. This confirms the idea that a few factors are able to capture the bulk of the dynamics within a group of (economic) indicators.

Secondly, adding more lags to the VAR regression of the factors did not lead to significant improvement of the models, especially not for the EA models. This also substantiates our

findings that the extracted factors showed significant autocorrelation at lag 1. Increasing the number of lags added more noise instead of useful information.
Thirdly, when the time series were properly transformed and made to be stationary, adding autoregressive terms of GDP did not enhance the nowcasting performance of the quarterly models. By contrast, it did lead to significant improvements in the monthly models and the nonstationary set-up. Although the first factor seemed to be able to catch up the common trend, the OA models persistently underestimated the GDP value. Adding autoregressive terms of GDP solved this problem to a large extent.
Concerning uncertainty, this paper confirms results found in previous nowcast experiments. When more information becomes available, uncertainty decreases. We can conclude that monthly data has an impact on the nowcast since the nowcast is updated when news becomes available. It shows the added value of using a Kalman filter to deal with jagged edges in the data set due to mixed publication frequencies. Besides this, the relevance of using soft information as indicators is shown.
Finally, combining nowcasts of several models improves the nowcasting performance and does in general not suffer from a bias. In addition, we found that first applying the MCS procedure to select the best performing models and thereby reduce the set of models leads to more accurate nowcasts than applying model averaging techniques to the entire set of models. This despite the fact that each individual nowcast could potentially add different pieces of information and thereby improve the nowcast. However, in order to test this more thoroughly and to investigate whether it also applies to other data sets, more research is needed.

Nevertheless, the model used has some limitations that could be further investigated in order to improve the nowcasting performance. A first limitation of the experiment conducted, is the assumption of constant parameters. Parameter uncertainty is partly taken account for by re-estimating them for every nowcast. However, we assumed them to be constant during one loop of the Kalman filter. This could be adjusted in further research.
Secondly, the model does not take into account revisions, which can be substantial. Although literature has shown that the impact is minimal, proper backtesting can only be done when using the real historical information that was available at a certain point in time. This requires using the information from before revision, which is for most time series not available at the moment, or using a separate time series technique to parametrically model such a break.
Thirdly, we found some substantial autocorrelation in the factors at lag twelve. This could be taken into account in the future, by adding in seasonal corrections to indicators. Another possibility could be to take yearly differences and provide year on year growth rates of GDP, thus the growth of GDP in a quarter with respect to the same quarter in the preceding year.
A fourth possible improvement of the model might be to use higher frequency data, such as weekly or even daily data. Now daily data was aggregated to monthly data. However, the state space model is appropriate for dealing with different frequencies.
Finally, some remarks concercing the data set used. The data set consisted of 88 indicator time series, which is quite a lot. According to various papers (see for example Banbura et al. (2010) and Doz et al. (2011)) this should not be detrimental to the performance of the model. It should be able to handle this. However, it could be that the full data set is somewhat oversampled or contains much noisy information. In particular when embedding this method into routine production of national statistics, the amount of effort needed to collect and process these data should ideally be no more than absolutely necessary. Boivin and Ng (2006) show that adding many time series does not necessarily lead to better performance (see for an empirical example Reijer (2005)). Our data set for example contains the time series of both the average retail production and the average retail revenues. The before mentioned papers all consider pure

factor models. Theoretically, more (timely) time series could contribute to the performance of the Kalman filter part. More research on the effect of different compositions of the predictor set is needed. One can imagine that it could be of added value to extend the number of financial time series, which were slightly underrepresented in this data set.

At the moment a quarterly nowcast is sufficient. Progress has been made since we are now able to provide an unbiased flash estimate together with a measure of the corresponding uncertainty. It is simply impossible to give a fully accurate growth rate after 45 days. Nonetheless, a future goal is to provide an accurate model that provides a monthly nowcast of the state of the economy throughout the quarter, potentially even at a higher frequency, that incorporates information as soon as it becomes available.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71(1)*, 135–171.

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

Banbura, M., D. Giannone, and L. Reichlin (2010). Nowcasting (working paper no. 1275). Technical report, ECB , Frankfurt.

Banbura, 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*, 333–346.

Bates, J. and C. Granger (1969). The combination of forecasts. *Operational Research Society 20(4)*, 451–468.

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

Bernardi, M. and L. Catania (2018). The model confidence set package for r. *Journal of Computational Economics and Econometrics 8(2)*, 144–158.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132(1)*, 169–194.

Bok, B., D. Caratelli, D. Giannone, A. Sbordone, and A. Tambalotti (2017). Macroeconomic nowcasting and forecasting with big data (staff reports no. 830). Technical report, Federal Reserve Bank of New York.

Buja, A. and N. Eyuboglu (1992). Remarks on parallel analysis. *Multivariate Behavioral Research 27(4)*, 509–540.

Caiado, J. and N. Crato (2005). Discrimination between deterministic and stochastic trend processes. In *Conference Applied Stochastic Models and Data Analysis, Brest, France*.

Cattell, R. (1966). The scree test for the number of factors. *Multivariate behavioural research 1*, 245–276.

Chernis, T. and R. Sekkel (2017). A dynamic factor model for nowcasting canadian gdp growth (staff working paper no. 2017(2)). Technical report, Bank of Canada.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*, 559–583.

Dagum, E. and S. Giannerini (2006). A critical investigation on detrending procedures for non-linear processes. *Journal of Macroeconomics 28(1)*, 175–191.

Diebold, F. and P. Pauly (1987). Structural change and the combination of forecasts. *Journal of Forecasting 6*, 21–40.

Diebold, F. and A. Senhadji (1996). The uncertain unit root in real gnp: comment. *The American Economic Review 86(5)*, 1291–1298.

Diks, C. and J. Vrugt (2010). Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment 24(6)*, 809–820.

Doz, C., D. Giannone, and L. Reichlin (2006). A two-step estimator for large approximate dynamic factor models based on kalman filtering. Université Libre de Bruxelles.

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*, 188–205.

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

Elder, J. and P. Kennedy (2001). Testing for unit roots: what should students be taught? *The Journal of Economic Education 32*, 137–146.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: identifiation and estimation. *The Review of Economics and Statistics 82*, 540–554.

Giannone, D., L. Reichlin, and L. Sala (2004). Monetary policy in real time. *NBER Macroeconomics Annual*, 161–200.

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

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

Hansen, P., A. Lunde, and J. Nason (2011). The model confidence set. *Econometrica 79(2)*, 453–497.

Henderson, R. (1916). Note on graduation by adjusted average. *Transactions of the Actuarial Society of America 17*, 43–48.

Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika 30(2)*, 179–185.

Jansen, J., X. Jin, and J. De Winter (2012). Forecasting and nowcasting real gdp: Comparing statistical models and subjective forecasts (working papers no. 365). Technical report, De Nederlandsche Bank.

Kaiser, H. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20(1)*, 141–151.

Lan, W., L. Ronghua, C. Tsai, H. Wang, and Y. Yang (2015). Testing the diagonality of a large covariance matrix in a regression setting. *Journal of Business & Economic Statistics 33(1)*, 76–86.

Mariano, R. 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.

Perron, P. (1988). Trends and random walks in macroeconomic time series. *Journal of Economic Dynamics and Control 12*, 297–332.

Perrucci, E. and F. Pijpers (2017). Filtering in the fourier domain: a new set of filters for seasonal adjustment of time series and its evaluation. Technical report, Statistics Netherlands.

Raîche, G., T. Walls, D. Magis, M. Riopel, B. Martin, and J. Blais (2013). Non-graphical solutions for cattell's scree test. *European Journal of Research Methods for the Behavioral and Social Sciences 9(1)*, 23–29.

Rapach, D. and J. Strauss (2005). Forecasting employment growth in missouri with many potential relevant predictors: an analysis of forecast combining methods. Technical report, Federal Reserve Bank of St. Louis. Regional Economic Development 1(1), pp. 97-112.

Rapach, D. and J. Strauss (2008). Forecasting us employment growth using forecast combining methods. *Journal of Forecasting 27(1)*, 75–93.

Reijer, A. d. (2005). Forecasting dutch gdp using large scale factor models (working papers 028). Technical report, De Nederlandsche Bank.

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

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*, 386–398.

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

Stock, J. and M. Watson (2011). Dynamic factor models. In M. Clements and D. Hendry (Eds.), *Oxford Handbook of Economic Forecasting*. Oxford University Press.

Tsay, R. (2010). *Analysis of financial time series*. John Wiley & Sons: New Jersey.

Watson, M. (2004). Comment on giannone, reichlin, and sala. *NBER Macroeconomics Annual 19*, 216–221.

Wei, X. (2009). Regression-based forecast combination methods. *Romanian Journal of Economic Forecasting*, 4.

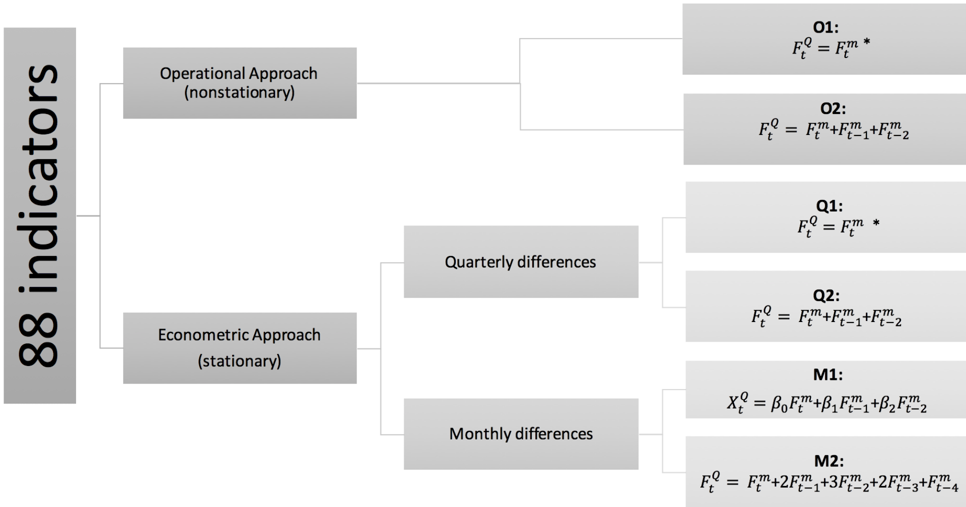
Wei, X. and Y. Yang (2012). Robust forecast combinations. *Journal of Econometrics 166(2)*, 224–236.

Zwick, W. and W. Velicer (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin 99(3)*, 432–442.

Appendix

I Overview models

Figure I.1 Overview of the models used and described in Section 2.4. The (*) means that every third element of the monthly factors is taken. For the first two months of a quarter there is no observation.



II Complete data set

Name	Frequency	Publ. lag	Group	Source
Number of self-employed	Q	3	L	SN
Labour participation	Q	3	L	SN
CPI general	M	0	CPI	SN
CPI goods	M	0	CPI	SN
CPI Industrial Goods	M	0	CPI	SN
CPI Durable Goods	M	0	CPI	SN
CPI Services	M	0	CPI	SN
CPI Electricity	M	0	CPI	SN
CPI Gas	M	0	CPI	SN
Service Price Index	Q	3	PP	SN
Oil Transport	M	3	GDP	SN
Number of Vacancies	M	3	L	SN
Unemployment (x1000)	M	1	L	SN
Number of bankruptcies	M	0	F	SN
Sentiment Production past 3 months	M	0	S1	SN
Sentiment Production coming 3 months	M	0	S1	SN
Sentiment Prices coming 3 months	M	0	S1	SN
Sentiment Order book position	M	0	S1	SN
Sentiment order book position abroad	M	0	S1	SN
Sentiment stock final product	M	0	S1	SN
Sentiment economy past 3 months	M	0	S1	SN
Sentiment revenue past 3 months	M	0	S1	SN

Continued on next page

Table II.1 – continued from previous page

Name	Frequency	Publ. lag	Group	Source
Sentiment revenue coming 3 months	M	0	S1	SN
Cons. conf.: gen. ec. situation past 3 months	M	0	S2	SN
Cons. conf.: gen. ec. situation coming 3 months	M	0	S2	SN
Cons. conf.: fin. situation past 3 months	M	0	S2	SN
Cons. conf.: fin. situation coming 3 months	M	0	S2	SN
Cons. conf.: big purchases	M	0	S2	SN
Domestic consumption	M	2	GDP	SN
Consumption food	M	2	GDP	SN
Consumption services	M	2	GDP	SN
Consumption durable goods	M	2	GDP	SN
Consumption other	M	2	GDP	SN
Daily production retail	M	2	GDP	SN
Daily revenues retail	M	3	GDP	SN
Daily production raw materials	M	2	GDP	SN
Daily revenues raw materials	M	3	GDP	SN
Production construction	M	2	GDP	SN
Gov. income from taxes	Q	3	G	SN
Gov. income social premiums	Q	3	G	SN
Gov. income goods and services	Q	3	G	SN
Gov. income from capital	Q	3	G	SN
Gov. intermediate consumption	Q	3	G	SN
Gov. net investments	Q	3	G	SN
Gov. social benefits paid	Q	3	G	SN
Gov. total subsidies paid	Q	3	G	SN
Gov. total income	Q	3	G	SN
Gov. total expenditures	Q	3	G	SN
Gov. net lending/borrowing	Q	3	G	SN
National debt	Q	3	G	SN
Number of mortgages granted (NL)	M	1	H	DNB
Total savings households EU	M	1	GDP	ECB
Loans granted to corporations NL	M	1	F	ECB
Loans housing purchase EU	M	1	H	ECB
Consumer credit households EU	M	1	GDP	ECB
Total import	M	1	GDP	SN
Total export	M	1	GDP	SN
PPI raw materials	M	1	PP	SN
PPI final products	M	1	PP	SN
PPI electricity	M	1	PP	SN
PPI water	M	1	PP	SN
Using price raw materials	M	1	PP	SN
Using price industrial products	M	1	PP	SN
PPI gas	M	1	PP	SN
Index added value construction	M	2	H	SN
Index revenues catering industry	Q	3	GDP	SN
Total production industry	M	1	GDP	SN
Total production raw materials	M	1	GDP	SN
Total production energy	M	1	GDP	SN
Total production water	M	1	GDP	SN
Total incapacity benefits	M	3	L	SN
Total social assistance benefits	M	2	L	SN
Total unemployment benefits	M	2	L	SN
Vacancy indicator industry	M	0	L	SN
Vacancy indicator construction	M	0	L	SN
Vacancy indicator commercial services	M	0	L	SN

Continued on next page

Table II.1 – continued from previous page

Name	Frequency	Publ. lag	Group	Source
CLI	M	1	S2	OECD
Production commercial services	Q	3	GDP	SN
Production other services	Q	3	GDP	SN
Total production construction	Q	3	GDP	SN
Total production	Q	3	GDP	SN
AEX closing price	D	0	F	AEX
Index lending consumption households EU	M	1	GDP	ECB
Index lending for housing purchase EU	M	1	H	ECB
Debt securities issued non MFI EU	M	1	F	ECB
Loans/securities MFI EU	M	1	F	ECB
Loans with central counterparts non MFI EU	M	1	F	ECB
Price index housing NL	M	0	H	SN

III Preliminary data analysis

Figure III.1 Linear filter output. Left: Season component of raw GDP series. Right: Noise component of raw GDP series

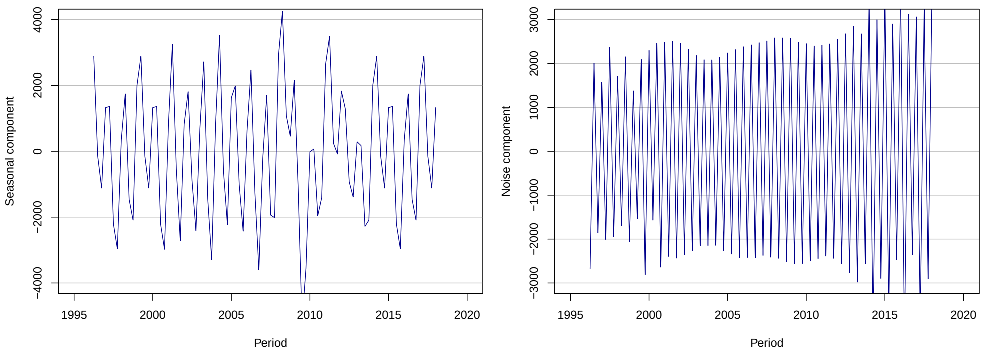


Figure III.2 Fourier Transform of growth series. Left: Fourier amplitude as a function of frequency in units of cycles per month. Right: the same Fourier amplitude shown as a function of period in units of months. Frequencies higher than 1/6 cycles per month, i.e. periods shorter than 6 months cannot be resolved with Fourier analysis of quarterly sampled series.

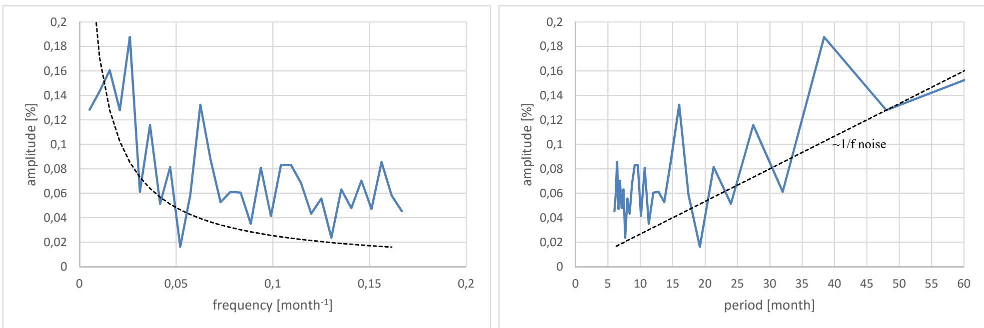


Figure III.3 Repeated application of linear filter on the raw GDP series, also suppressing the 16 month periodicity. Left: Season component, right: Noise component

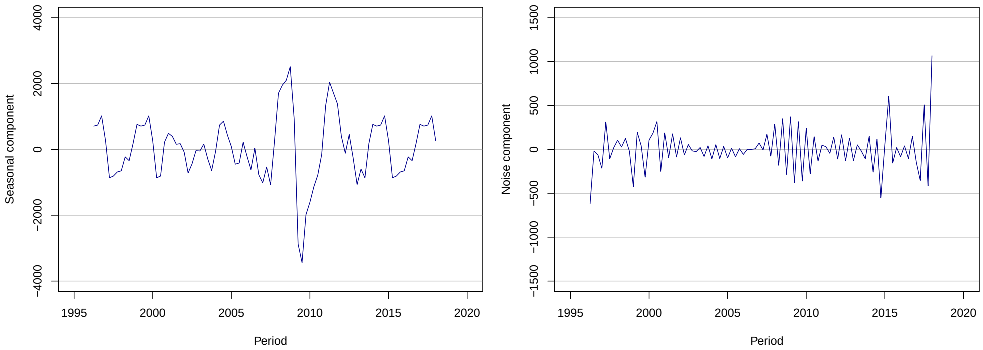
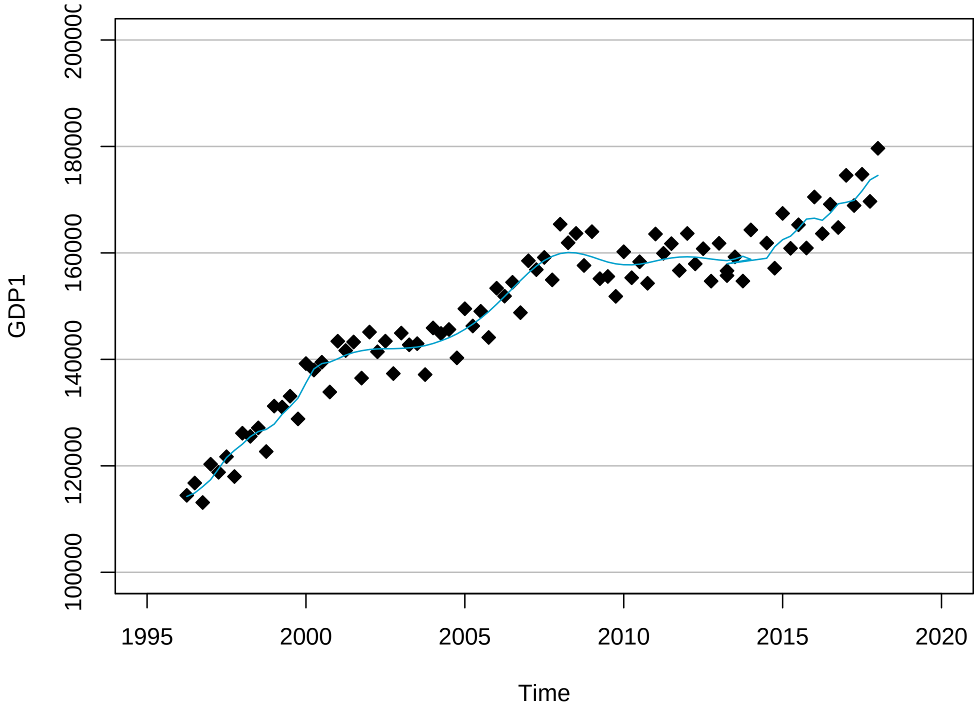


Figure III.4 Trend component of GDP series, together with GDP itself



IV PCA Analysis Monthly Series

Figure IV.1 Scree plot tests. Left panel: Scree plot monthly differences. Right panel: Scree plot with the four tests for monthly differenced series

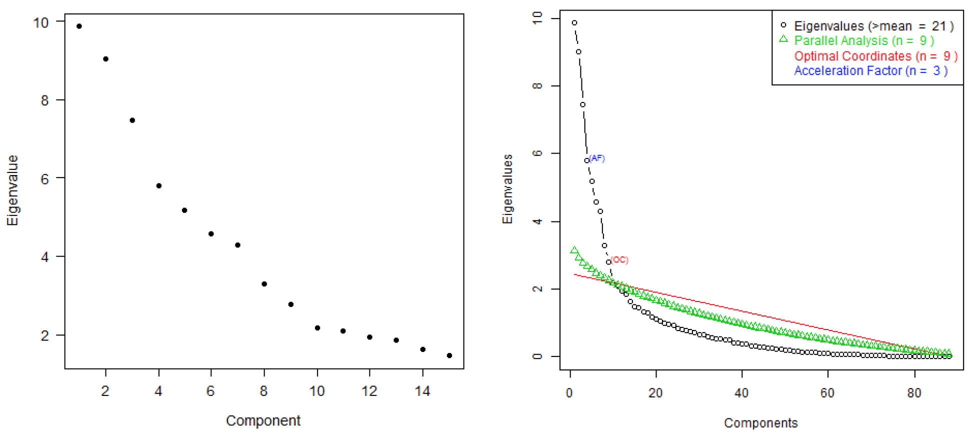


Figure IV.2 Eigenvalues and proportion of variance explained of first 15 components

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Eigenvalue	9.8608	9.0301	7.4595	5.7972	5.1796	4.5782	4.2938	3.2926	2.7822	2.1734	2.0897	1.9396	1.8507	1.6208	1.4629
Proportion	0.1121	0.1026	0.0848	0.0659	0.0589	0.052	0.0488	0.0374	0.0316	0.0247	0.0237	0.022	0.021	0.0184	0.0166
Cumulative	0.1121	0.2147	0.2994	0.3653	0.4242	0.4762	0.525	0.5624	0.594	0.6187	0.6425	0.6645	0.6855	0.704	0.7206

	1	2	3	4	5	6	7	8	9	10
Eigenvalue	9.8608	9.0301	7.4595	5.7972	5.1796	4.5782	4.2938	3.2926	2.7822	2.1734
IC 1	-	-0.0965	-0.1308	-0.149	-0.1664	-0.1807	-0.1973	-0.199	-0.1939	-0.177
IC 2	-	-0.1128	-0.1552	-0.1816	-0.207	-0.2295	-0.2543	-0.264	-0.2671	-0.2583

Figure IV.3 Values criteria with penalty functions g_1 and g_2 for up to 15 factors

V PCA Analysis Quarterly Series.

Figure V.1 Scree plot tests. Left panel: Scree plot quarterly differences. Right panel: Scree plot with the four tests for quarterly differenced series

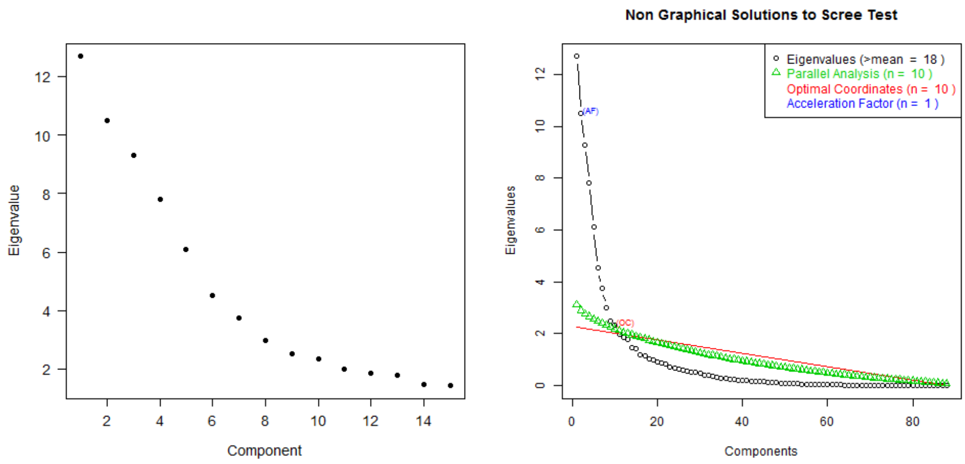


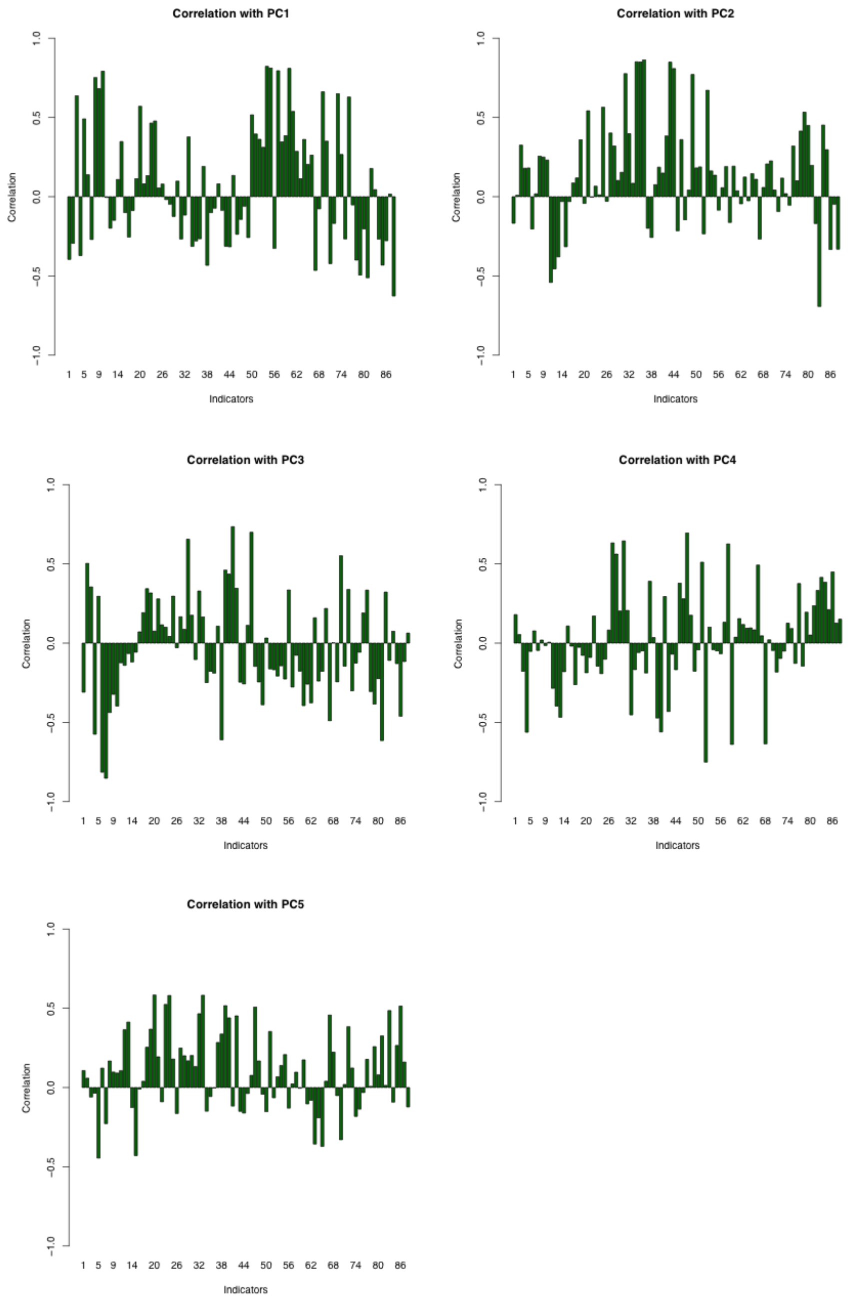
Figure V.2 Eigenvalues and proportion of variance explained of first 15 components

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Eigenvalue	12.6849	10.4887	9.2909	7.8104	6.1038	4.5222	3.7403	2.9881	2.5101	2.352	1.9964	1.8534	1.7917	1.4749	1.4308
Proportion	0.1441	0.1192	0.1056	0.0888	0.0694	0.0514	0.0425	0.034	0.0285	0.0267	0.0227	0.0211	0.0204	0.0168	0.0163
Cumulative	0.1441	0.2633	0.3689	0.4577	0.527	0.5784	0.6209	0.6549	0.6834	0.7101	0.7328	0.7539	0.7742	0.791	0.8073

Figure V.3 Values criteria with penalty functions g_1 and g_2 for up to 15 factors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Eigenvalue	12.6849	10.4887	9.2909	7.8104	6.1038	4.5222	3.7403	2.9881	2.5101	2.352	1.9964	1.8534	1.7917	1.4749	1.4308
IC 1	-	-0.1517	-0.2267	-0.2982	-0.3549	-0.3902	-0.4159	-0.43	-0.4352	-0.4437	-0.4455	-0.4475	-0.4537	-0.4506	-0.4515
IC 2	-	-0.1679	-0.2511	-0.3307	-0.3956	-0.439	-0.4728	-0.495	-0.5084	-0.525	-0.535	-0.5451	-0.5595	-0.5645	-0.5735

Figure V.4 Correlations indicators and first six factors. The figures reflect the correlation between the indicators and each of the first five factors. As expected, the correlations clearly decrease for higher factors. The indicators were sorted by their group, see table II.1 in the Appendix. Group L: 1-10, Group CPI: 11-17, Group PP: 18-25, Group GDP: 26-50, Group S1: 51-59, Group S2:60-65, Group H: 66-70, Group F: 71-76, Group G: 77-88.



VI PCA Analysis Operational Approach

Figure VI.1 Scree plot tests. Left panel: Scree plot Operational Approach. Right panel: Scree plot with the four tests for Operational Approach

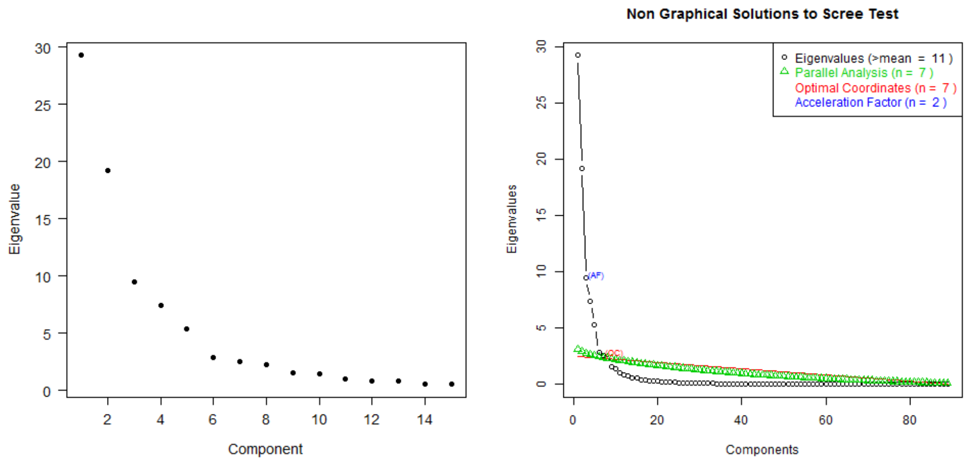


Figure VI.2 Eigenvalues and proportion of variance explained of first 15 components

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Eigenvalue	28.6932	19.013	9.3694	7.3623	5.3186	2.8528	2.5152	2.2546	1.5421	1.419	1.0026	0.8243	0.7603	0.5846	0.5232
Proportion	0.3261	0.2161	0.1065	0.0837	0.0604	0.0324	0.0286	0.0256	0.0175	0.0161	0.0114	0.0094	0.0086	0.0066	0.0059
Cumulative	0.3261	0.5421	0.6486	0.7322	0.7927	0.8251	0.8537	0.8793	0.8968	0.913	0.9244	0.9337	0.9424	0.949	0.9549

Figure VI.3 Values modified criteria with penalty functions g_1 and g_2 for up to 15 factors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Eigenvalue	28.6932	19.013	9.3694	7.3623	5.3186	2.8528	2.5152	2.2546	1.5421	1.419	1.0026	0.8243	0.7603	0.5846	0.5232
*IC 1**	-	0.6125	0.5866	0.5834	0.6031	0.6504	0.7018	0.7561	0.8183	0.8819	0.9502	1.0205	1.0916	1.1646	1.2383
*IC 2**	-	0.5966	0.5628	0.5516	0.5633	0.6027	0.6462	0.6925	0.7467	0.8024	0.8628	0.9251	0.9882	1.0533	1.119

Figure VI.4 Factor loadings

	F1	F2	F3	F4	F5
Number of self–employed	0.9636	0.1829	–0.0013	0.0822	–0.114
Labour participation	0.7504	–0.418	–0.3382	0.1628	0.1729
CPI general	0.97	0.1596	0.1358	–0.0153	–0.0538
CPI goods	0.9526	0.0374	0.0934	–0.12	0.0588
CPI Industrial goods	0.7981	–0.0629	0.0871	–0.2889	0.2476
CPI durable goods	–0.888	0.2057	0.1538	–0.0269	–0.0935
CPI services	0.9479	0.2166	0.1554	0.0413	–0.1205
CPI electricity	–0.3648	–0.2331	0.2009	–0.4179	0.2616
CPI gas	0.82	–0.2158	–0.1739	0.0082	0.1125
Service Price Index	0.9735	0.0699	–0.0854	0.1009	–0.0511
Oil transport	–0.0239	0.3043	–0.0773	0.1873	–0.1556
Number of vacancies	–0.491	0.5655	–0.4822	0.1675	0.1762
Unemployment	0.5337	–0.08	0.6011	–0.4609	–0.2678
Number of bankruptcies	–0.1566	–0.7459	0.2866	–0.3332	0.0472
Sentiment Production past 3 months	0.1204	0.7047	–0.0022	–0.2864	0.0578
Sentiment Production coming 3 months	–0.0662	0.6712	0.1996	–0.2884	0.2017
Sentiment prices coming 3 months	–0.3306	0.479	–0.4548	–0.4884	0.0824
Sentiment order position	–0.1855	0.8723	–0.0773	–0.3098	0.1668
Sentiment order position abroad	–0.105	0.7924	0.0347	–0.4929	0.2286
Sentiment stock final product	–0.2769	–0.43	–0.1088	0.451	–0.0857
Sentiment economy past 3 months	–0.2208	0.815	–0.1634	–0.27	0.2467
Sentiment revenue past 3 months	–0.3714	0.7355	–0.0303	–0.2847	0.0865
Sentiment revenue coming 3 months	–0.5497	0.6407	0.0681	–0.2687	0.2031
Consumer confidence: ec. situation past 3 month	0.0817	0.9348	–0.0786	0.026	–0.0435
Consumer confidence: ec. situation coming 3 months	–0.0058	0.6899	0.2227	0.0128	–0.1994
Consumer confidence: fin. situation past 3 months	–0.0189	0.4685	–0.6093	0.4579	0.0533
Consumer confidence: fin. situation coming 3 months	–0.2525	0.7133	–0.2173	0.2261	–0.0192
Consumer confidence: big purchases	0.3535	0.7189	–0.3108	0.2776	–0.0809
Domestic consumption	0.0738	0.2385	–0.7428	0.1856	0
Consumption food	0.1619	0.1602	–0.4433	0.0066	–0.1087
Consumption services	0.6838	0.4666	–0.24	0.348	0.1612
Consumption durable goods	–0.2788	0.1637	–0.628	0.2519	0.2107
Consumption other	–0.391	–0.199	–0.637	–0.2223	–0.3932
Daily production retail	0.4397	0.4289	–0.3542	–0.2579	0.0531
Daily revenues retail	0.6239	0.2401	–0.3053	–0.3419	0.4549
Daily production raw materials	–0.2308	–0.3547	–0.4459	–0.5126	–0.5593
Daily revenues raw materials	–0.0478	–0.5142	–0.4502	–0.5999	–0.2711
Production construction	0.0211	–0.0918	–0.5864	–0.5649	–0.4922
Gov. income from taxes	0.5882	0.4291	–0.3205	0.1866	–0.1523
Gov. income social premiums	0.7277	0.0935	–0.1897	–0.248	–0.3618
Gov. income goods and services	0.619	–0.0931	–0.5116	–0.1468	–0.3013
Gov. income from capital	–0.1116	–0.595	–0.3856	–0.3428	0.041
Gov. intermediate consumption	0.6894	–0.3845	–0.4601	0.0091	–0.0922
Gov. net investments	–0.5165	–0.159	–0.5425	0.3419	0.2408
Gov. social benefits	0.9466	0.0815	0.044	0.114	0.101
Gov. total subsidies paid	0.367	–0.4195	–0.5013	0.2268	–0.1152
Gov. total income	0.7701	0.1824	–0.4112	–0.0902	–0.2935
Gov. total expenditures	0.8987	–0.1295	–0.1993	0.1049	0.039
Gov. net lending/borrowing	–0.2065	0.3781	–0.2813	–0.1656	–0.1433
National debt	0.9565	–0.0515	0.2086	0.0157	–0.1429
Number of mortgages granted NL	–0.4779	0.5773	–0.1286	0.1623	–0.0952
Total savings households EU	0.4966	0.5833	–0.2645	0.4351	0.066
Loans granted to corporations Netherlands	0.7445	–0.5317	–0.2043	–0.0523	0.2398
Loans housing purchase EU	0.8928	–0.2269	–0.2522	0.0272	0.1696
Consumer credit households EU	–0.6823	–0.5794	–0.242	–0.0689	0.2749
Total import	0.879	0.3268	–0.0605	–0.1961	0.1225
Total export	0.8723	0.3487	–0.0508	–0.208	0.0533
PPI raw materials	0.1618	–0.5565	–0.2896	–0.4261	0.4981
PPI final products	0.7359	–0.1634	–0.0714	–0.4216	0.413
PPI electricity	–0.1827	–0.5521	–0.5574	–0.2517	0.4379
PPI water	–0.8572	–0.1348	–0.2635	0.1647	0.1149
Using price raw materials	0.2515	–0.4729	–0.1432	–0.5403	0.5585
Using price industrial products	0.8572	–0.0927	0.0235	–0.3547	0.2511
PPI gas	–0.1662	–0.5654	–0.5711	–0.2231	0.4374
Index added value construction	–0.1611	0.1718	–0.257	0.147	0.0305
Index revenues catering industry	0.5844	0.5059	0.0275	0.3674	0.2479
Total production industry	0.4242	0.4281	–0.3257	–0.2358	0.0815
Total production raw materials	–0.2289	–0.3527	–0.4473	–0.5129	–0.561
Total production energy	–0.1136	–0.0712	–0.4535	–0.3071	–0.7494
Total production water	0.2554	0.2212	0.0845	0.1114	0.0058
Total incapacity benefits	–0.895	0.0258	0.2838	–0.0489	–0.1795
Total social assistance benefits	0.6768	0.4845	0.4226	–0.1062	–0.2985
Total unemployment benefits	0.6108	0.2828	0.5647	–0.2213	–0.381
Vacancy Indicator industry	–0.0219	0.8674	0.0411	–0.3989	0.0514
Vacancy indicator construction	–0.0062	0.9075	0.08	–0.1941	–0.095
Vacancy indicator commercial services	–0.2473	0.9014	0.0646	–0.3048	0.0072
CLI	–0.2354	0.653	–0.0861	–0.5184	0.2882
GDP monthly	0.7627	0.4693	–0.3429	0.1611	0.026
Production commercial services	0.8101	0.4627	–0.1982	0.1752	0.0258
Production other services	0.8632	–0.0146	–0.1178	0.1068	0.0779
Total production construction	–0.273	0.2408	–0.4139	0.4439	–0.0119
Total production	–0.1794	0.1368	–0.5844	–0.3105	–0.4879
AEX closing price	–0.0247	0.9117	–0.0307	–0.1686	0.038
Index lending consumption households EU	–0.5837	0.7062	–0.0603	0.1726	–0.1409
Index lending for housing purchase EU	–0.8566	0.4124	0.0398	–0.1462	0.0148
Debt securities issues non MFI EU	–0.9296	0.0573	–0.101	–0.034	0.103
Loans/securities MFI EU	0.4466	–0.1706	–0.0234	0.4892	–0.355
Loans with central counterparts non MFI EU	–0.8314	0.3744	–0.2521	0.007	0.1358
Price Index housing NL	–0.4991	0.1188	–0.6596	0.4581	0.1465

VII model quality measures

VII.1 Econometric approach, $p = 1, ly = 0$

Table VII.1 Average standard deviation, $p = 1, ly = 0$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.7340	0.7704	0.8304	1.0013	1.0044	1.0375	1.0776	1.1322	1.1161
M2	0.6504	0.6087	0.5647	0.5549	0.5453	0.5494	0.5029	0.4899	0.4807
Q1	0.5253	0.5209	0.5217	0.5188	0.5278	0.5144	0.5074	0.5104	0.5142
Q2	0.4818	0.4688	0.4660	0.4635	0.4582	0.4514	0.4453	0.4448	0.4444

Table VII.2 Hitting times (out of 28) with $\alpha = 0.05, p = 1, ly = 0$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	28	28	27	28	28	28	28	28	28
M2	28	28	27	25	22	22	24	24	21
Q1	26	25	25	24	24	24	24	25	23
Q2	26	26	26	26	24	24	24	25	26

VII.2 Econometric approach, $p = 2, ly = 0$,

Table VII.3 Mean squared forecast error (MSFE), $p = 2, ly = 0$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	-								
M2	0.3259	0.3224	0.3035	0.5103	1.2276	1.8905	2.5948	87.91	380.3
Q1	0.9486	0.6503	1.079	1.223	0.979	1.3301	1.3484	1.6584	105.0
Q2	0.2636	0.3157	0.7032	0.5545	0.7104	0.5374	0.7539	5.668	2.144

Table VII.4 Average standard error, $p = 2, ly = 0$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	-								
M2	0.3259	0.3224	0.3035	0.5103	1.2276	1.8905	2.5948	87.91	380.3
Q1	1.0561	0.7661	1.0058	2.207	2.0808	1.7411	1.7472	2.7437	5.1815
Q2	0.5216	0.6052	1.9800	1.5093	1.8343	1.1902	1.3410	2.2027	2.5816

Table VII.5 Hitting times (out of 28) with $\alpha = 0.05, p = 2, ly = 0$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	-								
M2	28	28	27	28	28	28	28	28	28
Q1	27	26	28	28	27	28	28	28	28
Q2	26	28	28	28	28	28	28	28	28

VII.3 Econometric approach, $p = 1, ly = 1$

Table VII.6 Mean squared forecast error (MSFE), $p = 1, ly = 1$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.3005	0.3329	0.4391	0.4210	0.4683	0.4891	0.4596	0.5141	0.5003
M2	0.2179	0.2972	0.4083	0.4035	0.6916	0.7662	0.6672	0.6556	0.6654
Q1	0.3371	0.3484	0.3786	0.4441	0.4702	0.5337	0.4528	0.4770	0.6136
Q2	0.2856	0.3103	0.3235	0.375	0.4011	0.4080	0.3949	0.3998	0.4266

Table VII.7 Average standard error, $p = 1, ly = 1$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.4310	0.4967	0.5866	0.6593	0.6852	0.6954	0.7024	0.7363	0.7362
M2	0.6425	0.6009	0.5668	0.5583	0.5477	0.5516	0.5032	0.4721	0.4600
Q1	0.5383	0.5342	0.5366	0.5294	0.5371	0.5345	0.5174	0.5092	0.5130
Q2	0.4804	0.4640	0.4602	0.4572	0.4548	0.4384	0.4340	0.4323	0.4331

Table VII.8 Hitting times (out of 28) with $\alpha = 0.05, p = 1, ly = 1$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	24	23	26	28	28	28	28	26	27
M2	28	27	26	25	22	23	23	22	21
Q1	26	26	26	24	23	22	23	23	22
Q2	24	24	24	23	23	22	21	21	21

VII.4 Econometric approach, $p = 1, ly = 2$

Table VII.9 Mean squared forecast error (MSFE), $p = 1, ly = 2$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.3048	0.3208	0.4061	0.4296	0.4808	0.5105	0.4588	0.5310	0.5370
M2	0.2196	0.3235	0.4465	0.4667	0.7226	0.7513	0.7161	0.7531	1.506
Q1	0.3604	0.3866	0.4113	0.5076	0.55280	0.7089	0.6164	0.6234	0.7360
Q2	0.2922	0.3296	0.3430	0.4304	0.4555	0.5670	0.5616	0.5735	0.6215

Table VII.10 Average standard error, $p = 1, ly = 2$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.3674	0.4003	0.4642	0.5659	0.5885	0.6140	0.6201	0.6114	0.5830
M2	0.6526	0.6060	0.5663	0.5533	0.5469	0.5528	0.4943	0.4482	0.4755
Q1	0.5470	0.5439	0.5517	0.5423	0.5470	0.5673	0.5400	0.5270	0.5228
Q2	0.4859	0.4683	0.4655	0.4649	0.4547	0.4405	0.4304	0.4296	0.4287

Table VII.11 Hitting times (out of 28) with $\alpha = 0.05, p = 1, ly = 2$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	21	23	23	25	26	26	27	25	25
M2	28	27	26	24	21	21	23	20	20
Q1	25	25	25	23	22	21	21	22	21
Q2	24	24	24	23	22	21	21	21	20

VII.5 Econometric approach, $p = 1, ly = 3$

Table VII.12 Mean squared forecast error (MSFE), $p = 1, ly = 3$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.3063	0.3141	0.4076	0.4393	0.4902	0.5818	0.4801	0.5626	0.6612
M2	0.2883	0.3361	0.4739	0.5012	0.9815	0.9295	0.8725	0.8143	2.3767
Q1	0.3705	0.3886	0.4258	0.5176	0.5285	0.6786	0.6036	0.6164	0.7930
Q2	0.3014	0.3193	0.3380	0.4105	0.4387	0.5427	0.5181	0.5423	0.6053

Table VII.13 Average standard error, $p = 1, ly = 3$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.3269	0.3532	0.4134	0.5503	0.5384	0.5732	0.5806	0.5675	0.5522
M2	0.6469	0.6004	0.5586	0.5476	0.5495	0.5692	0.4958	0.4409	0.5029
Q1	0.5294	0.5280	0.5420	0.5356	0.5360	0.5634	0.5361	0.5239	0.5223
Q2	0.4635	0.4499	0.4458	0.4445	0.4321	0.4369	0.4234	0.4232	0.4227

Table VII.14 Hitting times (out of 28) with $\alpha = 0.05, p = 1, ly = 3$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	20	23	22	25	25	24	25	25	21
M2	28	26	26	24	19	19	18	20	17
Q1	25	24	24	24	24	21	21	21	22
Q2	24	24	24	23	22	21	20	20	21

VII.6 Econometric approach, $p = 1, ly = 4$

Table VII.15 Mean squared forecast error (MSFE), $p = 1, ly = 4$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.2874	0.2940	0.3899	0.4034	0.4207	0.4754	0.4064	0.5089	0.6241
M2	0.3412	0.3489	0.5850	0.6045	1.1561	1.0422	1.714	1.7469	6.0142
Q1	0.3935	0.4191	0.4644	0.5813	0.5856	0.7326	0.6773	0.7159	0.9210
Q2	0.3509	0.3398	0.3634	0.4468	0.4618	0.5350	0.5178	0.5493	0.6060

Table VII.16 Average standard error, $p = 1, ly = 4$

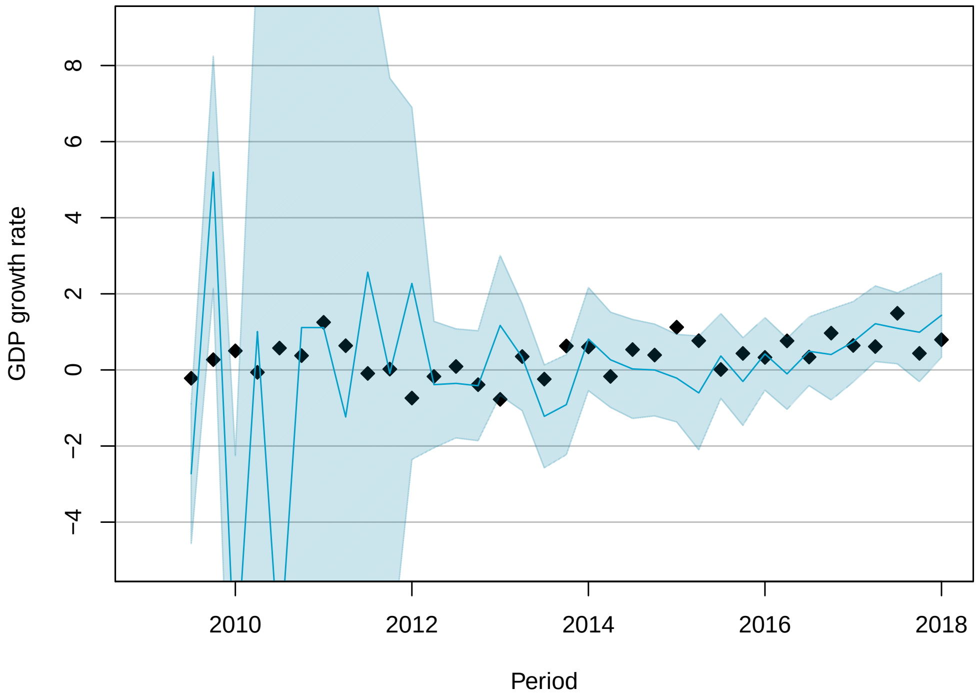
No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	0.6526	0.6378	0.6559	0.7243	0.7014	0.7104	0.7088	0.7010	0.7035
M2	0.6546	0.5994	0.5463	0.5321	0.5457	0.5862	0.5350	0.4754	0.5306
Q1	0.5375	0.5377	0.5590	0.5531	0.5563	0.5820	0.5632	0.5678	0.5984
Q2	0.4460	0.4375	0.4335	0.4338	0.4314	0.4363	0.4252	0.4279	0.4310

Table VII.17 Hitting times (out of 28) with $\alpha = 0.05, p = 1, ly = 4$

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
M1	28	28	27	27	27	27	28	27	27
M2	27	25	25	24	20	20	20	19	15
Q1	25	24	24	24	24	21	21	23	22
Q2	24	25	24	22	21	22	20	20	21

VII.7 Estimated GDP growth rate model M2

Figure VII.1 Nowcast M2 ($r = 8, p = 2, ly = 0$)



VII.8 Operational approach

Figure VII.2 Operational approach nowcast O2 ($r = 7, p = 2, ly = 0$)

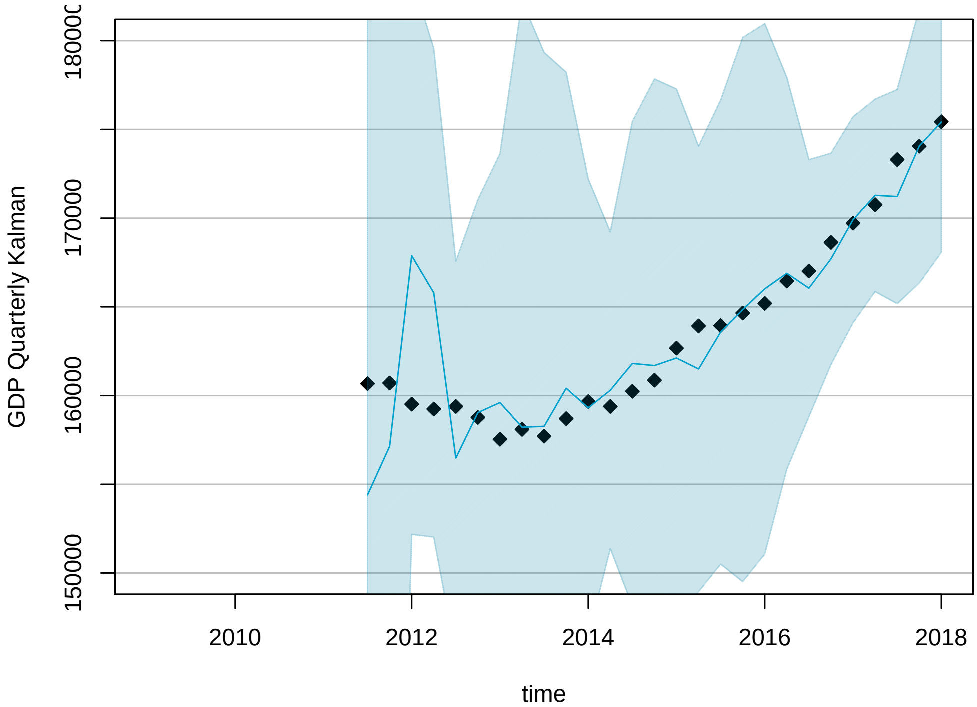


Table VII.18 Mean squared forecast error (MSFE) operational approach (x1000000), ($AR = 3.489$)

No. factors (r)	4	5	6	7	8	9	10	11	12
p = 1, ly = 0									
OA1	5.320	4.738	3.988	4.424	4.557	4.259	4.271	5.346	<u>3.737</u>
OA2	4.772	3.793	2.614	2.405	2.608	2.435	2.100	2.075	<u>1.978</u>
p = 2, ly = 0									
OA1	5.334	4.535	3.399	<u>2.959</u>	4.373	3.950	6.751	7.1320	213.1
OA2	3.923	2.826	2.124	2.786	2.236	<u>1.748</u>	2.527	3.441	227.7
p = 1, ly = 1									
OA1	1.367	<u>1.067</u>	1.205	1.261	2.152	2.253	3.098	2.893	2.403
OA2	1.282	<u>0.950</u>	1.013	0.996	1.224	1.286	1.454	1.401	1.474
p = 1, ly = 2									
OA1	1.424	<u>1.107</u>	1.244	1.273	2.281	2.415	3.230	2.517	2.202
OA2	1.270	<u>0.893</u>	0.972	.961	1.252	1.342	1.657	1.468	1.545
p = 1, ly = 3									
OA1	1.537	<u>1.047</u>	1.081	1.062	1.677	1.773	2.147	1.357	1.625
OA2	1.449	<u>0.947</u>	0.998	1.011	1.112	1.165	1.285	1.151	1.089

Table VII.19 Average standard error operational approach ($AR = 2515$)

No. factors (r)	4	5	6	7	8	9	10	11	12
p = 1, ly = 0									
OA1	1405	1315	1207	1183	1182	<u>1030</u>	1043	1129	1204
OA2	1078	1026	964	965	916	823	790	784	<u>781</u>
p = 2, ly = 0									
OA1	5132	7589	5072	4064	<u>2789</u>	3244	3628	4894	5503
OA2	<u>1838</u>	2508	2147	2665	3891	3675	4021	5042	18431
p = 1, ly = 1									
OA1	828	810	<u>801</u>	810	851	845	1064	1112	1174
OA2	782	757	755	750	739	728	746	<u>716</u>	721
p = 1, ly = 2									
OA1	843	834	<u>828</u>	850	873	869	1084	1100	1168
OA2	795	772	760	754	760	747	769	730	<u>732</u>
p = 1, ly = 3									
OA1	804	755	<u>700</u>	720	776	778	920	903	960
OA2	755	706	676	674	683	654	662	619	<u>600</u>

Table VII.20 Hitting times (out of 28) operational approach with $\alpha = 0.02$ ($AR = 1$)

No. factors (r)	4	5	6	7	8	9	10	11	12
p = 1, ly = 0									
OA1	**26**	24	**25**	23	22	21	23	23	23
OA2	23	24	**25**	**26**	23	20	21	21	23
p = 2, ly = 0									
OA1	**28**	**28**	**28**	**28**	**28**	**28**	**28**	**28**	**27**
OA2	**27**	**27**	**28**	**28**	**28**	**28**	**28**	**28**	**28**
p = 1, ly = 1									
OA1	24	**26**	**25**	**25**	24	24	24	**26**	**25**
OA2	24	**26**	**26**	**26**	24	24	23	22	23
p = 1, ly = 2									
OA1	**25**	**26**	**25**	**25**	24	24	**25**	**26**	**26**
OA2	24	**26**	**26**	**26**	24	24	23	23	24
p = 1, ly = 3									
OA1	**25**	**26**	**25**	**25**	24	23	24	**27**	**26**
OA2	24	**26**	24	24	23	24	24	23	24

VIII Model Confidence Set

Figure VIII.1 MCS output all models without the flash: $\alpha = 0.3, T_1, L^{(1)}, T = 39$

```
Model Q1_1010 eliminated 2018-06-20 09:40:18
Model M1_810 eliminated 2018-06-20 09:40:46
###########################################################################
Superior Set Model created     :
        Rank_M          v_M MCS_M Rank_R       v_R  MCS_R       Loss
Q1_410       2 -1.16908548     1      9  1.921325 0.4142  0.5410773
Q2_410       1 -1.19719007     1      1 -1.010321 1.0000  0.4742861
Q2_1010      5 -1.09773490     1      5  1.784926 0.5414  0.8216532
M1_411       3 -1.14370593     1      4  1.486709 0.8086  0.6022416
M2_411       4 -1.11908354     1      6  1.794286 0.5316  0.6984947
M2_412       8 -1.07667716     1      7  1.815019 0.5124  0.7945622
M2_810      11 -0.02620137     1      8  1.833363 0.4962  3.3541855
OA1_511     10 -0.85840897     1      3  1.416765 0.8478  1.2783292
OA1_513      9 -0.89226761     1     10  1.934639 0.4016  1.3379303
OA2_511      6 -1.09432517     1     12  2.054550 0.3014  0.7375796
OA2_512      7 -1.08711275     1     11  1.983174 0.3582  0.7287745
OA2_712     12  0.99233646     1      2  1.010321 0.9906 29.6884495
p-value :
[1] 0.3014

###########################################################################
```

Figure VIII.2 MCS output EA models: $\alpha = 0.2, T_1, L^{(2)}, T = 39$

```
Model M1_810 eliminated 2018-06-21 15:13:31
Model M2_810 eliminated 2018-06-21 15:13:40
###########################################################################
Superior Set Model created     :
        Rank_M        v_M  MCS_M Rank_R        v_R  MCS_R      Loss
Q1_410       2 -1.0458496 1.0000      3  0.4610784 0.9988 0.5480845
Q1_1010      7  1.2699107 0.5860      7  1.9662207 0.3418 0.6503379
Q2_410       1 -3.3050670 1.0000      1 -0.3620666 1.0000 0.5331687
Q2_1010      5  0.3444006 0.9976      6  1.3140686 0.8408 0.6008206
M1_411       3 -0.6892860 1.0000      2  0.3620666 0.9990 0.5508547
M2_411       6  0.4584673 0.9876      5  1.2865769 0.8580 0.6115571
M2_412       4  0.2056804 0.9996      4  1.0110536 0.9588 0.5961854
p-value :
[1] 0.3418

###########################################################################
```

Figure VIII.3 MCS output OA models: $\alpha = 0.2, T_1, L^{(2)}, T = 39$

```
Model OA1_513 eliminated 2018-06-21 15:14:38
###########################################################################
Superior Set Model created     :
        Rank_M        v_M  MCS_M Rank_R        v_R  MCS_R      Loss
OA1_511      3  0.2575103 0.9746      2  0.8130976 0.8546 0.6533294
OA2_511      1 -2.0465425 1.0000      1 -0.8130976 1.0000 0.6095079
OA2_512      2 -0.9571617 1.0000      3  0.9039476 0.8010 0.6263636
OA2_712      4  0.5086289 0.8782      4  0.9175677 0.7910 0.6695560
p-value :
[1] 0.791

###########################################################################
```

Figure VIII.4 MCS output complete set of models without flash estimate: $\alpha = 0.2, T_1$, $L^{(2)}, T = 28$

```
Model OA2_712 eliminated 2018-06-21 15:20:25
Model OA1_513 eliminated 2018-06-21 15:20:46
Model M1_810 eliminated 2018-06-21 15:21:05
Model OA1_511 eliminated 2018-06-21 15:21:24
Model OA2_511 eliminated 2018-06-21 15:21:37
Model OA2_512 eliminated 2018-06-21 15:21:50
Model Q2_1010 eliminated 2018-06-21 15:22:00
Model M2_810 eliminated 2018-06-21 15:22:08
###########################################################################
Superior Set Model created    :
        Rank_M        v_M  MCS_M Rank_R        v_R  MCS_R      Loss
Q1_410       5  0.9217780 0.7764      6  1.8720315 0.2508 0.4497277
Q1_1010      6  1.2547942 0.5368      5  1.6756403 0.3896 0.4998113
Q2_410       1 -2.3334586 1.0000      2  0.5544336 0.9934 0.3935464
M1_411       4  0.7881459 0.8568      4  1.5211433 0.5098 0.4528301
M2_411       3 -1.3147432 1.0000      3  0.7750035 0.9598 0.3726905
M2_412       2 -1.3296078 1.0000      1 -0.5544336 1.0000 0.3695383
p-value :
[1] 0.2508

###########################################################################
```

Figure VIII.5 MCS output complete set of models including flash estimate: $\alpha = 0.1$, $T_1, L^{(2)}, T = 28$

```
Model OA1_511 eliminated 2018-06-28 16:03:09
Model OA2_511 eliminated 2018-06-28 16:03:34
Model OA1_513 eliminated 2018-06-28 16:03:56
Model OA2_512 eliminated 2018-06-28 16:04:15
Model OA2_712 eliminated 2018-06-28 16:04:30
Model M2_810 eliminated 2018-06-28 16:04:44
Model M1_810 eliminated 2018-06-28 16:04:54
Model M1_411 eliminated 2018-06-28 16:05:05
Model Q1_410 eliminated 2018-06-28 16:05:12
Model Q2_1010 eliminated 2018-06-28 16:05:18
###########################################################################
Superior Set Model created    :|
        Rank_M        v_M  MCS_M Rank_R       v_R  MCS_R      Loss
Q1_1010      5  1.6724368 0.2014      5  2.155530 0.1082 0.4998113
Q2_410       4  0.6599571 0.8474      4  2.012126 0.1578 0.3935464
M2_411       3 -0.2578948 1.0000      3  2.012118 0.1578 0.3726905
M2_412       2 -0.3419660 1.0000      2  1.914257 0.2008 0.3695383
flash        1 -2.3565095 1.0000      1 -1.914257 1.0000 0.2662022
p-value :
[1] 0.1082

###########################################################################
```

IX Year on Year Growth

Figure IX.1 Historical and flash year on year growth 2001-2017.

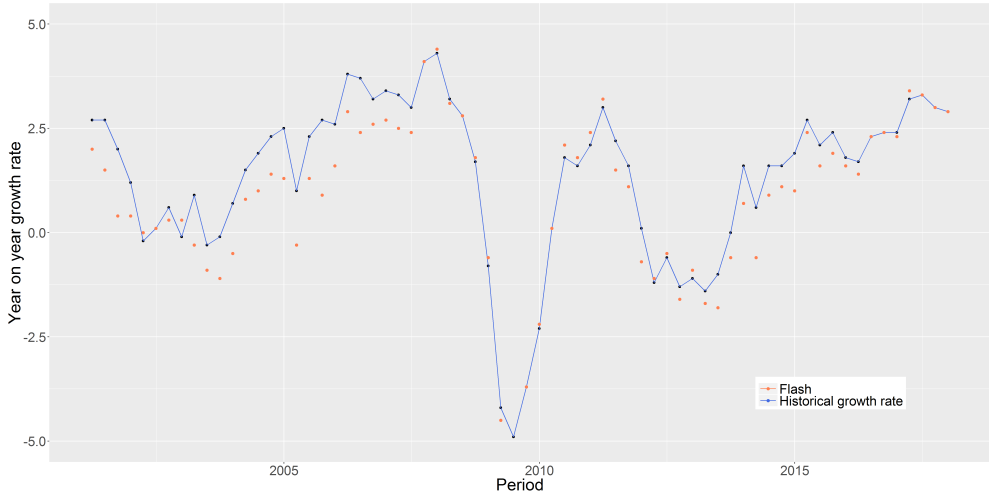


Figure IX.2 First seven principal components

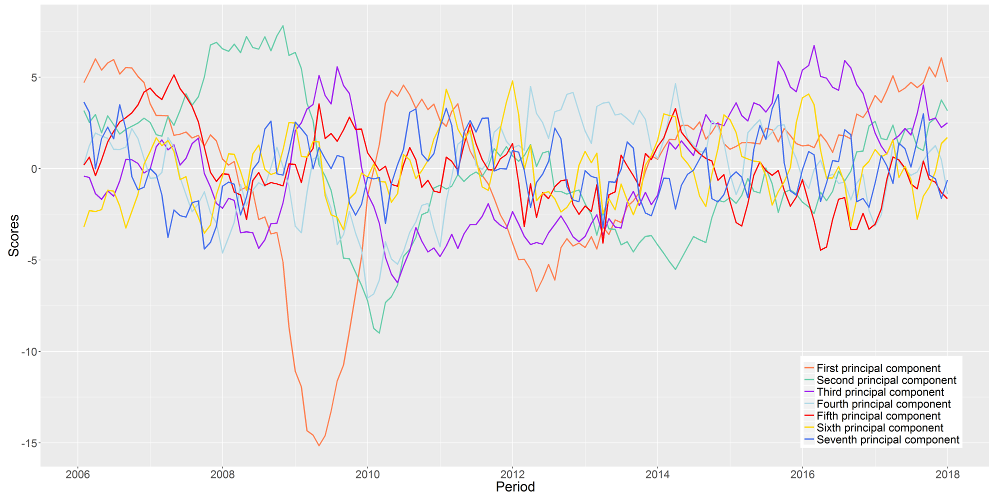
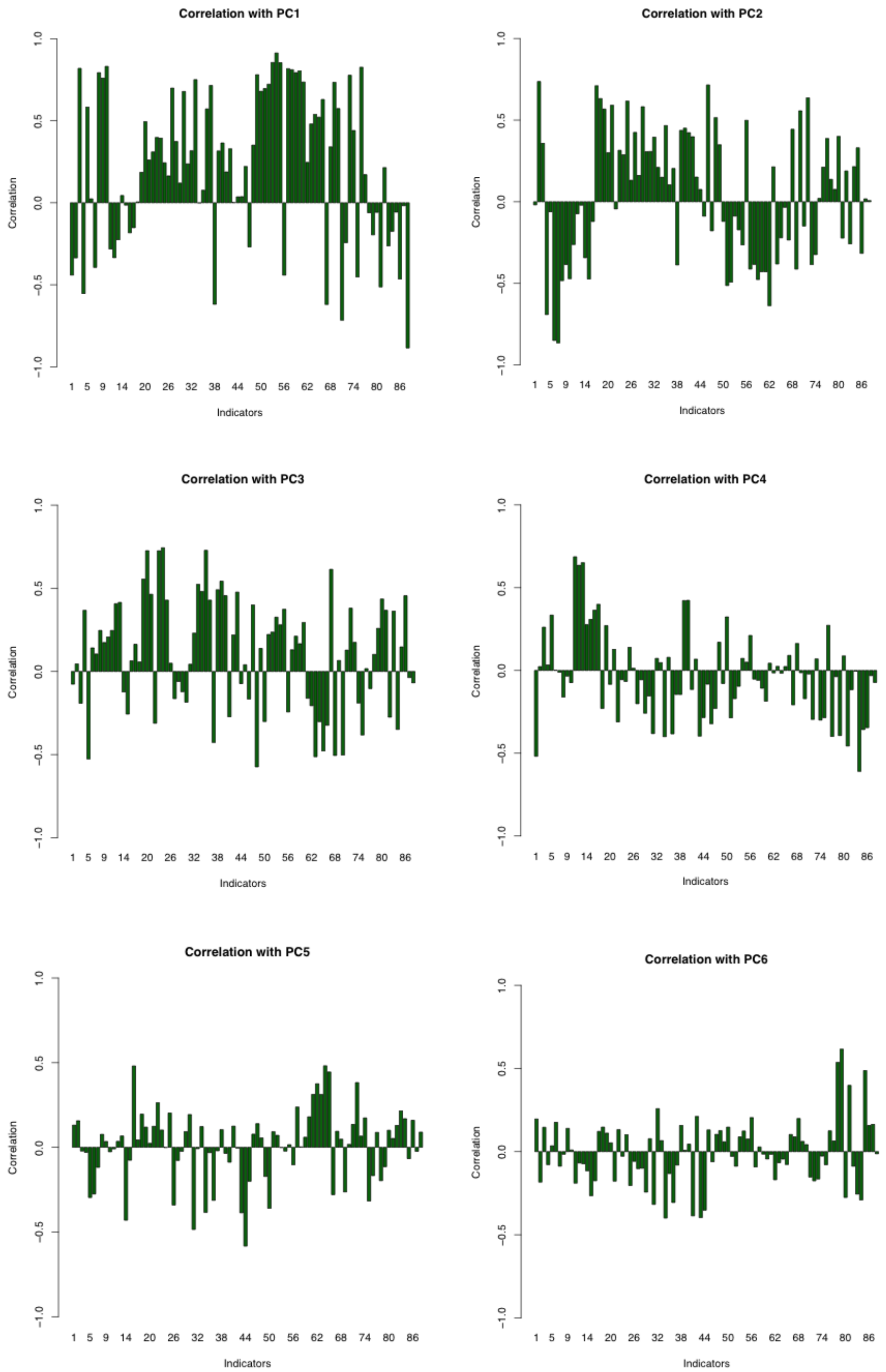


Table IX.1 Average standard error ($p = 1, ly = 1$). Over the period 2010-2017.

No. factors (r)	4	5	6	7	8	9	10	11	12
Method									
Y1	0.6766	0.6569	0.6368	0.6141	0.5874	0.5734	0.5707	0.5683	<u>0.5666</u>
Y2	0.5803	0.5624	0.5246	0.4912	0.4714	0.4639	0.4587	0.4693	<u>0.4438</u>

Figure IX.3 Correlations indicators and first six factors. The figures reflect the correlation between the indicators and each of the first six factors. As expected, the correlations clearly decrease for higher factors. The indicators were sorted by their group, see table II.1 in the Appendix. Group L: 1-10, Group CPI: 11-17, Group PP: 18-25, Group GDP: 26-50, Group S1: 51-59, Group S2: 60-65, Group H: 66-70, Group F: 71-76, Group G: 77-88.



Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information