Discussion Paper

Simultaneous Nowcasting of Netherlands' Macroeconomic Trends and Seasonal Patterns Based on Fourier Analysis

Frank P. Pijpers
Lucas Harlaar
Jan van den Brakel
Pim Ouwehand

May 27, 2025

An objective method is developed for providing fast initial estimates and corresponding uncertainty of large macroeconomic indicators, including the aggregate gross domestic product (GDP). Furthermore, a new approach to modelling trend and seasonal effects within time series based on Fourier analysis is explored. Underlying latent factors are extracted from a large panel of auxiliary time series, in a two-step procedure containing principal component analysis and dynamic factor state space analysis. This procedure is applied in a real-time framework to predict the present state of the target macroeconomic indicators, using the extracted factors. The information flow of the auxiliary time series is simulated in a real-time setting over the period 2013-2020, obtaining a target series estimate each quarter. These estimates, often called nowcasts, are compared with the realized figures of the target series to measure the performance of the present procedure. The results indicate an improvement on earlier studies and the current method used by Statistics Netherlands to provide so-called 'flash estimates'.

Keywords: Gross Domestic Product, state space model, dynamic factor model, Kalman filter, principal component analysis

Reviewer: Marc Smeets

1 Introduction

Statistical agencies like Statistics Netherlands (SN) provide information that is fundamental for economic and monetary policy-making. Important decisions however, are often based on incomplete data due to publication lags and revisions. One of the most important macroeconomic indicators is the quarterly gross domestic product (GDP) growth rate. Generally, SN publishes an initial estimate of this indicator 45 days after the reference quarter has ended. This is called the 'flash estimate'. The flash estimate is updated after another 45 days, when more information has become available. It takes up to 32 months until the final statistic is determined, while national and European policy-making is an ongoing process. Hence, policy-makers generally use the flash estimates to base their decisions on. In recent years, the flash estimates have shown significant deviations from the final statistic, indicating it might be beneficial to study different methods for providing early initial estimates. An extensive overview of this field of study is published by Eurostat (Bacchini et al. 2017). One of the techniques they mention is the use of (dynamic) factor models for nowcasting, which is what this paper revolves around.

Chamberlain and Rothschild (1983) introduce the factor structure with an application in finance, which is later on generalized into a dynamic framework with applications in macroeconomic forecasting (Forni, Hallin, et al. 2000; Forni and Lippi 2001; Stock and Watson 2002a; Stock and Watson 2002b). The general concept is to assume that the dynamics in a large panel of time series can be split into two orthogonal components. The common component captures cross-sectional correlations that are driven by a few unobserved common factors. The idiosyncratic component describes variable specific shocks, while permitting weak cross and serial correlations. Allowing the latent factors to change over time, generated by their own stochastic processes, is what defines the dynamic framework generalization of this concept.

In the following years, using dynamic factor models (DFMs) became the new standard for macroeconomic forecasting at central banks and other economic and statistical institutions. Many authors have proved them fruitful in this regard (Marcellino et al. 2003; Giannone, Reichlin, and Sala 2004; Boivin and Ng 2005; Forni, Hallin, et al. 2005). Consequently, Giannone, Reichlin, and Small (2008) pioneered the area of macroeconomic monitoring through real-time estimations using a DFM. They altered their approach from Giannone, Reichlin, and Sala (2004) and developed a statistical framework, which is broadly adopted in the literature nowadays. The framework exploits the information of more timely published auxiliary time series to estimate the variable of interest in real time. Furthermore, it can handle jagged edges in the data. This is a feature that arises when one wants to analyze the most recent releases of multiple time series simultaneously, since they have different reference periods and publication dates at agencies like SN or central banks. This estimation technique is called 'nowcasting', which refers to forecasting a variable that exhibits a publication lag in the current time.

After the introduction by Giannone, Reichlin, and Small (2008), their two-step estimation approach is developed more formally by Doz et al. (2011) and they prove the consistency properties. In the first step, the parameters of the DFM are estimated by OLS regression, where principal components obtained through principal component analysis (PCA) are treated as the true unobserved common factors. In the second step, the DFM is put into state space form, which is very useful when estimating unobserved components (Durbin and Koopman 2012). The principal components act as the initialization of the Kalman smoother, which is a recursive algorithm that estimates the unobserved factors using the estimated model parameters from the first step. These 'smoothed' factor estimates replace the PCA approximation. Finally, they fit a simple regression model to the target variable and the smoothed factor estimates, to obtain the nowcast of the former. Their technique shows improvement on earlier factor modelling techniques that only used the principal components as factor estimates without re-estimating them in the second step.

Other methods for estimation have been proposed, such as quasi maximum likelihood (QMLE) by Doz et al. (2012). Bräuning and Koopman (2014) alter the two-step method by simultaneously modelling the target variables, principal components and unobserved factors in a small scale 'collapsed' dynamic factor model as second step. The model parameters can be estimated by maximum likelihood, due to the dimension reduction.

In this paper another alteration is proposed, where the novelty of this alteration is threefold. The two-step approach of Doz et al. (2011) is followed, but it is applied separately on the trend and seasonal components of the time series in the information set. Moreover, the DFM parameter estimation is based on a combination of least squares regression and maximum likelihood. Finally, instead of identifying one target variable, the final optimization problem is extended to obtain multiple target variable nowcasts simultaneously. A stylised representation of the framework of modelling is presented in figure 1.1, showing how the standard DFM approach is duplicated, to account for the parallel treatment of trend and seasonal components, together with the new decompose and recombine steps at the beginning and end of that pipeline.

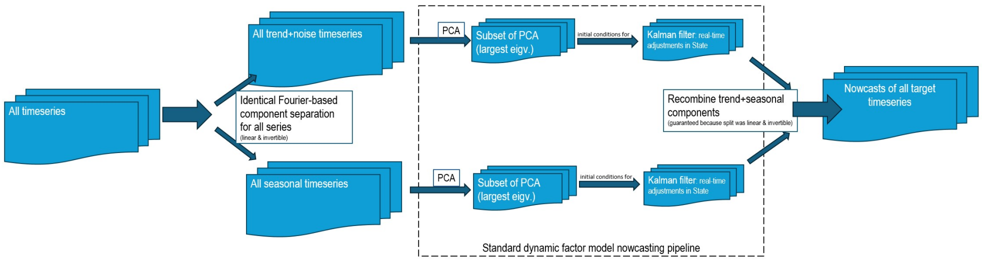


Figure 1.1 A stylised overview of the main steps in the modelling flow. All target and auxiliary time series are decomposed in an identical, linear and invertible way based on a filter designed in the Fourier domain, so that the unobserved components (trend, seasonal) can be put through the standard dynamic factor modelling pipeline for nowcasting, and can be recombined afterwards to obtain complete and consistent nowcasts for all target series.

Applying DFMs for nowcasting has become the cutting-edge method for central banks, statistical agencies and other economic policy institutions that are dealing with the publication lags of aggregated macroeconomic variables. See for example empirical studies at the American and Dutch central banks, (Hindrayanto et al. 2016; Bok et al. 2018; Dijk et al. 2024), Netherlands Bureau for Economic Policy Analysis (CPB) (Luginbuhl and Verstegen 2024) and also at SN (Schiavoni et al. 2021). This research contributes to the general consensus in macroeconomic nowcasting literature of the last decade, see Cascaldi-Garcia et al. (2024) or Stundziene et al. (2024) for an extensive overview. More specifically, it builds upon Kuiper and Pijpers (2020), which showed promising results for estimating GDP growth rate in a timely manner. They propose a DFM where PCA is applied to decompose many auxiliary time series in combination with a Kalman filter. Their approach is extended in this paper by adding more target variables to the model and initiating a different way of investigating component specific information within time series data. Moreover, an alternative method is suggested for dealing with the jagged edges. The investigation aims to ascertain whether this new approach outperforms the current procedure SN uses for providing 'flash estimates'.

The structure of the remainder of this paper is as follows. Section 2 describes the methodology, followed by section 3 which investigates the data on which the methods are applied. Section 4 documents the results from the empirical application and conclusions are in section 5.

2 Methods

The methodology in sections 2.1 and 2.5 distinguishes the current research from the discussion paper by Kuiper and Pijpers (2020), while revisiting and improving upon their methods in the remaining sections. In section 2.1 a different approach to time series modelling is proposed as well as a method for dealing with jagged edges in the data.

2.1 Time series decomposition

A general notion in time series analysis is to decompose a given time series into parts such as the trend, seasonal and noise component. More specifically, the identification and removal of the seasonal component is of great importance for statistical agencies like SN. Seasonality describes movements throughout a year that repeat with similar intensity in different years, making it predictable to a certain extent. Often, these movements are large enough to mask other characteristics of interest in the time series. They can arise with different amplitude and timing for different time series, complicating the analysis of current trends. Therefore, several procedures exist for the removal of the seasonal component in the time series, often based on the use of filters.

One can model these components explicitly using a structural time series (STS) model, as described in Harvey (1989) and Durbin and Koopman (2012). However, a different approach is taken here by making use of a linear filter proposed by Perrucci and Pijpers (2017). The design of the weights in the filtering moving average window are based on linear filters defined in the Fourier domain (see Appendix A for further details). The central idea is to define a broad bandwidth-limited filter in the Fourier domain so that also *quasi*-periodic signals with periods less than a year will be removed in seasonally corrected time series. The broad filter is described in Appendix A. The filter response function $R(f)$ selects the seasonal component, i.e. signals at all frequencies f within a band, and therefore $1 - R(f)$ selects the trend plus (slow) economic cycle plus stochastic components. The filter response $R(f)$ is specified to be $= 0$ or nearly so for $f < \frac{1}{year}$ and also for $f > \left(\frac{1}{2\Delta_t} - \frac{1}{year}\right)$. Here Δ_t stands for the cadence of the measurements. Note the distinction being made between 'cadence' and 'frequency': 'cadence' refers to the sampling strategy and the timing interval between the subsequent samples or measurements being taken and used for the analysis, whereas 'frequency' refers to the components intrinsic to signal being measured. For many official statistics this cadence is one month, which is strictly speaking slightly irregular sampling given the different lengths in time of months of the year. It could also be a week, or a fixed multiple of weeks. It could even be a quarter. For the purposes of calculating the filter weights a 'year' does not need to be an integer multiple of the cadence so one may choose e.g. 365.25 days as has been done here. A slightly different choice would have only a very minor influence on the value of the filter weights. In between these frequencies $\left[\frac{1}{year}, \frac{1}{2\Delta_t} - \frac{1}{year}\right]$ the $R(f) \approx 1$. The filter weights are obtained by performing an inverse Fourier transform on $R(f)$. It is worth noting that it is desirable to have a finite window width in the time domain, i.e. that the weights are identical to 0 for large positive and negative lags. The benefit of this for official statistics is that the part of a seasonally corrected time series that is further in the past from 'now' than the half-width of the window, by construction will no longer be modified by seasonal corrections and is therefore definitive. There is a close connection between the width of the window in the time domain and the shape of the filter in the Fourier domain. A top-hat shape, i.e. with very sharp edges at the frequencies where $R(f)$ changes from 0 to 1 and vice versa, will always require non-zero weights in the time domain over a very wide window. By allowing some 'softening' at the edges of the filter, so that the slope is high but not infinite as it would be for a top-hat shape, it is achieved that in the time domain the window over which the weights are non-zero is finite. Evidently, this implies some compromise between competing needs, which is discussed in more detail in Perrucci and Pijpers (2017). The filter weights need only be determined once: they do not depend on what time series the filter is applied to. The choice of weights given in Perrucci and Pijpers (2017) are considered to be a good pragmatic balance between a sharp edged filter in the frequency domain and a finite width window in the time domain. Applied in this way, the algorithm filters out the seasonal component in the data, essentially splitting each time series in a trend+noise component and a seasonal component. The operation is performed before modelling, as opposed to STS modelling. Main reason for adopting this new approach is that it allows extracting component-specific information throughout the analysis. Consequently, all steps are performed twice, separately for both components of the splitted series. Due to the linearity of the operations one can simply add the separate outcomes in the end to create nowcasts for the original target time series.

What is often referred to as seasonal effects in economic time series is only *quasi*-periodic: it does not repeat perfectly from one year to the next. However, many standard techniques start off with an assumption of exact periodicity and proceed from there with subsequent correction steps allowing for relaxation of that initial assumption. The linear filter of Perrucci and Pijpers (2017) does not make this initial assumption, and therefore can cope very well with quasi-periodic (stochastic) seasonality. The most important aspect for the present purpose of using this approach is that it is linear and invertible: it allows to split up the time series into the seasonal and trend+cycle components, perform all the nowcasting steps on those separate components, and then recombine the two again, in a well-behaved manner which also allows for straightforward propagation of all measurement and modelling uncertainties. As Perrucci and Pijpers (2017) demonstrate on some real time series, their proposed filter performs similarly to widely used methods for seasonal adjustments such as TRAMO/SEATS and X13-ARIMA. A more extensive explanation can be found in that paper.

Applying the Fourier transform directly to a time series is impractical in the present empirical application, because it would in principle need to be done again every single time a new data point becomes available, while in practice there is no need for that given the type of filtering that is to be applied. It is preferable to do the entire analysis in the time domain, which is straightforward using the linear weights presented in Perrucci and Pijpers (2017). This involves only a modest number of measurement points to calculate the filtered time series, and leaves most of the past of the resulting time series invariant. In the case of monthly and quarterly time series data, the sampling cadence of the time series is not regularly spaced since the number of days differs between the period hence some resampling is required for the Fourier transform to be applied. Resampling of time series is defined by changing the cadence of the observations, where upsampling refers to an increase in cadence of observations and downsampling to a decrease in cadence (Brownlee 2017).

Interpolation is a useful numerical method when upsampling of data is necessary. Conceptually, the underlying principle can be regarded as a two step operation. First, it creates a continuous reconstruction of the data, based on the discrete data points. Second, resampling takes places at a different sampling rate (Hou and Andrews 1978). In principle, a smooth continuous curve is to be found in the first step. Here a cubic spline is chosen. Once upsampled onto a regular cadence, the filter weights are applied. This combination of operations is linear and therefore measurement error trivially can be accounted for throughout. For an extensive discussion see Hou and Andrews (1978) and Maeland (1988).

Consequently, the cubic splines interpolation formula is applied to interpolate or (backward) extrapolate the data, depending on the periodicity and first publication date of each variable. This has the additional benefit of losing 'jagged edges' at the beginning of the sample. However, the data also exhibits jagged edges at the end of the sample due to differences in the publication lag of several variables. These forward jagged edges are extrapolated using a linear prediction algorithm, discussed by Makhoul (1975). Note that here 'linear' refers to the use of recent measurements; the extrapolation is not a straight line but it is based on autoregressive modelling and has proven to be better at extrapolating signals forward in time.

Combining these methods allows deriving a smooth continuous function $z = f(t)$ from which it is simple to sample at the desired cadence as the final step. Hence, a regularly spaced time series $\{z_t\}_{t=1}^{T}$ can be derived straightforwardly, i.e. the distance between the time steps (function points) is of equal length after this operation. Perrucci and Pijpers (2017) apply the linear filter to obtain a trend plus noise component from the regularly spaced series, denoted z_t^{τ}. Subsequently an estimate for the seasonal component is obtained as $z_t^{\varsigma} = z_t - z_t^{\tau}$.

2.2 Principal Component Analysis

The downside of parametric modelling is that when the dimensionality of the data increases, standard optimization techniques become unstable. For example, when the number of time series, say N, is large, it can be difficult to identify cointegrating relationships across all combinations in the cross-sectional dimension. These difficulties are examples of the curse of dimensionality, which refers to the collection of problems that arise when analysing data in high dimensional spaces. Fortunately, principal component analysis offers a non-parametric data combination technique to overcome these problems. It can serve as a proxy for missing cointegration information, by assuming most of the variation in the data is captured by a few unobserved components. An estimate of these components is obtained as a weighted summary of all dependencies between the variables in the data. Moreover, it reduces the variable and parameter space considerably, hence conventional estimation techniques can be applied (Bacchini et al. 2017).

In the context of factor models, Chamberlain and Rothschild (1983) prove that principal components provide a consistent estimate for latent factors as $N \to \infty$. Stock and Watson (2002a) draw the same conclusion for both $N, T \to \infty$. In order to achieve the desired dimension reduction, it is common to select the first r principal components, where $r << N$, and discard the remaining $L - r$ components without losing much of the information. Raîche et al. (2013) discuss multiple ways to determine the optimal r for principal component or factor analysis that exist in the literature. Classical ones like the Kaiser rule (Kaiser 1960) and parallel analysis (Horn 1965) rely on the analysis of the eigenvalues and numerical criteria. More specifically, let λ_i be an element on the diagonal of Λ, i.e. the i'th eigenvalue of the matrix of all cross-products of auxiliary time series. The Kaiser rule simply states that the optimal r equals the number of eigenvalues that are larger than the mean eigenvalue. The parallel analysis is based on bootstrapping correlation matrices from the matrix of auxiliary time series correlations and obtaining eigenvalues in each replication. These are used to formulate empirical eigenvalue distributions of which a location statistic (LS) can be compared with the eigenvalues corresponding to the actual observed data set. The LS could be the mean or a quantile, which is picked by the user. The number of eigenvalues that exceed the LS determines r. This produces

$$\begin{aligned} r_{kaiser} &= \textstyle\sum_i [\lambda_i \geq \bar{\lambda}] \\ r_{parallel} &= \textstyle\sum_i [\lambda_i \geq LS_i] \end{aligned} \tag{1}$$

where the brackets refer to the Iverson brackets notation, introduced by Iverson (1962). Note that square brackets will be used to define matrices throughout the rest of this paper unless explicitly stated otherwise.

Cattell (1966) offered a graphical alternative, based on plotting the ordered eigenvalues, called the scree test. Downside of the graphical nature of this test is that it can be unclear how many principal components/factors one needs to retain exactly. Therefore, Raîche et al. (2013) proposes non-graphical solutions to Cattell's scree test. The optimal coordinates (OC) index extrapolates the coordinates of the previous eigenvalue using linear regression. The predicted outcome is compared with the observed eigenvalue. The optimal r is determined by the number of observed eigenvalues that exceed the predicted eigenvalue. Either the Kaiser rule or parallel analysis criterion must be satisfied simultaneously in this non-graphical scree test solution. For robust decision making purposes the rounded average outcome of these three tests is used as optimal r in the present analysis.

2.3 State Space Model

The retrieved principal components are treated as proxy for the latent common factors that characterize the well-established DFM. This model is used to re-estimate the factors by applying the Kalman smoother algorithm, which is considered the second step in the estimation procedure. In order to perform this step the state space approach to time series analysis is adopted, see Durbin and Koopman (2012) for details of state space modelling and the Kalman filter.

Let $\{y_t\}_{t=1}^T$ be a series of multivariate observations. In the state space approach it is assumed that the development over time of the y_t's is determined by an unobserved series of vectors $\{\alpha_t\}_{t=1}^T$ known as the states. The linear Gaussian state space model (SSM) specifies the relation between the α_t's and y_t's and is given by,

$$\begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim \mathsf{N}(0, H_t), \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim \mathsf{N}(0, Q_t), \\ & & \alpha_1 &\sim \mathsf{N}(a_1, P_1), \end{aligned} \tag{2}$$

where y_t is the $(N \times 1)$ observation vector and α_t is the $(r \times 1)$ unobserved state vector. The error terms ε_t and η_t are assumed to be mutually and serially independent at all lags. The first equation in (2) is called the observation equation, which describes through the known design matrix Z_t how the observed series depend on the unobserved state variables α_t. The second equation is called the state equation, which describes through the known design matrix T_t how the state variables change over time according to a Markov structure. Furthermore, H_t and Q_t are covariance matrices of the measurement errors ε_t and state disturbance terms η_t. They define the dynamic properties of the SSM and their elements are often referred to as the hyperparameters. The Kalman filter assumes that the hyperparameters are known, which is generally not the case in practice. The hyperparameters are therefore estimated with maximum likelihood, see Durbin and Koopman (2012) for details. Finally R_t is a known selection matrix. To start the Kalman filter recursion, initial values for α_1 are required. They are obtained by a diffuse initialization, which implies that the starting values in a_1 are taken equal to zero with diagonal covariance matrix P_1 with large values on the diagonal.

The Kalman filter and smoother are applied to obtain optimal estimates for the state variables. The Kalman filter is a forward recursive algorithm that gives optimal estimates for the state vector α_t based on the observations obtained until period t. The Kalman smoother is a backward recursive algorithm that updates the Kalman filter estimates with the information that became available after period t.

2.4 Dynamic Factor Models

In this section, the DFM is specified. Joint modelling of the auxiliary and target series in a collapsed DFM, as described by Bräuning and Koopman (2014) for example, is not considered. Hence, the Kalman filter and smoother are only applied to the auxiliary series to update the common factor estimates.

Start with defining the classical factor model representation using matrix notation:

$$X = F\Lambda' + e = \chi + e \tag{3}$$

where $X = (x_1, \ldots, x_N)$ is a $(T \times N)$ matrix stacking the observed values of N auxiliary variables and $F = (f_1, \ldots, f_r)$ is the unknown $(T \times r)$ matrix of r latent factors. Furthermore, $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ is a $(N \times r)$ matrix of unknown factor loadings and $e = (e_1, \ldots, e_N)$ is a $(T \times N)$ matrix of idiosyncratic errors. Finally, $\chi = F\Lambda'$ is often referred to as the common component and is orthogonal to the idiosyncratic errors.

Note that equation (3) can also be written as an N-dimensional time series with T observations.

$$x_t^i = \Lambda^i f_t^i + e_t^i = \chi_t^i + e_t^i \quad e_t^i \sim \mathsf{N}(0, \Sigma_e^i) \quad i \in \{\tau, \varsigma\} \tag{4}$$

where the index i is used to explicitly show that this model is applied separately to the trend+noise and seasonal component and the index $t = 1, \ldots, T$ indicates the time dimension. In equation (4) $x_t^i = (x_{1t}^i, \ldots, x_{Nt}^i)'$ is an $(N \times 1)$ vector, Λ^i keeps its interpretation of equation (3), $f_t^i = (f_{1t}^i, \ldots, f_{r^i t}^i)'$ is an $(r^i \times 1)$ vector and $e_t^i = (e_{1t}^i, \ldots, e_{Nt}^i)'$ is an $(N \times 1)$ vector. Bai and Ng (2004) prove that when N and T are large, the common component $\chi_t = \Lambda f_t$ can be extracted without conforming to stationarity assumptions. Recall that x_t is decomposed in a trend+noise and seasonal component, i.e. $x_t = x_t^\tau + x_t^\varsigma$. Hence, it is expected to obtain a nonstationary model when $i = \tau$ and a stationary model when $i = \varsigma$ in equation (4).

When extending the classical factor model to a dynamic framework, one assumes that the latent factors follow their own dynamic process. Kuiper and Pijpers (2020) show that a VAR(1) process gives the best description of the dynamic transitioning of the latent factors for this empirical application, hence:

$$f_{t+1}^i = \Phi^i f_t^i + u_t^i \quad u_t^i \sim \mathsf{N}(0, \Sigma_u^i) \quad i \in \{\tau, \varsigma\} \tag{5}$$

where Φ^i is a $(r^i \times r^i)$ matrix containing autoregressive coefficients and u_t^i is a $(r^i \times 1)$ vector containing shocks to the common factors. Furthermore, e_t^i and u_t^i are assumed to be independent of each other and of their own lags. Testing a seasonal dummy model or trigonometric model instead of a VAR(1), which are described in Harvey (1989) and Durbin and Koopman (2012), as alternatives for the dynamic updating of the seasonal factors could be interesting for further research. Here it is decided to keep the transition equation identical for the trend+noise and seasonal factors, as shown in equation (5). This is done for convenience and consistency purposes in the rest of the current procedure.

Equations (4) and (5) form the DFM specification. Note that this fits the state space model framework of equation (2) by the following:

$$y_t = x_t^i, \quad \alpha_t = f_t^i, \quad \varepsilon_t = e_t^i, \quad \eta_t = u_t^i, \tag{6}$$

with the following system matrices,

$$Z_t = \Lambda^i, \quad H_t = \Sigma_e^i, \quad T_t = \Phi^i, \quad R_t = I_{r^i}^i, \quad Q_t = \Sigma_u^i \tag{7}$$

where $I_{r^i}^i$ reflects an $(r^i \times r^i)$ identity matrix. This shows that one can apply the Kalman filter and smoother to obtain the unobserved common factor estimates $\widehat{f}_t^i$. In order to do so, the model parameters within the system matrices need to be estimated. The focus here lies on 'approximate' dynamic factor models, i.e. N is large (Bacchini et al. 2017). This makes maximum likelihood estimation (MLE) infeasible as many authors have pointed out (Bai 2003; Doz et al. 2012; Bräuning and Koopman 2014). Two parameter vectors are initiated:

$$\begin{aligned} \theta_1^i &= (\sigma_{e_1}^{2\,i}, \dots, \sigma_{e_N}^{2\,i}, \phi_{11}^i, \phi_{21}^i, \dots, \phi_{(r^i-1)r^i}^i, \phi_{r^i r^i}^i)' \\ \theta_2^i &= (\sigma_{u_1}^{2\,i}, \dots, \sigma_{u_{r^i}}^{2\,i})' \end{aligned} \tag{8}$$

where $\sigma_e^{2\,i}$ and $\sigma_u^{2\,i}$ refer to diagonal elements of Σ_e^i and Σ_u^i respectively and ϕ_{jk}^i describes the auto-regressive coefficient in the j-th row and k-th column of Φ^i. An estimate for the elements in θ_1^i, say $\widehat{\theta}_1^i$, is obtained through least squares estimation, by treating $\widehat{\Lambda}^i$ and $\widehat{f}_t^i$ that were retrieved from the PCA step as the true unobserved factors and loadings. However, an estimate for θ_2^i, denoted by $\widehat{\theta}_2^i$, is obtained through maximum likelihood estimation, again using $\widehat{f}_t^i$. Since both error variance matrices Σ_e^i and Σ_u^i are diagonal by assumption, the latter can be estimated by maximum likelihood, because it only requires r^i parameters to be estimated.

2.5 Target series estimation

The final aim of this paper is to nowcast multiple time series of interest, i.e. target series, by exploiting the timeliness of a large set of auxiliary variables. Similar to the approach in Doz et al. (2011) and Kuiper and Pijpers (2020), the common factors that summarize the set of auxiliary series are, in a next step, used as explanatory variables in a linear regression on the target series. Here, the target series compose the national accounts of the production side of the economy which share GDP as their aggregate. A nowcast of GDP, and the uncertainty interval around its point estimate, can simply be obtained by aggregating respectively the point estimate nowcasts of the individual target series, or their uncertainties, following the approach proposed by Bollineni-Balabay et al. (2016). Alternatively it is possible to add GDP as additional target series to a model and add restrictions such that the sum of the nowcasts over the production categories equals the nowcast of the total GDP (Pandher 2002). As motivated by Bollineni-Balabay et al. (2016), adding total GDP does not add any new information and is less efficient compared to fitting a model to the underlying categories and estimating the total by aggregation over these categories.

The target time series are denoted with y^i_{jt}, with $i \in \tau, \varsigma$, where the index j runs over the amount of target series $j = 1, \dots, J$ and the index t indicates the successive sampling times of the time series $t = 1, \dots, T$. Furthermore, y^τ_{jt} now reflects the trend+noise component of target series j at time t and y^ς_{jt} reflects the corresponding seasonal component (cf. section 2.1 and Perrucci and Pijpers (2017)). The common factors are denoted with f^i_{lt} where the index $l = 1, \dots, \nu^i$ now runs over the optimal amount of factors r^i plus an extra term s^i, i.e. $\nu^i = r^i + s^i$ with $i \in \{\tau, \varsigma\}$. For the trend+noise component f^τ_{lt} a linear trend term is added to the set of factors, while an intercept term is added to the set of factors for both the trend+noise and seasonal components. Hence, $s^\tau = 2$ and $s^\varsigma = 1$. In the following, it is assumed that the target and factor series are all regularly sampled and all at the same instances. Then, their relation can be modelled as:

$$Y^i = F^i \beta^i + \varepsilon^i \tag{9}$$

where $Y^i = (y^i_1, \dots, y^i_J)$ is a $(T \times J)$ matrix containing the decomposed target series observations, $F^i = (f^i_1, \dots, f^i_{\nu^i})$ is a $(T \times \nu^i)$ matrix of common factors and additional intercept (and trend) term(s). The $(\nu^i \times J)$ matrix β^i relates the factors to the target series. Furthermore, $\varepsilon^i = (\varepsilon^i_1, \dots, \varepsilon^i_J)$ is a $(T \times J)$ matrix of idiosyncratic errors. The inverse variance weighted least squares estimator for the regression coefficients in the j-th column of β^i is:

$$\hat{\beta}^i_j = \left(F^{i\prime} W^i F^i\right)^{-1} F^{i\prime} W^i y^i_j, \tag{10}$$

where $W^i = \text{diag}(\mathbb{V}\text{ar}(y^i_{j1})^{-1}, \dots, \mathbb{V}\text{ar}(y^i_{jT})^{-1})$ and it is understood that the common factors are replaced by their estimates obtained from the Kalman smoother. Sometimes, if the $\mathbb{V}\text{ar}(y^i_{jt})$ are not well known, they are assumed equal for all sampling times t. In that case they drop out of the minimisation problem. They can be thought of as being set $= 1$ in equation (10). Point estimates and corresponding variance of the target series are obtained using the estimated regression coefficients $\hat{\beta}^i_j$ directly. Dividing the weighted sum of squared residuals by the number of observations T gives the corresponding variance of the target series estimates:

$$\begin{aligned} \hat{y}^i_{jt} &= \mathbb{E}(y^i_{jt}) = \hat{\beta}^{i\prime}_j \hat{f}^i_t \\ \mathbb{V}\text{ar}(\hat{y}^i_{jt}) &= \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\mathbb{V}\text{ar}(y^i_{jt})} \left(y^i_{jt} - \hat{\beta}^{i\prime}_j \hat{f}^i_t\right)^2 \end{aligned} \tag{11}$$

Now that all parametric components are specified of the estimation procedure that are part of the current empirical application, it is useful to fully specify the model here for convenience:

$$\begin{aligned} x^i_t &= \Lambda^i f^i_t + e^i_t, & e^i_t &\sim N\left(0, \Sigma^i_e\right), \\ f^i_{t+1} &= \Phi^i f^i_t + u^i_t, & u^i_t &\sim N\left(0, \Sigma^i_u\right), \\ y^i_t &= \beta^{i\prime} f^i_t + \varepsilon^i_t, & \varepsilon^i_t &\sim N\left(0, \Sigma^i_\varepsilon\right), \end{aligned} \tag{12}$$

with $i \in \{\tau, \varsigma\}$ and where x^i_t and e^i_t are $(N \times 1)$ vectors. Furthermore, f^i_t and u^i_t are $(r^i \times 1)$ vectors. Note however that f^i_t is a $(v^i \times 1)$ vector in the third equation, due to the manually added intercept (and linear trend) term(s). In addition, y^i_t and ε_t are $(J \times 1)$ vectors. The coefficient matrices Λ^i, Φ^i and β^i are $(N \times r^i)$, $(r^i \times r^i)$ and $(v^i \times J)$ respectively. The error variances matrices Σ^i_e, Σ^i_u and Σ^i_ε are diagonal $(N \times N)$, $(r^i \times r^i)$ and $(J \times J)$ matrices respectively. Finally, the estimate of the aggregated variable (GDP in the present case) is obtained by:

$$\hat{y}_{(J+1)t} = \sum_{j=1}^{J} \hat{y}^\tau_{jt} + \sum_{j=1}^{J} \hat{y}^\varsigma_{jt}, \tag{13}$$

which differs from Kuiper and Pijpers (2020) where GDP was directly used as the only target series in the model.

2.6 Real-time analysis

Ultimately, the goal is to obtain nowcasts of the target variables using the estimation procedure outlined above, while testing whether it can be implemented in the production process of official statistics at SN. Similar to forecasting, nowcasting predicts the current state of the target series, when $t = T$, since information is not available yet at the end of the sample. To obtain nowcasts over time, one can simulate the information flow in a real-time analysis, by progressing through the time dimension of the data, pretending each time step that the information set increases like it does in the concurrent time. More specifically, T increases each iteration with a bi-weekly observation, while the outlined estimation procedure is repeated. Note however that determining the optimal r^i, performing PCA and estimating $\hat{\theta}_2^i$ by MLE are only repeated after 26 bi-weekly iterations, i.e. when a year has passed. This has two main reasons. First, it is believed to be more realistic that these operations will only be re-executed once every while when the proposed procedure is used in production at SN. Second, it keeps the computational time at a reasonable length. All other procedures discussed above are repeated at each bi-weekly iteration.

It is important to note that this study is conducted in pseudo real-time due to the unavailability of preliminary historical data. When simulating the information flow from the past as if it progresses through the concurrent time, use is made of the final statistics. When an initial estimate is revised, the adjusted figure replaces the old estimate, while the latter is not stored properly in a different database. Preliminary figures are kept for 5 years only before they are dropped, which is insufficient for a proper time window in a real-time analysis. This problem arises for both statistics produced by SN or other sources used in this empirical application.

Let the analysis start by shrinking the number of observations to t' and iterate until the end of the original sample is reached, i.e. at T. This creates a nowcast window of size $\tilde{T} = T - t' + 1$. Each iteration, the estimates of the target series at the end of the sample are saved and a time series of nowcasts $\{\hat{y}_t\}_{t=t'}^{T}$ is obtained. The performance can be measured by comparing the nowcast time series with the published target series in the selected window. However, the published target series are on a quarterly basis. Hence, the nowcast cadence needs to be adjusted accordingly. Remember that the interpolated and resampled bi-weekly observations still reflect the unit measurements of the original cadence. This also means that the target series nowcast estimates reflect measurements on a quarterly level, but they are obtained with a bi-weekly cadence. The target series nowcasts and corresponding standard deviations are resampled to a quarterly cadence straightforwardly by the procedure described in Algorithm 1 (assuming t' corresponds to the last bi-weekly observation in a calendar year). Note that a quarter consists of 6.5 bi-weekly observations, which is why there are alternate jumps of 6 and 7 time steps to arrive at the closest approximation of quarterly border values within the bi-weekly sampled time series. Moreover, note that q', Q and $\tilde{Q}$ are the quarterly sampling cadence equivalents of t', T and $\tilde{T}$ respectively. Once nowcasts are obtained with quarterly cadence, they can be compared with the published target series values. Define the root mean squared forecast error (RMSFE) by,

$$RMSFE_j = \sqrt{\frac{1}{\tilde{Q}} \sum_{q=q'}^{Q} (\hat{y}_{jq} - y_{jq})^2} \tag{14}$$

and the mean error (ME) by,

$$ME_j = \frac{1}{\tilde{Q}} \sum_{q=q'}^{Q} (\hat{y}_{jq} - y_{jq}) \tag{15}$$

These statistics are used to measure the accuracy (RMSFE) and bias (ME) of each target series nowcast individually. The index j runs over the $J+1$ target series, including GDP in the assessment of the nowcast performance.

Additionally, the following mean absolute error (MAE) is introduced in such a way that it summarizes the size of revisions at the 90-day mark of the proposed nowcasting approach. Currently, SN publishes initial estimates 45 days after the reference quarter has ended, after which the first revision takes place when another 45 days have passed. The MAE is used to replicate this first flash estimate revision by comparing the current real-time target series estimate of $t-6$, for $t = t'+6, \ldots, T$, with the nowcast of six bi-weekly real-time iterations ago. The reason for this is that six bi-weekly iterations equals 84 days, which approximates the usual 90-day mark the best in the bi-weekly sampling. Consider again the bi-weekly nowcast $\hat{y}_{jt}$, which denotes the j-th target series estimate at time t, conditional on the auxiliary information set up until t. Let $\tilde{y}_{jt-6}$ denote the j-th target series estimate at time $t-6$, again conditional on the auxiliary information up until time t. Note that this differs from the nowcast at $t-6$. Then the MAE is defined as,

$$MAE_j = \frac{1}{\tilde{T}-6} \sum_{t=t'+6}^{T} |\tilde{y}_{jt-6} - \hat{y}_{jt-6}|, \tag{16}$$

and is calculated in the same bi-weekly nowcast window, losing six observations at the end of the sample.

Finally, a reduced χ^2 measure is introduced that reflects how much information is lost by selecting a set of auxiliary variables and reducing its cross-sectional dimension through PCA.

$$\begin{aligned}
\mathbb{V}\text{ar}_{sum} &\equiv \sum_{t=t'}^{T} \left[\mathbb{V}\text{ar}(y_{jt}^{\tau}) + \mathbb{V}\text{ar}(y_{jt}^{\varsigma}) + \mathbb{V}\text{ar}(\hat{y}_{jt}^{\tau}) + \mathbb{V}\text{ar}(\hat{y}_{jt}^{\varsigma}) \right] \\
\chi_j^2 &= \frac{RMSFE_j^2}{\left(\frac{1}{\tilde{T}-2} \mathbb{V}\text{ar}_{sum} \right)}
\end{aligned} \tag{17}$$

Algorithm 1 Resampling target series nowcasts and error margin

$q' \leftarrow \text{floor}((t')/6.5)$ ▷ No. of quarters at start nowcast window
$Q \leftarrow \text{floor}((T)/6.5)$ ▷ No. of quarters at end nowcast window
$t \leftarrow t' + 7$
for q in $q' : Q$ **do**
    if q is even **then**
        $\hat{y}_q \leftarrow \hat{y}_t$
        $\sigma(\hat{y}_q) \leftarrow \sigma(\hat{y}_t)$
        $t \leftarrow t + 6$
    else if q is odd **then**
        $\hat{y}_q \leftarrow \frac{1}{2}(\hat{y}_{t+1} + \hat{y}_t)$
        $\sigma(\hat{y}_q) \leftarrow \frac{1}{2}(\sigma(\hat{y}_{t+1}) + \sigma(\hat{y}_t))$
        $t \leftarrow t + 7$
    end if
end for

It compares the variance of the nowcasts and the published target series. Note however that SN does not publish an uncertainty measure corresponding to published figures. Therefore, a decision is made to propagate margins for y^{τ}_{jt}, y^{ς}_{jt}, x^{τ}_{kt} and x^{ς}_{kt}, where the index k runs over $1, \ldots, N$, which is more realistic than no uncertainty at all. A large χ^2 measure indicates that differences between the published target series and nowcasts are too large to be explained by the propagated margins of the auxiliary series. Given that these margins are approximations of the measurement uncertainty variances of several time series, the usual interpretation of the reduced χ^2 as a probability of exceedance goodness-of-fit measure is somewhat compromised. A value of up to roughly 4 or 5 in this reduced χ^2 may well arise due to these issues. This means that at present, only values well in excess of 4 would indicate that essential information is still missing in the set of auxiliary time series. Publication of robust estimates of the measurement uncertainty variance of all time series could improve this situation in future.

It is common to compute growth rates of macroeconomic variables, besides the unit level measurements, and measure the performance based on that. Therefore it is useful to describe the relation between the quarterly target series nowcasts and the different growth rates that can be constructed. For example, SN publishes seasonally adjusted quarter-on-quarter growth rates of macro economic indicators, which is to be replicated by taking the trend+noise component nowcast, while neglecting the seasonal component, and calculating the following:

$$\hat{g}^{\tau}_{jq} = \frac{\hat{y}^{\tau}_{jq} - \hat{y}^{\tau}_{jq-1}}{\hat{y}^{\tau}_{jq-1}} = \frac{\hat{y}^{\tau}_{jq}}{\hat{y}^{\tau}_{jq-1}} - 1. \tag{18}$$

The variance of (18) can be approximated with a first-order Taylor approximation as:

$$\begin{aligned} \mathbb{V}\text{ar}\left(\hat{g}^{\tau}_{jq}\right) &= \mathbb{V}\text{ar}\left(\frac{\hat{y}^{\tau}_{jq}}{\hat{y}^{\tau}_{jq-1}} - 1\right) \\ &\approx \left(\frac{y^{\tau^2}_{jq}}{y^{\tau^2}_{jq-1}}\right)\left(\frac{\mathbb{V}\text{ar}(\hat{y}^{\tau}_{jq})}{y^{\tau^2}_{jq}} + \frac{\mathbb{V}\text{ar}(\hat{y}^{\tau}_{jq-1})}{y^{\tau^2}_{jq-1}}\right). \end{aligned} \tag{19}$$

Furthermore, growth rates of the original quarterly series (containing both the trend+noise and seasonal component) are published compared to the same quarter of the previous year, which are obtained as follows:

$$\hat{g}_{jq} = \frac{\hat{y}_{jq} - \hat{y}_{jq-4}}{\hat{y}_{jq-4}} = \frac{\hat{y}_{jq}}{\hat{y}_{jq-4}} - 1, \tag{20}$$

with corresponding variance approximation based on a first-order Taylor linearisation,

$$\begin{aligned} \mathbb{V}\text{ar}\left(\hat{g}_{jq}\right) &= \mathbb{V}\text{ar}\left(\frac{\hat{y}_{jq}}{\hat{y}_{jq-4}} - 1\right) \\ &\approx \left(\frac{y^2_{jq}}{y^2_{jq-4}}\right)\left(\frac{\mathbb{V}\text{ar}(\hat{y}_{jq})}{y^2_{jq}} + \frac{\mathbb{V}\text{ar}(\hat{y}_{jq-4})}{y^2_{jq-4}}\right). \end{aligned} \tag{21}$$

Both these growth rates can be used in measuring the performance of the proposed estimation procedure as well, by plugging them into equations (14) and (15).

3 Data

The target variable GDP can be determined by three types of approaches: the income, expenditure and production approach. In this paper we model the four series that make up the production approach, i.e.:

$$\begin{aligned}\text{GDP} &= \text{(Agriculture \& Industry Goods)} + \text{(Commercial Services)} \\ &\quad + \text{(Non-Commercial Services)} + \text{(Product Taxes - Product Subsidies)}\end{aligned} \tag{22}$$

This is the first layer of the production approach as defined by SN. As a result, the target variable y_t is a vector that contains the series of these four production categories of GDP. An overview of the target series is shown in Table E1 in Appendix E. Furthermore, it shows a complete overview of the auxiliary time series used in this study. Care has been taken when selecting the variables. They should have some predictive power on the variables of interest, i.e. some correlation should exist, in order to contribute to this nowcasting procedure rather than hurting it. Moreover, the timeliness of these variables is also important to be able to exploit this predictive power. This means that variables with a yearly cadence are excluded, since the variables of interest here are published quarterly.

3.1 Exploratory Data Analysis

The observations in the dataset run between 1-1-2000 and 1-7-2022. However, a subset is selected between 1-1-2001 and 1-1-2020 for preprocessing purposes, which is discussed in the following subsection. This means that the auxiliary time series $\{x_t\}_{t=1}^{T}$ can be gathered in a (497×139) matrix consisting of $T = 497$ bi-weekly observations of $N = 139$ auxiliary time series. Similarly, the target series data set is a (497×4) matrix containing the $J = 4$ production series that make up GDP. As mentioned before, GDP is not part of the linear regression to obtain target series estimates in the final step. Nowcasts are obtained simply by aggregating the estimates of the target series of the production approach. Figure 3.1 shows the quarterly values of the national accounts production series and GDP published by SN, in unit measurements and quarterly year-on-year growth expressed as percentage. Furthermore, all series show a (small)

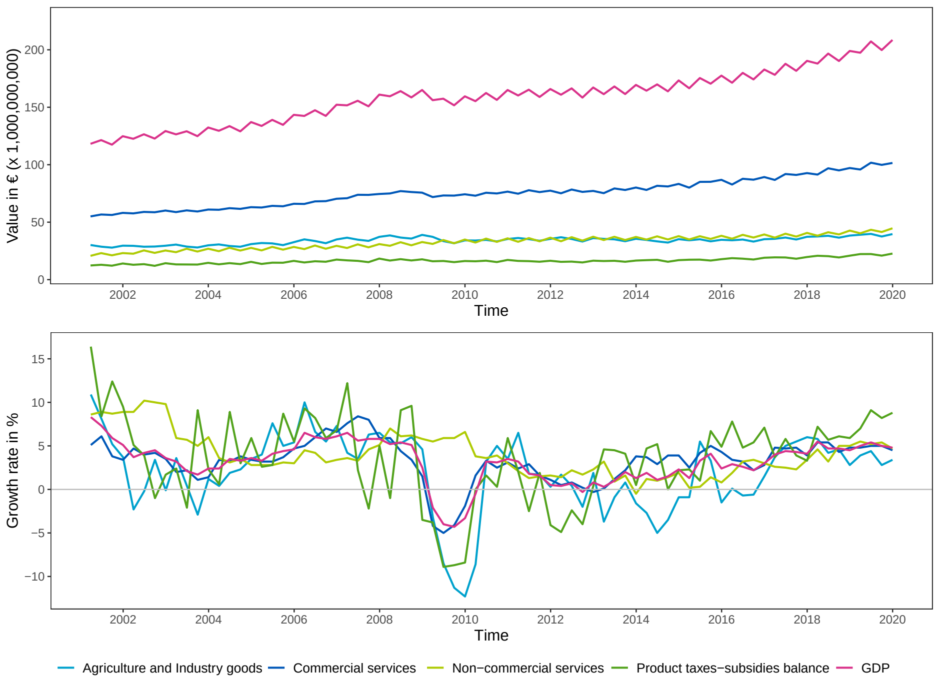


Figure 3.1 Unit level measurements (top) and growth rates (bottom) of the quarterly target time series over the selected time window (2001-2020).

upward trend, apart from the production of goods in agriculture and industry. This is a general development within service oriented economies of developed countries. Another remarkable observation is that the production of non-commercial services is the only variable that keeps growing during the financial crisis of 2008-2010. One could think of healthcare, education and government as major parts within Non-Commercial Services, which are sectors that are generally unaffected by the business cycle. As expected, it is clear that the Commercial Services sector is the biggest contributor to the added-value of the knowledge-based Dutch economy. Apart from the trend, Figure 3.1 also shows some seasonality present in the series. It is most persistent in the Non-Commercial Services sector, but is present in all of the target variables.

To investigate whether other selected auxiliary time series are indeed correlated with the target time series, their correlations are computed, see Table E2. Table 3.1 summarizes it by reporting the share of auxiliary time series that have a correlation larger than 0.1 in absolute value. Since it is well above 80% for all target series, it shows that the set of auxiliary variables is indeed suited for nowcasting these production series.

Table 3.1 The share of auxiliary variables that show a correlation of at least 0.1 in absolute value per target variable.

Production agriculture and industry	82.73%
Production commercial services	89.93%
Production non-commercial services	84.89%
Product taxes/subsidies balance	89.93%
GDP	90.65%

3.2 Preprocessing

The methods described in Chapter 2 require a substantial amount of preprocessing steps on the selected data set before they can be applied. Initially, the cubic spline interpolation technique is applied individually on all 139 auxiliary time series and 4 target time series. Each series is resampled from the resulting smooth continuous function at a bi-weekly cadence. The AEX-index is the only series with an original daily cadence (see Table E1), hence no interpolation is required for bi-weekly resampling purposes.

Once all time series are resampled the Fourier transform-based linear filter is applied, again on each series individually. It allows extracting the trend+noise component and seasonal component and storing them separately in the matrices $X^\tau, X^\varsigma, Y^\tau$ and Y^ς. Recall that the time series are trimmed to the time domain of 1-1-2001 to 1-1-2020. The main reason for this is because the linear filter that is used to decompose the time series is inappropriate at the edges since it would need non-existent measurements. Therefore, the edges of the decomposed parts of each y_{jt} and x_{kt} are dropped.

Recall that Bai and Ng (2004) state that factor analysis can be applied on both I(0) and I(1) time series. Consequently, before the real-time analysis is started, three well-known integration order/unit root tests are applied on the bi-weekly auxiliary series (Dickey and Fuller 1979; Kwiatkowski et al. 1992; Phillips and Perron 1988). It is understood that these tests are applied to the original series, not to the decomposed parts X^τ and X^ς. An overview is given in Table 3.2. Since Table 3.2 only shows 9 time series that have an integration order above I(1), while being the result of only one of the three different tests, the conclusion is that these auxiliary series are

Table 3.2 Integration order test results of the different tests on the auxiliary time series.

	I(0)	I(1)	I(2)
ADF	89	50	0
KPSS	11	119	9
PP	57	82	0

eligible for the proposed estimation procedure without the need of any other alterations like differencing or log transformations.

Subsequently, the cubic spline method serves another preprocessing purpose. It is used, together with a linear prediction algorithm to extrapolate individual time series that exhibit backward or forward jagged edges respectively. The former occurs when the initial publishing date at SN or another data source differ between time series, while the latter arises when their publication lag differ. This leads to differences in the length of the time dimension, hence jagged edges when stacking these time series in the matrix X^i. PCA is based on the eigendecomposition of $X^{i\prime}X^i$ in order to compute the principal components of X^i. The eigendecomposition requires a square matrix, which is achieved by extrapolating the jagged edges in X^i. Moreover, all auxiliary time series are centered and re-scaled. These are the final preprocessing steps after which PCA can be performed. Figure 3.2 shows Unemployment as an example auxiliary time series when all preprocessing steps are done. Note that the target series in Y^i are not preprocessed, such that the regression of the target series on the factor estimates result in target series estimates that can be compared directly with the unit measurements or growth rates published by SN.

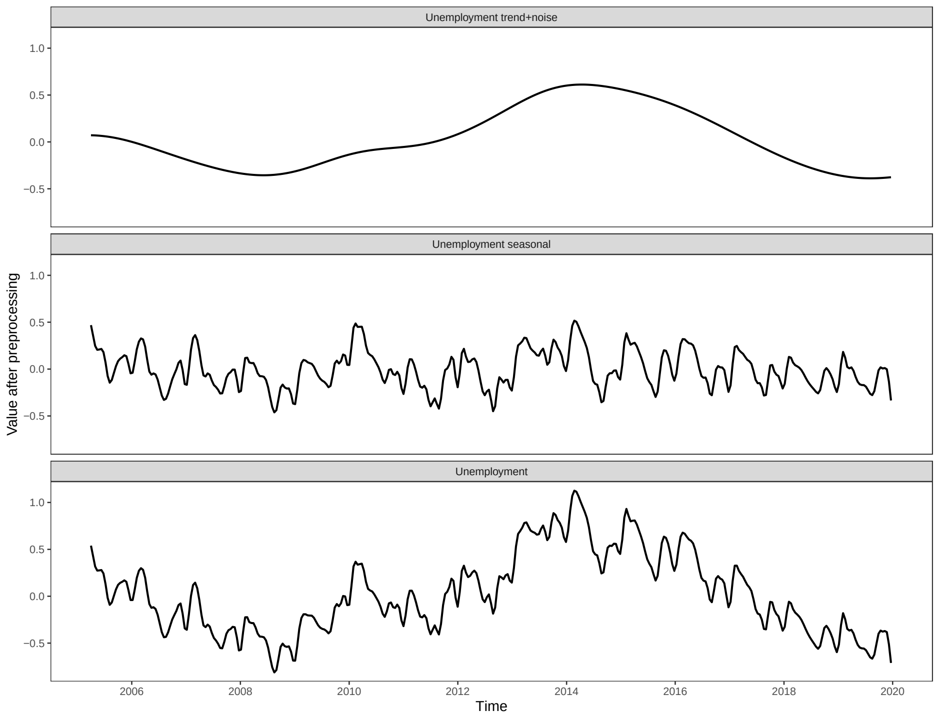


Figure 3.2 An example auxiliary time series after preprocessing. The cadence is bi-weekly. The trend+noise (top) and seasonal (middle) components are extracted from the original series (bottom). The series is centered around zero and rescaled to a range between -1 and 1.

4 Empirical Results

The results of intermediate steps within the estimation procedure are outlined in sections 4.1 and 4.2. These steps are performed repeatedly in a real-time simulation, creating intermediate results in each iteration. It is not feasible to demonstrate these results for all 184 iterations. Consequently, intermediate results corresponding to the final real-time iteration are presented, when the time dimension is increased up until the start of 2020, as an example. However, the factor selection tests and PCA are only re-executed at the start of each calendar year, meaning their results are occasionally presented over the simulated time instead. It is made clear when that is the case. In all other cases, the intermediate results refer to the results of the last iteration.

4.1 Principal Component Analysis

The first step in the latent factor estimation procedure is performing PCA on the matrix of auxiliary time series X^i. The resulting eigenvalues and eigenvectors play an important role in this section. Table 4.1 shows the first 20 eigenvalues and the proportion of the variation in X^i that is captured by each corresponding principal component. This proportion is determined by dividing each eigenvalue over the sum of all eigenvalues. It gives an initial indication of how many components one can neglect without losing much information. Furthermore, it indicates that the seasonal patterns require more principal components to explain the same variation compared to its trend+noise counterpart. The first trend+noise principal component explains 48% of the comovement in the set of auxiliary variables, while this is merely 12% for the first seasonal principal component.

Table 4.1 Proportion of the variance explained by the first r^i principal components.

	Trend+noise			Seasonal		
r^i	Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative
1	2081.18	0.4761	0.4761	177.02	0.1242	0.1242
2	1108.03	0.2535	0.7295	162.30	0.1138	0.2380
3	464.15	0.1062	0.8357	138.00	0.0975	0.3355
4	225.93	0.0517	0.8874	107.87	0.0757	0.4111
5	197.40	0.0452	0.9326	92.11	0.0646	0.4757
6	115.27	0.0264	0.9589	74.74	0.0524	0.5282
7	65.33	0.0149	0.9739	49.23	0.0345	0.5627
8	37.63	0.0086	0.9825	42.76	0.0300	0.5927
9	31.36	0.0072	0.9896	40.90	0.0287	0.6214
10	18.69	0.0043	0.9939	35.79	0.0251	0.6465
11	11.15	0.0026	0.9965	30.26	0.0212	0.6677
12	5.14	0.0012	0.9977	27.30	0.0191	0.6868
13	3.23	0.0007	0.9984	25.93	0.0182	0.7050
14	2.84	0.0006	0.9990	25.08	0.0176	0.7226
15	1.71	0.0004	0.9994	22.11	0.0155	0.7381
16	1.19	0.0003	0.9997	21.98	0.0154	0.7535
17	0.48	0.0001	0.9998	20.94	0.0147	0.7682
18	0.23	0.0001	0.9999	17.84	0.0125	0.7807
19	0.19	0.0000	0.9999	16.41	0.0115	0.7922
20	0.14	0.0000	0.9999	15.93	0.0112	0.8034

Gaining this insight is one of the benefits of splitting the time series in separate parts, using the Fourier transform. It allows for a more tailor-made procedure which seems preferable, as the results in Table 4.1 raise the suspicion that the optimal r^i, i.e. amount of principal components to retain, clearly differs between the trend+noise and seasonal component.

4.1.1 Factor selection

One of the main advantages of PCA is the dimension reduction it entails. This requires a decision on how many principal components are kept. The non graphical solutions based on the scree plot proposed by Raîche et al. (2013) are applied to answer this question. Figure 4.1 shows the scree plot and how the ordered eigenvalues visually relate to the optimal component solution found by the different tests. The elbow of the scree plot seems to evolve more gradually and lie further away from the y-axis for the seasonal series. This is confirmed by the generally higher test results stated in the legend.

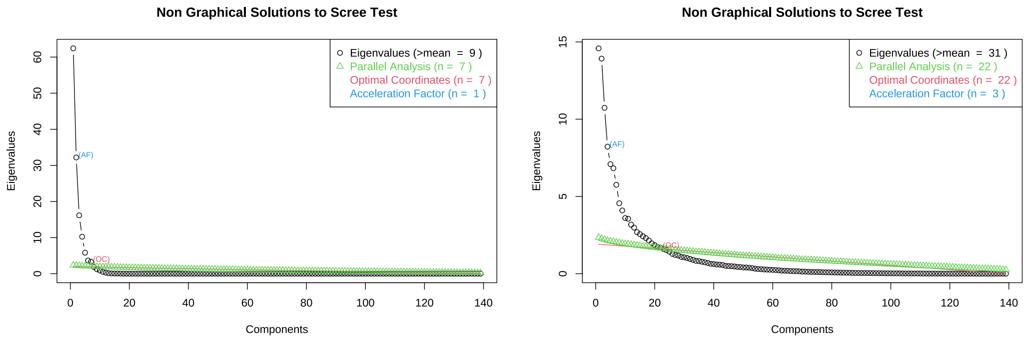


Figure 4.1 Optimal number of factors to retain, according to the four tests, based on the trend+noise series (left) and seasonal series (right).

Table 4.2 presents these test results over time. Note that an extra iteration of PCA and factor selection is added at the start of the second quarter in 2014. This is done to incorporate initial observations of the important auxiliary (European) import and export time series, for which only backward extrapolated values existed before that starting date.

The optimal r^i deviates slightly when more data has become available. No drastic deviations are observed, confirming the expectation that re-determining the optimal r^i is not necessary during *each* real-time iteration. On average over the entire real-time window, $r^{\tau} = 7$ and $r^{\varsigma} = 23$. Since the differences in Table 4.2 are marginal, it is decided to take these total averages and keep them fixed in each iteration throughout the rest of the analysis for simplicity. Note that the outcome is in line with the presumption that more principal components should be kept for the seasonal component than for the trend+noise component. Finally, recall that the principal components serve as initialization when extracting the unobserved common factors of the DFM using the Kalman smoother in the following step. Hence, the results in Table 4.2 determine both the optimal principal component amount r^i as well as the length of the vector of common factors f_t^i in the DFM equations 4 and 5. Moreover, the obtained eigenvectors act as estimate of the factor loading matrix Λ^i.

4.1.2 Factor interpretation

The factor loadings in Λ^i can be used to derive a certain economic interpretation of the latent factors f_t^i. The interpretation of the factors is also facilitated by the fact that the preprocessing of the auxiliary variables did not require differencing or log transformations of any kind. The factor

Table 4.2 The results of the non-graphical scree tests throughout the real-time analysis.

	Trend+noise				Seasonal			
	Kaiser	Parallel	OC	$r^{\mathcal{T}}$	Kaiser	Parallel	OC	$r^{\mathcal{S}}$
2013Q1	7	6	6	6	24	16	16	19
2014Q1	7	7	7	7	27	19	3	22
2014Q2	7	7	7	7	27	19	19	22
2015Q1	7	7	7	7	28	20	20	23
2016Q1	8	6	6	7	29	19	19	22
2017Q1	8	7	7	7	29	19	19	23
2018Q1	9	7	7	8	29	21	21	24
2019Q1	9	7	7	8	29	21	21	24
2020Q1	9	8	8	8	29	20	18	25

loadings explain how the auxiliary variables are related to these underlying common factors. To make this clear, all eigenvectors are re-scaled to the interval $[-1, 1]$, dividing each element by the largest element in absolute value. The re-scaled values in the first three eigenvectors are selected and analyzed in more detail. Figure 4.2 presents the re-scaled factor loadings, for each of the 139 auxiliary time series. If the re-scaled factor loading value is close to zero, the corresponding time series does not contribute much to the establishment of that factor, while a value close to one in absolute value indicates strong similarities between the factor and the auxiliary series. A negative value indicates their movement is contrary to each other, i.e. they are negatively correlated. The figure gives the impression that the first principal component captures

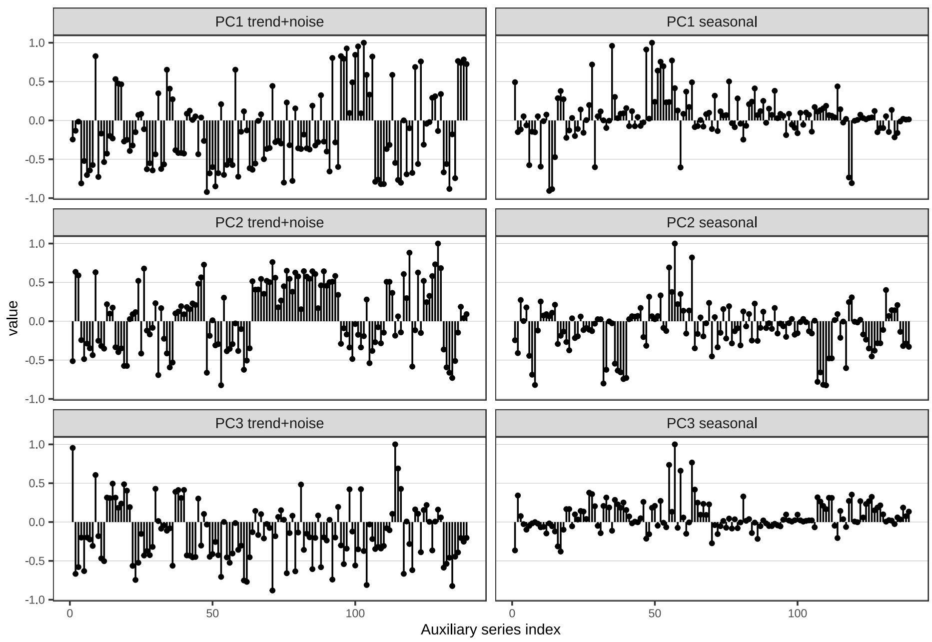


Figure 4.2 Re-scaled factor loading values showing which auxiliary time series contributes the most to a principal component. The first 3 PC's are visualized as an example.

a general trend within the data, since the factor loading values are relatively high in absolute value across all auxiliary variables. To explore this impression more carefully, the top 10% contributing series are selected for the first three trend+noise and seasonal principal components. Table 4.3 gives an overview of the series that contribute the most to a trend+noise principal component (PC), while Table 4.4 is the seasonal PC equivalent. Finally, Figure 4.3 helps interpreting these principal components by plotting them over time.

Table 4.3 Top 10% contributors to the first three trend+noise principal components.

PC1 trend+noise	PC2 trend+noise	PC3 trend+noise
PPI sales price electricity	Consumptive credit taken by households	Unemployment benefits
PPI usage price electricity, gas, heating	Catering revenue volume index	Unemployment
PPI sales price electricity, gas, heating	Gov. income other	Industry sales prices bal. next 3 mths
Gov. debt	Industry sales prices bal. next 3 mths	Securities MFIs EU
Securities non-MFI fin. institutions EU	Loans to households	PPI sales price coal
Gov. income social charges	Securities MFIs EU	Gov. spending other capital
PPI usage price industrial production	Gov. income/expenditures balance	Gov. spending other
PPI sales price mineral extraction	Bankruptcies	Housing price index
CPI durable goods	Borrowing for housing by households	Other sales prices bal. next 3 mths
PPI sales price gas	Building permits residential	Gov. income other
Export value index	Gov. debt	Social welfare benefits
Export volume index	Securities non-MFI fin. institutions EU	Vacancies total
Self-employed	Transport sales prices bal. next 3 mths	Construction added value index
Other sales prices bal. next 3 mths	Real-estate purchase orders bal. next 3 mths	Transport sales prices bal. next 3 mths

Table 4.4 Top 10% contributors to the first three seasonal principal components.

PC1 seasonal	PC2 seasonal	PC3 seasonal
Gov. income total	Gov. spending wages	Gov. spending wages
Mineral extraction revenue	Export volume index	Gov. consumption
Gov. income/expenditures balance	Gov. consumption	Gov. spending total
Gas price, consumption < 20 GJ	CPI industrial goods	Gov. spending social benefits
Gas price, consumption 20-200 GJ	Export value index	Cons. confidence, econ. situation past 12 mths
Catering revenue volume index	Industrial production	Gas price, consumption 100-1000 TJ
Gov. spending intermediate goods and services	Import value index	Consumption households index
Gov. income production activities	Import goods EU	Unemployment
Catering revenue value index	Export goods EU	Consumption households goods
Consumption households goods	Gov. spending total	Catering revenue volume index
Gov. income other	CPI goods	Vacancies total
Gov. income social charges	Import volume index	Real-estate sales prices bal. next 3 mths
Gov. spending social benefits	Export goods total	CLI NL
Consumption households services	Import goods total	Import value index

The top contributing auxiliary variables for the first trend+noise principal component confirm the belief that this component captures a general economic trend. It contains a mixture of different types of series, e.g. it exhibits international trade series, price indices, financial series from the ECB and government budget series. Generally, macroeconomic time series exhibit a steady growing trend over time. This is confirmed by Figure 4.3 even though it shows a decreasing trend at PC 1. This is a general feature of PCA, as the eigenvectors are defined up to a sign. Recall from Figure 4.2 that the majority of the auxiliary series are indeed negatively linked to this first trend+noise PC. Furthermore, the second trend+noise principal component seems to capture a financial trend. A large share of the top contributors consists of financial series from the ECB. Most of these series are positively linked to the principal component, indicating the component summarizes the financial downfall and recovery in the years during and after the financial and euro crises. The graph of the third principal component shows a more cyclical trend however. It induces the impression that this component captures the business cycle, which describes the alternating phases of economic growth, i.e. expansions versus recessions. The corresponding contributors confirm this impression, since it contains variables related to the job and real-estate markets. Variables such as unemployment, vacancies or housing prices are known of being sensitive to fluctuations in economic growth, which have a lagged impact on these variables. Again vacancies and housing prices are negatively correlated with PC 3, while unemployment is positively correlated. This fits the expectation that the third trend+noise principal component captures the business cycle progression. The further one goes, the harder it gets to find a clear distinct intuition behind the observed principal components, as they reflect an orthogonal summary of the cohesion between all N time series. That does not mean those components are unusable in the analysis however, but it is why only a selection of the first couple of PCs are presented and interpreted here.

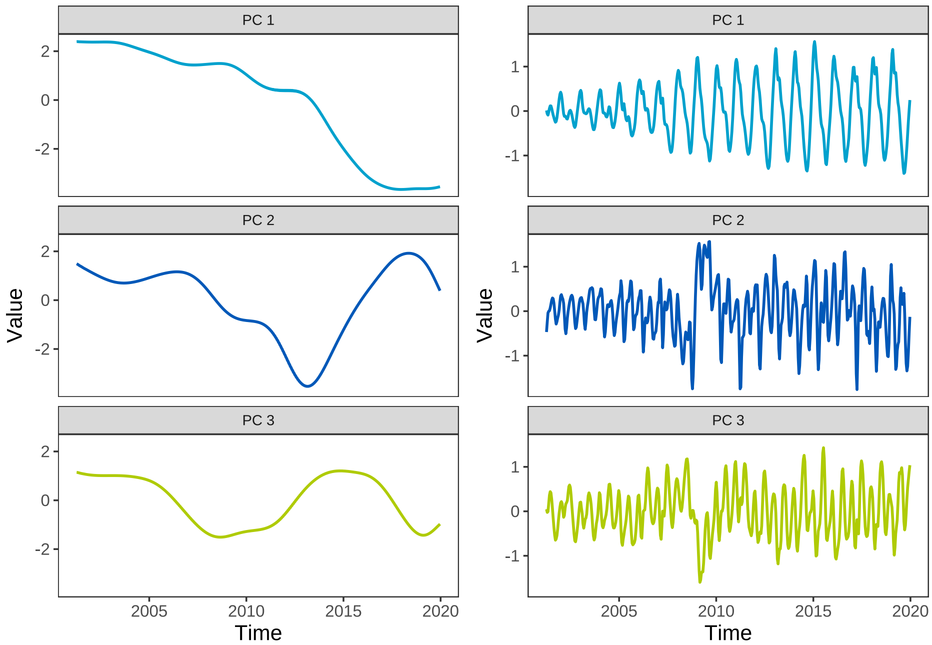


Figure 4.3 The first three principal components over time, obtained from both the trend+noise and seasonal PCA.

For the seasonal component it is harder to identify such economic intuitive interpretations. Nevertheless, some intuition can be gleaned on these principal components based on the time series in Table 4.4. The first component seems to capture a general seasonal pattern, since the first column contains another mixture of variables, such as government, gas, catering and consumption series. In the second column there exists a clear group of variables that dominates the list, namely the import/export series. Furthermore, the majority of the series correspond to goods rather than services. The third principal component contains government, job market and consumption as dominating contributors.

4.2 Kalman smoother

After PCA the Kalman smoother is applied as a second step by treating the principal components as the latent factors f_t^i in model (12) and updating them using (4) and (5) as observation and state equation respectively. Figure 4.4 shows the PCA and Kalman smoother estimates of f_{1t}^{τ} as an example. Note that PCA produces a good initial estimate, as it is quite similar to the Kalman smoother estimate. As expected, the trend+noise factors seem to follow a random walk process, which is confirmed by their estimated VAR coefficient matrix:

$$\hat{\Phi}^{\tau} = \begin{bmatrix} 1.002 & 0.005 & -0.006 & 0.008 & 0.000 & -0.006 & 0.005 \\ -0.007 & 0.999 & 0.020 & -0.021 & 0.000 & 0.014 & -0.008 \\ 0.002 & -0.011 & 0.999 & 0.011 & 0.008 & -0.002 & -0.008 \\ -0.004 & 0.006 & -0.009 & 1.004 & 0.016 & -0.006 & 0.028 \\ 0.000 & -0.001 & -0.004 & -0.016 & 0.999 & -0.029 & -0.027 \\ 0.002 & 0.001 & 0.004 & 0.002 & 0.020 & 0.990 & -0.024 \\ 0.000 & -0.001 & 0.000 & -0.011 & 0.009 & 0.022 & 0.996 \end{bmatrix}$$

The diagonal elements are close to 1 in absolute value, while the off-diagonal elements are relatively small. Recall equation (5). If $\hat{\Phi}^{\tau}$ is close to an identity matrix it indicates that the trend+noise factors are an accumulation of random shocks over time, hence a random walk.

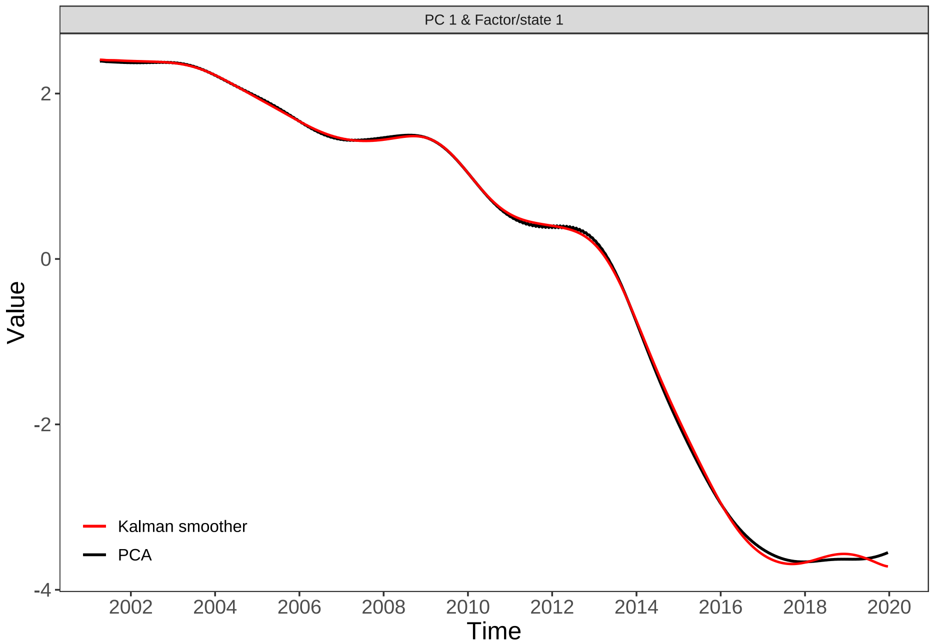


Figure 4.4 The first trend+noise principal component and factor (also referred to as state in state space formulation), retrieved by PCA and the Kalman smoother respectively.

The latent factors for the seasonal series, obtained with PCA, are smoothed with the Kalman smoother as well. For the seasonal factors however, their dynamic progression is expected to be generated by a stationary process. Recall Figure 3.2 from Chapter 3. Contrary to the trend

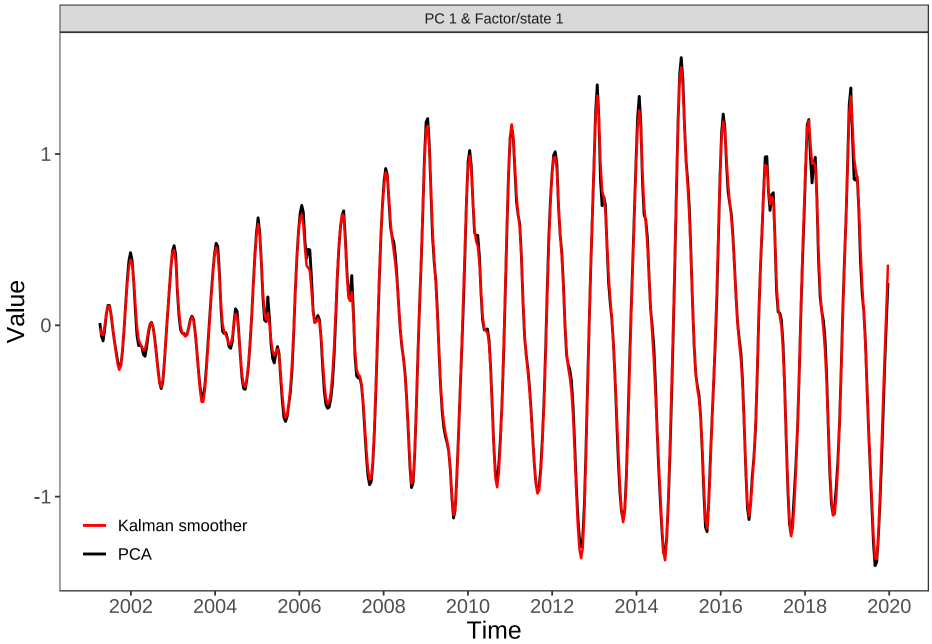


Figure 4.5 The first seasonal principal component and factor (also referred to as state in state space formulation), retrieved by PCA and the Kalman smoother respectively.

component within a time series, the seasonal pattern itself is reoccurring with a constant mean and variance over time. However visually, Figure 4.5 indicates that the variance might be time-varying for f^S_{1t}. Nevertheless, equation (B.1) in Appendix B shows a subset of $\hat{\Phi}^S$ by focusing on the first twelve factors and it clearly shows that the diagonal elements are smaller than 1 in absolute value, including the element in the top-left most corner of the matrix shown in (B.1), confirming the stationarity expectation of the seasonal factors. Note that this pattern holds for all 23 seasonal factors but the remaining estimates are not presented here for readability purposes.

Again, the initial PCA estimate seems very accurate already. One should keep in mind however that the difference between PCA and the Kalman smoother increases for higher factors. This is made clear in Figures B1 and B2 in Appendix B , which graph the PCA and Kalman smoother estimates of the last factors instead of the first. As expected, the estimates deviate more in these plots. Higher factors represent a smaller portion of the variance in the data, hence they exhibit more erratic behavior over time. Another way of looking at this is that as the eigenvalues become smaller, the information contained in the associated principal component becomes smaller. The behaviour over time must therefore be more and more dominated by random variations. This makes it harder to follow for any estimation algorithm, like PCA or the Kalman smoother, which explains the bigger difference in their predictions.

Recall (8) and note that the maximum likelihood estimation of θ^i_2, like PCA and the optimal factor tests, is only re-executed once every year in the real-time analysis. Reason for this is that it is computationally more costly than the least squares method. However, MLE is more robust compared to LS when it is not guaranteed that the factor innovations are normally distributed, hence it is selected nonetheless. The R package 'KFAS' is used, see Helske (2017) for a comprehensive discussion on this package, in the state space modelling part. The likelihood function is numerically optimized using the BFGS optimization algorithm. Figure B3 in Appendix B shows these maximum likelihood estimates over the simulated time. Again, no drastic deviations between the years are observed, indicating this approach is stable in time.

5 Nowcasting

First, the nowcasts for the individual series of the four production categories that make up GDP are analysed, and together with GDP are presented in Figures 5.1 - 5.5. Nowcasts reflect point estimates at time T, retrieved directly as the fitted values of the final linear regression of the target series on the re-estimated factors. The sum of the squared residuals of this regression is used to determine the variance of these point estimates. Different ways of relating the target series to the estimated factors, possibly using the factor variances produced by the Kalman smoother to create nowcast variances, are left for further research. The point estimates are denoted with a red dot. Their error margin is denoted with an error bar, reflecting a deviation of one standard error. The black line indicates the quarterly target series directly obtained from the database of SN (statline). The blue line represents the same figures, but after upsampling, filtering, merging and downsampling again. The series can exhibit slight deviations between them due to these operations, but the figures show nothing alarming. It would impose an

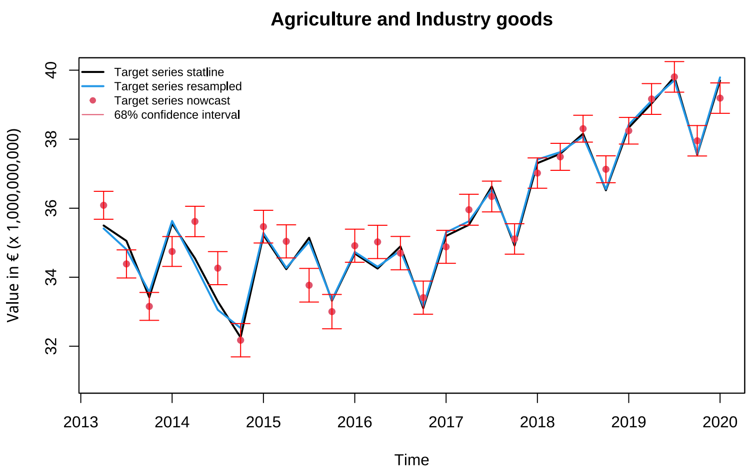


Figure 5.1 Nowcasts and realized figures of the first target series in unit measurements.

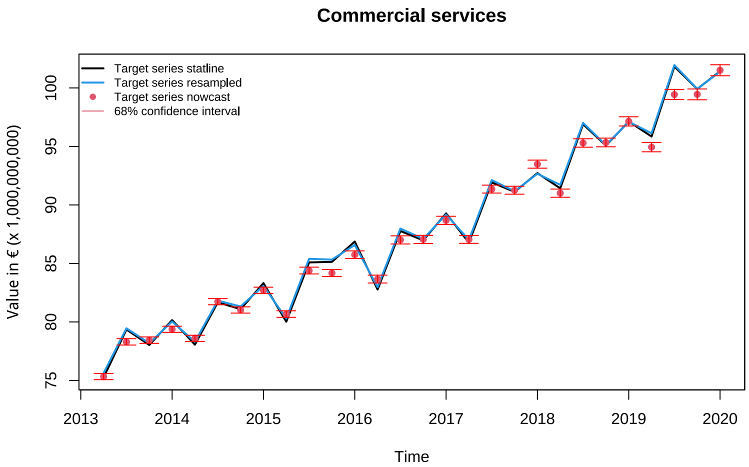


Figure 5.2 Nowcasts and realized figures of the second target series in unit measurements.

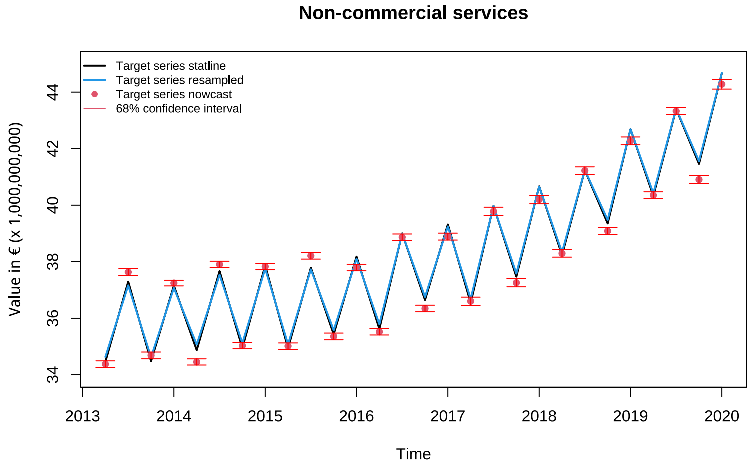


Figure 5.3 Nowcasts and realized figures of the third target series in unit measurements.

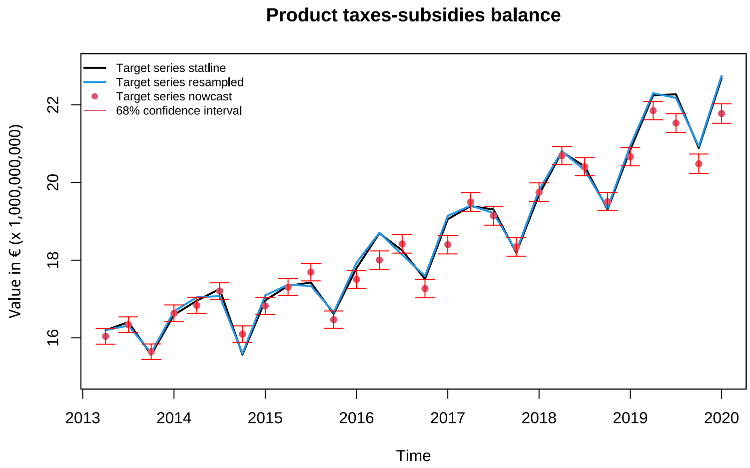


Figure 5.4 Nowcasts and realized figures of the fourth target series in unit measurements.

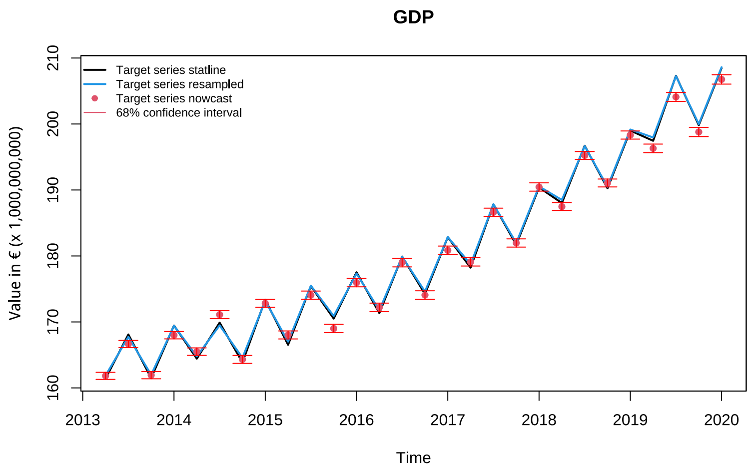


Figure 5.5 Nowcasts and realized figures of the fifth target series (GDP) in unit measurements.

Table 5.1 Performance measures for accuracy and bias, in level and relative to statline unit measurements. The RMSFE and ME in level are published in € (x1,000,000,000).

	RMSFE	ME	RMSFE rel.	ME rel.
Agr. & Ind. Goods	0.55	0.06	1.57%	0.17%
Commerc. Services	0.80	-0.31	0.88%	-0.33%
Non-Commerc. Services	0.27	-0.11	0.7%	-0.26%
Product Tax/Subidies bal.	0.35	-0.14	1.76%	-0.66%
GDP	1.25	-0.50	0.68%	-0.25%

artificial bias in the estimation procedure if they deviate a lot, because the resampled bi-weekly target series are used in the linear regression on the factor estimates but in the end the quarterly statline equivalents are used in the performance measures.

Overall, the proposed nowcasting procedure seems to perform satisfactory when visually analyzing the consecutive graphs. Surprisingly, the uncertainty surrounding the nowcasts is bigger for the Agriculture and Industry Goods time series compared to the rest. Furthermore, a persistent underestimation is observed in 2019 in Figures 5.2, 5.4 and 5.5. The most plausible cause for this is a structural raise of the VAT rate in the Netherlands at the start of 2019. These kind of policy changes influencing the underlying definition of certain time series should be added manually to the model. This would be similar to adding the slope and intercept terms to the vector of factors for example, but is left for further research.

Besides the visual analysis, the performance of the procedure is measured by calculating the RMSFE and ME. Table 5.1 reports these results. Recall equations (14) and (15). $Q = 28$ since there are 28 quarters in the selected real-time time window. The first two columns are in euros, i.e. the RMSFE of GDP equals €1.25 billion. This is an improvement on the first operational approach in Kuiper and Pijpers (2020). Note that they provide the MSFE instead of the RMSFE. Moreover, keep in mind that their performance is measured in a different time window, so a one-to-one comparison is not completely accurate but still insightful. The third and fourth column reflect the same measures relative to the statline unit measurements, expressed in percentages. While GDP has the largest RMSFE in absolute terms, it has the smallest relative RSMFE. It seems that the Product Taxes-Subsidies balance is the hardest time series to nowcast for the present model with a relative RMSFE of 1.76%. That said, part of this should be explained by the structural change of the VAT rate. Table C1 in Appendix C presents the results when 2019 is excluded from the time window, to get an indication on the potential gain in performance when correcting for this issue. It shows large improvements in the performance for all target series apart from the Agriculture and Industry Goods. Besides, Table 5.1 shows that the ME is about 2-10 times smaller than the RMSFE, indicating there is hardly any bias present in the nowcasts.

The growth rate nowcasts are obtained by differencing the fitted values of the final linear regression, instead of nowcasting them directly. Due to the uninterpretable outcome of the integration order tests on the decomposed parts, taking differences of these separate parts is avoided within the present model, preventing directly estimating target series growth rates. As mentioned before, this could be investigated in further research. Generally, statistical agencies publish quarterly growth rates compared to the same quarter of the previous year and sometimes also compared to the previous quarter. Seasonal effects will interfere with the latter however, hence they are typically based on seasonally adjusted time series of the production categories that compose GDP. Note that the National Accounts team at SN uses a different method of adjusting a time series for seasonal effects than proposed here in Section 2.1, see Bikker et al. (2019). This difference causes an artificial bias in the nowcasts when comparing them to the seasonally adjusted figures on statline.

Nevertheless, the present trend+noise nowcasts are compared with the quarter-on-quarter growth rates, see Figures D1 and D2 in Appendix D. Note that the blue and black line deviate a lot. The filter removes all fluctuations from a time series, leaving a smooth trend line, while the seasonal adjusted series published by SN still fluctuates, as shown in Figure D1. The deviation in 2019 will again be partly explained by the raise of the VAT rate, but the current seasonal adjustment method in use at SN produces preliminary figures during the most recent years, that will require quite some revisions when more information has become available. This is a downside of the current method and is present to a lesser extent in the proposed method. However, the National Accounts team are bound to the Eurostat criteria for seasonal adjustments, making it hard to implement a different seasonal filtering method.

The growth rates compared to the same quarter of the previous year, i.e. quarterly year-on-year growth rates, are presented in Figures 5.6 - 5.10. These growth rates are usually picked up by the media and are frequently subject of the public debate. Again the model seems to perform quite well in general. However, small deviations between the nowcast time series and the realized series in unit measurements are enlarged when differencing. Furthermore, the underestimation in 2019 persists in Figures 5.7, 5.9 and 5.10.

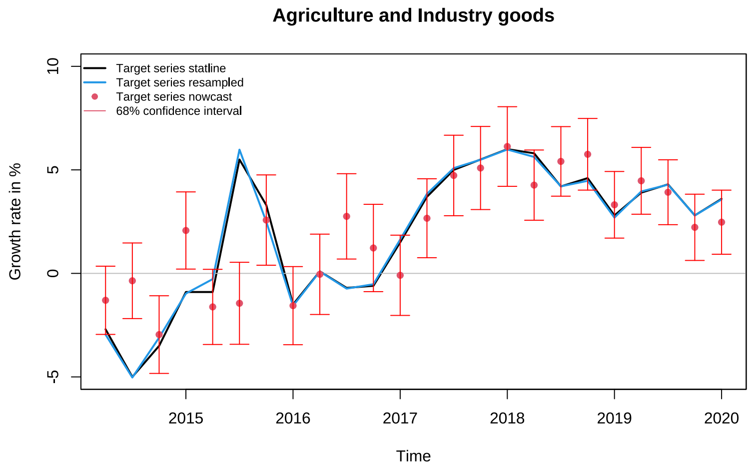


Figure 5.6 Nowcasts and realized figures of the first target series in growth rates: Agriculture and Industry Goods.

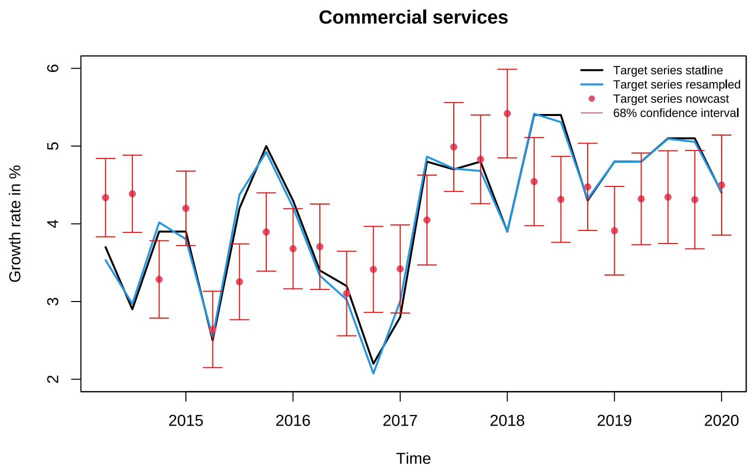


Figure 5.7 Nowcasts and realized figures of the second target series in growth rates: Commercial Services.

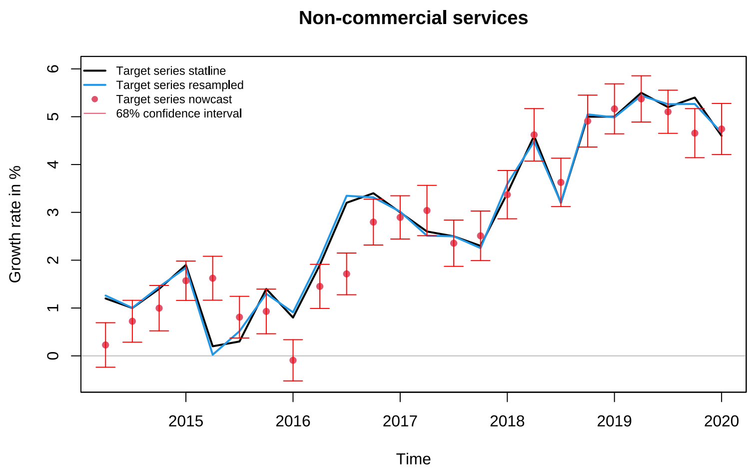


Figure 5.8 Nowcasts and realized figures of the third target series in growth rates: Non-Commercial Services.

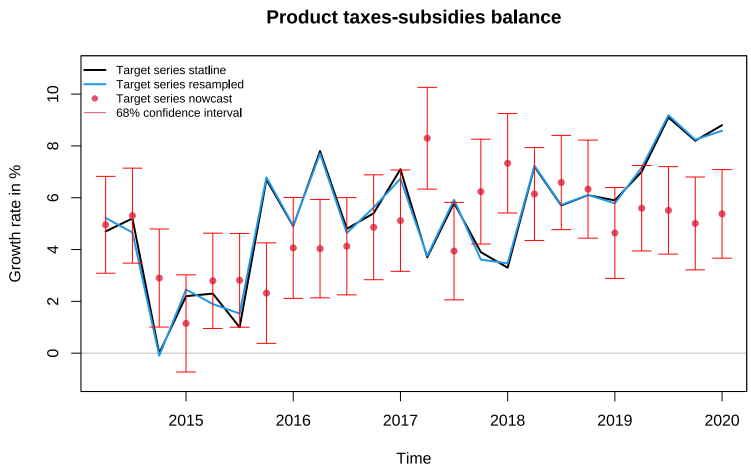


Figure 5.9 Nowcasts and realized figures of the fourth target series in growth rates: balance of Taxes and Subsidies.

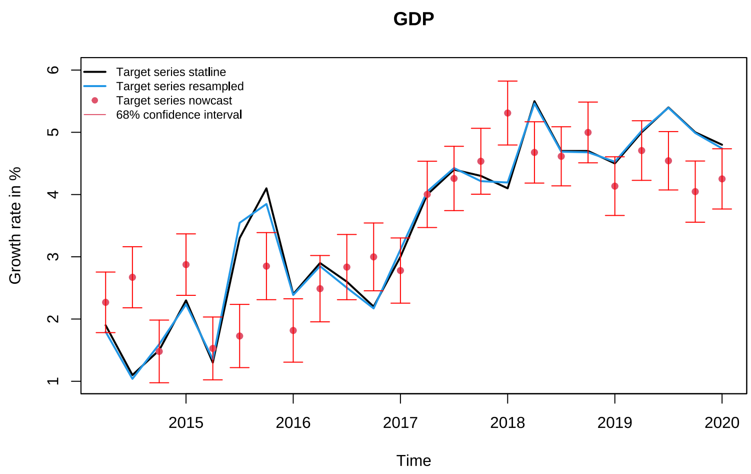


Figure 5.10 Nowcasts and realized figures of the fifth target series in growth rates: GDP.

Table 5.2 contains the performance measures for both types of growth rates, with lag 1 (q-on-q growth rates) and with lag 4 (y-on-y growth rates). Compared to the approach of Kuiper and Pijpers (2020), the performance of the present method improved for the q-on-q growth rates. Their Q1 econometric approach resembles the current q-on-q procedure the most. It is good to note once again however that the real-time window differs in both studies. Kuiper and Pijpers (2020) report an MSFE of 0.30%, which corresponds to an RMSFE of 0.55%. This is slightly higher than the RMSFE of GDP presented in Table 5.2, hence the current procedure improves upon their method. Furthermore, they mention that the currently employed flash estimate method at SN has a MSFE of 0.23%, which equals 0.48% after taking the square root, which is in turn slightly lower than the value of the data-driven alternative presented here. Keep in mind that the q-on-q growth rates suffer from both the change in VAT rate and the artificial bias caused by the difference in seasonal adjustment method, hence there is reason to believe that the present model will outperform the current method when correcting for this. Table C2 in Appendix C shows the potential gain in performance when correcting for the VAT rate problem, by comparing it with Table 5.2. Apart from GDP, which now even outperforms the flash estimate method of SN, there is no gain observed in the nowcasting performance of the q-on-q growth rates. However,

Table 5.2 Performance measures for accuracy and bias, quarterly growth rates with lag 4 and lag 1. Quarterly growth with a lag of 1 is published as seasonal adjusted time series, hence the trend+noise component nowcasts are used to calculate those performance measures, without adding the seasonal component nowcasts.

	RMSFE l=4	ME l=4	RMSFE l=1	ME l=1
Agr. & Ind. Goods	2.13%	0.12%	1.4%	0.02%
Commerc. Serv^s	0.78%	-0.09%	0.63%	-0.12%
Non-Commerc. Serv^s	0.59%	-0.16%	0.57%	-0.01%
Product Tax/Sub^s bal.	2.4%	-0.47%	2.01%	-0.13%
GDP	0.73%	-0.11%	0.52%	-0.06%

Table D1 in Appendix D presents the results when comparing the nowcasts to the the seasonally filtered target series based on Fourier analysis (blue line) instead of the seasonally adjusted series on statline (black line). This gives an indication on the potential gain in performance when there is no difference in seasonal adjustment method present, providing grounds to believe the presently proposed estimation procedure outperforms the current method in use at SN and the proposed method of Kuiper and Pijpers (2020) quite substantially.

For the quarterly y-on-y growth rates, the performance is also better compared to Kuiper and Pijpers (2020). Their Y1 econometric approach resembles the present quarterly y-on-y procedure, for which they report an RMSFE of 0.84%. Table 5.2 shows an RMSFE for GDP of 0.73%. Finally, it is good to mention that the ME is much smaller than the RMSFE for the growth rates as well in all scenarios.

The Mean Absolute Error (MAE), defined in (16), is calculated to give insight in the size of the first revision of the flash estimate using the model. The MAE is calculated for the nowcast of six bi-weekly real-time iterations ago, compared to the target series estimate of six observations before the end of the sample in the current real-time iteration. This comparison in absolute value is made relative to the nowcast of six bi-weekly real-time iterations ago, i.e. $\hat{y}_{jt-6}$, and expressed as percentage. Results are presented in Table 5.3. Furthermore, the χ^2 measure from

Table 5.3 Performance measure for the 90 days revision, in level and relative to statline unit measurements. The MAE in level is published in € (x1,000,000,000). A χ^2 measure $>> 1$ indicates that the nowcasting procedure is insufficient.

	MAE	MAE rel.	χ^2
Agriculture and Industry Goods	0.12	0.33%	1.49
Commercial Services	0.12	0.34%	5.72
Non-Commercial Services	0.08	0.23%	4.61
Product Taxes-Subsidies balance	0.07	0.18%	2.29
GDP	0.23	0.65%	4.08

equation (17) is presented. A result $>> 1$ indicates that the nowcasting procedure is insufficient, i.e. the variation in the target series is not captured by the nowcasting procedure. The visual impression from Figures 5.6 - 5.10 suggests that, except for the Commercial Services, the movement of the nowcasts is indeed similar to the target time series. This indicates that the somewhat high values for the reduced χ^2 in Table 5.3 for some series could be due to the issue that the true measurement errors of all auxiliary series are perhaps slightly underestimated at present, as alluded to in Section 2.6. For the nowcast of Commercial Services (Figure 5.7) it is possible that additionally some mechanism influencing the movement of this time series is still not fully captured by the current set of auxiliary time series. Including more PCs than the set presented here was attempted, but did not improve the situation. Overall, these results are sufficiently encouraging to conclude that the present method is suitable for nowcasting multiple target series in a data-driven way, without the need of large revisions. Currently, large revisions occur occasionally, which is part of the motivation for this current study.

6 Conclusion

A new procedure is introduced in this paper for analysing and nowcasting macroeconomic time series. Decomposing a time series in its trend+noise and seasonal component, and modelling them separately allowed estimating seasonally adjusted time series and original time series simultaneously. Using a two-step estimation procedure containing principal component analysis and fitting a dynamic factor model to a large panel of auxiliary time series, it is possible to extract underlying latent factors that are assumed to drive the general progression of macroeconomic time series. Furthermore, multiple target series are estimated using these factors in a multivariate framework.

This study extends the earlier work of Kuiper and Pijpers (2020) and the model is now able to match the performance of the current flash estimate method in use at SN in terms of the root mean squared forecast error. An objective data-driven method is provided that produces unbiased point estimates including an uncertainty measure for these estimates. Additionally, it is a much faster way of obtaining flash estimates compared to the current method in use at SN, since it does not require much human interference and deals with publication lags of auxiliary variables using forward extrapolation. It is possible to make an initial estimate right after the quarter has ended, while currently this happens after 45 days. This could then be repeated at the 45 or 90 day mark for example when some of these extrapolations are replaced by official figures. It seems that the initial estimate does not require large revisions when using the proposed method, contrary to the flash estimates in recent years. On top of that, it is tested whether the variation in the target series is captured by the proposed procedure, which appears to hold true.

Using dynamic factor models and nowcasting have become very popular tools for macroeconomic analysis and forecasting during the last decade. This study contributes to the corresponding literature by providing a new approach for nowcasting decomposed parts of the target series separately, which deemed successful in an empirical setting. Testing whether the model assumptions hold theoretically in this new format would be an interesting topic for further research, as well as testing different seasonal models, found in the STS modelling literature, for the dynamic updating of the seasonal factors. Moreover, simultaneously modelling the set of factors with the target series in a collapsed (or extended) dynamic factor model could be tested as an alternative for the elementary target series estimation step. Finally, correcting for structural policy changes can be achieved through adding very specific (dummy or intervention) variables to the set of auxiliary variables, because they are announced in advance.

Acknowledgements
The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. The authors thank the statisticians at the SN National Accounts team. Their expert knowledge on macroeconomic time series helped determine the set of

auxiliary variables used. We are also grateful to Prof. S.J. Koopman of the Vrije Universiteit Amsterdam for his valuable advice and feedback, and Marc Smeets of CBS for his careful review.

References

Bacchini, Fabio et al. (2017). *Handbook on Rapid Estimates*. Luxembourg: Publications Office of the European Union.

Bai, Jushan (2003). "Inferential theory for factor models of large dimensions". In: *Econometrica* 71.1, pp. 135–171.

Bai, Jushan and Serena Ng (2004). "A PANIC attack on unit roots and cointegration". In: *Econometrica* 72.4, pp. 1127–1177.

Bikker, R. et al. (2019). "Consistent Multivariate Seasonal Adjustment for Gross Domestic Product and its Breakdown in Expenditures". In: *Journal of Official Statistics* 35.1, pp. 9–30. DOI: `10.2478/jos-2019-0002`.

Boivin, Jean and Serena Ng (2005). *Understanding and Comparing Factor-Based Forecasts*. Working Paper 11285. National Bureau of Economic Research. DOI: `10.3386/w11285`. URL: `http://www.nber.org/papers/w11285`.

Bok, Brandyn et al. (2018). "Macroeconomic Nowcasting and Forecasting with Big Data". In: *Annual Review of Economics* 10.1, pp. 615–643. DOI: `10.1146/annurev-economics-080217-053214`. URL: `https://doi.org/10.1146/annurev-economics-080217-053214`.

Bollineni-Balabay, Oksana, Jan van den Brakel, and Franz Palm (2016). "Multivariate state space approach to variance reduction in series with level and variance breaks due to survey redesigns". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 179.2, pp. 377–402.

Bräuning, Falk and Siem Jan Koopman (2014). "Forecasting macroeconomic variables using collapsed dynamic factor analysis". In: *International Journal of Forecasting* 30.3, pp. 572–584.

Brownlee, Jason (2017). *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future*. Machine Learning Mastery.

Cascaldi-Garcia, Danilo, Matteo Luciani, and Michele Modugno (2024). "Lessons from Nowcasting GDP across the World". In: *Handbook of Research Methods and Applications in Macroeconomic Forecasting*. Edward Elgar Publishing, pp. 187–217.

Cattell, Raymond B (1966). "The scree test for the number of factors". In: *Multivariate behavioral research* 1.2, pp. 245–276.

Chamberlain, Gary and Michael Rothschild (1983). "Arbitrage, factor structure, and mean-variance analysis on large asset markets". In: *Econometrica* 51.5, pp. 1281–1304.

Dickey, David A and Wayne A Fuller (1979). "Distribution of the estimators for autoregressive time series with a unit root". In: *Journal of the American statistical association* 74.366a, pp. 427–431.

Dijk, Dorinth van, Mick van Rooijen, and Jasper de Winter (2024). *DFROG: A Nowcasting Model for GDP Growth*. Working Paper 819. Available at SSRN: `https://ssrn.com/abstract=5032731`. De Nederlandsche Bank. DOI: `10.2139/ssrn.5032731`. URL: `https://ssrn.com/abstract=5032731`.

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin (2011). "A two-step estimator for large approximate dynamic factor models based on Kalman filtering". In: *Journal of Econometrics* 164.1. Annals Issue on Forecasting, pp. 188–205. ISSN: 0304-4076. DOI: `https://doi.org/10.1016/j.jeconom.2011.02.012`. URL: `https://www.sciencedirect.com/science/article/pii/S030440761100039X`.

– (Nov. 2012). "A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models". In: *The Review of Economics and Statistics* 94.4, pp. 1014–1024.

Durbin, James and Siem Jan Koopman (2012). *Time series analysis by state space methods*. Oxford: Oxford university press.

Forni, Mario, Marc Hallin, et al. (2000). "The generalized dynamic-factor model: Identification and estimation". In: *Review of Economics and statistics* 82.4, pp. 540–554.

– (2005). "The generalized dynamic factor model: one-sided estimation and forecasting". In: *Journal of the American Statistical Association* 100.471, pp. 830–840.

Forni, Mario and Marco Lippi (2001). "The generalized dynamic factor model: representation theory". In: *Econometric theory* 17.6, pp. 1113–1141.

Giannone, Domenico, Lucrezia Reichlin, and Luca Sala (2004). “Monetary policy in real time”. In: *NBER macroeconomics annual* 19, pp. 161–200.
Giannone, Domenico, Lucrezia Reichlin, and David Small (2008). “Nowcasting: The real-time informational content of macroeconomic data”. In: *Journal of Monetary Economics* 55.4, pp. 665–676.
Harvey, Andrew C (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge university press.
Helske, Jouni (2017). “KFAS: Exponential family state space models in R”. In: *Journal of Statistical Software* 78.10, pp. 1–39.
Hindrayanto, Irma, Siem Jan Koopman, and Jasper de Winter (2016). “Forecasting and nowcasting economic growth in the euro area using factor models”. In: *International Journal of Forecasting* 32.4, pp. 1284–1305. ISSN: 0169-2070. DOI: `https://doi.org/10.1016/j.ijforecast.2016.05.003`. URL: `https://www.sciencedirect.com/science/article/pii/S0169207016300632`.
Horn, John L (1965). “A rationale and test for the number of factors in factor analysis”. In: *Psychometrika* 30.2, pp. 179–185.
Hou, Hsieh and H Andrews (1978). “Cubic splines for image interpolation and digital filtering”. In: *IEEE Transactions on acoustics, speech, and signal processing* 26.6, pp. 508–517.
Iverson, Kenneth E. (1962). *A programming language*. New York, NY, USA: John Wiley & Sons, Inc.
Kaiser, Henry F (1960). “The application of electronic computers to factor analysis”. In: *Educational and psychological measurement* 20.1, pp. 141–151.
Kuiper, Mariska and Frank P. Pijpers (2020). *Nowcasting GDP growth rate: a potential substitute for the current flash estimate*. Tech. rep. The Hague: Statistics Netherlands. URL: `https://www.cbs.nl/en-gb/background/2020/20/nowcasting-gdp-growth-rate-a-potential-substitute-for-the-current-flash-estimate`.
Kwiatkowski, Denis et al. (1992). “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” In: *Journal of econometrics* 54.1-3, pp. 159–178.
Luginbuhl, Rob and Loes Verstegen (2024). *Nowcasting GDP Growth*. Tech. rep. Accessed: 2025-01-23. CPB Netherlands Bureau for Economic Policy Analysis. URL: `https://www.cpb.nl/en/nowcasting-gdp-growth`.
Maeland, E. (1988). “On the comparison of interpolation methods”. In: *IEEE Transactions on Medical Imaging* 7.3, pp. 213–217. DOI: `10.1109/42.7784`.
Makhoul, John (1975). “Linear prediction: A tutorial review”. In: *Proceedings of the IEEE* 63.4, pp. 561–580.
Marcellino, Massimiliano, James H Stock, and Mark W Watson (2003). “Macroeconomic forecasting in the euro area: Country specific versus area-wide information”. In: *European Economic Review* 47.1, pp. 1–18.
Pandher, Gurupdesh S (2002). “Forecasting multivariate time series with linear restrictions using constrained structural state-space models”. In: *Journal of Forecasting* 21.4, pp. 281–300.
Perrucci, Eugenio and Frank P. Pijpers (2017). *Filtering in the Fourier domain: a new set of filters for seasonal adjustment of time series and its evaluation*. Tech. rep. The Hague: Statistics Netherlands. URL: `https://www.cbs.nl/nl-nl/achtergrond/2017/25/filtering-in-the-fourier-domain`.
Phillips, Peter CB and Pierre Perron (1988). “Testing for a unit root in time series regression”. In: *Biometrika* 75.2, pp. 335–346.
Raîche, Gilles et al. (2013). “Non-graphical solutions for Cattell’s scree test”. In: *Methodology*.
Schiavoni, Caterina et al. (2021). “A dynamic factor model approach to incorporate Big Data in state space models for official statistics”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184.1, pp. 324–353.
Stock, James H and Mark W Watson (2002a). “Forecasting using principal components from a large number of predictors”. In: *Journal of the American statistical association* 97.460, pp. 1167–1179.
– (2002b). “Macroeconomic forecasting using diffusion indexes”. In: *Journal of Business & Economic Statistics* 20.2, pp. 147–162.
Stundziene, Alina et al. (2024). “Future directions in nowcasting economic activity: A systematic literature review”. In: *Journal of Economic Surveys* 38.4, pp. 1199–1233.

Appendix

Appendix A: Filtering

For a discrete (regular) sampling of the time series with the sampling cadence Δ_t, its Fourier transform is also discretely sampled with a frequency spacing of $\frac{1}{T_L}$ where T_L is the total length of the sampled series, and limited from frequency $f = 0$ up to the Nyquist frequency:

$$f_{Nyq} = \frac{1}{2\Delta_t} \tag{A.1}$$

It can be demonstrated that a weighted moving window average of a time series in the time domain is a convolution process, which in the Fourier domain corresponds to multiplying the Fourier transform of the time series, with a Fourier transform of the weights in the moving average window. This latter Fourier transform is also known as the filter response function $R(f)$.

A top-hat filter response function in the frequency domain, suppressing all Fourier components in the time series with frequencies $f < \frac{1}{year}$ and also suppressing all frequencies $f > f_{Nyq} - \frac{1}{year}$ is:

$$R(f) = \begin{cases} 0 & f < 1/\text{year} \\ 1 & 1/\text{year} < f < f_{Nyq} - \frac{1}{year} \\ 0 & f_{Nyq} - \frac{1}{year} < f \end{cases} \tag{A.2}$$

In essence this filter selects all variation in the time series which is periodic or quasi-periodic with periods less than a year, but greater than a period which would be considered stochastic variations. Everything else is suppressed. Multiplying instead by $1 - R(f)$ suppresses all such intra-year periodicities. An inverse Fourier transform of $R(f)$ thus yields appropriate weights. The inverse Fourier transform of $R(f)$ is:

$$w(t) = \frac{4\sin(at)}{\pi t} \cos\left(\frac{\pi t}{2\Delta_t}\right) \tag{A.3}$$

where $t = 0$ corresponds to the centre of the window for the moving average, and in which the constant a:

$$a = \pi\left(f_{Nyq} - \frac{2}{year}\right) \tag{A.4}$$

In practice, such a sharp edged filter is impractical because it implies that the window over which to perform the weighted average in the time domain becomes very large or even infinite. It is more practical to modify $R(f)$ so that at the transition frequencies the changeover from 0 to 1 is not discontinuous but it is as steep as feasible. By smoothing out the abrupt edges of the filter, it is achieved that the width of the moving average window is reduced; the weights can be set to 0 for lags that (in absolute value) are greater than a desired window half-width. More details can be found in Perrucci and Pijpers (2017).

Appendix B : Factor estimation of the separated components

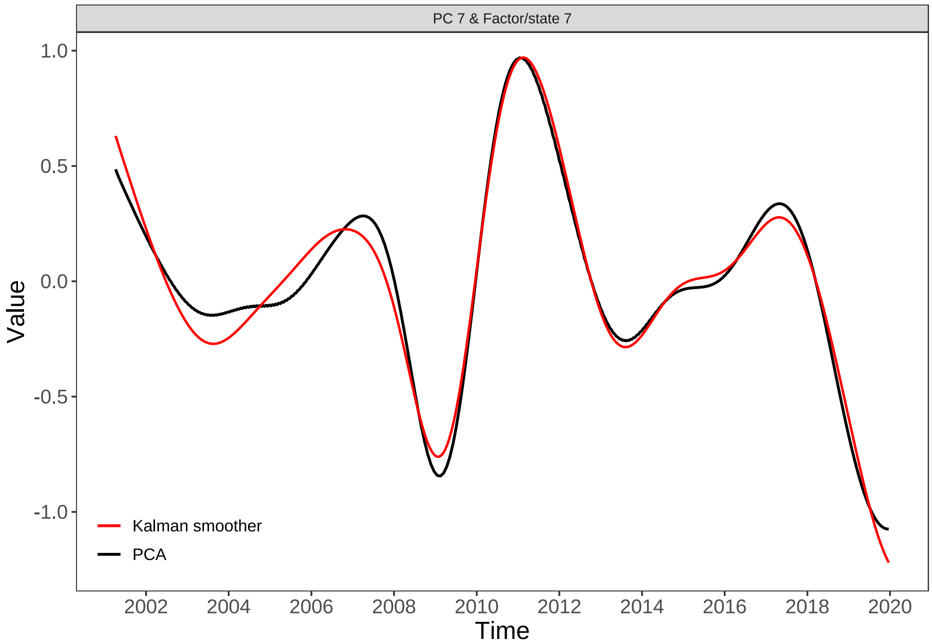


Figure B1 The last trend+noise principal component and factor, retrieved by PCA and the Kalman smoother respectively.

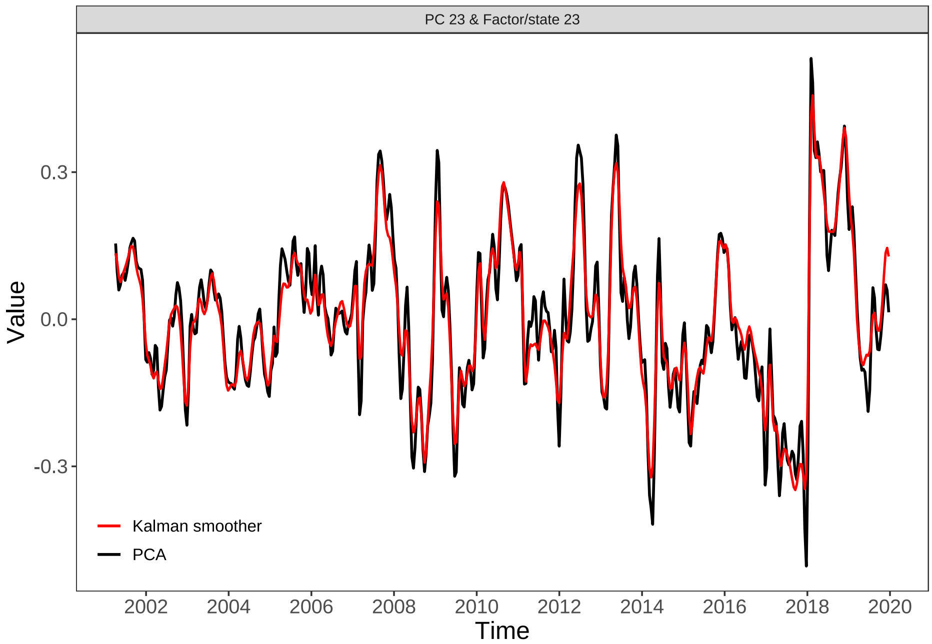


Figure B2 The last seasonal principal component and factor, retrieved by PCA and the Kalman smoother respectively.

$$\hat{\Phi}^{\varsigma} = \begin{bmatrix} 0.96 & -0.04 & -0.04 & 0.25 & 0.05 & 0.09 & -0.03 & -0.11 & -0.03 & -0.03 & 0.09 & -0.12 \\ 0.04 & 0.87 & 0.23 & 0.08 & -0.26 & 0.12 & -0.05 & -0.15 & -0.21 & -0.01 & 0.11 & 0.15 \\ -0.01 & -0.18 & 0.90 & -0.08 & -0.10 & 0.17 & -0.10 & -0.12 & -0.15 & 0.00 & 0.07 & -0.36 \\ -0.16 & -0.01 & 0.01 & 0.95 & 0.00 & -0.01 & 0.02 & -0.07 & 0.03 & -0.01 & 0.00 & -0.12 \\ -0.01 & 0.06 & 0.14 & 0.03 & 0.94 & 0.06 & 0.04 & -0.07 & 0.00 & -0.01 & -0.01 & 0.16 \\ -0.05 & 0.02 & -0.13 & -0.03 & 0.02 & 0.94 & -0.02 & -0.04 & 0.03 & 0.01 & -0.05 & -0.15 \\ 0.00 & 0.02 & 0.03 & -0.02 & -0.02 & 0.01 & 0.98 & 0.00 & -0.04 & 0.00 & 0.04 & -0.02 \\ 0.04 & 0.00 & 0.05 & 0.04 & 0.00 & 0.06 & -0.01 & 0.93 & 0.00 & 0.02 & -0.03 & 0.03 \\ 0.01 & 0.03 & 0.06 & 0.01 & -0.02 & 0.01 & 0.03 & -0.01 & 0.95 & 0.03 & 0.01 & 0.04 \\ 0.00 & 0.02 & -0.01 & -0.01 & 0.01 & -0.02 & -0.01 & -0.02 & -0.02 & 0.97 & 0.05 & 0.00 \\ -0.01 & -0.03 & -0.02 & 0.01 & 0.01 & 0.03 & -0.02 & 0.03 & 0.00 & -0.03 & 0.95 & 0.03 \\ 0.00 & 0.02 & 0.03 & -0.01 & -0.01 & 0.02 & -0.02 & -0.03 & -0.02 & -0.04 & -0.01 & 0.83 \end{bmatrix} \tag{B.1}$$

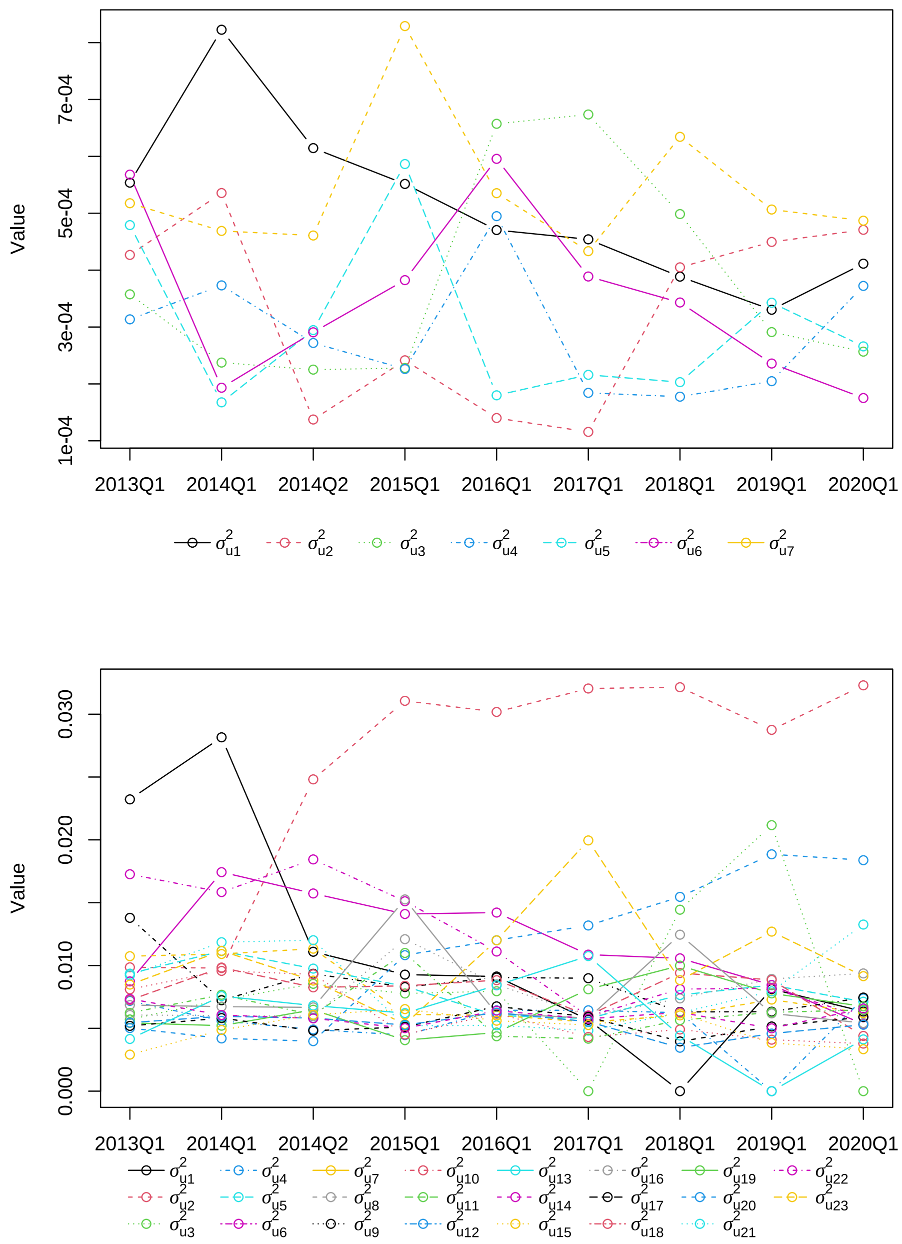


Figure B3 Maximum likelihood estimates of Σ_u^{τ} (top) and of Σ_u^{ς} (bottom) throughout the real-time analysis.

Appendix C : VAT rate suspicion

Performance measures excluding 2019

Table C1 Performance measures for accuracy and bias, in level and relative to statline unit measurements. The RMSFE and ME in level are published in € (x1,000,000,000). The four quarters in 2019 are excluded from the calculation, leaving 24 quarters in the window of 2013-2018. Measures that have decreased due to this adjustment are marked grey.

	RMSFE	ME	RMSFE rel.	ME rel.
Agr. & Ind. Goods	0.58	0.07	1.66%	0.2%
Commerc. Services	0.68	-0.21	0.79%	-0.24%
Non-Commerc. Services	0.26	-0.08	0.68%	-0.21%
Product Tax/Subidies bal.	0.26	-0.06	1.47%	-0.31%
GDP	1.08	-0.29	0.62%	-0.15%

Table C2 Performance measures for accuracy and bias, quarterly growth rates with lag 4 and lag 1. Quarterly growth with a lag of 1 is published as seasonal adjusted time series, hence the trend+noise component nowcasts are used to calculate those performance measures, without adding the seasonal component nowcasts. The four quarters in 2019 are excluded from the calculation, leaving 20/23 quarters in the window of 2013-2018 after first/seasonal differencing to compute growth rates. Measures that have decreased due to this adjustment are marked grey.

	RMSFE l=4	ME l=4	RMSFE l=1	ME l=1
Agr. & Ind. Goods	2.31%	0.22%	1.49%	0.12%
Commerc. Services	0.81%	-0.01%	0.65%	-0.06%
Non-Commerc. Services	0.63%	-0.15%	0.61%	0.01%
Product Tax/Subisidies bal.	2.26%	0.01%	2.02%	0.03%
GDP	0.73%	0%	0.47%	0%

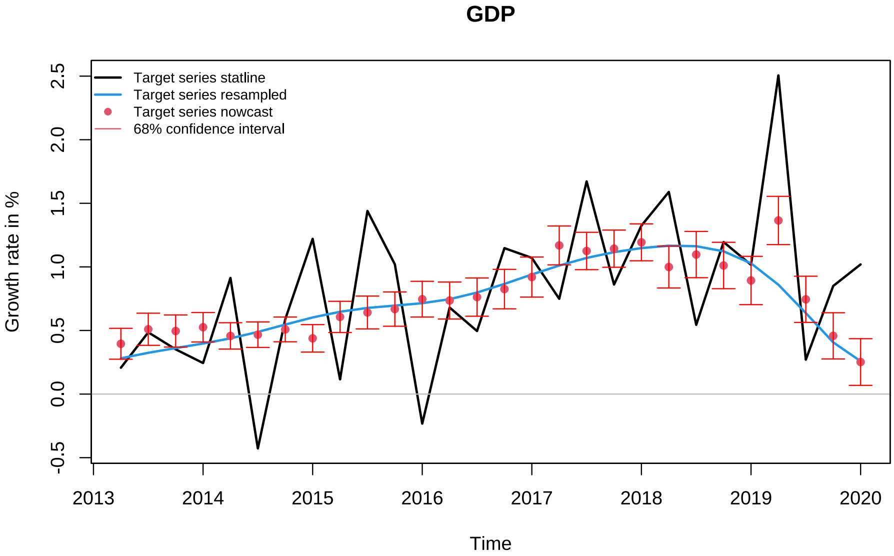
Appendix D : Q-on-Q growth rates

Performance measures excluding bias

Table D1 Performance measures for accuracy and bias, quarterly growth rates with lag 1. The performance of the nowcast is measured compared to the seasonal adjusted target series based on the Fourier transform, i.e. the 'blue' line in all nowcast plots, to overcome the artificial bias in the q-on-q growth rates. Measures that have decreased due to this adjustment are marked grey.

	RMSFE l=1	ME l=1
Agriculture and Industry Goods	0.24%	0.01%
Commercial Services	0.14%	0.03%
Non-Commercial Services	0.13%	0.03%
Product Taxes-Subsidies balance	0.31%	0%
GDP	0.13%	0.02%

Nowcasts



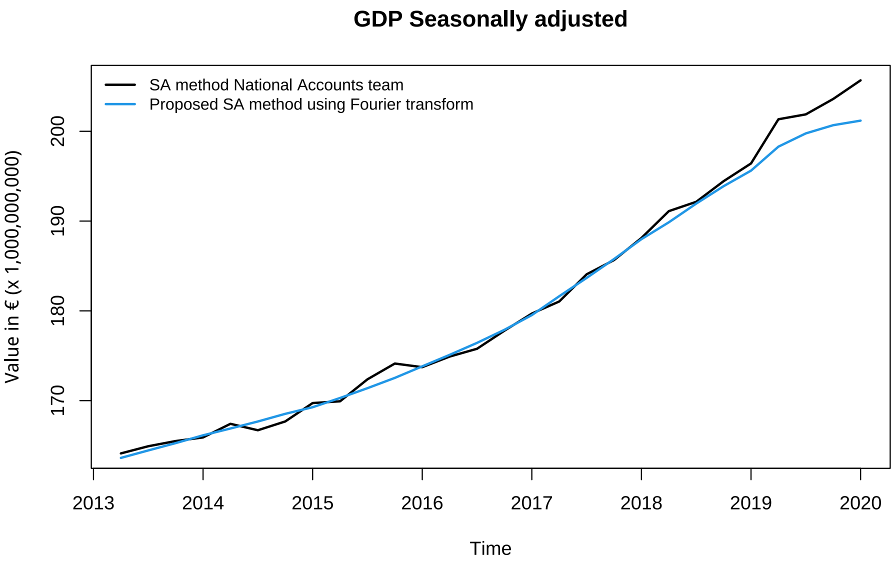


Figure D1 top panel: Nowcasts and realized figures of the fifth target series (GDP) in growth rates. bottom panel: Trend+noise component of the fifth target series in level, directly from statline and using the proposed seasonal filter method.

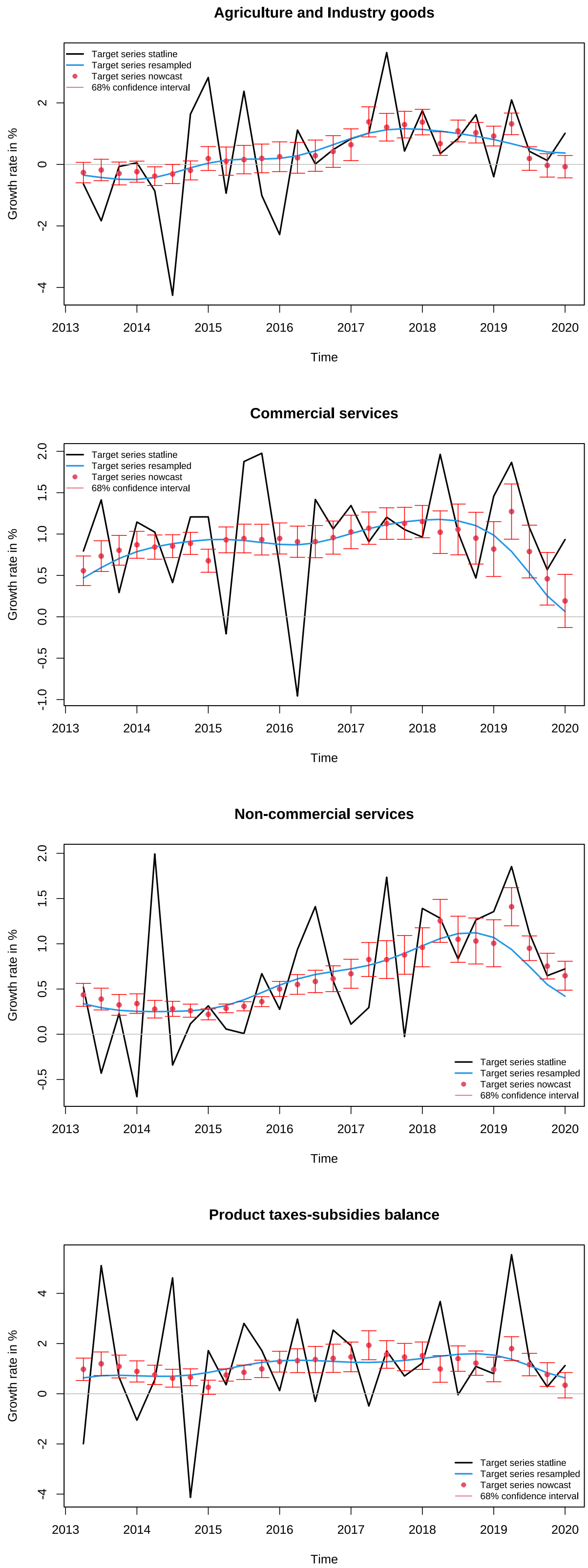


Figure D2 Nowcasts and realized figures of the first to fourth target series in growth rates (from top to bottom).

Appendix E : Data set overview

Table E1 The sets of 4 target variables and 139 auxiliary variables that are assumed to be correlated with the target variables and GDP, i.e. have some predictive power for nowcasting. Lag refers to the time span in days between the last published data point and today. Cadence refers to the original periodicity of publishing (Q=quarterly, M=monthly, D=daily). Source mentions the data source.

	Lag	Cadence	Source
Target variables (GDP production)			
Agriculture and Industry Goods	74	Q	SN
Commercial Services	74	Q	SN
Non-Commercial Services	74	Q	SN
Product Taxes/Subsidies balance	74	Q	SN
Auxiliary variables			
Unemployment	43	M	SN
Vacancies total	74	Q	SN
Vacancies gov.	74	Q	SN
Self-employed	74	Q	SN
Gross labour participation	74	Q	SN
CPI total	15	M	SN
CPI goods	15	M	SN
CPI industrial goods	15	M	SN
CPI durable goods	15	M	SN
CPI services	15	M	SN
CPI electricity	15	M	SN
CPI gas	15	M	SN
Gas price, consumption $<$ 20 GJ	166	Q	SN
Gas price, consumption 20-200 GJ	166	Q	SN
Gas price, consumption 1-10 TJ	166	Q	SN
Gas price, consumption 10-100 TJ	166	Q	SN
Gas price, consumption 100-1000 TJ	166	Q	SN
Gas price, consumption $\geq$ 1000 TJ	166	Q	SN
Fuel price, gasoline	8	M	SN
Fuel price, diesel	8	M	SN
DPI (services price index)	74	Q	SN
Housing average sales price	43	M	SN
Housing price index	43	M	SN
Housing newly build properties	43	M	SN
Building permits total	74	M	SN
Building permits residential	74	M	SN
Consumption households index	74	M	SN
Consumption households goods	74	M	SN
Consumption households services	74	M	SN
Companies	0	Q	SN
Bankruptcies	15	M	SN
Industrial production	43	M	SN
Industrial revenue	43	M	SN
Mineral extraction production	43	M	SN
Mineral extraction revenue	43	M	SN
Total industry production	43	M	SN

Time series overview (contd.)

	Lag	Cadence	Source
Import goods total	74	M	SN
Export goods total	74	M	SN
Import goods EU	74	M	SN
Export goods EU	74	M	SN
Import services total	439	Q	SN
Export services total	439	Q	SN
Import services EU	439	Q	SN
Export services EU	439	Q	SN
Oil transport	166	M	SN
Construction revenue index	74	M	SN
Gov. income/expenditures balance	166	Q	SN
Gov. debt	166	Q	SN
Gov. income total	166	Q	SN
Gov. income taxes	166	Q	SN
Gov. income social charges	166	Q	SN
Gov. income production activities	166	Q	SN
Gov. income other	166	Q	SN
Gov. income investments	166	Q	SN
Gov. exp. total	166	Q	SN
Gov. exp. interm. goods & services	166	Q	SN
Gov. exp. wages	166	Q	SN
Gov. exp. interest	166	Q	SN
Gov. exp. social benefits	166	Q	SN
Gov. exp. subsidies	166	Q	SN
Gov. exp. other	166	Q	SN
Gov. exp. other capital	166	Q	SN
Gov. consumption	166	Q	SN
Cons. conf., econ. past 12 mths	15	M	SN
Cons. conf., econ. next 12 mths	15	M	SN
Cons. conf., fin. past 12 mths	15	M	SN
Cons. conf., fin. next 12 mths	15	M	SN
Cons. conf., right time for big purchases	15	M	SN
Industry production bal. past 3 mths	15	M	SN
Industry production bal. next 3 mths	15	M	SN
Industry sales prices bal. next 3 mths	15	M	SN
Industry purchase orders bal. next 3 mths	15	M	SN
Industry foreign orders bal. next 3 mths	15	M	SN
Transport production bal. past 3 mths	15	M	SN
Transport production bal. next 3 mths	15	M	SN
Transport sales prices bal. next 3 mths	15	M	SN
Transp. purch. orders bal. next 3 mths	15	M	SN
Transp. for. orders bal. next 3 mths	15	M	SN
Catering sales prices bal. next 3 mths	15	M	SN
Catering purchase orders bal. next 3 mths	15	M	SN
Real-est. sales prices bal. next 3 mths	15	M	SN
Real-est. purch. orders bal. next 3 mths	15	M	SN
Com. services production bal. past 3 mths	15	M	SN
Com. services production bal. next 3 mths	15	M	SN
Com. services sales prices bal. next 3 mths	15	M	SN
Com. services purch. orders bal. next 3 mths	15	M	SN

Time series overview (contd.)

	Lag	Cadence	Source
Com. services for. orders bal. next 3 mths	15	M	SN
Culture sales prices bal. next 3 mths	15	M	SN
Culture purchase orders bal. next 3 mths	15	M	SN
Other production bal. past 3 mths	15	M	SN
Other production bal. next 3 mths	15	M	SN
Other sales prices bal. next 3 mths	15	M	SN
Other purchase orders bal. next 3 mths	15	M	SN
Other foreign orders bal. next 3 mths	15	M	SN
PPI sales price mineral extraction	43	M	SN
PPI sales price industrial production	43	M	SN
PPI sales price electricity, gas, heating	43	M	SN
PPI sales price water&waste distribution	43	M	SN
PPI usage price mineral extraction	43	M	SN
PPI usage price industrial production	43	M	SN
PPI usage price electricity, gas, heating	43	M	SN
PPI usage price water&waste distribution	43	M	SN
PPI sales price electricity	43	M	SN
PPI sales price coal	43	M	SN
PPI sales price crude oil	43	M	SN
PPI sales price gas	43	M	SN
Import value index	74	M	SN
Import volume index	74	M	SN
Export value index	74	M	SN
Export volume index	74	M	SN
Vacancies indicator total	15	M	SN
Vacancies indicator com. services	15	M	SN
Disability benefits	43	M	SN
Unemployment benefits	43	M	SN
Social welfare benefits	43	M	SN
State pension benefits	43	M	SN
Construction added value index	74	M	SN
Catering revenue value index	74	Q	SN
Catering revenue volume index	74	Q	SN
Mortgages outstanding	380	M	DNB
Mortgages outstanding interest	380	M	DNB
Mortgages new contracts	380	M	DNB
Mortgages new contracts interest	380	M	DNB
AEX index	1	D	AEX
CLI OECD countries	380	M	OECD
CLI NL	380	M	OECD
Loans to non-financial corporations	380	M	ECB
Loans to households	380	M	ECB
Consumptive credit househ.	380	M	ECB
Borrowing for housing by househ.	380	M	ECB
Securities non-MFI fin. institutions NL	408	M	ECB
Securities MFIs NL	408	M	ECB
Securities non-MFI fin. institutions EU	408	M	ECB
Securities MFIs EU	408	M	ECB
National stock index long term securities	408	M	ECB

Time series overview (contd.)

	Lag	Cadence	Source
Lending for house purchase NL	380	M	ECB
Loans to non-financial institutions NL	380	M	ECB
Lending for house purchase EU	380	M	ECB
Loans to non-financial institutions EU	380	M	ECB

Correlation between $X_{1:T}$ and $Y_{1:T}$

Table E2 The correlation between each auxiliary variable and each target variable, where $y_1, \ldots, y_4$ refer to the partial variables and y_5 refers to the aggregated GDP. In the time series names $+3m$ refers to the next three months and $-3m$ refers to past three months, and similar reasoning applies for the notations $+12m$ and $-12m$.

	y_1	y_2	y_3	y_4	y_5
Unemployment	-0.22	-0.07	0.06	-0.23	-0.08
Vacancies total	0.47	0.39	0.18	0.54	0.37
Vacancies gov.	0.28	0.24	0.07	0.38	0.22
Self-employed	0.72	0.94	0.93	0.83	0.93
Gross labour participation	0.78	0.79	0.81	0.69	0.81
CPI total	0.75	0.96	0.97	0.85	0.96
CPI goods	0.77	0.93	0.95	0.83	0.94
CPI industrial goods	0.78	0.86	0.88	0.76	0.87
CPI durable goods	-0.77	-0.87	-0.91	-0.73	-0.88
CPI services	0.73	0.96	0.97	0.85	0.96
CPI electricity	0.60	0.51	0.49	0.54	0.54
CPI gas	0.89	0.91	0.92	0.85	0.93
Gas price, cons. $<$ 20 GJ	-0.10	0.23	0.16	0.06	0.16
Gas price, cons. 20-200 GJ	-0.07	0.25	0.18	0.11	0.18
Gas price, cons. 1-10 TJ	-0.10	0.26	0.19	0.21	0.20
Gas price, cons. 10-100 TJ	-0.12	-0.37	-0.35	-0.25	-0.33
Gas price, cons. 0.1-1 PJ	-0.22	-0.54	-0.46	-0.48	-0.49
Gas price, cons. $>$ 1 PJ	-0.16	-0.45	-0.39	-0.41	-0.41
Fuel price, gasoline	0.12	0.28	0.36	0.14	0.27
Fuel price, diesel	0.17	0.30	0.37	0.17	0.29
DPI (servs. price index)	0.34	0.66	0.63	0.60	0.63
Housing average sales price	0.80	0.84	0.74	0.89	0.85
Housing price index	0.73	0.63	0.53	0.74	0.65
Housing newbuilt	0.15	0.14	0.08	0.24	0.14
Building permits total	0.08	-0.01	0.04	-0.04	0.02
Building permits residential	0.11	0.19	0.04	0.31	0.16
Consumption househ. index	0.83	0.98	0.95	0.94	0.98
Consumption househ. goods	0.84	0.87	0.87	0.88	0.90
Consumption househ. servs.	0.77	0.98	0.93	0.92	0.97
Companies	0.17	0.54	0.48	0.52	0.49
Bankruptcies	-0.18	-0.39	-0.20	-0.45	-0.33
Industrial production	0.36	0.54	0.45	0.51	0.51
Industrial revenue	0.44	0.58	0.54	0.50	0.56
Mineral extraction prod.	-0.37	-0.66	-0.54	-0.67	-0.61
Mineral extraction revenue	0.05	-0.32	-0.22	-0.27	-0.25
Total industry prod.	0.09	-0.03	0.00	-0.02	-0.01

Correlations (contd.)

	y_1	y_2	y_3	y_4	y_5
Import goods total	0.24	0.50	0.44	0.47	0.46
Export goods total	0.23	0.51	0.45	0.48	0.47
Import goods EU	0.29	0.53	0.45	0.53	0.50
Export goods EU	0.25	0.49	0.43	0.47	0.46
Import servs. total	0.26	0.22	0.16	0.26	0.22
Export servs. total	0.31	0.23	0.16	0.32	0.24
Import servs. EU	0.36	0.37	0.27	0.43	0.36
Export servs. EU	0.35	0.32	0.23	0.41	0.32
Oil transport	0.05	0.17	0.12	0.18	0.14
Construction revenue index	0.10	0.14	0.05	0.22	0.12
Gov. inc./exp. balance	0.29	0.19	0.04	0.42	0.20
Gov. debt	0.59	0.80	0.89	0.63	0.80
Gov. inc. total	0.88	0.93	0.92	0.95	0.96
Gov. inc. taxes	0.77	0.87	0.83	0.94	0.88
Gov. inc. social charges	0.76	0.86	0.86	0.79	0.87
Gov. inc. prod. activities	0.84	0.93	0.96	0.88	0.95
Gov. inc. other	0.33	-0.07	0.06	-0.03	0.02
Gov. inc. investments	0.35	0.54	0.47	0.54	0.51
Gov. exp. total	0.82	0.91	0.98	0.83	0.94
Gov. exp. intermediate	0.84	0.88	0.93	0.83	0.91
Gov. exp. wages	0.74	0.86	0.94	0.81	0.89
Gov. exp. interest	-0.58	-0.89	-0.85	-0.83	-0.86
Gov. exp. social benefits	0.77	0.94	0.98	0.82	0.94
Gov. exp. subsidies	0.70	0.69	0.74	0.66	0.72
Gov. exp. other	0.36	0.05	0.07	0.12	0.11
Gov. exp. other capital	0.41	0.34	0.45	0.32	0.39
Gov. cons.	0.83	0.95	0.99	0.86	0.97
Cons. conf., econ. $-12m$	0.40	0.56	0.41	0.59	0.52
Cons. conf., econ. $+12m$	0.11	0.30	0.25	0.30	0.27
Cons. conf., fin. $-12m$	0.00	-0.01	-0.17	0.05	-0.04
Cons. conf., fin. $+12m$	-0.28	-0.21	-0.37	-0.12	-0.26
Cons. conf., big purchases	0.40	0.54	0.40	0.55	0.50
Industry prod. bal. $-3m$	0.11	0.27	0.20	0.31	0.25
Industry prod. bal. $+3m$	0.05	0.20	0.07	0.24	0.15
Industry sales prc. bal. $+3m$	-0.04	-0.21	-0.27	-0.04	-0.19
Indust. purch.ord. bal. $+3m$	0.05	0.21	0.07	0.27	0.17
Indust. for. orders bal. $+3m$	0.13	0.28	0.22	0.29	0.25
Transp. prod. bal. $-3m$	0.06	0.18	0.15	0.19	0.16
Transp. prod. bal. $+3m$	0.30	0.52	0.43	0.50	0.48
Transp. sales prc. bal. $+3m$	0.08	-0.01	0.00	0.07	0.01
Transp. purch.ord. bal. $+3m$	0.17	0.34	0.20	0.42	0.30
Transp. for.ord. bal. $+3m$	0.23	0.43	0.36	0.42	0.40
Cater. sales prc. bal. $+3m$	0.08	-0.05	-0.11	0.15	-0.03
Cater. purch.ord. bal. $+3m$	0.15	0.23	0.13	0.34	0.21
Real-est. sales prc. $+3m$	-0.02	0.05	0.07	0.04	0.05
Real-est. purch.ord. $+3m$	0.12	0.23	0.09	0.33	0.20
Com. servs. prod. bal. $-3m$	0.22	0.37	0.25	0.47	0.34
Com. servs. prod. bal. $+3m$	0.27	0.37	0.26	0.48	0.35
Com. servs. sales prc. $+3m$	0.09	0.01	-0.01	0.14	0.03
Com.servs. purch.ord. $+3m$	0.18	0.36	0.20	0.46	0.32

Correlations (contd.)

	y_1	y_2	y_3	y_4	y_5
Com.servs. for.ord.bal. $+3m$	0.19	0.38	0.31	0.38	0.35
Culture sales prc. bal. $+3m$	0.07	-0.04	-0.03	0.07	-0.01
Culture purch.ord. bal. $+3m$	0.13	0.23	0.11	0.36	0.21
Other prod. bal. $-3m$	0.26	0.34	0.27	0.39	0.33
Other prod. bal. $+3m$	0.30	0.56	0.44	0.56	0.51
Other sales prc. bal. $+3m$	-0.05	-0.19	-0.17	-0.10	-0.16
Other purch.ord. bal. $+3m$	0.16	0.31	0.18	0.39	0.27
Other for.ord. bal. $+3m$	0.22	0.49	0.41	0.51	0.45
PPI sales mineral extr.	-0.38	-0.74	-0.64	-0.69	-0.68
PPI sales industrial prod.	-0.22	-0.56	-0.52	-0.45	-0.51
PPI sales electr., gas, heat.	-0.45	-0.81	-0.73	-0.74	-0.76
PPI sales water&waste dist.	-0.37	-0.41	-0.30	-0.52	-0.40
PPI usage mineral extr.	-0.16	-0.45	-0.34	-0.47	-0.40
PPI usage industrial prod.	-0.28	-0.62	-0.59	-0.51	-0.57
PPI usage electr., gas, heat.	-0.44	-0.80	-0.73	-0.74	-0.76
PPI usage water&waste dist.	-0.36	-0.41	-0.30	-0.52	-0.40
PPI sales electricity	-0.44	-0.80	-0.72	-0.73	-0.75
PPI sales coal	-0.12	-0.40	-0.42	-0.29	-0.37
PPI sales crude oil	-0.05	-0.26	-0.15	-0.31	-0.22
PPI sales gas	-0.31	-0.66	-0.56	-0.62	-0.60
Import value index	0.81	0.95	0.93	0.88	0.95
Import volume index	0.78	0.97	0.91	0.92	0.96
Export value index	0.80	0.95	0.92	0.86	0.94
Export volume index	0.77	0.95	0.91	0.89	0.94
Vacancies indic. total	0.29	0.35	0.20	0.43	0.32
Vacancies indic.com.servs.	0.28	0.31	0.16	0.40	0.29
Disability benefits	-0.43	-0.63	-0.66	-0.48	-0.62
Unemployment benefits	-0.12	0.21	0.28	0.07	0.18
Social welfare benefits	0.20	0.62	0.60	0.51	0.57
State pension benefits	0.36	0.74	0.74	0.60	0.70
Construction added value	0.19	0.12	0.00	0.23	0.12
Catering revenue value	0.36	0.74	0.63	0.65	0.67
Catering revenue volume	0.05	0.24	0.08	0.26	0.18
Mortgages outstanding	0.73	0.77	0.81	0.65	0.79
Mortgages outstanding intrst.	-0.66	-0.93	-0.84	-0.90	-0.90
Mortgages new	0.28	0.50	0.38	0.57	0.46
Mortgages new intrst.	-0.36	-0.75	-0.70	-0.69	-0.70
AEX index	0.23	0.38	0.19	0.44	0.33
CLI OECD countries	-0.12	-0.05	-0.12	-0.03	-0.08
CLI NL	-0.02	-0.01	-0.10	0.05	-0.03
Loans to non-fin. corps.	0.04	-0.18	-0.32	0.00	-0.17
Loans to households	-0.25	-0.22	-0.31	-0.11	-0.25
Cons. credit househ.	-0.01	0.09	-0.01	0.17	0.06
Borrow. housing by househ.	-0.29	-0.25	-0.34	-0.15	-0.28
Securities non-MFI fin. inst. NL	0.82	0.92	0.95	0.83	0.93
Securities MFIs NL	0.72	0.73	0.82	0.62	0.77
Securities non-MFI fin. inst. EU	0.72	0.86	0.94	0.72	0.88
Securities MFIs EU	0.61	0.41	0.53	0.33	0.47
Nat. stock idx. (long term securs.)	0.78	0.91	0.96	0.78	0.92

Correlations (contd.)

	y_1	y_2	y_3	y_4	y_5
Lending house purch. NL	-0.37	-0.75	-0.71	-0.69	-0.71
Loans to non-fin. inst. NL	-0.39	-0.75	-0.80	-0.61	-0.72
Lending house purch. EU	-0.42	-0.80	-0.80	-0.70	-0.77
Loans to non-fin. inst. EU	-0.36	-0.73	-0.77	-0.61	-0.70

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

© Statistics Netherlands, The Hague/Heerlen/Bonaire 2024.
Reproduction is permitted, provided Statistics Netherlands is quoted as the source