ISSN 1977-978X

eurostat

2/2015

EURONA

EUROSTAT REVIEW ON NATIONAL ACCOUNTS AND MACROECONOMIC INDICATORS

SPECIAL ISSUE
NEW TECHNIQUES AND TECHNOLOGIES FOR STATISTICS (NTTS) 2015

eurostat

eurostat
2/2015

EUROSTAT REVIEW ON NATIONAL ACCOUNTS AND MACROECONOMIC INDICATORS

SPECIAL ISSUE NEW TECHNIQUES AND TECHNOLOGIES FOR STATISTICS (NTTS) 2015

eurostat

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the Internet (http://europa.eu).

Luxembourg: Publications Office of the European Union, 2015

ISSN 1977-978X
Cat. No: KS-GP-15-002-EN-N

Theme: Economy and finance
Collection: Statistical books

Printed in Luxembourg

Contents

Aims and scope

EURONA is an open access, peer-reviewed, scholarly journal dedicated to National Accounts and Macroeconomic Indicators. EURONA aims at providing a platform for researchers, scholars, producers and users of macroeconomic statistics to exchange their research findings, thereby facilitating and promoting the advancement of National Accounts and Macroeconomic Indicators.

EURONA publishes empirical and theoretical articles within the scope of National Accounts and Macroeconomic Indicators, as well as articles on important policy uses of these statistics. They may relate to both users' and producers' interests, present subjects of general relevance or investigate specific topics.

EURONA is non-partisan and applies the highest standards to its content, by emphasising research integrity, high ethical standards, validity of the findings and cutting edge results. EURONA gives room to all viewpoints.

The articles published in EURONA do not necessarily reflect the views or policies of the European Commission.

Website: http://ec.europa.eu/eurostat/web/national-accounts/publications/eurona

Contact: ESTAT-EURONA@ec.europa.eu

Editors

Editorial board

(1) Paulus.Konijn@ec.europa.eu

(2) sartore@unive.it

Editorial

Every second year, Eurostat organizes an international conference on New Techniques and Technologies for Statistics (NTTS). The event is a success story. The 2015 edition set a record participation with statisticians, researchers, economists and policy makers from all continents, in some 40 to 50 sessions. National accounts had its own session and many presentations from other sessions were covering methodological topics with interest for this domain. This issue of EURONA is dedicated to papers that were presented or are linked to presentations that were given at the NTTS 2015 conference. The conference aims to stimulate exchanges of ideas, experiences and tools in the network of official statistics producers and so does EURONA.

The four articles of this issue are presenting a palette of findings from methodological work by David Antonio de Liedo in nowcasting, to empirical experience in benchmarking techniques by Geoffrey Brent, Alex Stuckey and Tom Davidson, a proposal for visualizing the business cycle by Gian Luigi Mazzi and finally with research by Rosa Ruggeri Cannata, Dario Buono and Ferdinando Biscosi on how to improve data coverage and thus the quality of indicators for policy makers. The four of them have in common the search for better, more timely, more complete and more accessible statistics to provide a better basis for economic policy making.

The methodological work of David de Antonio Liedo uses a joint state-space model for nowcasting both the euro area and Belgian GDP growth rates, taking into account the intra-quarterly data flow as an input, to construct early estimates of GDP growth and update them in real time. His proposal is particularly relevant for national accounts because it allows users to read macroeconomic news at any point in time and weight them in a transparent manner in order to extract the underlying growth signal. Thanks to the timely availability of qualitative survey data, those nowcasts are shown to be informative already three months before the publication of the official flash estimates.

In their article, Geoffrey Brent, Alex Stuckey and Tom Davidson explore different benchmarking and forecasting combinations, applied to both the 'forward' and 'back' series, on a range of Australian economic series and on synthesized data. All combinations were assessed according to the criteria of magnitude and timeliness of revisions of estimates, apparent bias of the estimates and how well movements in the indicator series were preserved in the benchmarked estimates. The results they obtained are improving the estimates of the quarterly national accounts by using different benchmarking and forecasting combinations anticipating the availability of additional annual benchmarks.

Gian Luigi Mazzi discusses the use of composite indicators versus dashboards and scoreboards of indicators as alternative ways to present economic data to users. The Business Cycle Clock is presented as one example of a graphical approach to the presentation of complex data, allowing to display messages on hidden signals the naked eye does not detect.

Rosa Ruggeri Cannata, Dario Buono and Ferdinando Biscosi underline the importance of official statistics in the Macroeconomic Imbalances Procedure (MIP) and, consequently, of their quality with respect to the data coverage. The availability of data on the various MIP indicators is heterogeneous, depending both on country and indicator. It is quite difficult to assess policies on a medium to long-term period, if the available time series is not long enough. The authors have shown how the backcalculation techniques on the MIP indicators have enlarged the data coverage.

Have a fruitful reading!

Domenico Sartore

Editor of EURONA

1

Nowcasting Belgium

David de Antonio Liedo [1]

Abstract: This paper proposes a dynamic factor model that takes into account the calendar of European and Belgian intraquarterly data releases to automatically update GDP growth expectations or 'nowcasts' in real-time. Those updates can be decomposed in terms of all the forecast errors observed in every data release. This paper contributes to the literature that exploits qualitative surveys to anticipate macroeconomic variables, but this is the first analysis that isolates 'quality' from 'timeliness' as independent properties that can be expressed in function of the model parameters. The modeling framework allows for the incorporation of an heterogeneous information set including different kinds of survey data directly in levels. It also allows for the use of flash GDP and subsequent revisions as separate indicators without imposing strict assumptions regarding the rationality of the statistical agency. The empirical results emphasize the quality of survey data, which allows the model to produce accurate real GDP growth nowcasts for Belgium three months prior to the publication of the official flash estimate.

Keywords: news, dynamic factor models, EM algorithm.

JEL codes: C32, C53, E37.

(1) National Bank of Belgium.

1. Introduction

The meteorological term 'nowcasting' has become increasingly popular in economics over the last few years following the success of statistical methods in formalizing the mixture of judgment and expert knowledge involved in the calculation of early estimates of economic activity. Unlike nowcasting in meteorology, where forecasters base their decisions on the current weather along with forecasts for a period of zero to six hours ahead, institutions responsible for economic policy need to make important decisions without directly observing the current state of the economy. The so-called 'flash' estimate of Gross Domestic Product (GDP) for Belgium is published by the Belgian National Accounts Institute (NAI) about 30 days after the end of the quarter, while Eurostat publishes the aggregate euro area flash figure with an approximate delay of 45 days. This implies that not only the current quarterly growth of the economy is actually unknown, but also the one corresponding to the previous quarter is available with a significant delay. However, many other economic indicators and surveys become available within each quarter.

In this paper, I propose a joint state-space model for the euro area and Belgian economies formalizing the role of the intra-quarterly data flow as an input to construct early estimates of GDP growth and update them in real time. Those updates will be decomposed in terms of surprises embodied in each one of the macroeconomic releases. The model is estimated with maximum likelihood using the adaptation of the Expectation-Maximization (EM) algorithm proposed by Banbura and Modugno (2010). Thus, nowcasts are defined as forecasts conditional on the available information set, which may have gaps at the beginning, in the middle, and at the end of the sample, but which expands forward with every new data release. Evans (2005) and Giannone et al. (2008) are the first papers that formalize the so-called nowcasting process and also emphasize the real-time data flow as an essential element in short-term forecasting. Nevertheless, as argued by Banbura et al. (2011), nowcasting also requires understanding the mapping from new data releases onto forecasting revisions. This second dimension of the nowcasting problem cannot be addressed with partial models such as bridge equations, and requires a joint model for all indicators.

Like many of the existing tools available for nowcasting GDP growth in real time, the method presented here takes into account the presence of strong co-movements in macroeconomic data by incorporating restrictions inspired by the literature on dynamic factor models. Factor models are relatively restrictive representations allowing GDP growth to be expressed as the sum of two orthogonal components: one driven by pervasive factors that spread throughout the economy, and a measurement error component that is idiosyncratic. Such restrictions have also been successful in nowcasting US and euro area data, as shown by Giannone et al. (2008) or Camacho and Perez-Quiros (2010), respectively.

In this paper, I focus on the accuracy of the forecasts for the real GDP growth rates in Belgium in a framework that allows for the co-existence of both Belgium specific and euro area wide shocks. The empirical results underline the importance of survey data such as the Business Confidence Index constructed by the National Bank of Belgium (NBB).This is not necessarily a surprising feature given the popularity of the Belgian Business Survey as a leading indicator of the euro area economy, which was highlighted by the Wall Street Journal (1999). I show that the release corresponding to the first month of each quarter plays a particularly big role in updating GDP growth expectations. Other indicators that have a large impact on growth forecasts are the Markit Economics PMI (Manufacturing) release for the euro area, 3-month Euribor and real house prices in Belgium, which turn out to contribute mainly at longer horizons. This is consistent with my finding that three months prior to the publication of the Belgian flash, the nowcast turns out to be as accurate as the flash release itself. Given the information available in real time, the flash release for Belgium does not provide a significant gain in estimation accuracy regarding the state of the economy. This paper goes further than the literature in understanding whether the importance of survey data can be accounted for by their timeliness or rather their quality. In a counterfactual exercise, I show that the weights associated with survey data do not deteriorate when all hard data is published with the same degree of timeliness. This result underlines the quality of survey data. The importance of survey data in Belgium is also discussed in detail by Piette and Langenus (2014).

This paper complements the large literature that has investigated the role of qualitative survey data at forecasting macroeconomic variables. For example, Martinsen et al. (2014), Abberger (2007), Claveria et al. (2007) or Lui et al. (2011), exploit disaggregated survey data at forecasting macroeconomic variables, while Giannone et al. (2009) conclude for the euro area that qualitative surveys at a disaggregate level do not contain any information beyond the aggregate sentiment index. Others papers such as Gayer et al. (2014) or Angelini et al. (2012) obtain the same conclusion as I do in this paper. However, their evidence in favour of surveys is based on their ability to improve forecasting accuracy, throughout a given subsample of the data, relative to a benchmark model that does not include those surveys. Their approach is different from the methodology followed in this paper, which exploits the whole sample by monitoring how the Kalman filtering weights the news in the data. Finally, there are other papers that test for the usefulness of aggregate surveys in a simpler way without taking into consideration the real-time data flow, e.g. D'Agostino and Schnatz (2012), Koenig (2002) or Barhoumi et al. (2009). The last group of papers is also not able to quantify the marginal contribution with respect to the information set already available at the time of their publication. From the methodological point of view, my paper is closest to Camacho and Perez-Quiros (2010), or Banbura et al. (2011).

This paper is structured as follows. Section 2 defines the model and compares it to the state of the art. Section 3 presents the data and the particularities of GDP revisions. Section 4 studies the precise role of all data releases in the process of updating the real GDP growth rate. In addition I briefly discuss why such role depends crucially on timeliness and quality, which are desirable characteristics of macroeconomic data releases. Finally, section 5 presents out-of-sample forecasts that would have been obtained by the model since the last quarter of 2007 using the information that was available in real-time. The section 6 concludes.

The model estimation, news analysis and forecasting simulation results presented in this paper can be easily reproduced and extended by installing a nowcasting plug-in in the JDemetra+ software (results presented in this paper can be easily reproduced and extended by installing a nowcasting plug-in in the JDemetra+ software, https://www.nbb.be/en/jdemetra).

2. Modeling framework

All monthly and quarterly variables are represented as a parametric dynamic factor model, which can be expressed in state-space form. Within this very specific framework, I outline the most common approach to link GDP, which is a quarterly variable, with the unobserved factors, which are specied at a monthly frequency.

2.1. A state-space representation

I describe here the particularities of the joint model for the Belgian economy and the aggregate euro area data. The following expression links the monthly growth rates of the variables to the vector of underlying monthly factors:

(1) $y_t \; = \; \bar{y} + \Lambda_y f_t + \Theta_y b_t + \psi_t$ *Measurement Belgium*

(2) $x_t \; = \; \bar{x} + \Lambda_x f_t + \chi_t$ *Measurement euro area*

Because the model has been designed for the conduct of short-term analysis, it makes sense to represent all these series, including GDP, in terms of monthly growth rates or monthly differences. Belgian time series will be denoted by y_t, while euro area series are represented by x_t. The factor f_t represents the latent monthly growth rate of the area economy, which is also relevant for Belgium, and b_t formalizes the part of Belgian economy's monthly growth rate not captured by f_t. The so-called Belgium specific factor, b_t, will be able to capture the possible leading behavior of Belgian indicators ([2]), but it does not load contemporaneously on the euro area data, i.e. $\Theta_x = 0$, which makes the model more parsimonious. By considering these two blocks, the model can be used in coordinated forecasting exercises in order to forecast the Belgian economy conditional on the ECB's views about the short-term evolution of the euro area as a whole.

The error terms χ_t and ψ_t are assumed to be uncorrelated with the factors at all leads and lags. They are also assumed to be independently and identically distributed following a normal distribution: $\chi_t \sim N(0, R_\chi)$ and $\psi_t \sim N(0, R_\psi)$. Both covariance matrices are assumed to be diagonal, which implies that the factors will account for 100 % of the co-movements implicit in the model. As suggested by Doz et al. (2012), this assumption is not very restrictive. They show that Quasi-ML estimation of the factors is consistent even in the presence of weak cross-correlation patterns in the error term.

Because the monthly growth rates of official GDP figures are not published, equations 1 and 2 need to be modified. Thus, GDP growth rates published by the statistical agencies (i.e. y_t^Q for Belgium and the x_t^Q for the euro area) are linked to the quarterly growth rates of the underlying factors, which can be expressed as a moving average of their monthly growth rates:

(3) $y_t^Q \; = \; \bar{y}^Q + \Lambda_y^Q f_t^Q + \Theta^Q b_t^Q + \chi_t^Q, \quad t = 3, 6, 9, \ldots$ *Belgian GDP*

(4) $x_t^Q \; = \; \bar{x}^Q + \Lambda_x^Q f_t^Q + \psi_t^Q, \qquad t = 3, 6, 9, \ldots$ *euro area GDP*

([2]) If that was the case, as suggested for example by Vanhaelen et al. (2000), a model with a single factor common to both datasets would be misspecified.

where

$$
\begin{aligned}
f_t^Q &= \frac{1}{3}f_t + \frac{2}{3}f_{t-1} + f_{t-2} + \frac{2}{3}f_{t-3} + \frac{1}{3}f_{t-4} \\
b_t^Q &= \frac{1}{3}b_t + \frac{2}{3}b_{t-1} + b_{t-2} + \frac{2}{3}b_{t-3} + \frac{1}{3}b_{t-4}
\end{aligned}
$$

As mentioned above, f_t and b_t represent monthly growth rates of the latent factors. The last expressions for f_t^Q and b_t^Q are based on the technical assumption that the quarterly level of the factors can be represented by the geometric mean of the latent monthly levels ([3]). This assumption makes it possible to obtain a simple expression for the quarterly growth rate of the factors as a moving average of the latent monthly growth rates. Because I apply the Mariano and Murasawa (2003) approximation to the factors alone, and not to the observables, the error terms χ_t^Q and ψ_t^Q are assumed to be iid normally distributed and uncorrelated with all factors at all leads and lags.

So far, I have described the so-called measurement equation, which defines the link between the unobserved factors and the two types of observable time series: monthly variables and quarterly variables (e.g. GDP). Specifying the joint dynamics of all variables in both the euro area and Belgium requires a second equation representing the factors as a vector autoregressive (VAR) process with a non-diagonal covariance matrix for the error term. Thus, even if Belgian specific factors do not load contemporaneously on euro area data (see equation 2), they can be correlated to the euro area factors. To sum up, representation given by equations 5 and 6 conforms to the so-called state-space representation of this model and determines the joint dynamics of both the euro area and the Belgian business cycles:

$$
(5) \quad
\begin{pmatrix} x_t^Q - \bar{x}^Q \\ x_t - \bar{x} \\ \hline y_t^Q - \bar{y}^Q \\ y_t - \bar{y} \end{pmatrix}
=
\left(\begin{array}{ccccc|ccccc}
\Lambda_x^Q & 2\Lambda_x^Q & 3\Lambda_x^Q & 2\Lambda_x^Q & \Lambda_x^Q & 0 & 0 & 0 & 0 & 0 \\
\Lambda_x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\Lambda_y^Q & 2\Lambda_y^Q & 3\Lambda_y^Q & 2\Lambda_y^Q & \Lambda_y^Q & \Theta_y^Q & 2\Theta_y^Q & 3\Theta_y^Q & 2\Theta_y^Q & \Theta_y^Q \\
\Lambda_y & 0 & 0 & 0 & 0 & \Theta_y & 0 & 0 & 0 & 0
\end{array}\right)
\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \hline b_t \\ b_{t-1} \\ b_{t-2} \\ b_{t-3} \\ b_{t-4} \end{pmatrix}
+
\begin{pmatrix} \chi_t^Q \\ \chi_t \\ \hline \psi_t^Q \\ \psi_t \end{pmatrix}
$$

([3]) The approximation proposed by Mariano and Murasawa (2003) is applied to the factors. Let F_t be the monthly level of the economy and let $f_t=lnF_t-lnF_{t-1}$ be its monthly growth rate. Now, define F_t^Q as the geometric mean of the last three levels. This implies that $lnF_t^Q=⅓(lnF_t+lnF_{t-1}+lnF_{t-2})$. The resulting quarterly growth rate of the factors, which we denote as f_t^Q, can be expressed as $lnF_t^Q - lnF_{t-3}^Q$. By substituting both terms by the geometric mean approximation we obtain $f_t^Q = ⅓(lnF_t + lnF_{t-3}) + ⅓(lnF_{t-1} + lnF_{t-4}) + ⅓(lnF_{t-2} + lnF_{t-5})$. Finally, a simple expression for the quarterly growth rate of the factors in terms of their monthly growth rates can be obtained as follows: $f_t^Q = ⅓(f_t + f_{t-1} + f_{t-2}) + ⅓(f_{t-1} + f_{t-2} + f_{t-3}) + ⅓(f_{t-2} + f_{t-3} + f_{t-4})$. Rearranging terms yields the expression $f_t^Q = ⅓f_t + ⅔f_{t-1} + f_{t-2} + ⅔f_{t-3} + ⅓f_{t-4}$ presented above.

$$(6) \quad \begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \hline b_t \\ b_{t-1} \\ b_{t-2} \\ b_{t-3} \\ b_{t-4} \end{pmatrix} = \left(\begin{array}{ccccc|ccccc} A_{11} & 0 & 0 & 0 & 0 & A_{12} & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline A_{21} & 0 & 0 & 0 & 0 & A_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \end{array}\right) \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ f_{t-5} \\ \hline b_{t-1} \\ b_{t-2} \\ b_{t-3} \\ b_{t-4} \\ b_{t-5} \end{pmatrix} + \begin{pmatrix} u_t^f \\ 0 \\ 0 \\ 0 \\ 0 \\ u_t^g \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

where the innovations to the Belgian and external factors are allowed to be cross-correlated:

$$\begin{pmatrix} u_t^f \\ u_t^g \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} Q_1 & \Omega_{12} \\ \Omega_{12} & Q_2 \end{bmatrix}\right)$$

These error components are also uncorrelated with all measurement error terms, in line with the literature on factor models. For simplicity, and in contrast to the model built by Mariano and Murasawa (2003), I do not incorporate autocorrelation in the measurement errors. This helps to keep the size of the state vector as small as possible without restricting the extent to which the factors can account for the business cycle co-movements.

2.2. Estimation in the context of missing observations

Once the building blocks of the model have been described, we need to tackle the problem of estimation. The alternative versions of the model, which will be described in detail in the next section, could be misspecified if the innovations do not follow a normal distribution or if the covariance of the noise component is not diagonal. The Quasi-maximum likelihood procedure of Doz et al. (2012) is used here with the aim of achieving a consistent estimation even in the presence of weak correlation patterns in the measurement errors. Thus, the model is estimated under the restriction that the off diagonal elements of the measurement error covariance matrix are equal to zero. This has the practical implication that one hundred percent of the cross-correlation patterns generated by the model will be fully accounted for by the factors.

The model is estimated at monthly frequency with maximum-likelihood even in the presence of missing observations. For example, survey data for the euro area is often not available prior to 2000, while some of the Belgian series date back to 1980. The presence of quarterly data also generates additional missing observations, since they are treated as indicators that are observed every third month of the quarter, i.e. y_t^Q as a missing variable for $t \neq 3, 6, \ldots$ Finally, as in most macro-economic forecasting applications, the relevant information set is based on indicators that arrive gradually throughout the quarter and with important delays with respect to the period of time to which they refer, i.e. the real-time data flow. Thus, in practice, it is unavoidable to have missing values at the end of the sample. Banbura and Modugno (2010) provide a detailed overview of the estimation method used in this paper. Below, I summarize the most important concepts underlying the approach with special emphasis on the aspects that are relevant in our nowcasting framework:

- **Maximum-likelihood**. In this application, the state-space model represented by equations 5 and 6 is estimated with the Expectation-Maximization (EM) algorithm. The Kalman (1960) filter and smoothing recursions, however, need to be slightly modified so that only the actual observations can be taken into account in the estimation of the factors and the evaluation of the likelihood. The EM algorithm was derived by Shumway and Stoffer (1982) only for the case where the factor loadings multiplying the factors in the measurement equation are known. Banbura and Modugno (2010) are the first ones to apply this algorithm to the current set-up, where the loadings need to be estimated in the context of missing observations. They show that their method is consistent and computationally feasible even in the case where the number of variables is large. Alternatively, Camacho and Perez-Quiros (2010) propose the use of standard optimization routines to maximize the likelihood of a model of the same class, but based on a smaller number of variables (4).
- **Identification of the factors**. The strongest assumption, which is key for identification, is that the measurement errors in expression 5 are uncorrelated with the factor innovations in the transition equation 6. This allows for a clear-cut separation of the measurement errors and the signal provided by the factors. In the absence of the restrictions I impose in the factor loadings, the model would be identified only up to an invertible linear transformation. That is, applying the following transformation, $g_t = Gf_t$, the transition equation $g_t = GA_1G^{-1}g_{t-1} + \ldots + GA_pG^{-1}g_{t-p} + Gu_t$ would be observationally equivalent to the one represented by equation 6. Nevertheless, Dempster, Laird and Rubien (1997) suggest that the EM algorithm is not affected by this lack of identification. The space generated by the factors, and thereby the projections on such space, are obviously unaffected by the choice of G. This identification issue is well known in factor analysis and does not distort any of the results presented in this paper.

3. Data and forecasting models

3.1. Data selection

The goal established in the data selection stage was to approximate the information set of professional forecasters and market analysts, incorporating six variables representing balances of qualitative surveys for Belgium and the euro area. Data selection algorithms such as the one proposed by Piette and Langenus (2014) for Belgian data are thereby ruled out in this paper even if they could potentially help to achieve further forecast accuracy gains (5). In addition, the use of real-time data for model validation enables me to

(4) Numerical optimization of the likelihood, which is feasible for parsimonious models, has the advantage that it does not require the Kalman smoother. Moreover, the multithreading ability of most software packages is able to reduce the execution time by exploiting multiple processors. For example, the current estimation of dynamic factor models in JDemetra+ is feasible without the need of applying the EM algorithm, and it turns out to be much faster in the current application.

(5) De Mol et al. (2008) suggest that variable selection methods are unlikely to help in the presence of strong co-movements in the data.

simulate the actual environment of professional forecasters and, as suggested by Croushore and Stark (2002), avoid any misleading conclusion that may be obtained when the models are estimated and used on the basis of latest available data.

This paper also represents, together with work by Camacho and Perez-Quiros (2010), one of the few attempts in Europe to reproduce the real-time data availability ([6]). The first dimension of real-time data is the presence of data revisions, which I consider for GDP growth rates alone. Fortunately, the variables that have the largest impact in my analysis, i.e. surveys, are not subject to revision. The second dimension of real-time data is the existence of a very specific publication schedule for each indicator. A visual representation of such calendar is provided in table 1. Since I consider the exact release dates for all series, my analysis precisely reproduces the actual work environment of professional forecasters in real time.

In this section, I define the variables incorporated in the model's information set and the release calendar. The publication schedule for the Belgian National Accounts, which were published on a quarterly basis for the first time in 1998, has been gradually adapted to conform to changes in European recommendations, but continues to rely heavily on the annual national accounts. The first part of table 2 lists the most relevant releases for the euro area and Belgium. The first real GDP growth estimate, i.e. the so-called 'flash' release, is published only 30 days after the end of the quarter (two weeks earlier than the flash release for the euro area). This estimate is based on disaggregated data on industrial production and VAT returns for the first two months of the quarter. Given the incompleteness of such information set, the flash is subject to a first revision around 70 days after the end of the reference quarter. However, the revision process does not stop there, but continues indefinitely. Thus, the flash estimate is released together with revisions for previous quarters. The current revision schedule for the euro area is slightly more complex, since data needs to be sent from the national statistical agencies to Eurostat. Eurostat often releases the figures without having received the data from all countries. Figure 1 illustrates the different figures published for the growth rate of 2008Q3, which is the quarter where GDP starts reflecting the recession in Belgium. The main message from this figure is that the first scheduled revisions can be meaningless compared to the continuous revision process that keeps on revising the history year after year. Annex B discusses in more detail some of the properties of data revisions.

The most relevant survey data for Belgium is described in the second part of table 2. The first indicator is the so-called Business Confidence Index (BCI), which results from the overall NBB business confidence survey results. The second indicator corresponds to the results obtained by a subset of questions of the NBB survey, which aims to measure the forward-looking component of the manufacturing sector. The Consumer Confidence Index has been selected not only because it is published around 11 days before the end of each month, but also because consumption represents a significant fraction of total output in the Belgian economy. The last three survey indicators in table 2 refer to the euro area.

The flash release of the Purchasing Manager Indices (PMI) for both the manufacturing and services sectors in the euro area, which is published as promptly as the NBB survey are also incorporated in the analysis together with the Economic Sentiment Indicator (ESI).

([6]) Giannone, Reichlin and Simonelli (2014), for example, use real-time data only for the analysis of the third quarter of 2009.

Table 1: The news flow

		Quarter 1			Quarter 2		
		Month 1	Month 2	Month 3	Month 1	Month 2	Month 3
Financial	Euronext100 (euro area)						
	Brussels all shares						
	Spreads on Belgian 10-year government bonds relative to the German bund						
	3-month Euribor (euro area)						
	German 10-year government bond yield minus 3-month Euribor						
Soft	NBB consumer survey						
	NBB business confidence survey						
	Markit purchasing manager index (flash, euro area)						
	Eurostat economic sentiment indicator (euro area)						
Hard	National CPI						
	Registration of new private cars						
	HICP (euro area)						
	Adjusted harmonised unemployment rate						
	Work volume of temporary workers						
	Retail Sales (euro area)						
	Total turnover (*chiffre d'affaires*)						
	Total receipts (*recettes*)						
	Trade balance in goods						
	Industrial production						
	Industrial production (euro area)						
	Real GDP at market prices (flash, Belgium)					Q1	
	Permits for new residential buildings						
	Real GDP at market prices (flash, euro area)					Q1	
	Real house prices					Q1	
	New residential building starts						

Source: Author's calculations

Note that financial variables are available at a daily basis and without delays. The remaining indicators, however, are typically released with some delay. As represented in the picture, survey data corresponding to a given month is published before the end of this month. By using colors to represent the months (or quarters) to which data refers, I am able to provide a precise description of the news flow from the timeliness viewpoint.

Table 2: GDP and Surveys

Real GDP time series	Source	Unit	Start	Linked to	100xlog	Diff	Average delay (days after the end of the quarter)	Group (region)
Real GDP at market prices (seasonally adjusted (SA), flash)	NBB	Volume	2002Q1	$f^Q(t)$	1	1	35	GDP (BE)
Real GDP at market prices (SA, first)	NBB	Volume	2003Q4	$f^Q(t)$	1	1	71	
Real GDP at market prices (SA, second)	NBB	Volume	1999Q1	$f^Q(t)$	1	1	124	
Real GDP at market prices (SA, 2 years)	NBB	Volume	1999Q1	$f^Q(t)$	1	1	2 years	
Real GDP at market prices (SA, NBB history)	NBB	Volume	1980Q1	$f^Q(t)$	1	1	–	
Real GDP at market prices (SA, flash, euro area)	Eurostat	Volume	2001Q1	$f^Q(t)$	1	1	45	GDP (EA)
Real GDP at market prices (SA, first, euro area)	Eurostat	Volume	2001Q1	$f^Q(t)$	1	1	78	
Real GDP at market prices (SA, second, euro area)	Eurostat	Volume	2001Q1	$f^Q(t)$	1	1	118	
Real GDP at market prices (SA, 2 years, euro area)	Eurostat	Volume	2001Q1	$f^Q(t)$	1	1	2 years	
Real GDP at market prices (SA, Area Wide Model (AWD) history, euro area)	ECB	Volume	1980Q1	$f^Q(t)$	1	1	–	

'Soft data'	Source	Unit	Start	Linked to	100xlog	Diff	Average delay (days after the end of the quarter)	Group (region)
NBB business confidence survey (overall)	Business Survey NBB	Index	2002Q1	$f^Y(t)$	0	0	– 7	Survey (BE)
NBB demand expectations (manufacturing)	Business Survey NBB	Index	2003Q4	$f^Y(t)$	0	0	– 7	
Consumer confidence indicator	Consumer Survey NBB	Index	1999Q1	$f^Y(t)$	0	0	– 11	
Purchasing Manager Index (flash, manufacturing, euro area)	Markit Economics	Index	1999Q1	$f^Q(t)$	0	0	– 6	Survey (EA)
Purchasing Manager Index (flash, services, euro area)	Markit Economics	Index	1980Q1	$f^Q(t)$	0	0	– 6	
Economic sentiment indicator (euro area)	Eurostat Survey	Index	2001Q1	$f^Y(t)$	0	0	2	

Source: Author's calculations

Note: None of the surveys are transformed or filtered in any way. The column 'linked to' defines how the variables are linked to the factors. Surveys from the European Commission or the National Bank of Belgium are linked to the cumulative sum of the factors over the last 12 months (f_t^Y), as in Camacho and Perez-Quiros (2010). The European Commission (2006) explicitly states that the guiding principle for the selection of questions in the different surveys is the aim to achieve as high as possible coincident correlation of the confidence indicator with year-on-year growth of the reference series. De Greef and Van Nieuwenhuyze (2009) also emphasize the coincident correlation of the NBB Business Survey with year-on-year GDP growth rates. Althernatively, the PMI indicators are linked to the factors through the Mariano-Murasawa filter (f_t^Q), in exactly the same way as GDP growth.

Table 3: Activity indicators

'Hard data'	Source	Unit	Start	Linked to	100xlog	Diff	Average delay	Group (region)
Trade balance in goods based on the national concept	NAI	Millions of euro	1995M1	f(t)	0	0	52	External (BE)
Industrial production (excluding construction)	FPS (*) Economy	Volume index 2005 = 100)	2000M1	f(t)	1	1	54	Idustrial Production Index (IPI (BE)
Industrial production (excluding construction, euro area)	FPS Economy	Volume index (2005 = 100)	2000M1	f(t)	1	1	54	IPI (EA)
Registration of new private cars (transit not included)	FEBIAC (**)	Number of units	1980M1	f(t)	1	1	1	Cars (BE)
Total turnover	FPS Economy and NAI	Millions of euro	1980M1	f(t)	1	1	51	Sales (BE)
Total receipts	NBB	Millions of euro	1980M2	f(t)	1	1	51	Sales (BE)
Retail sales (deflated turnover in the retail trade, excluding motor vehicles, euro area)	Eurostat	Volume index 2005 = 100)	1995M1	f(t)	1	1	35	Sales (EA)
Adjusted harmonised unemployment rate (Eurostat definition)	National Employment Office (NEO)	Percentages of the active population	1983M1	f(t)	0	1	21	Labour (BE)
Work volume of temporary workers	Federgon (***)	Thousands of hours	1992M1	f(t)	1	1	24	Labour (BE)

Source: Author's calculations

Note: The column 'linked to' defines how the variables are linked to the factors, which represent the underlying monthly growth rates of the economy. As opposed to surveys, all those indicators are directly linked to the factors. (*) stands for Federal Public Service, (**) is the 'Fédération Belge des Industries de l'Automobile et du Cycle' and (**) a federation of HR service providers.

Table 4: Extending the information set

'Hard data'	Source	Unit	Start	Linked to	100xlog	Diff	Average delay	Group (region)
National CPI	FPS Economy	Price index 1996 = 100	1996M1	f(t)	0	1	– 2	Prices (BE)
HICP (euro area)	Eurostat	Price index 1996 = 100	1996M1	f(t)	0	1	17	Prices (EA)
Euronext100 (euro area)	Thomson Reuters Datastream	Price index in euros (monthly average)	01-Jan-00	f(t)	1	1	0	Financial (EA)
Brussels all shares	Thomson Reuters Datastream	Price index in euros (monthly average)	01-Jan-80	f(t)	1	1	0	Financial (BE)
Spreads on Belgian 10-year government bonds relative to the German bund	ECB	Basis points (monthly average)	05-Jun-89	f(t)	0	1	0	Financial (BE, DE)
3-month Euribor (euro area)	ECB	Percentage points (monthly average)	01-Jan-00	f(t)	0	1	0	Financial (EA)
German 10-year government bond yield minus 3-month Euribor	ECB	Percentage points (monthly average)	01-Jan-80	f(t)	0	1	0	Financial (EA,DE)
Permits for new residential buildings (SA)	FPS Economy	Volume in cubic metre	1990M1	f(t)	1	1	105	Housing (BE)
New residential building starts (SA)	FPS Economy	Volume in cubic metre	1980M1	f(t)	0	0	135	Housing (BE)
Real house prices (SA)	NBB	Index	1980Q1	$f^{Q}(t)$	1	1	0	Housing (BE)

Source: Author's calculations

Note: The column 'linked to' defines how the variables are linked to the factors, which represent the underlying monthly growth rates of the economy. As opposed to surveys, most of those indicators are directly linked to the factors. Note that financial variables are available at daily or even higher frequency. Incorporating them at the highest frequency simply requires the factors to defined accordingly. See Banbura et al. (2013) for a detailed discussion.

The variables listed in table 3 and table 4 describe the remaining series introduced into the model. Most of the series in table 3 are well-known indicators, such as industrial production, sales, car registrations, employment, unemployment or the balance of trade. They are typically used to assess the short term evolution of the economy even if they are published with a significant delay. The fifth indicator in the table is the total turnover of Belgian firms according to the VAT declarations. It should be mentioned that, together with the industrial production data, this is a key indicator in the construction of the GDP flash estimate at the NAI.

Table 4 incorporates a new set of variables that economists and policy makers have been closely monitoring since the beginning of the crisis even if they are not often incorporated in nowcasting applications for GDP growth. Those variables include stock prices and consumer prices indices for both the euro area and Belgium, housing market variables for Belgium, and several interest rates (3-month Euribor, term spread, sovereign risk). To sum up, the indicators described represent relatively aggregate quantities and values that approximate the information set used by forecasters. Using more disaggregated data could be useful in order to mimic the information sets available to the statistical agency. Luciani (2014) shows that the so-called 'non-pervasive' shocks, which affect a group of variables within a given sector without spreading towards the rest of the economy, do not significantly distort the estimation of the aggregate factors. Finally, most of the variables are transformed in order to obtain stationarity. The link of that transformed data with the factors is a modeling choice, which explicitly represented in the column 'linked to' of tables 2, 3, 4, and appropiately explained in each table's description.

3.2. Forecasting models

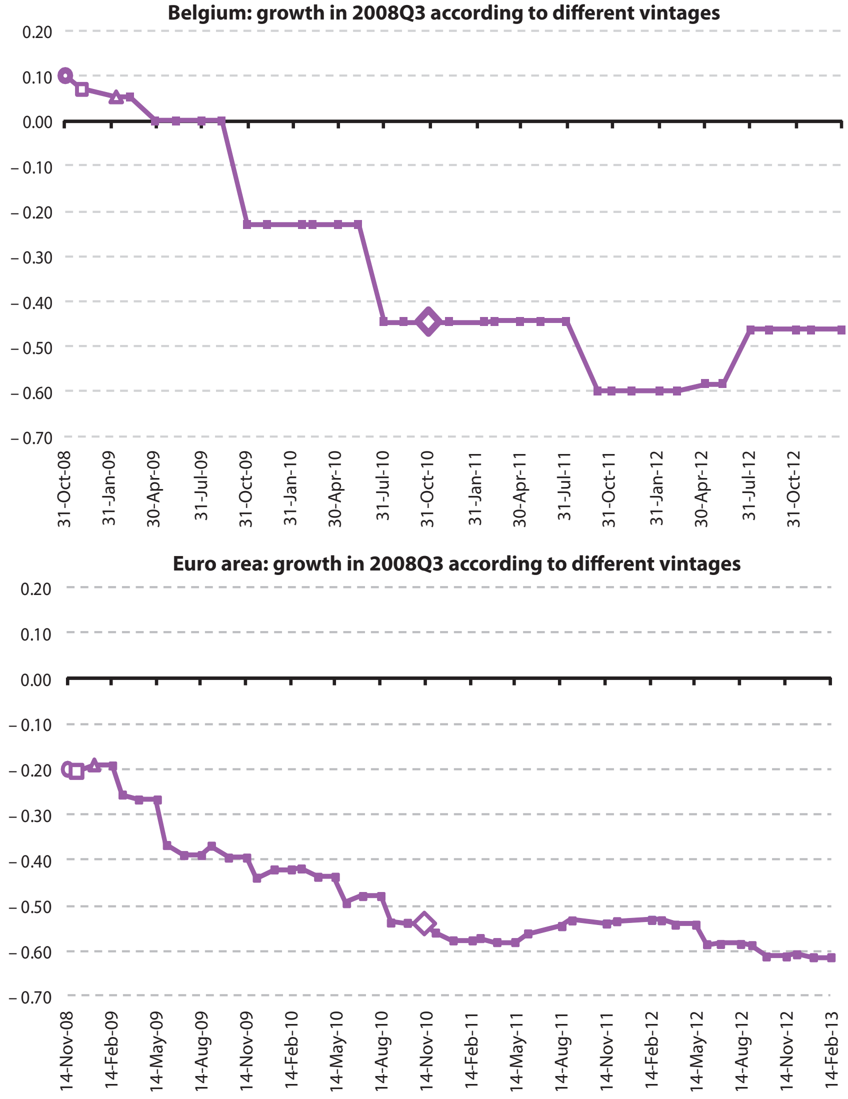
Two alternative models will be estimated to conduct the forecast evaluation exercise that will be described in section 5:

1. The **Benchmark** model, represented in the measurement equation 5, will be estimated using the information set described in tables 2 and 3 in section 3.1. As discussed in the description of our modeling approach (section 2.1.), Belgian variables will load on two factors, while euro area variable will load only on one. However, note that because some surveys are linked to the year-on-year growth rates of the factors, as written in the tables describing the data. This implies we will need twelve lags of the factors and not just five, as in the simplistic representation provided in equations 5-6.
2. The **Financial & Housing** model, however, aims to also take into account the Financial and Housing market variables described in table 4. This case study is particularly relevant, since accounting for specific developments in the Belgian housing market may help to better pinpoint the medium and long-term growth in the Belgian economy. With this in mind, introducing a new block of factors, τ_t, that loads on the housing market variables h_t turns out to be a simple solution to make sure we capture housing market specific developments that are not accounted for by the remaining factors. As described in figure 2, housing market variables in Belgium are assumed to be driven not only by the so-called external (f_t) and specifically Belgian (b_t) factors, but also by the housing factor (τ_t). Since the transition equation species a vector autoregressive structure for all latent factors, unexpected movements in

housing market factors can also affect the forecast for the other two factors (7), and thereby for all variables in the system, which are linked to the factors through the measurement equation.

Figure 1: The continuous revision process

(%)



Source: Author's calculations

Note: The circle, the square and the triangle are used to mark the flash, first, and second revision, respectively. The model will use the Belgian flash (30 days), second revision (120 days) and the figures available two years after the reference quarter. For the euro area, given the small size of the scheduled revisions, only the flash and the release available two years after the reference quarter will be considered in the model's information set.

(7) Because the focus of this paper is on forecasting accuracy, I am not discussing the options for conducting structural analysis for the purpose of identifying the shocks, as typically done in the literature on vector autoregressive modeling.

Figure 2: State-space representation of Financial & Housing model

$$
\left(\begin{array}{c}
x_t^Q-\bar{x}^Q\\
x_t^S-\bar{x}^S\\
x_t-\bar{x}\\ \hline
y_t^Q-\bar{y}^Q\\
y_t^S-\bar{y}^S\\
y_t-\bar{y}\\ \hline
h_t-\bar{h}
\end{array}\right)
=
\left(\begin{array}{cccccccccccc|cccccccccccc|c}
\Lambda_x^Q & 2\Lambda_x^Q & 3\Lambda_x^Q & 2\Lambda_x^Q & \Lambda_x^Q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & \Lambda_x^S & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\Lambda_x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ \hline
\Lambda_y^Q & 2\Lambda_y^Q & 3\Lambda_y^Q & 2\Lambda_y^Q & \Lambda_y^Q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Theta_y^Q & 2\Theta_y^Q & 3\Theta_y^Q & 2\Theta_y^Q & \Theta_y^Q & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Lambda_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & \Theta_y^S & 0\\
\Lambda_y & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Theta_y & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ \hline
\Lambda_h & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Theta_h & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Gamma_h
\end{array}\right)
\left(\begin{array}{c}
f_t\\ f_{t-1}\\ f_{t-2}\\ f_{t-3}\\ f_{t-4}\\ f_{t-5}\\ f_{t-6}\\ f_{t-7}\\ f_{t-8}\\ f_{t-8}\\ f_{t-10}\\ f_{t-11}\\ \hline
h_t\\ h_{t-1}\\ h_{t-2}\\ h_{t-3}\\ h_{t-4}\\ h_{t-5}\\ h_{t-6}\\ h_{t-7}\\ h_{t-8}\\ h_{t-9}\\ h_{t-10}\\ h_{t-11}\\ \hline
\tau_t
\end{array}\right)
+
\left(\begin{array}{c}
\chi_t^Q\\ \chi_t\\ \chi_t^\pi\\ \hline
\psi_t^Q\\ \psi_t\\ \psi_t^\pi\\ \hline
\eta_t
\end{array}\right)
$$

$$
\left(\begin{array}{c}
f_t\\ f_{t-1}\\ f_{t-2}\\ f_{t-3}\\ f_{t-4}\\ f_{t-5}\\ f_{t-6}\\ f_{t-7}\\ f_{t-8}\\ f_{t-8}\\ f_{t-10}\\ f_{t-11}\\ \hline
h_t\\ h_{t-1}\\ h_{t-2}\\ h_{t-3}\\ h_{t-4}\\ h_{t-5}\\ h_{t-6}\\ h_{t-7}\\ h_{t-8}\\ h_{t-9}\\ h_{t-10}\\ h_{t-11}\\ \hline
\tau_t
\end{array}\right)
=
\left(\begin{array}{cccccccccccc|cccccccccccc|c}
A_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_{13}\\
I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ \hline
A_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_{23}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 & 0\\ \hline
A_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_{32} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_{33}
\end{array}\right)
\left(\begin{array}{c}
f_{t-1}\\ f_{t-2}\\ f_{t-3}\\ f_{t-4}\\ f_{t-5}\\ f_{t-6}\\ f_{t-7}\\ f_{t-8}\\ f_{t-8}\\ f_{t-10}\\ f_{t-11}\\ f_{t-12}\\ \hline
h_{t-1}\\ h_{t-2}\\ h_{t-3}\\ h_{t-4}\\ h_{t-5}\\ h_{t-6}\\ h_{t-7}\\ h_{t-8}\\ h_{t-9}\\ h_{t-10}\\ h_{t-11}\\ h_{t-12}\\ \hline
\tau_{t-1}
\end{array}\right)
+
\left(\begin{array}{c}
u_t^f\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ \hline
u_t^g\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ \hline
u_t^h
\end{array}\right)
$$

with $\begin{pmatrix} u_t^f\\ u_t^g\\ u_t^h \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} Q_1 & \Omega_{12} & \Omega_{13}\\ \Omega_{12} & Q_2 & \Omega_{23}\\ \Omega_{13} & \Omega_{13} & Q_3 \end{bmatrix}\right)$ *uncorrelated with the idiosyncratic measurement errors.*

Note: The superscript Q refers to the quarterly variables. The EC and the NBB surveys, which have the 'superscript', will be directly linked to the cumulative sum of the factors over the past 12 months. The filter applied to the the PMI indicators is the same one as for the quarterly variables even if not explicitly mentioned here.

3.3. Details of our modeling approach

Figure 2 displays the state-space representation of the 'Financial & Housing' model, which is based on three blocks of unobserved factors (f_t, b_t, τ_t) and their lags. Such representation, however, remains at a highly abstract level. First of all, note that the vector of variables with upper index Q incorporate real GDP growth rates of alternative indicator variables: flash release, the so-called second release, the figures available two years after the end of each quarter, and the historical series, which dates back to 1980. By allowing the factor loadings to be different for all releases, I acknowledge the possibility of them all referring to different concepts and that the methodology used in their construction can be different. Alternatively, Camacho and Perez-Quiros (2010) or Evans (2005) assume that revisions can be formalized in terms of pure noise (see Mankiw and Shapiro (1986)). In such case, the difference between the GDP flash release and the subsequent revision is assumed to be uncorrelated with the underlying measure of economic activity.

However, this specification has the counterfactual implication that the variance of the flash is larger than the variance of the subsequent revisions. As shown in the annex, the noise hypothesis is overwhelmingly rejected within our sample period for both Belgium and the euro area. More flexible models for the data revisions such as those pioneered by Kishor and Koenig (2012) or Jacobs and Van Norden (2011) are not explored in this paper.

A key issue that needs to be clarified is the total number of factors and the number of lags in the VAR representation. Although the transition equation of figure 2 is a VAR(1), all the empirical results presented in the paper have been obtained with a VAR(4) representation. This choice, which was meant to capture some of the complex interactions between the business cycle and the housing market, does not affect the precision of GDP forecast over out-of-sample evaluation even if the number of parameters increases. In order to render the model as parsimonious as possible, I decided to introduce only one factor in each of the three blocks represented in the figure. Thus, one can think of f_t, b_t and τ_t as three factors and not as three blocks of factors as suggested above. This implies that the number of factor loadings to estimate remains as small as possible.

4. Analysis of news

The 'news' associated with a given release is represented by the discrepancy of the published figure with respect to the forecast of the model. Thus, I will use the words 'news', 'innovations' or 'forecast errors' interchangeably (see Durbin and Koopman (2001), section 4.8). Once the concept of news is clearly defined, I will show how the so-called 'Kalman gain' induces the model to update the forecast path for GDP or any other variable of interest after a given piece of news becomes available. The role played by the news that gradually enters the forecaster's information set is given not only by their quality, but also by their timeliness, which crucially depends on the release calendar. The last part of this section quantifies the precise role of all data releases in forecasting Belgian GDP growth rates.

Let's consider a generic recursive representation, which encompasses any of the model specifications discussed before for the observable indicators:

(7) $\mathrm{y_t} = \Lambda \mathrm{f_t} + \mathrm{e_t}$

(8) $\mathrm{f_{t+1}} = A\mathrm{f_t} + \eta_t$

with normally distributed measurement errors and shocks to the factors:

$e_t \sim N(0, R_t),\ \eta_t \sim N(0, Q_t).$

Defining the information sets

The concept of news can be formalized by specifying information sets that enter the model:

$\mathscr{F}^{old}$ contains all time series available right before the publication of the news. Consider for the sake of simplicity that all observations are available until time t.

$\mathscr{F}^{new}$ includes the previous information set, $\mathscr{F}^{old}$, plus new data corresponding to a given macroeconomic release. Again for the sake of simplicity, one can assume that the release extends by one month, $(t+1)$, two of the indicators, i.e. the PMI release for services ($\mathrm{y_{t+1}^{PMI_s}}$) and for manufactures ($\mathrm{y_{t+1}^{PMI_m}}$).

The forecast for the whole vector of variables y_{t+h} is formulated in our framework in terms of model consistent conditional expectations:

(9) $E_\theta[y_{t+h}|\mathscr{F}^{new}] = E_\theta[y_{t+h}|\mathscr{F}^{old}, \{V_{t+1}\}]$

where the expression on the right-hand side decomposes the new conditioning information set in two orthogonal parts. In this particular example, $V_{t+1} \equiv \mathscr{F}^{new} - \mathscr{F}^{old} = [\mathrm{v_{t+1}^m}\ \ \mathrm{v_{t+1}^s}]'$ incorporates two innovations or news, which are defined as the difference between the released figures and the model's forecasts:

$$\mathrm{v_{t+1}^m} = \mathrm{y_{t+1}^{PMI_m}} - E_\theta[\mathrm{y_{t+1}^{PMI_m}}|\mathscr{F}^{old}]$$

$$\mathrm{v_{t+1}^s} = \mathrm{y_{t+1}^{PMI_s}} - E_\theta[\mathrm{y_{t+1}^{PMI_s}}|\mathscr{F}^{old}]$$

Thus, one could state that, even if the released figures have declined with respect to the recent past, the model could interpret them as good news as long as they are above the values that the model was expecting.

The Kalman filter gain

This news is exploited by the Kalman filter 'gain' in order to update GDP forecasts together with the remaining variables. If we could observe $\mathrm{f_{t+h}}$, obtaining the forecast would be straightforward: $E_\theta[y_{t+h}|\mathscr{F}^{new}] = \Lambda A^{h-1} f_{t+1}$. But unfortunately, the factor $\mathrm{f_{t+1}}$ cannot be observed because only two data releases for $t+1$ are available and they are assumed to contain measurement errors. Thus, the conditional expectation in expression 9 must be developed further:

$$
\begin{aligned}
E_\theta[y_{t+h}|\left\{\mathscr{F}^{old}, V_{t+1}\right\}] &= \Lambda A^{h-1} E_\theta[\mathrm{f}_{t+1}|\left\{\mathscr{F}^{old}, V_{t+1}\right\}] \\
&= \underbrace{\Lambda A^{h-1} E_\theta[\mathrm{f}_{t+1}|\mathscr{F}^{old}]}_{old\,forecast} \\
&+ \Lambda A^{h-1} \underbrace{E_\theta[\mathrm{f}_{t+1}V'_{t+1}] E_\theta[V_{t+1}V'_{t+1}]^{-1}}_{Gain\,(quality,timeliness)} \underbrace{V_{t+1}}_{news}
\end{aligned} \tag{10}
$$

All the expectations are calculated with the smoothed covariance, which is given by the precision of the filter. Interestingly, the product of expectations shown in the expression above defines how the 'news' induces an update ([8]) of the state of the economy, which is represented by f_{t+1}. The precise weight of each one of the innovations at updating the expectations about the state of the economy depends on the quality and the timeliness of the indicators, as will be shown in detail in the annex. By quality, we refer to the correlation of the factor with the innovations $E_\theta[\mathrm{f}_{t+1}V'_{t+1}]$. The role of timeliness, which determines the pattern of missing observations, is also crucial at defining the weights. Thus, it can be easily understood that variables that arrive and enter first the model's information set will receive a larger weight than in the case where they are part of a larger group of data releases. The reason is that in the presence of strong collinearity where all variables incorporate the same signal, only one variable is enough.

Defining the standard impact of news

In this simple example with only two data releases and one factor, the last term of expression 10 can be written in terms of parameters $\sigma_m^2 = var_\theta(v_{t+1}^m)$, $\sigma_s^2 = var_\theta(v_{t+1}^s)$, $\sigma_{ms}^2 = cov_\theta[v_{t+1}^m, v_{t+1}^s]$, $\beta_m = cov_\theta[\mathrm{f}_{t+1}, v_{t+1}^m]$, $\beta_s = cov_\theta[\mathrm{f}_{t+1}, v_{t+1}^s]$ (see annex A).

Thus:

$$
\begin{aligned}
E_\theta[y_{t+h}|\left\{\mathscr{F}^{old}, V_{t+1}\right\}] - E_\theta[y_{t+h}|\left\{\mathscr{F}^{old}\right\}] &= \underbrace{\Lambda A^{h-1} \frac{\beta_m \sigma_s^2 - \beta_s \sigma_{ms}^2}{\sigma_m^2 \sigma_s^2 - \sigma_{ms}^2 \sigma_{ms}^2}}_{impact\ of\ PMI\ manufacturing} v_{t+1}^m \\
&+ \underbrace{\Lambda A^{h-1} \frac{\beta_s \sigma_m^2 - \beta_m \sigma_{ms}^2}{\sigma_m^2 \sigma_s^2 - \sigma_{ms}^2 \sigma_{ms}^2}}_{impact\ of\ PMI\ services} v_{t+1}^s
\end{aligned} \tag{11}
$$

This illustration has served as a vehicle to underline that the whole set of news, i.e. the vector of innovations V_{t+1}, induce an update of the path for all variables in y_t. The extent to which all the individual pieces of news induce change expectations for GDP growth rates in Belgium depends on all the different factors and on the particularities of the calendar of data releases. Quantifying the precise role of all the news is the goal of the next subsection. By multiplying the impacts defined in the equations above by the standard deviation of the news associated with each data release, I obtain a measure that allows me to compare the average informative content of the different indicators. Such

([8]) This update takes the form of a simple OLS regression of the factors on the innovations. Note that the size of the news vector V_{t+1} may be large in practical applications. For example when many variables can be released at the same time, or many observations for the same variable are made available simultaneously.

a measure will be referred to as the 'standard impact' of a release, e.g. PMI, on a given variable of interest, e.g. GDP growth.

4.1. The 'standard impact' of macroeconomic releases

I have defined the 'standard impacts' associated with each one of the news releases as the product of the impact coefficients defined in equation 11 and the standard deviation of the respective innovations, i.e. the RMSE associated with the release of each series. The flow of information within the quarter can be represented by table 1 in section 3. As discussed above, timely releases will tend to have a higher weight, so it is important to take into account the exact release calendar. Given that the calendar is relatively stable over the last seven years, the results from an exercise taking the 2008 calendar as a reference can be considered to be representative enough.

The resulting standard impacts for all data releases are displayed in descending order in table 5. The element in the first row and third column of the table can be read as follows: 'Purchasing Manager Index' releases corresponding to the last month of each quarter, which have a standard deviation of 1.288 units with respect to the model forecast (see the last column), will typically induce revisions in GDP growth forecasts, for the next quarter, of 0.189 percentage points in absolute terms. While GDP growth in a given quarter is actually affected by the PMI releases corresponding to all months in the previous quarter, the impact of the Business Confidence Index published by the National Bank of Belgium (BCI) concentrates in the publications referring to the first month of the current quarter and the last month of the previous quarter. It is remarkable that the 3-month Euribor is the third most important variable for this model, even if we are only taking it into account at the end of each month ([9]).

The second block of table 5 displays the impact of quarterly releases on the real GDP growth rate in Belgium. The most remarkable facts are the relatively large impact of house prices, which contribute to updating GDP growth forecasts several quarters ahead and the economically insignificant impact of the flash releases. As formalized in annex B.1, the latter result implies that the flash release does not help the model to significantly update the state of the economy.

The most puzzling result in table 5 is the relatively small impact of hard data. This will be further investigated in the next subsection. This fact is also visible when I plot the standard impacts of news for more distant forecast horizons (see figure 3). Although the figure analyzes the standard impact of data releases on the last quarter of 2008, the results for other years and quarters remain unchanged if the publication schedule for all indicators does not change. Again, the releases corresponding to the last six months of the year for both the BCI and the PMI releases stand out as the most relevant pieces of information in forecasting growth. In addition to them, innovations in the 3-month Euribor and on the quarterly releases of real house prices, which can have an impact on the forecasts of Belgian growth about more than two years ahead, also have a significant impact.

([9]) Banbura et al. (2013) propose a state-space model with daily factors that allows for the treatment of daily data.

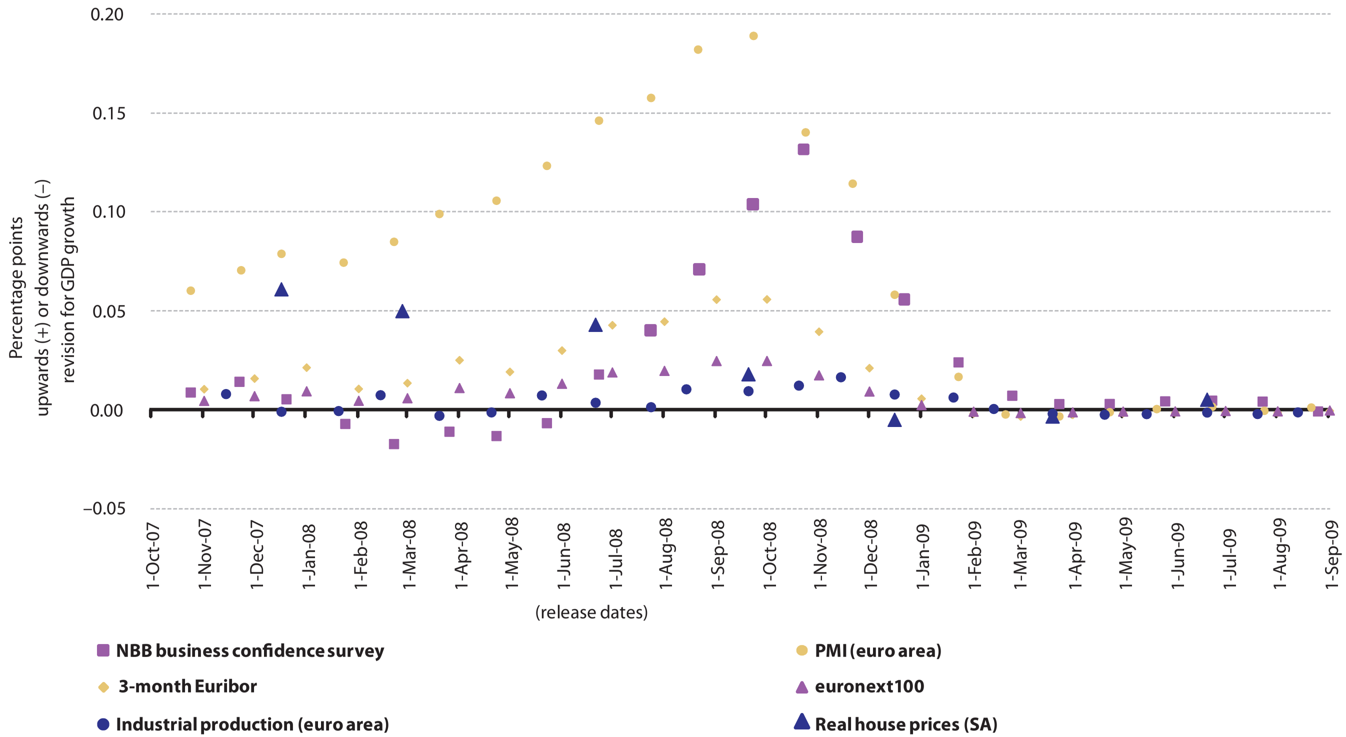
Table 5: Standard impact of news in Belgian GDP growth

	Standard impact = weight x (standard deviation (stdev))						Stdev
	Previous quarter			Current quarter			
	M1	M2	M3	M1	M2	M3	
Purchasing Manager Index (flash, manufacturing, euro area)	0.158	0.182	0.189	0.140	0.114	0.058	1.288
NBB business confidence survey (overall)	0.040	0.071	0.104	0.132	0.087	0.056	2.000
3-month Euribor (euro area)	0.045	0.056	0.056	0.040	0.021	0.006	0.154
Consumer confidence indicator	0.017	0.023	0.027	0.013	0.019	0.008	4.544
Purchasing Manager Index (flash, services, euro area)	0.019	0.023	0.024	0.017	0.015	0.008	1.672
Euronext100 (euro area)	0.020	0.025	0.025	0.018	0.009	0.003	4.650
NBB demand expectations (manufacturing)	0.005	0.012	0.021	0.024	0.020	0.013	4.323
Economic sentiment indicator (euro area)	0.018	0.021	0.022	0.017	0.011	0.004	3.323
HICP (euro area)	0.016	0.022	0.020	0.016	0.007	0.002	0.181
German 10-year government bond yield minus 3-month Euribor (euro area)	– 0.015	– 0.019	– 0.019	– 0.013	– 0.007	– 0.002	0.222
Industrial production (excluding construction, euro area)	0.010	0.012	0.017	0.008	0.006	0.001	0.949
Brussels all shares	0.004	0.009	0.013	0.014	0.009	0.003	4.601
Work volume of temporary workers	0.007	0.009	0.023	0.007	0.003	0.000	9.934
National CPI	0.008	0.011	0.010	0.008	0.004	0.001	0.263
Industrial production (excluding construction)	0.000	0.000	0.007	0.010	0.008	0.002	1.573
Adjusted harmonised unemployment rate	– 0.011	– 0.009	– 0.005	0.000	0.005	0.001	0.002
Total turnover	0.001	0.002	0.005	0.005	0.003	0.001	2.428
Retail sales (deflated turnover in the retail trade, euro area)	0.002	0.003	0.004	0.002	0.002	0.000	0.703
Permits for new residential buildings (SA)	0.002	0.000	0.001	0.001	0.002	0.001	8.861
New residential building starts (SA)	0.000	0.000	0.001	0.001	0.000	0.000	6.346
Registration of new private cars	0.000	0.000	0.000	0.000	0.000	0.000	7.835
Trade balance in goods	– 0.002	– 0.001	– 0.001	0.001	0.001	0.001	916.786
Spreads on Belgian 10-year government bonds	0.005	0.004	0.000	– 0.002	– 0.004	– 0.004	53.967
Total receipts	– 0.001	– 0.001	0.000	0.000	0.001	0.001	4.700

	Standard impact = weight x (standard deviation (stdev))										Stdev
	Q-9	Q-8	Q-7	Q-6	Q-5	Q-4	Q-3	Q-2	Q-1	current Q	
Real house prices (SA)	0.063	0.067	0.062	0.060	0.061	0.050	0.043	0.018	– 0.005	– 0.003	0.930
Flash real GDP growth	0.000	0.000	0.000	– 0.001	– 0.001	0.000	– 0.003	– 0.007	– 0.003	0.011	0.210
Flash real GDP growth (euro area)	0.000	0.000	0.000	– 0.001	– 0.001	0.000	– 0.002	– 0.007	– 0.002	0.012	0.300

Source: Author's calculations

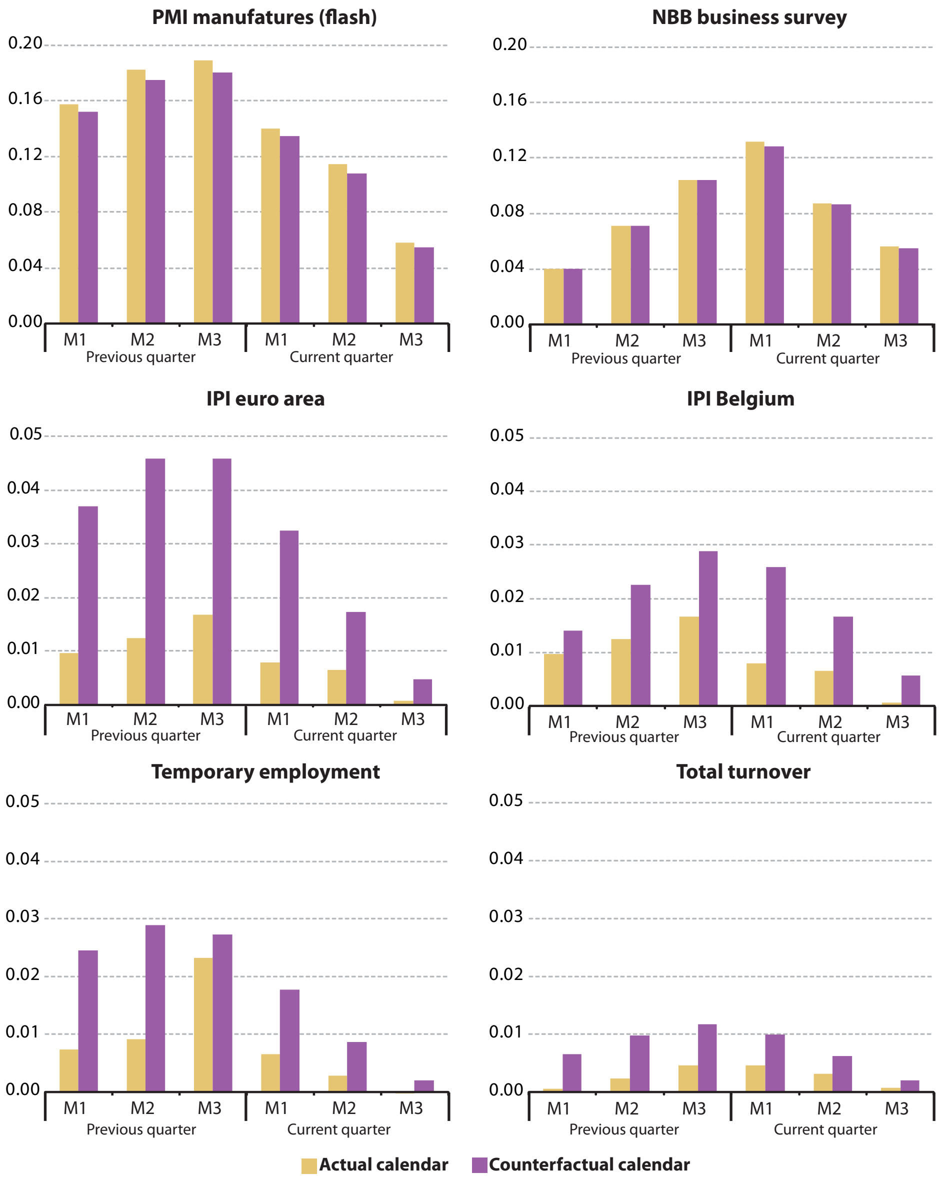
Figure 3: Standard Impact of news on Belgian growth for 2008Q4
(%)



Source:: Author's calculations

Note: This figure represents the impact of a better than expected release of the standard size for several indicators at different points in time. The releases corresponding to the last six month of the year are highlighted with a thicker marker. The survey released by the NBB corresponding to the month of October is expected to have a larger impact than any other release for this indicator on the forecast for the last quarter of the year. Given that the calendar is relatively stable, the same conclusion could be made about any other quarter of the year. The pattern for the PMI releases by Markit Economics is slightly different. Their release corresponding to the month of September has the largest impact. Finally, quarterly releases of real house prices start to have a relatively large impact on Belgian growth forecasts about one year ahead.

Figure 4: Counterfactual 'standard impact' of news on Belgian growth for 2008Q4
(%)



Source: Author's calculations

4.2. Counterfactual analysis of timeliness

We have seen that hard data releases, notably the industrial production index or total turnover, do not seem to have a large contribution in forecasting revisions for Belgian GDP. This is surprising because those series are used in the construction of the official GDP growth releases. Table 1 could solve the puzzle, since it shows that hard data are published with a significant delay. However, large publication lags are only a small part of the story.

In order to understand whether publication lags of hard data, e.g. industrial production, total returns or temporary employment, determine their low impact on GDP forecast revisions, I will compute their impact under the counterfactual assumption that the figures corresponding to a given month are published at the end of each month without any delay. This experiment will enable us to assess whether news associated with the most relevant hard data releases are sufficiently correlated with the factors of the model.

The results displayed in figure 4 reveal that two survey releases with the largest impact, PMI and BCI, continue to have a very large impact at updating growth expectations in Belgium when they have to compete with hard data releases. Interestingly, the standard impact of industrial production innovations in the euro area is about four times larger in the counterfactual absence of publication lags. The figure also exhibits an increase in the standard impacts of industrial production, total turnover and temporary employment in Belgium, but the role of those variables remains very small relative to the surveys.

To conclude, even if timeliness remains an important property of survey data, it is quality that seems to help the most in the current set-up. The role of surveys is well known in the nowcasting literature since Giannone, Reichlin and Small (2008), but this is the first paper that attempts to separate the effects of those of quality from timeliness. The work by Camacho and Perez-Quiros (2010) for the euro area is particularly relevant for our current application, since they place a particular emphasis on the timeliness of the BCI published by the National Bank of Belgium, which receives 100 % of the weight when it is published ([10]). However, they do not clarify the precise role of quality.

Note that this analysis is very different from other studies that investigate the role of qualitative surveys, since they use forecasting accuracy comparisons over a given subsample with and without surveys, without exploiting the full sample information.

5. Real-time forecasting results

We recognize that the analysis of news presented in the previous section, which highlights the crucial role of surveys, may partly be driven by the particular choice of the model. In this section, we test whether such modeling choice is reasonable from the point of view of forecasting accuracy. Thus, the forecasting accuracy of the Financial & Housing model described in section 3.2 will be compared with the Benchmark specifica-

([10]) Rather than focusing on the contribution of the news in updating the forecast, they calculate the weight of the variables on the factor driving GDP. The problem with this approach, which has also been proposed by Banbura and Runstler (2011), is that the contribution of an indicator can increase at the release date even if it does not incorporate any news, i.e. the released figure is in line with expectations. As shown by Banbura and Modugno (2010), by focusing on the news one can obtain a more refined analysis of the sources of forecast revisions.

tion, which does not take into account Financial and Housing market variables.

5.1. Out-of-sample evaluation

Detail of this exercise

A bird's eye view on the most recent real-time forecasting performance of the models is given in figures 5 and 6. In this evaluation setting, the models produced GDP growth forecasts for a given quarter using all the information available at three different points in time:

1. 90 days before the end of the quarter;
2. 30 days before the end of the quarter;
3. 15 days after the end of the quarter, which roughly corresponds to 15 days before the GDP flash estimate is released in Belgium.

The model parameters are estimated exploiting all information available at the first forecasting round and it is not re-estimated again until next quarter. The nowcasts produced by the models are based on the concept of revised GDP growth rate, which is available only two years after the end of the quarter, and can be compared to both the official flash release and the subsequently revised values. The bars in figures 5 and 6 visualize both GDP growth concepts and show that the decline of the real GDP third quarter of 2008, first such decline since the onset of the Great Recession, could be anticipated by both models prior to the official releases.

The performance of the model is often evaluated in terms of the (squared) root of the average mean squared out-of-sample forecast errors (RMSE) over the period 2008Q1–2012Q2, and compared to competing forecasts. The error is defined here as the difference between the nowcast and the GDP figures available two years after the initial flash release. The hypothesis of equal forecasting accuracy among different models will be tested using the Diebold and Mariano (DM henceforth) (1995) test. Although many alternative tests have been proposed and some of them are particularly suitable for real-time forecasting applications, we emphasize that our goal remains to compare forecasts and not models. Thus, the conditions under which the test is valid remain relatively weak, as noted by Diebold (2012).

Results

Given the real-time nature of this forecasting exercise, I need to make such a comparison at different points in time. The graphs in figure 7 represent the decreasing pattern of the RMSE as more and more information is incorporated into the models' information sets. As shown in the DM test results in the second column of table 6, both models perform significantly ([11]) better than the Belgian Prime News forecast, which is distributed among the participants almost 30 days before the end of each quarter together with expectations for the annual growth rate. More strikingly, the third column of table 6

([11]) DM test rejects the equal forecast accuracy hypothesis with a 80 % confidence level. Defining the time-t loss differential between two given forecasts 1 and 2 as $L_t^{12} = e_{1,t}^{\ 2} - e_{2,t}^{\ 2}$, the Null hypothesis of equal forecast accuracy can be expressed as $h0 : E[L_t^{12}] = 0$. Under this assumption, the Diebold-Mariano test statistic, $DM^{12} = \bar{L}^{12}/\hat{\sigma}_{\bar{L}^{12}}$, follows a standard normal distribution. The sample mean loss differential is defined as $\bar{L}^{12} = \frac{1}{T}\sum_{t=1}^{T} L_t^{12}$ and $\hat{\sigma}_{\bar{L}^{12}}$ is a consistent estimate of its standard deviation.

shows that both models' nowcasts are as precise as the flash estimate itself, even when the former are obtained two months before the end of the quarter. It should be recalled from the news analysis that surveys released during the first month of the quarter have a very strong influence on the forecasts. The RMSE results reveal that such information is enough to produce a reliable estimate of growth for the quarter. As a matter of fact, the second test displayed in table 6 actually shows that the hypothesis that the nowcasts are as accurate as the flash can only be rejected at more distant horizons that the 165 days horizons represented in the first two rows of the table. Finally the third test displayed in the last column of the figure suggests that the Financial & Housing specification is more accurate than the Benchmark, but only at very short-term horizons.

Table 6: Diebold-Mariano test

	Delay in days	RMSE	DM test 1 (Prime N)	DM test 2 (flash)	DM test 3 (Pairwise)
Benchmark	– 165	0.96	1.96 (***)	1.93 (**)	– 1.71 (**)
Financial & Housing	– 165	1.01	2.21 (***)	2.15 (***)	
Benchmark	– 90	0.62	0.61	1.19	– 0.74
Financial & Housing	– 90	0.64	0.75	1.23	
Benchmark	– 30	0.37	– 1.33 (*)	– 0.06	1.53 (*)
Financial & Housing	– 30	0.33	– 1.47 (*)	– 0.57	
Benchmark	+ 15	0.40	– 1.09	0.39	1.29 (*)
Financial & Housing	+ 15	0.35	– 1.4 (*)	– 0.37	
Prime News	– 25	0.57			
Flash	+ 30	0.37			

Source: Author's calculations

Note: The DM test statistic rejects the null hypothesis of equal forecast accuracy at significance levels of 95 % (***), 90 % (**) and 80 % (*). The RMSE of the different models is displayed in the first column of the table. The second column contains the DM test comparing the models with the Prime News forecasts, while the third column compares the models forecasts errors with the ones obtained when the flash is interpreted as a nowcast. Finally, the last column implements pairwise comparisons between the Benchmark and the Financial & Housing models at given horizons.

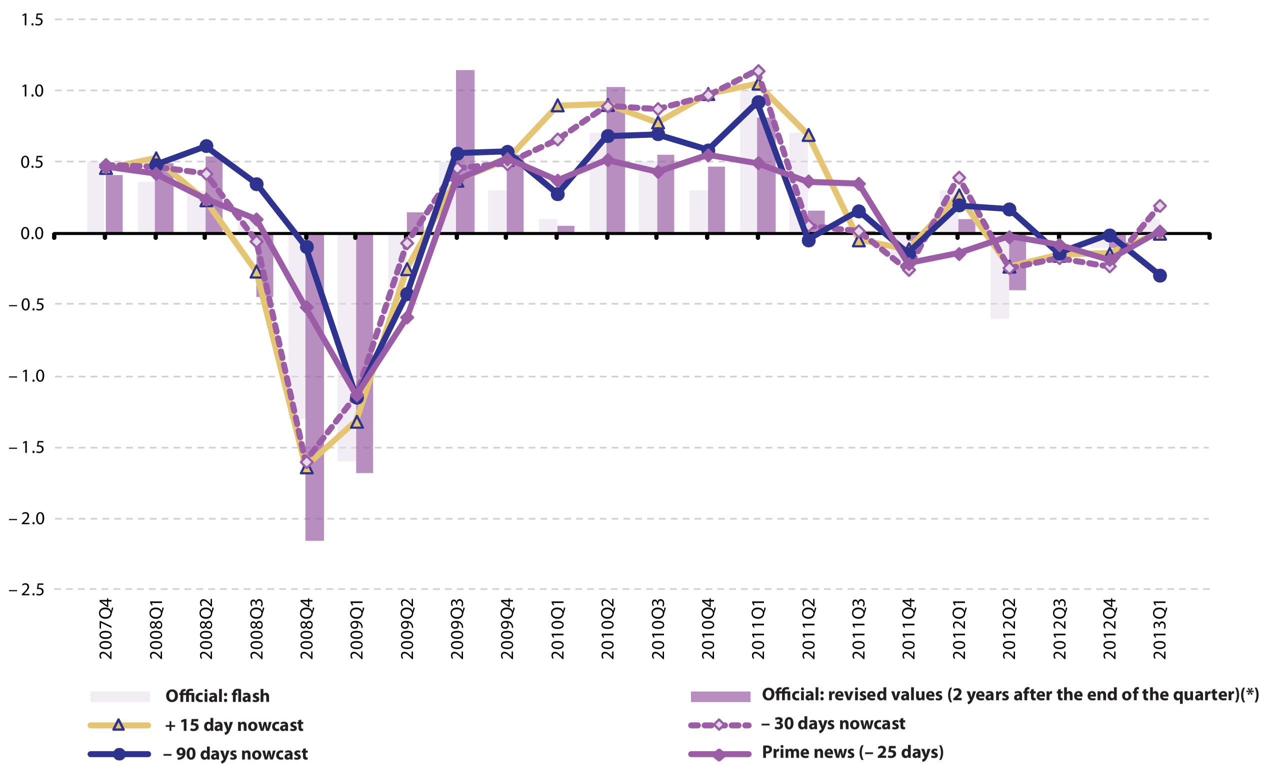
5.2. Evaluation of fixed events

Here, I show that any of the two models can be used to update the forecasts as an increasing amount of information becomes available. We have seen that both the Benchmark and the Financial & Housing models turn out to offer a very similar performance (on average). However, the presence of real house prices and 3-month Euribor in the second model induces a higher sensitivity to news at the cost of more volatile forecasts for 2012Q2, as depicted in figure 8.

Because fluctuations in house prices have an effect only on medium term growth expectations ([12]), both models can offer very similar forecasts at short term horizons. Figure 8 illustrates the relative performance of both models at forecasting 2008Q3, which eventually turned out to be the first quarter of negative growth rate, and 2012Q2, which was a negative surprise both for the National Bank of Belgium and for the European Commission.

([12]) see standard impacts in table 5.

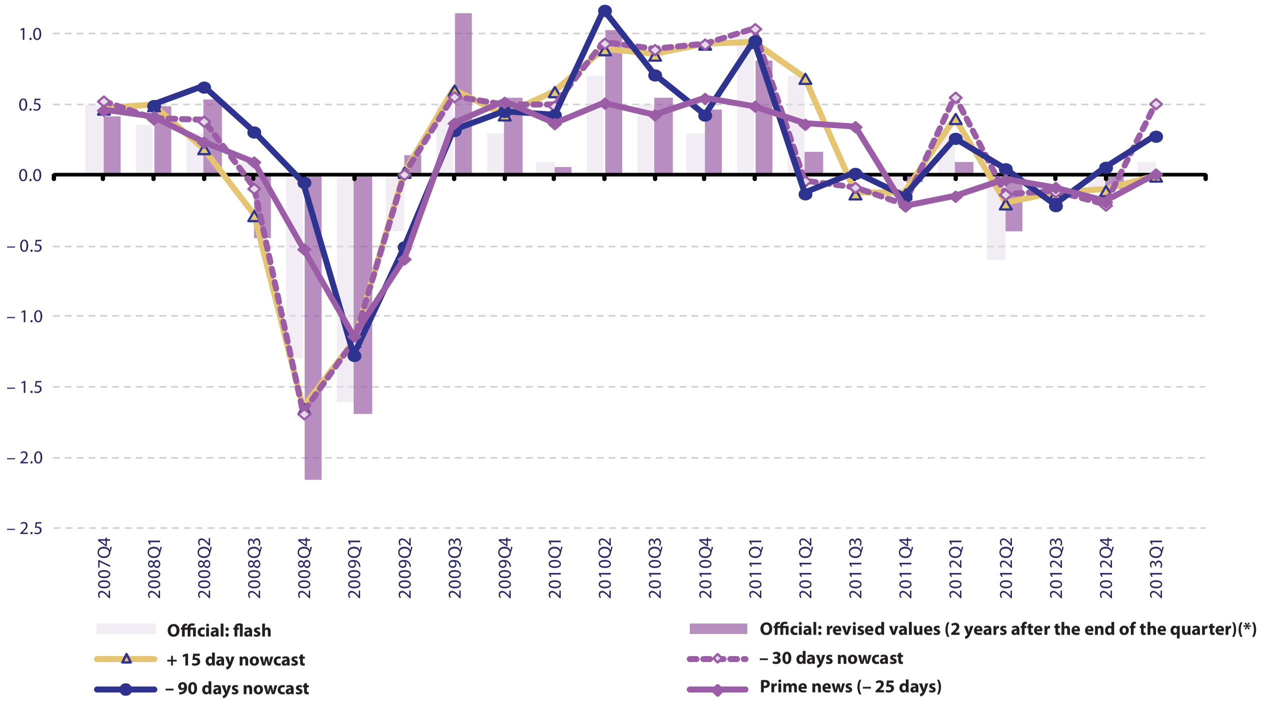
Figure 5: Visualizing the real-time forecasts at different horizons (benchmark model)



Source: Author's calculations

(*) From 2011Q1 onwards, I take the last available vintage.

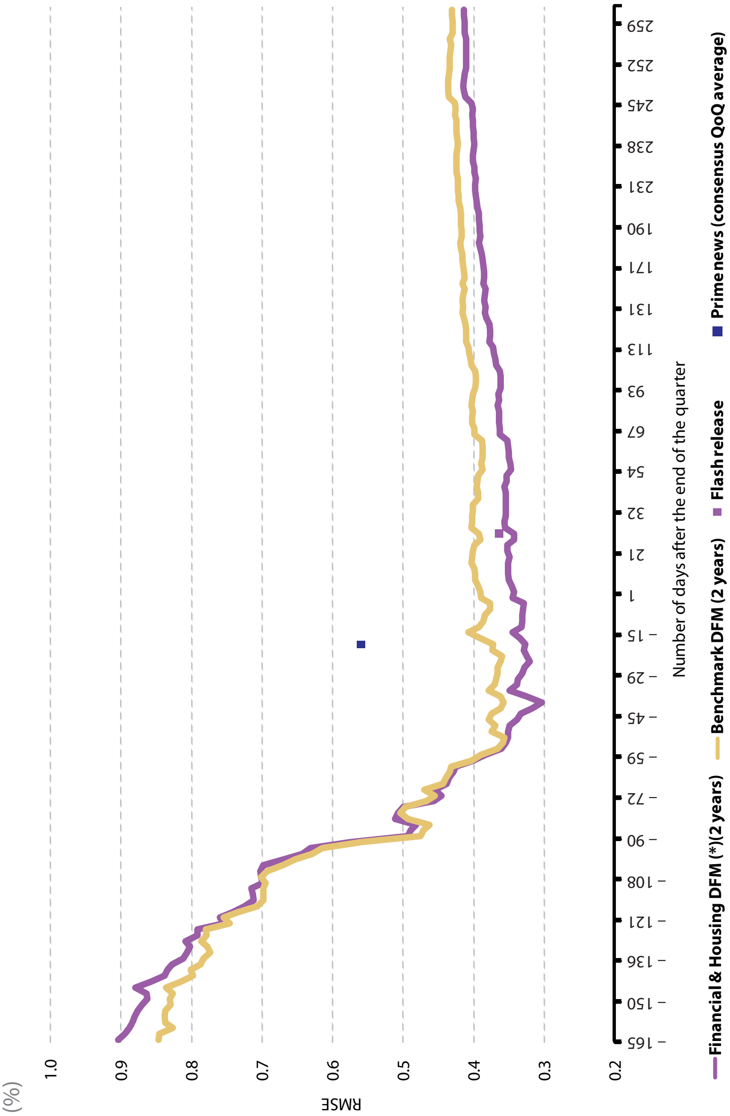
Figure 6: Visualizing the real-time forecasts at different horizons (Financial & Housing model)
(%)



Source: Author's calculations

(*) From 2011Q1 onwards, I take the last available vintage.

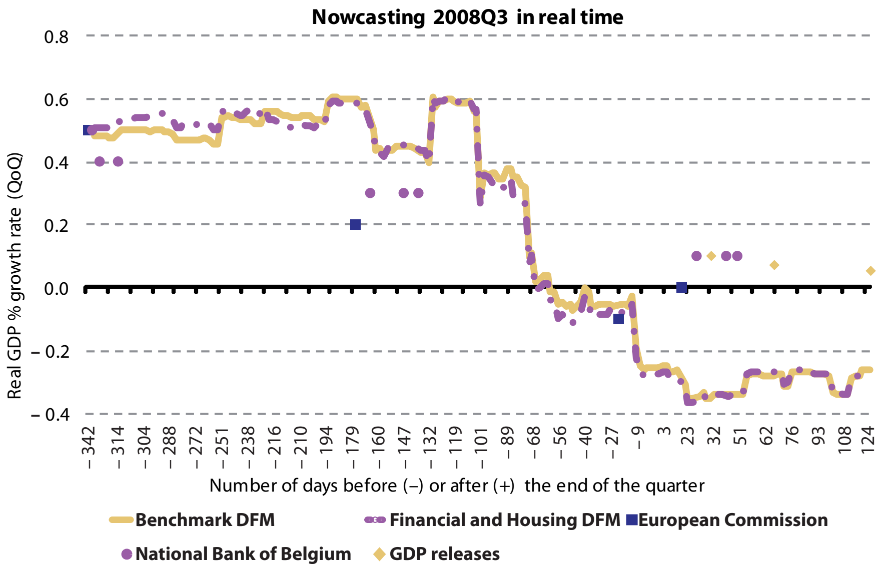
Figure 7: RMSE over time for the two models (evaluation sample: 2008Q1–2012Q2)
(%)

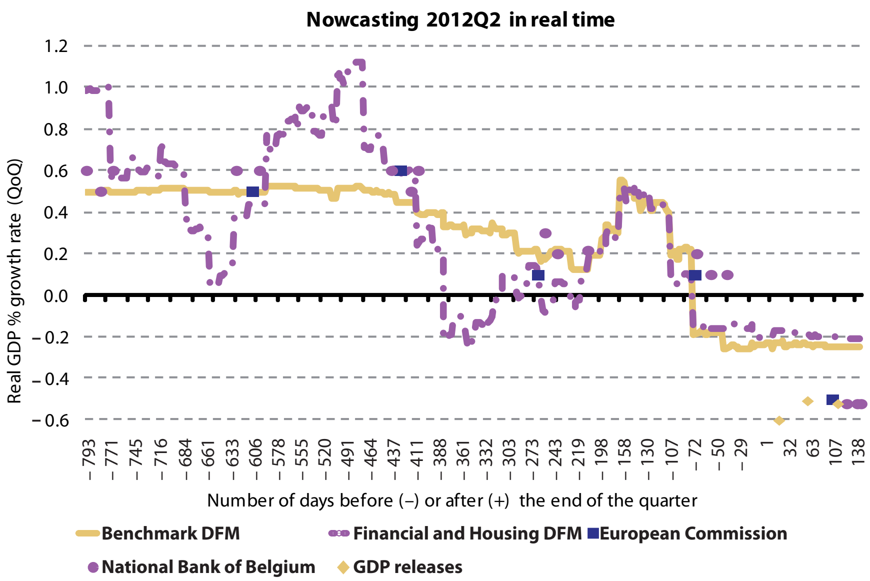


Source: Author's calculations

(*) Dynamic Factor Model

Figure 8: Comparing the two models
(%)

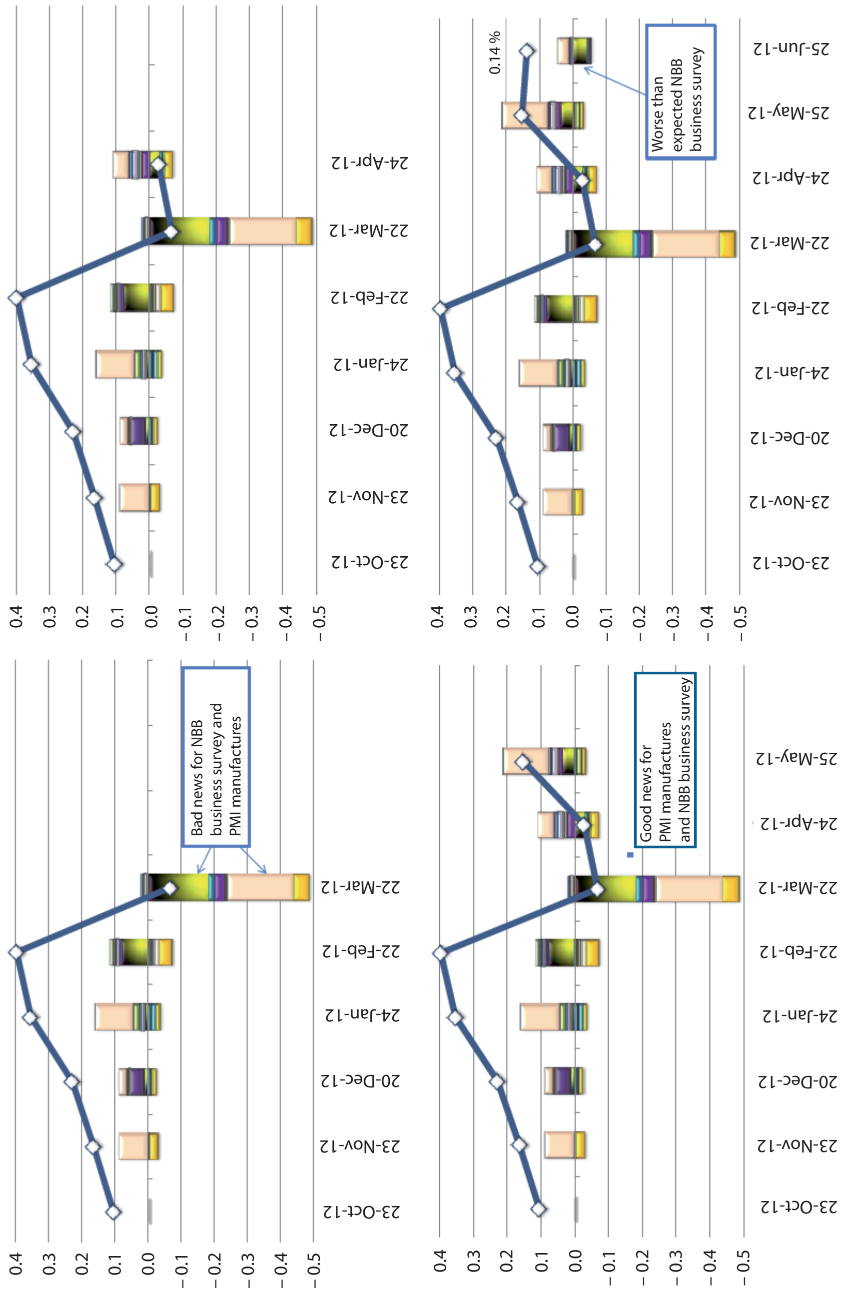




Source: Author's calculations

Note that the x-axis represents the actual date when the forecast was made. The NBB forecasts are always constructed in three rounds, which are not publicly available in terms of quarter-on-quarter growth rates. The EC forecasts are reported on the European Commission's website.

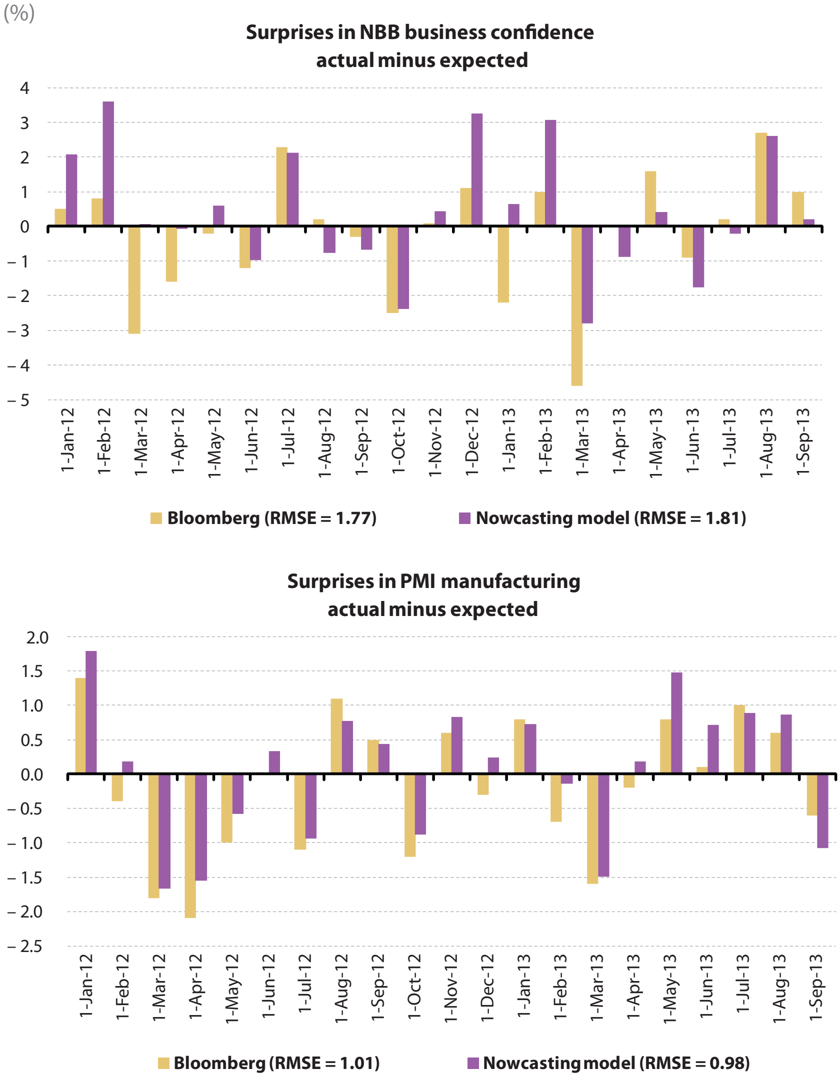
Figure 9: Anticipating the GDP growth for 2013Q2
(%)



Source: Author's calculations

Figure 10: Financial & Housing model versus Bloomberg

(%)



Source: Author's calculations

5.3. Nowcasting in practice

Given the supremacy of survey data documented in the previous section, we could consider one of the simplest possible uses of the model: running it only after the most relevant surveys have been released. This subsection illustrates how the Financial & Housing model works at updating the forecasts by the end of each month, when the most relevant survey data is released.

The model, which is re-estimated only once a year, reads the information set available each month and updates the forecasts for growth in Belgium. The target variable is the real GDP growth release that will be made available after two years and not the flash release.

The first graph in figure 9 shows that the information available on March 22 is read as bad news with respect to the information set available one month ago. Given the large impact of surveys, which I have documented in the previous section, it is not surprising to have such a significant revision in the forecast. By looking at the four panels of figure 9, we can see the news interpretation provided by the model month by month.

The underlying news that we obtained in real time over the most recent months, which have been defined in section 4 as the difference between the released figures and the values expected by the model, i.e. forecast errors, can actually be compared to the errors produced by Bloomberg analysts. According to figure 10, most of the times that Bloomberg reads good news, the model will provide the same interpretation. There are, however, some important exceptions.

The BCI released by the National Bank of Belgium for March and April 2012 was interpreted by Bloomberg as a negative surprise of 3 units, while the factor model was interpreting it as neutral or no news.

6. Conclusion

In this paper, I propose a joint state-space model for the euro area and Belgian economies formalizing the role of the intra-quarterly data flow as an input in constructing early estimates of GDP growth and updating them in real-time. Those updates are given by the unexpected component included in each one of the macroeconomic releases. The impact that those surprises have on the forecasts is precisely determined by important properties of the statistical releases, such as timeliness and quality, which can be explicitly expressed as a function of the model parameters. Those objective weights that the model gives to each data release insures the analysts against the human tendency to favour information that confirms their beliefs or hypotheses.

The empirical results underline the importance of survey data such as the Business Confidence Index constructed by the National Bank of Belgium, the Markit Economics PMI (Manufactures) release for the euro area, 3-month Euribor and real house prices in Belgium, which turn out to contribute mainly at long-term horizons. The large impact of the survey releases that refer to the first month of a given quarter is consistent with the empirical finding that three months prior to the publication of the Belgian flash, the nowcast turns out to be as accurate as the flash release itself. The paper goes further than the literature in understanding whether the importance of survey data can be attributed to their relative timeliness or rather their quality. In a counterfactual exercise, I show that the weights associated with survey data do not deteriorate when all hard data is artificially published with the same degree of timeliness. This result underlines quality as a relevant property of survey data.

Acknowledgements

This work can be replicated with JDemetra+, thanks to Jean Palate and my colleagues at the R & D cell. The nowcasting plug-in that we have developed for JDemetra+ allows anyone to reproduce and extend the model estimation, news analysis and forecasting simulation results presented in this paper (Downloads: https://github.com/nbbrd/jdemetra-nowcasting/wiki). This paper has benefited from Domenico Giannone's advice during his frequent visits to the National Bank of Belgium and from Raf Wouters' guidance throughout the entire project cycle. I have also benefited from the interaction with Andrea Silvestrini, Philippe Jeanfils, Michele Modugno, Geert Langenus and Isabelle Brumagne, Christophe Piette and from all suggestions from seminar participants at ECARES, CORE, National Bank of Belgium, Shanghai University of Finance and Economics (9th Dynare Conference), and the NTTS 2015.

References

Abberger, K. (2007), 'Qualitative business surveys and the assessment of employment a case study for Germany', *International Journal of Forecasting*, 23, pp. 249–258.

Angelini, E., G. Camba-Mendez, D. Giannone, G. Runstler, and L. Reichlin (2011), 'Short-Term Forecasts of Euro Area GDP Growth', *The Econometrics Journal*, 1, C25–C44.

Banbura, M., D. Giannone, M. Modugno and L. Reichlin (2013), 'Nowcasting and the Real-Time Data Flow', *Handbook of Economic Forecasting*, in G. Elliot and A. Timmermann, 2, Elsevier.

Banbura, M., D. Giannone and L. Reichlin (2011), 'Nowcasting', *Oxford Handbook on economic Forecasting*, in Michael P. Clements and David F. Hendy, 2011.

Banbura, M. and M. Modugno (2010), 'Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data', *ECB Working Paper Series*, 1189.

Banbura, M. and G. Runstler (2011), 'A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP', *International Journal of Forecasting*, 27, pp. 333–346.

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

Camacho, M. and G. Perez-Quiros (2010), 'Introducing the Euro-Sting', *Journal of Applied Econometrics*, 25, 2010, pp. 663–694.

Claveria, O., E. Pons and R. Ramos (2007), 'Business and consumer expectations and macroeconomic forecasts', *International Journal of Forecasting*, 23, pp. 47–69.

Croushore, D. and T. Stark (2002), 'Forecasting with a Real-Time Data Set for Macroeconomists', *Journal of Macroeconomics*, 24, pp. 507–531.

D'Agostino, A. and B. Schnatz (2012), 'Survey-based nowcasting of US growth: a real-time forecast comparison over more than 40 years', *ECB Working Paper Series*, 1455.

De Greef, I. and Ch. Van Nieuwenhuyze (2009), 'The National Bank of Belgium as new business survey indicator', *Economic Review*, NBB, June, pp. 35–51.

Diebold, F.X. and R.S. Mariano (1995), 'Comparing Predictive Accuracy', *Journal of Business and Economic Statistics*, 13, pp. 253–265.

Diebold, F.X. (2012), 'Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests', mimeo, University of Pennsylvania.

De Mol, C., D. Giannone and L. Reichlin (2008), 'Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components?', *Journal of Economerics*, 146, pp. 318–328.

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

Durbin, J. and S.J. Koopman (2001), 'Time Series Analysis by State Space Methods', Oxford University Press.

Evans, M.D.D. (2005), 'Where are we now? Real-Time Estimates of the Macroeconomy', *International Journal of Central Banking*, 1.

Fixler, D.J. and J.J. Nalewaik (2007), 'News, Noise, and Estimates of the "True" Unobserved State of the Economy', *Finance and Economics Discussion*, Series 2007-34, Board of Governors of the Federal Reserve System (U.S.).

Gayer, A., A. Girardi and A. Reuter (2014), 'The role of survey data in nowcasting euro area GDP growth', *Economic Papers*, 538, December.

Giannone, D., L. Reichlin and S. Simonelli (2009) 'Nowcasting Euro Area Economic Activity in Real Time: The Role of Confidence Indicators', *National Institute Economic Review*, 210, pp. 90–97.

Giannone, D., L. Reichlin and D. Small (2008), 'Nowcasting: The Real-Time Informational Content of Macroeconomic Data Releases', *Journal of Monetary Economics*, 55, pp. 665–676.

Jacobs, J.P.A.M. and S. Van Norden (2011), 'Modeling data revisions: Measurement error and dynamics of true values', *Journal of Econometrics*, 161, pp. 101–109.

Kalman, R.E (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering*, 82, pp. 35–45.

Kishor, N.K., and E.F. Koenig (2012), 'VAR Estimation and Forecasting When Data are Subject to Revision', *Journal of Business and Economic Statistics*, 30, pp. 182–190.

Koenig, E. F. (2002), 'Using the Purchasing Managers a Index to Assess the Economy as Strength and the Likely Direction of Monetary Policy', Federal Reserve Bank of Dallas Economic and Financial Policy Review, 1, pp. 2–14.

Luciani, M. (2014), 'Forecasting with Approximate Dynamic Factor Models: the role of non-pervasive shocks', *International Journal of Forecasting*, 30, pp. 20–29.

Lui, S., J. Mitchell and M. Weale (2011), 'Qualitative business surveys: signal or noise?', *Journal of the Royal Statistical Society*, Series A, 174, 2011, pp. 327–348.

Mankiw, N.G. and M.D. Shapiro (1986), 'News or Noise? An Analysis of GNP Revisions', *Survey of Current Business*, May, pp. 20–25.

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

Martinsen, K., F. Ravazzolo and F. Wulfsberg (2014), 'Forecasting Macroeconomic Variables Using Disaggregate Survey Data', *International Journal of Forecasting*, 30, pp. 65–77.

Piette, C. and G. Langenus (2014), 'Using BREL to now-cast the Belgian business cycle: the role of survey data', *Economic Review*, forthcoming, National Bank of Belgium.

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

The Wall Street Journal (1999), 'Euroland discovers a surprise indicator: Belgian confidence', 14 July.

Vanhaelen, J.J., L. Dresse and J. De Mulder (2000), 'The Belgian industrial confidence indicator: Leading indicator of economic activity in the euro area?', NBB Working Paper, 12.

Annex A — The role of quality and timeliness

The process of updating the forecasts in response to news is formalized by equation 10 above. In state-space modeling terminology, the innovations are weighted by the so-called 'gain' of the Kalman filter in order to have an update of the state vector, i.e. the unobserved factors. In addition to this, the factor loadings Λ and the matrix accounting for the transition of the factors A play an important role in determining the sign, the impact, and the smoothness in the propagation of macroeconomic news into the forecast.

The focus of this subsection is on the gain, since it enables the forecast updates to be expressed in terms of the different types of news. In line with the nowcasting practice, this framework allows data releases with high quality and timeliness to have a large weight in the forecast revisions. To understand why, I will express the gain in equation 10 as a function of parameters implicit in the definition of the generic state-space representation given by equations 7 and 8:

$$
\begin{aligned}
E_\theta[\mathrm{f}_{t+1}V'_{t+1}]E_\theta[V_{t+1}V'_{t+1}]^{-1}V_{t+1} = &\begin{pmatrix} cov_\theta[\mathrm{f}_{t+1}, v^m_{t+1}] & cov_\theta[\mathrm{f}_{t+1}, v^s_{t+1}] \end{pmatrix} \\
&\times \begin{pmatrix} var_\theta[v^m_{t+1}] & cov_\theta[v^m_{t+1}, v^s_{t+1}] \\ cov_\theta[v^m_{t+1}, v^s_{t+1}] & var_\theta[v^s_{t+1}] \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{v}^{\mathrm{m}}_{\mathrm{t+1}} \\ \mathrm{v}^{\mathrm{s}}_{\mathrm{t+1}} \end{pmatrix}
\end{aligned} \tag{12}
$$

One can develop this expression further under the simplifying assumption that there is only one factor. As a result, a very simple formula for the weights associated with the two pieces of news can be obtained. It is worth simplifying the notation: $\sigma^2_m = var_\theta(v^m_{t+1})$, $\sigma^2_s = var_\theta(v^s_{t+1})$, $\sigma^2_{ms} = cov_\theta[v^m_{t+1}, v^s_{t+1}]$, $\beta_m = cov_\theta[\mathrm{f}_{t+1}, v^m_{t+1}]$, $\beta_s = cov_\theta[\mathrm{f}_{t+1}, v^s_{t+1}]$. The resulting expression has a very simple form:

$$
\begin{aligned}
E_\theta[\mathrm{f}_{t+1}V'_{t+1}]E_\theta[V_{t+1}V'_{t+1}]^{-1}V_{t+1} &= \\
&= \frac{\beta_m\sigma^2_s - \beta_s\sigma^2_{ms}}{|cov_\theta[V]|} v^m_{t+1} + \frac{\beta_s\sigma^2_m - \beta_m\sigma^2_{ms}}{|cov_\theta[V]|} v^s_{t+1} \\
&= \underbrace{\frac{\beta_m\sigma^2_s - \beta_s\sigma^2_{ms}}{\sigma^2_m\sigma^2_s - \sigma^2_{ms}\sigma^2_{ms}}}_{w_1} v^m_{t+1} + \underbrace{\frac{\beta_s\sigma^2_m - \beta_m\sigma^2_{ms}}{\sigma^2_m\sigma^2_s - \sigma^2_{ms}\sigma^2_{ms}}}_{w_2} v^s_{t+1}
\end{aligned} \tag{13}
$$

I will now describe how those weights are associated with both the quality and timeliness of the indicators. Note that in the event that the two innovations are perfectly correlated ($\frac{\sigma^2_{ms}}{\sigma_m\sigma_s} = 1$), the determinant $|cov_\theta[V]|$ is equal to zero. This has the practical implication that the weights defining the Kalman gain are not uniquely identified, i.e. giving all the weight to one indicator to the detriment of the alternative yields the same forecasts as giving the same weight to both indicators.

Quality

Consider the possibility that the manufacturing index, $\mathrm{y}^{\mathrm{PMI_m}}_{\mathrm{t+1}}$, and the services index, $\mathrm{y}^{\mathrm{PMI_s}}_{\mathrm{t+1}}$, are published simultaneously. In order to better understand the importance of relative quality, let us rewrite expression 13 by writing the two pieces of news in terms of their driving forces, i.e. the factor innovations and the measurement errors:

(14) $v_{t+1}^{m} = \beta_m(f_{t+1} - E_\theta[f_{t+1}|\mathscr{F}^{old}]) + \sigma_{\varepsilon^m}\varepsilon_{t+1}^{m}$

(15) $v_{t+1}^{s} = \beta_s(f_{t+1} - E_\theta[f_{t+1}|\mathscr{F}^{old}]) + \sigma_{\varepsilon^s}\varepsilon_{t+1}^{s}$

where f_t, ε_t^s and ε_t^m are uncorrelated to each other and their variance is normalized to one. Thus, σ_{ε^m} and σ_{ε^s} dene the standard deviations of the measurement errors present in the two series. The weights that expression 13 would yield have a very simple form in this case:

(16) $$w_1 = \frac{\beta_m\sigma_{\varepsilon^s}^2}{\beta_m^2\sigma_{\varepsilon^s}^2 + \beta_s^2\sigma_{\varepsilon^m}^2 + \sigma_{\varepsilon^m}^2\sigma_{\varepsilon^s}^2}$$

(17) $$w_2 = \frac{\beta_s\sigma_{\varepsilon^m}^2}{\beta_m^2\sigma_{\varepsilon^s}^2 + \beta_s^2\sigma_{\varepsilon^m}^2 + \sigma_{\varepsilon^m}^2\sigma_{\varepsilon^s}^2}$$

This parameterization enables us to observe, more clearly than in expression 13, that the weight associated with one indicator not only depends on its quality, but also on the quality of the competing data releases. Thus, the weight (w_1) associated with the innovation in the manufacturing sector increases with the size of the measurement errors in the services sector (σ_{ε^s}).

In order to understand the role played by quality, we can consider two extreme cases:

Case 1: The manufacturing release does not contain measurement errors, i.e. $\sigma_{\varepsilon^m}^2 = 0$. This implies that the corresponding innovation is perfectly correlated with the factor. In this case all the weight is attached to the news in the manufacturing sector, leaving the services release with zero weight:

$$\begin{aligned} w_1 &= \frac{1}{\beta_m} \\ w_2 &= 0 \end{aligned}$$

The weight associated with the manufacturing release is equal to one divided by the standard deviation of the news.

Case 2: The variance of the measurement errors for the manufacturing and the services releases tends to zero, which implies that both innovations tend to be perfectly correlated with the factor. This is a slightly more complex case because, in the limit, the covariance matrix of the innovations is not invertible, which implies that the weights in expression 13 cannot be obtained. The limit of the weights when both $\sigma_{\varepsilon^m}^2$ and $\sigma_{\varepsilon^s}^2$ tend to zero can easily be derived as follows:

(18) $$\begin{aligned} \lim_{\sigma_{\varepsilon^m}^2,\sigma_{\varepsilon^s}^2\to 0} w_1 &= \left[\frac{1}{\lim_{\sigma_{\varepsilon^m}^2,\sigma_{\varepsilon^s}^2\to 0} w_1}\right]^{-1} = \frac{\beta_m}{\beta_m^2+\beta_s^2} \\ \lim_{\sigma_{\varepsilon^m}^2,\sigma_{\varepsilon^s}^2\to 0} w_2 &= \left[\frac{1}{\lim_{\sigma_{\varepsilon^m}^2,\sigma_{\varepsilon^s}^2\to 0} w_2}\right]^{-1} = \frac{\beta_s}{\beta_m^2+\beta_s^2} \end{aligned}$$

The results derived from our simple parameterization of the model suggest that the weights associated with the two innovations are determined by their covariance with the factor, i.e. β_m and β_s, respectively.

Timeliness

The impact of $\mathrm{y}_{t+1}^{\mathrm{PMI_m}}$ would be very different if it was published one day earlier than the services index. This implies that only the news component v_{t+1}^{m} will be incorporated in $\mathscr{F}^{new}$

$$(19)\quad E_\theta[y_{t+h}|\mathscr{F}^{new}] = E_\theta[y_{t+h}|\mathscr{F}^{old}, v_{t+1}^{m}]$$

Applying equation 10 to the case where v_{t+1}^{m} alone enters the new information set defines a slightly simpler Kalman gain than the one shown in expression 12:

$$(20)\quad \begin{aligned} E_\theta[y_{t+h}|\mathscr{F}^{new}] &= E_\theta[y_{t+h}|\mathscr{F}^{old}, v_{t+1}^{m}] \\ &= \underbrace{\Lambda A^{h-1}E_\theta[\mathrm{f}_{t+1}|\mathscr{F}^{old}]}_{old\,forecast} + \Lambda A^{h-1}\underbrace{E_\theta[\mathrm{f}_{t+1}v_{t+1}^{m}{}'] E_\theta[v_{t+1}^{m}v_{t+1}^{m}{}']^{-1}}_{Gain\,(quality,timeliness)}\underbrace{v_{t+1}^{m}}_{news} \\ &= \Lambda A^{h-1}E_\theta[\mathrm{f}_{t+1}|\mathscr{F}^{old}] + \Lambda A^{h-1}\underbrace{\frac{\beta_m}{\sigma_m^2}}_{w_1^*} v_{t+1}^{m} \end{aligned}$$

As before, the natural parameterization of the innovations represented in expressions 14 and 15 yields to a simple expression for the weight associated to the 'manufacturing' news release:

$$(21)\quad w_1^* = \frac{\beta_m}{\beta_m^2 + \sigma_{\varepsilon m}^2}$$

My previous focus on quality suggests the size of the measurement errors is an important determinant of weight associated to the news. Now, I emphasize the importance of timeliness. By computing expressions 21 and 16 for the case where both releases have the same quality $\beta_m = \beta_s = \beta$ and $\sigma_m^2 = \sigma_{\varepsilon^s}^2 = \sigma^2$, i.e. both the measurement errors associated to the two releases and the covariance of the factor with both innovations are identical, we obtain:

$$\begin{aligned} w_1 &= \frac{\beta\sigma_\varepsilon^2}{2\beta^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2\sigma_\varepsilon^2} \\ w_1^* &= \frac{\beta}{\beta^2 + \sigma_\varepsilon^2} = \frac{\beta\sigma_\varepsilon^2}{\beta^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2\sigma_\varepsilon^2} \end{aligned}$$

The denominator of the first expression is always larger, which implies the weight associated with the 'manufacturing' index is smaller when the 'services' index release occurs simultaneously. The intuition is that when the news in manufacturing does not have to 'compete' with the services release, they receive more attention. By comparing expressions 16 and 21, it can easily be shown that in the case where there is no measurement error in the manufacturing data, i.e. $\sigma^2_{\varepsilon m} = 0$, timeliness does not matter, i.e. $w_1 = w_1^*$.

While timeliness and quality can be defined as independent properties of the data releases, the interaction of those two properties is an essential determinant of their relevance, which I represent here by the weights derived from the Kalman gain.

Annex B — The flash release and subsequent GDP revisions

The flash release for GDP is considered by the model as an individual time series, together with both the so-called second release and the figures available two years after the end of each quarter. By allowing the factor loadings to be different for all releases, I acknowledge the possibility that they all refer to different concepts and that the methodology used in their construction can be different.

B.1. Improving the signal about the state of the economy

An important advantage of incorporating the flash release as a variable in the model is that one can easily quantify how much value is added to the information set previously available, $\mathscr{F}^{old}$, which does not contain the flash. Let us define $V_t = y_t^{flash} - E_\theta[y_t^{flash}|\mathscr{F}^{old}]$ as the innovation associated with the flash release or the difference between the GDP growth expected just before the flash release and its realized value.

Following the usual decomposition of information, the relevance of the official flash release in updating the estimate of growth will depend on the gain and on the magnitude of the innovation V_t:

$$
\begin{aligned}
y_t^{flash} &= \lambda E_\theta[\mathrm{f_t}|\left\{\mathscr{F}^{old}, V_t\right\}] + E_\theta[e_t^{flash}|\left\{\mathscr{F}^{old}, V_t\right\}] \\
&= \underbrace{\lambda E_\theta[\mathrm{f_t}|\mathscr{F}^{old}]}_{expected\ flash} + \underbrace{\lambda E_\theta[\mathrm{f_t}V_t']E_\theta[V_tV_t']^{-1}V_t}_{improvement\ of\ the\ signal} + \underbrace{E_\theta[e_t^{flash}V_t']E_\theta[V_tV_t']^{-1}V_t}_{estimated\ noise\ in\ flash\ release}
\end{aligned}
\tag{22}
$$

where f_t here stands without loss of generality for the whole vector of factors associated with the 'flash' release and the lags of those factors (see measurement equation 5) and λ represents the factor loadings. This expression suggests that both the underlying factors and the measurement error can account for the discrepancy between the flash release and what the model expected. The first component improves the signal about the state of the economy thanks to the so-called Kalman gain, which 'interprets' the flash release. The empirical results for Belgium suggest that the Kalman gain associated with the flash release is negligible for this particular model. The bottom of table 5 shows that this result also holds for the euro area. In other words, forecast errors for the flash release do not induce updates of the state with respect to the expectations given by $\mathscr{F}^{old}$. Section 4 provided the framework to formalize this important idea in terms of the impact that a given release has in the updating process of a given series.

B.2. 'News' versus 'Noise' hypothesis

GDP releases revising the flash estimate define a more accurate and comprehensive picture of growth. On one hand, they exploit a larger information set that was not available when the flash had to be published. Second, measurement errors may be removed. These two sources of data revisions were coined by Mankiw and Shapiro (1986) as 'news' and 'noise' respectively. This strict taxonomy for data revisions has implications for the model. If the flash release was a noisy estimate or a rational nowcast of the subsequently revised values, I would have to incorporate the assumption that the subsequently revised value is uncorrelated with the revision itself, e.g. Camacho and Perez-Quiros (2010). I will argue that the first hypothesis can be rejected for real GDP growth in both the euro area and Belgium:

- **The Noise Hypothesis**. Consider the flash estimate of real GDP growth, y_t^Q and its revised value after two years, y_t^{Q*}. Assuming that the revision error r_t is independent from the y_t^{Q*} has implications for the relative variance of the two estimates.

  (23) $y_t^Q = y_t^{Q*} + r_t$

  (24) $E[y_t^{Q*} r_t] = 0 \Rightarrow var(y_t^Q) = var(y_t^{Q*}) + var(r_t)$

  Thus, the extra variance added to the flash does not incorporate information about the values that will be obtained after two years. This hypothesis can actually be tested with the simple regression approach suggested by Mankiw and Shapiro (1986).

  $$\begin{aligned} & y_t^Q = c + \beta y_t^{Q*} + r_t^{noise} \\ (25) \qquad & H0: c = 0, \beta = 1 \\ & H1: c \neq 0, \beta \neq 1 \end{aligned}$$

- **The News Hypothesis**. Here, the revision error r_t is correlated with the revised data y_t^{Q*}, but independent from the flash.

  (26) $y_t^{Q*} = y_t^Q - r_t$

  (27) $E[y_t^Q r_t] = 0 \Rightarrow var(y_t^{Q*}) = var(y_t^Q) + var(r_t)$

  This implies that the revision is actually adding information instead of removing noise. This hypothesis assumption conforms the rational expectations hypothesis.

  $$\begin{aligned} & y_t^{Q*} = c + \beta y_t^Q + r_t^{news} \\ (28) \qquad & H0: c = 0, \beta = 1 \\ & H1: c \neq 0, \beta \neq 1 \end{aligned}$$

In our case, the results do not leave much room for discussion, since the noise hypothesis is clearly rejected both in Belgium and in the euro area. Figure 11 summarizes the main findings resulting from the estimation of the noise regression (equation 25) and the news regression (equation 28) over the sample period 2002Q10–2010Q3. The test suggest that the null hypothesis of noise is clearly rejected for both Belgian and aggregate euro area GDP growth rates. Thus, I argue that the extreme assumption that data revisions are uncorrelated to the final release should not be used in my current application.

Figure 11: 'Noise' versus 'news' regression analysis

Belgium

News hypothesis — News regression for Belgium

	Coef	Std. E	t-Stat	p-value
c	– 0.01	0.06	– 0.12	0.91
β	1.15	0.10	11.34	0.00

R-squared	0.80	Wald Test
Mean dependent var	0.35	H0: $c = 0, \ \beta = 1$
S.D. dependent var	0.67	$X_2 = 2.69$ (p-value = 0.26)

Noise hypothesis — Noise regression for Belgium

	Coef	Std. E	t-Stat	p-value
c	0.07	0.05	1.50	0.14
β	0.69	0.06	11.34	0.00

R-squared	0.80	Wald Test
Mean dependent var	0.31	H0: $c = 0, \ \beta = 1$
S.D. dependent var	0.52	$X_2 = 26.07$ (p-value = 0.00)

Euro area

News hypothesis — News regression for the euro area

	Coef	Std. E	t-Stat	p-value
c	0.00	0.03	0.06	0.95
β	1.05	0.04	26.50	0.00

R-squared	0.96	Wald Test
Mean dependent var	0.24	H0: $c = 0, \ \beta = 1$
S.D. dependent var	0.69	$X_2 = 1.84$ (p-value = 0.40)

Noise hypothesis — Noise regression for the euro area

	Coef	Std. E	t-Stat	p-value
c	0.01	0.02	0.35	0.73
β	0.91	0.03	26.50	0.00

R-squared	0.96	Wald Test
Mean dependent var	0.23	H0: $c = 0, \ \beta = 1$
S.D. dependent var	0.64	$X_2 = 7.19$ (p-value = 0.03)

Source: Author's calculations

Note: Under the null hypothesis, the Wald statistic has an asymptotic n distribution, where n is the number of restrictions consistent with such null hypothesis. The p-values associated with the noise hypothesis suggest that such null hypothesis is overwhelmingly rejected, with significance levels of 0 % and 3 % in Belgium and in the euro area respectively. The test has been conducted over the same sample period 2002Q1-2010Q3 for both countries. In the light of those results, assuming data revisions are uncorrelated to the final release is at least questionable.

Reducing forward-series errors for benchmarked Quarterly National Accounts

2

Geoffrey Brent, Alex Stuckey and Tom Davidson (1)

Abstract: Producing Quarterly National Account estimates (QNAs) typically involves benchmarking previous years' estimates, where an annual benchmark is available. In the absence of a current annual benchmark, extrapolation is required based on a modelled relationship with an indicator series. As additional benchmarks become available, published QNAs are revised. We seek methods to improve these revisions. We explore different benchmarking and forecasting combinations applied to both the 'forward' and 'back' series on Australian economic series and on synthesised data. We introduce metrics to assess different methods based on magnitude and timeliness of revisions, apparent bias, and preservation of movements from the indicator. These metrics are normalised according to the variability of the series to be predicted, allowing comparisons over multiple data series with differing levels of volatility, and may be useful for other researchers.

The results indicate that using a random-walk-with-drift (RWD) model to forecast benchmark/indicator relationships in the forward series can achieve a noticeable reduction in bias, with very small increase in net revisions. For multiplicative data series, the best results came from combining RWD forecasting with benchmarking via Denton-Cholette, Cholette-Dagum, or Enhanced Denton (Di Fonzo/Marini) methods. For additive data series, the best results came from Denton-Cholette benchmarking with RWD forecasts.

Keywords: quarterly national accounts, revisions, temporal benchmarking, forecasting.

JEL codes: C32, C53, E37.

(1) Australian Bureau of Statistics, Methodology and Data Management Division.

1. Introduction

National statistical offices publish Quarterly National Accounts (QNA) estimates that are required to mimic the movements of a quarterly indicator series whilst maintaining consistency with an annual benchmark series. The problem is complicated by lag in the availability of the benchmark series (Australian benchmarks are often not available for approximately 2 years after the start of the year to which they refer) and by imperfect correlation between the quarterly indicator series and the variable of interest.

For the 'back series' (years where a benchmark is available), the usual process is to divide the annual benchmark figure between quarters in a way that as closely as possible preserves movements observed in the indicator series. A variety of benchmarking methods are available to achieve this, with all having different assumptions of the relationship between the estimates and the indicator series. Among the most prominent are proportional Denton-Cholette, proportional Cholette-Dagum regression-based, and Chow-Lin regression-based (Eurostat (2013), International Monetary Fund (2014), Australian Bureau of Statistics (2000)). Another variation on the Denton method (Enhanced Denton) has recently been proposed (Bloem et al. (2001), Di Fonzo and Marini (2012)). Currently the Australian Bureau of Statistics uses Proportional Denton-Cholette as its primary benchmarking method.

There are several studies of the relative merits of different benchmarking methods in different scenarios. Some comparative studies Ciammola et al. (2005), Chen (2007), and Reber and Pack (2014) compare performance on metrics such as preservation of movements of the indicator series, ability to recover high frequency simulated series and detection of turning points. In contrast to these papers, the focus of our study is the revisions that are imposed on the quarterly estimates due to the delay in availability of the annual benchmarks.

For the 'forward series' (recent years, where the annual benchmark has not yet been compiled, so only the quarterly indicator is available), producing QNA estimates requires some way to predict the item of interest from the indicator series. Regression-based benchmarking methods offer an obvious way to do this: take the same modelled benchmark/indicator relationship that was observed for the back series, and apply it to the forward-series indicators.

However, a model that gives good results when applied to the domain where it was fitted (the back series) under benchmarking constraints may not necessarily perform well when extrapolated beyond that domain. In particular, many benchmarking methods assume the relationship between annual benchmark and indicator series is static over time. If the relationship is in fact evolving, this may lead to bias in the forward series, requiring large revisions once benchmarks become available. (Ideally we would use an indicator that provides a stable Benchmark to Indicator (BI) ratio, but such indicators are not always available.)

An alternative approach is to handle forward-series extrapolation as a separate modelling process, using benchmarked estimates as input for a model that forecasts benchmark-indicator relationships for the forward series, e.g. International Monetary Fund (2014) pp. 15–18. Previous work by our colleagues of Australian Bureau of Statistics (Joymungul Poorun et al.) indicated that these BI ratios may have some structure that could be captured by common time series models.

Another concern in benchmarking is the possibility of an unforeseen break in the BI ratio. No method can be expected to give good forward estimates immediately after such a break. However, a BI break has the potential to harm quality of forward estimates in later years, even after it becomes identifiable in the benchmarks. The choice of benchmark/forecasting methods may affect how quickly the forward series recovers after such a break.

In this investigation, we assess the performance of different combinations of several benchmarking and forecasting methods. We do this by re-benchmarking Australian economic data, using only data that would have been available when estimates were generated, and we compare performance on a range of metrics.

One complication in benchmarking is that the short-term movements we observe in the indicator series do not all represent movements in the true series. Some of the observed movements will be spurious, caused by measurement errors or scope differences between the indicator and the benchmark. A perfect benchmarking method would preserve the component of the movements that represents genuine movements in the underlying quantity, and eliminate those caused by errors/scope discrepancies — hence, exact preservation of movements would not be the desired goal.

In practice it is impossible to separate these two components perfectly, but some forecasting methods may do better than others. To investigate this question we tested the same benchmark/forecast methods on synthetically generated data series chosen to represent cases of interest: we generate the 'true' quarterly series according to a specified model, then combine this with an error term also generated from a specified model, and test how each method performs at matching movements in the true series as distinct from matching the indicator.

2. Methods

We tested five benchmarking methods available in the R 'tempdisagg' package (Sax and Steiner (2013)) (Chow-Lin-maxlog, Fernandez, Litterman, Denton-Cholette, original Denton). We also tested our own implementations of Enhanced Denton (Di Fonzo and Marini (2012)) and of proportional Cholette-Dagum (2006) (with autoregressive parameters set at 0.84 and 0.93. The Cholette-Dagum results are abbreviated as 'CD0.84' and 'CD0.93' respectively in some graphs.

The Denton method preserves the movements of an indicator series. The Denton-Cholette is a variant of this method with improved behaviour at the beginning of the benchmarked series. The Chow and Lin method is based on regressing the annual benchmarks on the annualised indicator using Generalized Least Squares. The residual term in the regression is assumed to be a stationary AR(1) process for the Chow and Lin method. Alternative error structures have also been proposed such as a random walk process (Fernandez (1981)) and an ARIMA(1,1,0) process (Litterman (1983)). We have used the default maximum likelihood parameter estimation options for the Chow-Lin regression, as recommended by Sax and Steiner (2013).

We combined these with several forecasting methods, including random-walk (RW) and random-walk with drift (RWD) (drift estimated by fitting a linear model to the period-to-period change), the automated ARIMA model selection algorithm (auto.arima), Holt and automated exponential smoothing state space model selection (ETS). All of these methods were implemented by calling the functions available from the 'forecast' package in R (Hyndman (2011)). ETS was used with the full model space and also reduced to only consider nonseasonal models with additive errors. We also included simple linear regression over time (LM), with and without an intercept. The forecasting methods were applied to the quarterly BI ratios or, in the case of the enhanced Denton method, the annual BI ratios.

Forecast method 'none' indicates the default forward-estimate calculation for the benchmarking method in question. For Denton-Cholette and Denton-Enhanced this is equivalent to random-walk forecasting. Some benchmark/forecast combinations were excluded because of incompatibility.

In this study we simulate the timing of benchmarking in the Australian Quarterly National Accounts. Benchmarking is done over a 5-year window of the data and extrapolation is required for up to eight quarters ahead of the most recent annual benchmark (Australian Bureau of Statistics (2000)). The default method for ABS benchmarking is proportional Denton-Cholette with default (RW) forecasting. Forecast methods are fit over the most recent 5 benchmarked years (i.e. 20 quarters). Many of these methods are available in both proportional and additive versions, suited to multiplicative or additive time series models respectively. In the results below we consider both cases.

3. Inputs

We tested multiplicative methods on a range of Australian QNA series: 13 seasonally-adjusted industry series and 44 public capital series containing seasonal effects. Several of these series show evidence of long-term structure in the BI ratio, with either gradual or abrupt changes. We also tested additive methods on 21 public capital series that contained zero or negative values and hence were not suited to multiplicative methods.

We used simulated data to explore how various characteristics of data might affect model performance, for example, whether an abrupt change in BI ratios can cause large revisions to the back series. For this investigation we created 16 different model classes and randomly generated 10 data sets for each of these classes. Selected models are described in section 5 'Results'.

4. Metrics

We defined four metrics to assess benchmarking/forecasting performance. These are based on 'mean absolute scaled error' (MASE) (Hyndman (2006)), including the idea of normalising relative to the variability of the series to be predicted, but modified to emphasise prediction of movements. The function of these four metrics is:

1. Measure how closely final benchmarked estimates emulated quarter-to-quarter movements of the indicator series. A value of 0 indicates perfect agreement between estimator and indicator movements; 1 indicates that the discrepancies between estimator and indicator movements are comparable in magnitude to the typical differences in movement between successive quarters; more than 1 indicates that discrepancies between estimator and indicator are larger than quarter-to-quarter movements. Note that this measure is only based on the final benchmarked results, not initial estimates (forward series); we included it as a check on movement preservation and not on forecasting methods.
2. Measure how closely final benchmarked estimates emulated quarter-to-quarter movements of the true quarterly series (only possible for synthetic data where the true series is known). As above, a value of 0 indicates perfect agreement, and a value greater than 1 indicates that discrepancies between estimator and true movements are greater than quarter-to-quarter movements. As above, this measure does not include the forward series.
3. Measure the magnitude and timeliness of revisions. A value of 0 indicates zero revisions; 1 indicates that the magnitude of a revision, multiplied by the delay before making that revision, is comparable to the typical differences in movement between successive quarters, with greater values representing larger and/or later revisions.
4. Measure bias in initial estimates, relative to final benchmarked estimates. A value of 0 indicates no net positive or negative behaviour in revisions; 1 indicates a positive or negative bias in initial estimates that has comparable magnitude to the quarter-to-quarter movement, and a value greater than 1 indicates that the magnitude of bias in initial estimates (relative to final) is greater than typical quarter-to-quarter movement.

Details for these metrics are given in annex A below.

5. Results

For each combination of multiplicative benchmarking and forecasting methods, results were averaged across all 13 Industry series and (separately) across all 44 public capital series. For synthetic series, results were averaged for each class (10 trials per class).

5.1. Synthetic data

We used a range of synthetic data series to explore how varying behaviour of the true series and of BI ratios might affect benchmarking/forecasting performance. The simulated 'true' quarterly series was generated with a trend with a randomly-generated noise component; we then generated a true BI ratio series through a similar model. Benchmarking and forecasting were based on the observed indicator series (true series divided by true BI ratio) and benchmarks (annual totals of the true series). For each class of model we generated ten replicates, differing only in random-number outcomes. Annex D describes the data generation process in more detail.

Results from synthetic data can give valuable insight into behaviours of these methods that cannot be observed from real data, because only synthetic data allows us to observe the exact quarterly series and BI ratios that the benchmarking/forecasting methods attempt to predict. This can shed light on the scenarios where a given method might be expected to work well or badly, and to identify metrics that may be misleading.

However, we have not attempted to quantify overall performance of these methods for synthetic data. To do this would require careful consideration of what scenarios to include, and of how to weight them to give a result likely to be representative for real data.

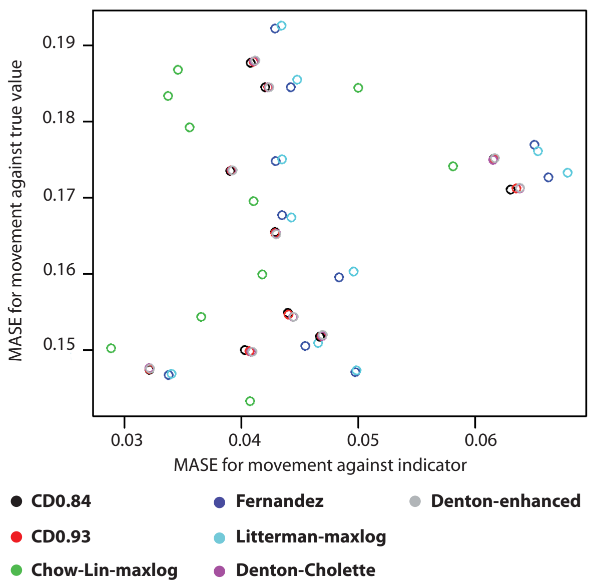
Indicator and true series movements

When analysing the synthetic series it became apparent that a benchmarking method's ability to replicate indicator movement was *not* a good predictor of ability to replicate the true series; figure 1 shows example results for one model class. In particular, there were several synthetic models where Chow-Lin benchmarking gave the best results for matching the indicator while showing poor performance on the true series.

In some cases, performance against the true series was actually better than performance against the indicator. This happens, for example, when there is a strong trend but little noise in the BI ratio, introducing a trend in the indicator series that doesn't reflect true-series movements; benchmarking to annual totals removes this spurious trend, which accounts for most of the movement seen in the indicator.

These results suggest that ability to reproduce movement is not by itself a reliable criterion for choosing between benchmarking methods.

Figure 1: Example of performance matching true series and indicator movements, with colour representing benchmarking methods



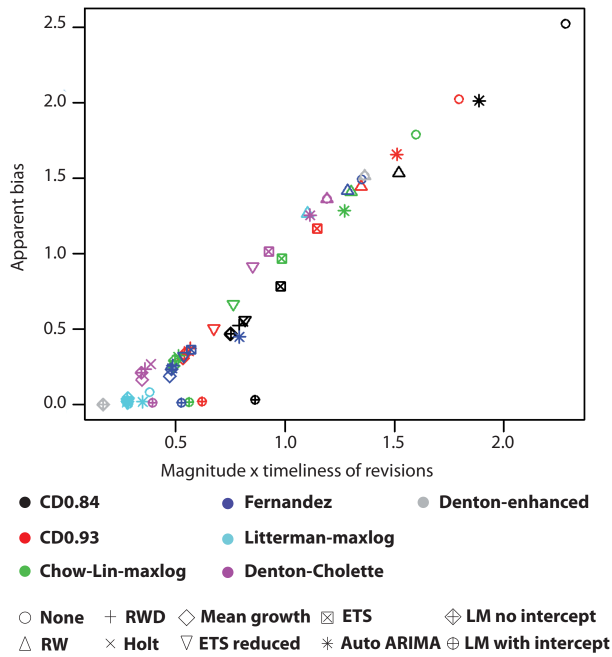
Source: Author's calculations

Bias and revisions

In this section, two series that provide some insight into bias and revisions for different benchmarking and forecasting combinations are presented.

The first series has a moderately smooth true series with a small amount of noise, and a strong linear trend in the BI ratio (steady increase) with no noise. The results highlight that benchmarking and forecasting combinations that allow forecasted trends in the BI ratio can decrease both bias and revisions, see figure 2 for results.

Figure 2: Revisions and bias for smooth series with linear BI ratio, with colour representing benchmarking methods and symbols forecasting approaches



Source: Author's calculations

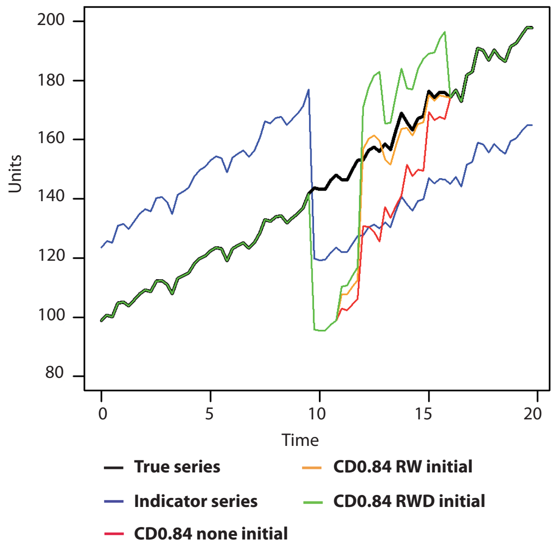
Most of the benchmarking and forecasting combinations fail to anticipate the steady increase in BI ratios. They either use the last benchmarked BI ratio (which will underestimate future BI ratios) or a regression model that predicts based on BI relationships over the benchmarking period (which may not be suitable for extrapolating outside that period).

An exception is the Litterman method which uses an ARIMA(1,1,0) model for residuals, allowing a trend in series residuals that effectively simulates the trend in BI ratios. Using an appropriate forecasting method, e.g. RWD, greatly improves performance: by anticipating the increasing BI ratios, bias in initial estimates are reduced and hence so is the need for revisions.

The second series discussed in this section explores effects of a break in BI ratio. Discontinuities in the BI ratio can arise where there is an abrupt change in the level of the quarterly indicator series but not in the annual benchmark series. The level change in the indicator series could come from a real-world phenomenon that is out of scope of the annual benchmark measure and therefore the quarterly benchmarked series. The level change could also come from a change in method or a change in source data. In either case, the level change in the indicator series can distort the QNA series. We simulate this situation by a well-behaved true series and an abrupt break in an otherwise-constant BI ratio, causing a large break in the indicator series that is not reflected in benchmarks.

Figure 3 presents results for one parameterisation of the Cholette-Dagum method under three different forecasting methods. Black and blue lines show the true and indicator series; other colours show the initial estimate for each quarter under each forecasting method. These results demonstrate that allowing forecasting of the BI ratio with trend can lead to overshooting and undershooting after a BI trend break.

Figure 3: Effect of level change in BI ratio on benchmarking



Source: Author's calculations

Before the first post-break benchmark becomes available, all methods underpredict, since they're tracking the movement in the indicator and there is no information available yet to signal that the cause of the indicator movement is a shift in BI ratio rather than in the true series.

Once post-break benchmarks become available, RW forecasting is very effective in removing this error from the forward series, since it effectively ignores pre-break years in forecasting. Two years after the break (i.e. as soon as post-break benchmarks are available) the RW forward series estimates closely track the true series.

In contrast, RWD forecasting overshoots when the post-break benchmarks arrive, because it's attempting to fit a linear drift across a period that spans the trend break. BI ratio was 0.8 before the break and 1.2 afterwards; hence, RWD fitted over a 5-year benchmarking window will predict BI ratios of approximately 1.36 and 1.28 for the two most recent (non-benchmarked) years. This error persists until the trend break falls out of the benchmarking window, at which point the overshoot error disappears.

Default forecasting continues to underpredict for several years, but (unlike RWD) this error becomes smaller over time. Whereas RWD fits based on the first and last point of the series, the default method is a regression based on all points within the benchmarking window. As time goes by there are more post-break and fewer pre-break years in that window, so the regression becomes a more accurate prediction of post-break behaviour.

The combination of initial underestimation followed by later overestimation for RWD means that the net bias metric is close to zero. The bias metric for RW is higher (but revisions are lower) because it has the same underestimation but doesn't follow with overestimation.

Other simulation models with a more gradual change in BI ratio ('ramp' rather than 'step' behaviour) resulted in similar over/underestimation behaviour for forecasting methods that allow a trend, resulting in lower bias but higher revisions than RW. Clearly, bias metrics should not be evaluated in isolation; if a method shows low bias but high revisions, it may indicate two systematic errors in opposite directions at different times.

5.2. Australian multiplicative QNA series

Overall, results for Industry and public capital series were consistent with behaviour observed in synthetic models. The Denton-Cholette, Cholette-Dagum, and enhanced Denton methods generally performed better than the others considered, and discussion will focus on these 'DCD' methods.

Industry series

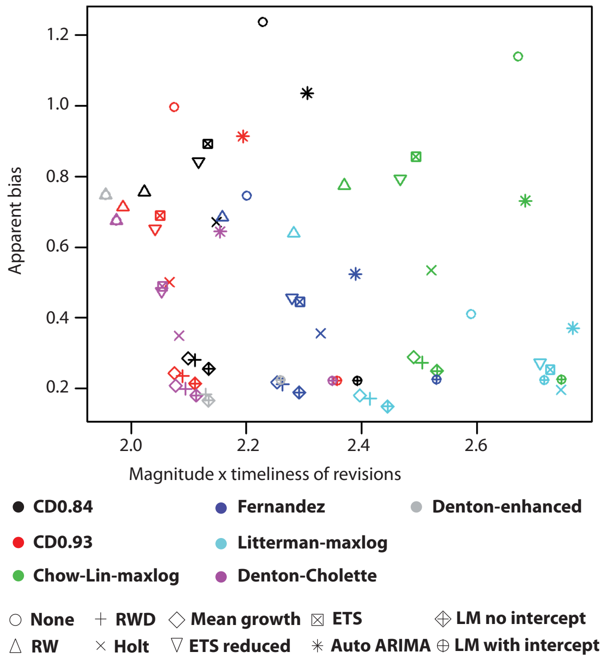
Metric values varied from series to series, but the normalisation approach used in calculation of metrics prevented the appearance of outliers.

All methods showed similar results for replication of indicator movement, with average discrepancy metric ranging from about 0.218 (Chow-Lin) to 0.236 (Litterman). Given the closeness of these results and Chow-Lin's tendency to perform better for the indicator than for the true series in synthetic data, there seems little reason to prefer any particular method on this ground.

As shown in figure 4, enhanced Denton with random-walk forecasting gave the smallest revisions, with Cholette-Dagum and Denton-Cholette showing very similar performance. However, combining these methods with RWD forecasting resulted in a marked decrease in bias for only a very small increase in revisions.

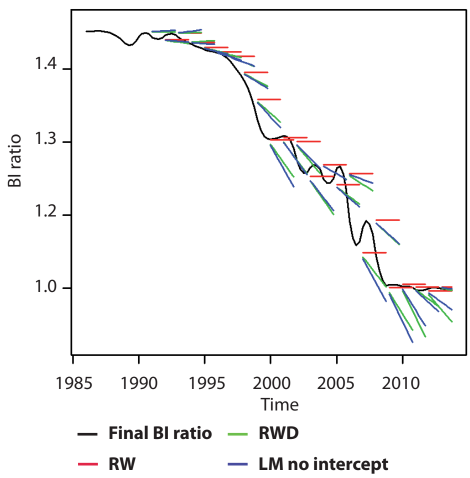
More sophisticated forecasting methods (automated exponential smoothing and ARIMA) performed poorly relative to RW and RWD. This may be because the benchmarking period only contains 5 years/20 quarters of data, which may not be enough to fit for a model with more degrees of freedom. A short series span also led Di Fonzo and Marini (Chen (2007) p. 15) to choose simple forecast methods in their study. Examination of individual series showed some evidence of overfitting (figure 5 shows examples).

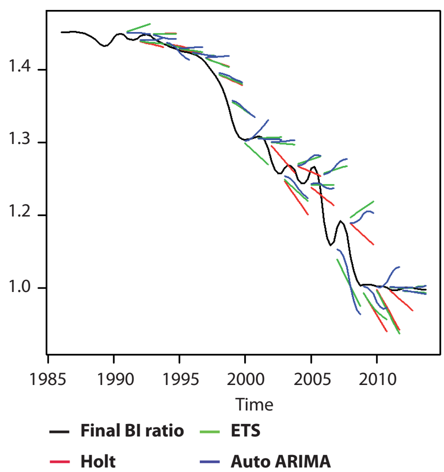
Figure 4: Bias and revisions metrics for industry series
(averaged metrics over all series)



Source: Author's calculations

Figure 5: Forecasts of BI ratios for a particular Industry series under various methods





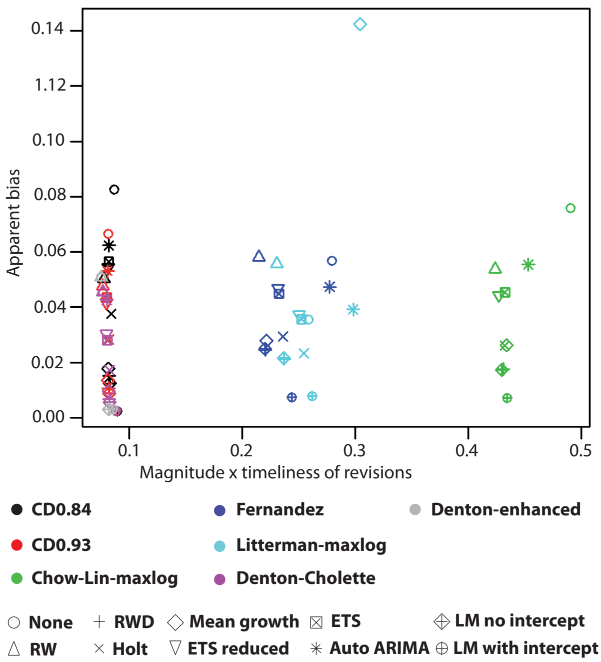
Source: Author's calculations

Public capital series

Public capital data showed some outliers in terms of revisions for the Chow-Lin method. These are related to a series where the indicator dropped very close to zero for one quarter before rebounding, resulting in very large proportional movements; a small error in estimation of level for this quarter could have a large effect on movements and hence on metrics. However, eliminating the outliers would not have substantially changed the following aggregated results.

The 'DCD' methods performed better than others for replication of movements (metric approx. half that for other methods) and showed the same superiority for revisions. As for Industry series, RWD forecasting caused a marked reduction in bias for only a very small increase in revisions, most clearly seen in the table in annex C. Revision metrics for Chow-Lin methods and one Litterman method on one particular series were very large; these outliers contribute to the averages shown in figure 6.

Figure 6: Bias and revisions metrics for public capital series
(averaged metrics over all series)



Source: Author's calculations

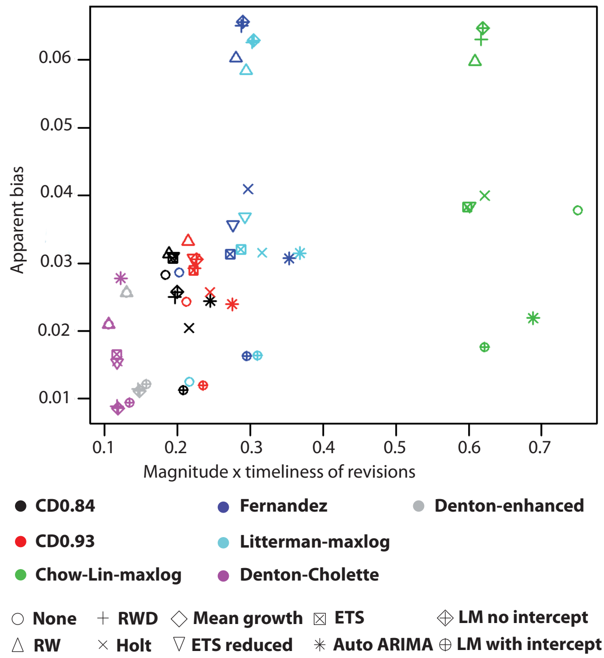
Additive series

In addition to the series analysed above, twenty-one problematic public capital series were investigated. These series were deemed problematic as they contained annual benchmarks and/or indicator series values that were negative or at least close to zero.

For this type of series, the proportional first difference Denton-Cholette and the proportional version of the Cholette-Dagum method are not practical. However, additive first difference Denton-Cholette and the additive Cholette-Dagum can be utilised along with the Chow-Lin-type methods which are already additive in nature. We implemented an additive analogue of the enhanced-Denton which is given in annex B. In regards to forecasting the forward series, additive differences between the benchmark and indicator are forecasted as opposed to the usual BI ratio. All the above forecasting methods were included in this investigation, apart from mean growth as a change in sign or zero value will lead to unreliable/unusable results. A plot of the revisions metric versus apparent bias is provided for each benchmarking and forecasting combination averaged over all series.

Similar to the public capital data, there were some outliers for Chow-Lin relating to one series. This series is the same outlier series detailed in section 5.2.2; it was hoped forecasting additive corrections when using the Chow-Lin method on this series, would result in marked improvements in the timeliness and magnitude of the revisions. This however, was not the case. Please note that eliminating these outliers does not change the assertions made about the benchmark and forecasting combinations.

Figure 7: Bias and revisions metrics for problematic public capital series (averaged metrics over all series)



Source: Author's calculations

Denton-Cholette showed superior performance to the other benchmarking methods in terms of revisions for these 21 series. On reviewing some synthetic series, it was apparent that the additive first difference Denton-Cholette method didn't necessarily have the lowest revisions, but most of the revising happens relatively quickly when compared to the other benchmarking methods. This offers an explanation to the above results. Random walk with drift, at least for the Denton-Cholette and both parameterisations of Cholette-Dagum method, has a distinct benefit in regards to the apparent bias when compared to the usual random walk method. One possible explanation for this is that these series (both the annual benchmarks and indicator series) are increasing in variability over time, meaning that the additive corrections will also increase. The drift term allows for a systematic increase in the additive corrections which naturally decreases the apparent bias.

6. Conclusions

Explorations on synthetic data suggest that there is a trade-off between RW and RWD forecasting. RWD can reduce bias by predicting a trend in BI ratios, but also risks increasing error by over- or under-stating trends; RW gives more conservative estimates that may incur bias but reduce other errors in the forward series, and hence reduces revisions.

Overall, the results from both industry and public capital data suggest that RWD forecasting has potential for a reduction in forward-series bias with very little increase in revisions. This method appears to be particularly effective when combined with enhanced Denton, Denton-Cholette, or Cholette-Dagum benchmarking methods. Testing on a wider range of benchmarking series would be desirable, as well as exploring other metrics of performance.

We propose further study into prior correcting quarterly indicator series to remove level change effects that are not evident in the corresponding annual benchmark series. This could avoid unnecessary distortion to the resulting QNA estimates. Simple regression-ARIMA models with appropriate regressors are one way of estimating the magnitude of the level shift that is not reflected in the annual benchmark series.

7. Limitations and potential future work

Our work concentrated on quarterly Australian series where the benchmarks can be expected to have a 2-year lag behind the indicator series, and where the series generally show an increasing trend over time. Our analysis concentrates on performance at a seasonal level, both in movement preservation and in magnitude of revisions, with an emphasis on growth rates.

In many European countries benchmarks are available with a shorter lag. It would be interesting to compare methods with a four-quarter benchmark lag, to see if results varied from those we observed on an eight-quarter lag. It would also be valuable to test how well these methods compare when the quarterly predictions are aggregated to annual estimates, and how they perform at predicting and preserving turning points.

One limitation on our work was the speed of the programs used to generate the benchmark data. Because of the requirement to re-benchmark each series for each quarter, for several different combinations of benchmarking and forecasting methods, it took many hours to run; this was particularly an issue for the synthetic data series where we generated ten data sets from each of sixteen models.

Acknowledgements

We thank the reviewers who made many helpful suggestions on the first draft of this work. We thank our ABS colleagues in the National Accounts Branch for their assistance.

References

Australian Bureau of Statistics (2000), 'Australian National Accounts: Concepts, Sources and Methods', chapter 13, p. 167.

Bloem, A.M., R. Dippelsman and N. Ø. Mæhle (2001), 'Quarterly national accounts manual: concepts, data sources, and compilation', *International Monetary Fund*, 2001.

Chen, B. (2007), 'An empirical comparison of methods for temporal distribution and interpolation at the national accounts', *Bureau of Economic Analysis*.

Ciammola, A., F. Di Palma and M. Marini (2005), 'Temporal disaggregation techniques of time series by related series: A comparison by a Monte Carlo experiment'. Eurostat, Working Paper.

Dagum, E. and P. Cholette (2006), 'Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series', *Lecture Notes in Statistics*, Vol. 186, pp. 80–81.

Di Fonzo, T. and M. Marini (2012), 'On the Extrapolation with the Denton Proportional Benchmarking Method', *IMF Working Paper*, June.

Eurostat (2013), 'Handbook on quarterly national accounts', chapter 5, pp. 121–171.

Hyndman, R.J. (2006), 'Another Look At Forecast-Accuracy Metrics For Intermittent Demand', Foresight Issue 4, p. 46, http://robjhyndman.com/papers/foresight.pdf.

Hyndman, R.J. (2011), 'forecast: Forecasting functions for time series', R package version 3.04, http://CRAN.R-project.org/package=forecast.

International Monetary Fund (2014), 'Update of the Quarterly National Accounts Manual', draft, chapter 6, January.

Joymungul Poorun, Y., R. Mathews, J. Chien and P. Gould, 'Improving Extrapolation Method for National Accounts', Australian Bureau of Statistics, unpublished work.

Reber, R.L. and S.J. Pack (2014), 'Methods of Temporal Disaggregation for Estimating Output of the Insurance Industry', Bureau of Economic Analysis, Working paper 0115.

Sax, C. and P. Steiner (2013), 'tempdisagg: Methods for Temporal Disaggregation and Interpolation of Time Series', R package version 0.22, http://CRAN.R-project.org/package=tempdisagg.

Annex A — Metrics (detail)

We use the following notation:

n indicates the total number of quarters for which we have final benchmarked data.

q indicates the quarter to which an estimate refers.

l indicates the lag of an estimate: l=0 is the first estimate made for that quarter, l=1 is the first revision to that estimate, etc.

k indicates the lag at which an estimate is finalised and will not experience further revisions.

$y_{I,q}$ indicates the value of the indicator series for quarter q.

$y_{E,q,l}$ indicates the value of the lag-l estimator for quarter q; hence, $y_{E,q,k}$ is the final benchmarked estimate for that quarter.

r indicates quarter-to-quarter movement, either final or at some specified lag:

$$r_{E,q} = \frac{y_{E,q,k}}{y_{E,q-1,k}} - 1; \;\; r_{I,q} = \frac{y_{I,q}}{y_{I,q-1}} - 1; \;\; r_{E,q,l} = \frac{y_{E,q,l}}{y_{E,q-1,l+1}} - 1.$$

Metric for preservation of movements relative to indicator series:

$$M_{preservation} = \frac{\frac{1}{n-1}\sum_{q=2}^{n}\left|r_{E,q,k} - r_{I,q,k}\right|}{\frac{1}{n-2}\sum_{q=3}^{n}\left|r_{I,q} - r_{I,q-1}\right|}.$$

i.e. mean absolute value of difference between estimator and indicator movements, normalised by mean absolute value of difference between successive quarterly movements in the indicator.

The metric for preservation of movements relative to the true series (where available) is calculated as above, but replacing indicator movements with true movements.

Note that the metrics for movement preservation are based on the final benchmarked results. Hence, they are not affected by choice of forecasting method.

Metric for revisions:

$$M_{revision} = \frac{\frac{1}{n-1}\sum_{l=0}^{k}\sum_{q=2}^{n}\left|r_{E,q,l} - r_{E,q,k}\right|}{\frac{1}{n-2}\sum_{q=3}^{n}\left|r_{E,q,k} - r_{E,q-1,k}\right|}.$$

i.e. sum over all lags of difference between lagged and final movement estimates, normalised by quarter-to-quarter changes in final estimates.

When designing this metric we considered normalising by the number of periods involved, i.e. including a factor of $1/\mathrm{k}$. The reason we decided against this is that our aim was to measure not only the size of the discrepancy between initial and final movement estimates, but also how quickly that discrepancy is removed by intermediate revisions.

For example, suppose we wish to compare two methods with different revisions periods: method 1 gives final values after 2 years (k = 8) and method 2 is finalised after 3 years (k = 12). Both show similar discrepancies between the initial and final movement estimates, but method 2 takes 50 % longer to converge to the final estimate. In this case, the metric for method 2 will be 50 % larger than that for method 1. Normalising by $1/k$ would eliminate that difference and represent both methods as equally good, which we consider undesirable.

However, for the results shown in this paper, all methods had the same revisions period i.e. the same value of k. Hence, normalising by $1/k$ would simply have scaled this metric by a common factor across all methods, without changing relative performance.

This metric considers both the forward series (extrapolated benchmarks) and the period where benchmarks are available but the estimate is not yet finalised. In practice, once benchmark data is available for a given quarter, subsequent revisions for that quarter are usually relatively small — hence in most cases the main contributor here is extrapolation error.

Metric for bias:

$$M_{bias} = \frac{\frac{1}{n}\left|\sum_{q=1}^{n}\left(\frac{y_{E,q,0}}{y_{E,q,k}} - 1\right)\right|}{\frac{1}{n-1}\sum_{q=2}^{n}\left|r_{E,q}\right|}.$$

i.e. absolute value of mean proportional revision between initial and final estimates, normalised by average magnitude of quarter-to-quarter differences.

Where used with additive benchmarking models, these metrics were modified to quantify movements and revisions as additive rather than multiplicative changes, e.g. $r_{E,q} = y_{E,q,k} - y_{E,q-1,k}$ rather than $r_{E,q} = \frac{y_{E,q,k}}{y_{E,q-1,k}} - 1$.

Annex B — Incorporating forecasting methods into matrix form of enhanced Denton

In this section we provide the matrix forms for the simple forecasting methods that can be plugged directly into the matrix form of the Enhanced Denton already provided by Di Fonzo and Marini (2012). As we are using the matrix form of Di Fonzo and Marini we adopt the notation of that paper exactly. That is, let $\boldsymbol{y} = (y_1, \dots, y_{sN})'$ be the high frequency benchmarked series to be estimated which must sum over the year to the annual benchmarks $\mathbf{y_0} = \left(y_{0,1}, \dots, y_{0,N}\right)'$ via the relationship $\mathbf{Jy} = \mathbf{y}_0$ where the temporal aggregation matrix is constructed by $\mathbf{J} = \mathbf{I}_N \otimes \mathbf{1}_s'$. The indicator (or preliminary) series $\mathbf{p} = (p_1, \dots, p_{s(N+h)})'$ is used in a diagonal form $\widehat{\mathbf{p}} = \mathrm{diag}(\mathbf{p})$ and has annual sums $\mathbf{p}_0 = (p_{0,1}, \dots, p_{0,N+h})'$. Other matrices required are:

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, E = \begin{bmatrix} \mathbf{0}_{sh,sN} & \mathbf{0}_{sh\times sh} \\ \mathbf{I}_{sN} & \mathbf{0}_{sN\times sh} \end{bmatrix}, \widehat{\mathbf{p}}_{0E}^{-1} = \mathrm{diag}(\mathbf{E}\mathbf{p}_0^{\star-1}),$$

and $\mathbf{p}_0^\star = \begin{bmatrix} \mathbf{p}_0 \otimes \mathbf{1}_s \\ \mathbf{1}_{sh} \end{bmatrix}$.

The solution to the enhanced Denton method is given in equation (4) of Di Fonzo and Marini (2012) as:

$$\begin{bmatrix} \boldsymbol{y} \\ \lambda \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{p}} & 0 \\ 0 & \mathbf{I}_{sN} \end{bmatrix} \begin{bmatrix} \mathbf{D'D} & \widehat{\mathbf{p}}\mathbf{J}^{\star\prime} \\ \mathbf{J}^\star\widehat{\mathbf{p}} & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \boldsymbol{y}_0^\star \end{bmatrix}.$$

Using the notation from that paper, we can incorporate several simple forecasting methods into the matrix solution directly. This could be advantageous when coding the method into various production systems. Here there are N annual benchmarks available and we need to extrapolate h years past the last available annual benchmark. The benchmark-to-indicator ratios are denoted b_n for year n.

$$\mathbf{J}^\star = \begin{bmatrix} [\mathbf{I}_N \otimes \mathbf{1}_s' \quad \mathbf{0}_{N\times sh}] \\ \cdots \\ \mathbf{k} \end{bmatrix}, \mathbf{k} = [\mathbf{R}\widehat{\mathbf{p}}^{-1}\widehat{\mathbf{p}}_E\widehat{\mathbf{p}}_{0E}^{-1}], \boldsymbol{y}_0^\star = \begin{bmatrix} \boldsymbol{y}_0 \\ \cdots \\ \mathbf{b}^\star \end{bmatrix},$$

$$\mathbf{R} = [\mathbf{0}_{sN}' \quad \mathbf{1}_{sh}'], \mathbf{b}^\star = \begin{bmatrix} b_{N+1} \\ b_{N+2} \\ \vdots \\ b_{N+h} \end{bmatrix} = \widetilde{\mathbf{X}} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} = \widetilde{\mathbf{X}}\mathbf{b} = \widetilde{\mathbf{X}}\widehat{\mathbf{p}}_0^{-1}\boldsymbol{y}_0.$$

The form of $\tilde{X}$ depends on the approach taken to generate forecasts of the annual BI ratios:

Random walk	Random walk with drift	Linear fit with intercept	Linear fit without intercept
$\tilde{\mathbf{X}} = [\mathbf{0} \; \vdots \; \mathbf{1}]_{h\times N}$	$\tilde{\mathbf{X}} = \frac{1}{N-1}\begin{bmatrix} -1 & & & & N \\ -2 & \vdots & \mathbf{0} & \vdots & N+1 \\ \vdots & & & & \vdots \\ -h & & & & N+h-1 \end{bmatrix}_{h\times N}$	$\tilde{\mathbf{X}} = \mathbf{X}^*(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ where $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & N \end{bmatrix}$, and $\mathbf{X}^* = \begin{bmatrix} 1 & N+1 \\ 1 & N+2 \\ \vdots & \vdots \\ 1 & N+h \end{bmatrix}$.	This requires shifting the BI series to the origin before fitting $\boldsymbol{b}^* = \mathbf{1}_h b_N + \tilde{\mathbf{X}}(\mathbf{b} - \mathbf{1}_N b_N)$ $\tilde{\mathbf{X}} = \mathbf{X}^*(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ where $\mathbf{X} = \begin{bmatrix} 1-N \\ 2-N \\ \vdots \\ 0 \end{bmatrix}$, $\mathbf{X}^* = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ h \end{bmatrix}$.

For an additive version of the Enhanced Denton method we deal with the annual benchmark-indicator differences, rather than ratios, defining $b_n = y_{0,n} - p_{0,n}$ for year n. Alternative forms of the required matrices are:

$$\mathbf{b}^* = \begin{bmatrix} p_{0,N+1} \\ p_{0,N+2} \\ \vdots \\ p_{0,N+h} \end{bmatrix} - \hat{\mathbf{b}}\frac{y_{0,N}}{b_N}, \mathbf{k} = \left[\mathbf{0}_{h\times s(N-1)} \;\vdots\; \mathbf{1}_s' \otimes \frac{-\hat{\mathbf{b}}}{b_N} \;\vdots\; \mathbf{I}_h \otimes \mathbf{1}_s'\right] \text{and } \hat{\mathbf{b}} = \begin{bmatrix} b_{N+1} \\ b_{N+2} \\ \vdots \\ b_{N+h} \end{bmatrix}$$

are the annual benchmark-indicator difference forecasts which can be obtained by a variety of forecasting procedures.

The solution is:

$$\begin{bmatrix} \mathbf{y} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{D}'\mathbf{D} & \mathbf{J}^{*'} \\ \mathbf{J}^* & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}'\mathbf{D} & \mathbf{0} \\ \mathbf{J}^* & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ y_0^* - \mathbf{J}^*\mathbf{p} \end{bmatrix}.$$

This form has an obvious problem when $b_N = 0$ — that is, when the most recent annual benchmark available is equal to the annual sum of the indicator series for that year. In our implementation, when $b_N \approx 0$ we set $\mathbf{b}^* = \mathbf{0}$.

Annex C — ABS results summary

		Public capital		Industry		Additive series	
Benchmarking	**Forecasting**	**Revision**	**Bias**	**Revision**	**Bias**	**Revision**	**Bias**
Denton-Cholette	RW	0.077	0.046	1.975	0.675	0.106	0.021
	RWD	0.083	0.007	2.094	0.199	0.116	0.009
	mean.growth	0.082	0.01	2.078	0.208		
	LM no int.	0.083	0.006	2.113	0.18	0.118	0.009
	LM with int.	0.089	0.003	2.35	0.221	0.134	0.009
	Holt	0.083	0.017	2.083	0.349	0.118	0.016
	ETS.reduced	0.08	0.03	2.053	0.475	0.117	0.015
	ETS	0.081	0.028	2.055	0.489	0.116	0.016
	auto.ARIMA	0.081	0.043	2.155	0.644	0.127	0.032
Denton-enhanced	RW	0.076	0.051	1.956	0.748	0.13	0.026
	RWD	0.083	0.003	2.129	0.183	0.146	0.011
	LM no int.	0.083	0.003	2.134	0.166	0.148	0.011
	LM with int.	0.086	0.003	2.259	0.224	0.157	0.012
Cholette-Dagum0.84	none	0.087	0.083	2.229	1.237	0.184	0.028
	RW	0.078	0.05	2.023	0.756	0.188	0.031
	RWD	0.083	0.015	2.11	0.281	0.196	0.025
	mean.growth	0.082	0.018	2.099	0.285		
	LM no int.	0.083	0.012	2.135	0.256	0.199	0.026
	LM with int.	0.09	0.002	2.393	0.222	0.208	0.011
	Holt	0.084	0.038	2.149	0.671	0.216	0.02
	ETS.reduced	0.082	0.054	2.117	0.843	0.194	0.031
	ETS	0.082	0.056	2.133	0.892	0.201	0.029
	auto.ARIMA	0.082	0.062	2.306	1.036	0.235	0.029
Cholette-Dagum0.93	none	0.082	0.067	2.075	0.997	0.212	0.024
	RW	0.077	0.048	1.985	0.713	0.215	0.033
	RWD	0.082	0.011	2.09	0.235	0.224	0.029
	mean.growth	0.082	0.013	2.076	0.243		
	LM no int.	0.083	0.009	2.111	0.214	0.226	0.031
	LM with int.	0.089	0.002	2.357	0.222	0.235	0.012
	Holt	0.082	0.028	2.067	0.501	0.245	0.026
	ETS.reduced	0.08	0.042	2.041	0.652	0.222	0.031
	ETS	0.08	0.043	2.05	0.689	0.232	0.027
	auto.ARIMA	0.081	0.053	2.194	0.913	0.273	0.03

		Public capital		Industry		additive series	
Benchmarking	**Forecasting**	**Revision**	**Bias**	**Revision**	**Bias**	**Revision**	**Bias**
Fernandez	none	0.28	0.057	2.201	0.746	0.202	0.029
	RW	0.215	0.058	2.158	0.684	0.28	0.06
	RWD	0.221	0.025	2.263	0.211	0.288	0.065
	mean.growth	0.222	0.028	2.253	0.217		
	LM no int.	0.221	0.025	2.292	0.189	0.29	0.066
	LM with int.	0.244	0.007	2.53	0.225	0.295	0.016
	Holt	0.236	0.029	2.33	0.357	0.297	0.041
	ETS.reduced	0.232	0.047	2.279	0.456	0.276	0.036
	ETS	0.233	0.045	2.293	0.445	0.28	0.031
	auto.ARIMA	0.278	0.047	2.389	0.524	0.349	0.033
Litterman-maxlog	none	0.259	0.035	2.59	0.411	0.216	0.013
	RW	0.231	0.056	2.282	0.639	0.294	0.058
	RWD	0.237	0.021	2.415	0.171	0.303	0.063
	mean.growth	0.305	0.142	2.397	0.18		
	LM no int.	0.237	0.022	2.445	0.149	0.305	0.063
	LM with int.	0.262	0.008	2.718	0.224	0.31	0.016
	Holt	0.255	0.023	2.747	0.196	0.317	0.032
	ETS.reduced	0.251	0.037	2.71	0.272	0.293	0.037
	ETS	0.253	0.036	2.728	0.253	0.298	0.032
	auto.ARIMA	0.299	0.039	2.767	0.371	0.363	0.034
Chow-lin-maxlog	none	0.491	0.076	2.672	1.14	0.75	0.038
	RW	0.424	0.054	2.37	0.774	0.608	0.06
	RWD	0.431	0.018	2.505	0.273	0.617	0.063
	mean.growth	0.434	0.026	2.491	0.289		
	LM no int.	0.43	0.017	2.531	0.249	0.619	0.065
	LM with int.	0.435	0.007	2.747	0.225	0.622	0.018
	Holt	0.433	0.026	2.522	0.534	0.623	0.042
	ETS.reduced	0.427	0.044	2.468	0.793	0.601	0.039
	ETS	0.433	0.045	2.494	0.856	0.602	0.032
	auto.ARIMA	0.453	0.055	2.685	0.731	0.685	0.028

Source: Author's calculations

Annex D — Generation of synthetic data

Synthetic series were generated with a 20-year span (80 quarters). The true series and the BI ratio are each created as a combination of a specified long-term trend component and an autoregressive transient component generated randomly according to a specified model. The trend is then divided by BI ratio to give the quarterly indicator series, and aggregated to give annual benchmarks for analysis.

In more detail, the process for the BI ratio was:

1. Specify trend and determine the BI trend time series, according to one of four models:
    - 'Flat': fixed at 0.9 throughout.
    - 'Smooth': linear increase from 0.8 to 1.2 over the duration of the series.
    - 'Ramp': fixed at 0.8 up to t = 8, then linear increase to 1.2 at t = 12, then fixed at 1.2 after t = 12.
    - 'Step' (i.e. trend break): fixed at 0.8 up to t = 10, then fixed at 1.2 from t = 10.
2. Specify one of four models for transient BI component and randomly generate the BI transient time series:
    - 'None': no transient component.
    - 'Small': normally-distributed with standard deviation 0.02 and autocorrelation 0.5.
    - 'White noise': normally-distributed with standard deviation 0.05 and autocorrelation 0.
    - 'High autocorrelation': normally-distributed with standard deviation 0.1 and autocorrelation 0.95 — close to random-walk.
3. Add the trend and transient components to give the BI series.

    The true series follows the same process, except that we aimed to create data appropriate for a multiplicative/exponential model. To achieve this, we generated data through a similar process to that described for the BI ratio (resulting in flat, linear or piecewise linear trend plus homoscedastic transients) then applied an exponential transform, resulting in a flat, exponential, or piecewise exponential trend with heteroscedastic transients proportional to the series magnitude.

    Parameter options for the true trend series:

    - 'Flat': fixed at 100 throughout.
    - 'Smooth': exponentially increasing from 100 to 200 over the duration of the series.
    - 'Ramp': slow exponential increase from 100 to 120 in the first 8 years, then faster exponential increase from 120 to 180 between years 8 and 12, then slow exponential from 180 to 200 out to year 20.

- 'Step': slow exponential increase from 100 at year 1 to 120 at year 10, then abrupt increase to 180, then slow exponential increase from 180 to 200 between years 10 and 20.

Parameter options for the log-true-transient series:

- 'None': no transient component.
- 'Small': normally-distributed with standard deviation = ln(1.02) and autocorrelation 0.5.
- 'White noise': normally-distributed with standard deviation = ln(1.05) and autocorrelation 0.
- 'High autocorrelation': normally-distributed with standard deviation = ln(1.1) and autocorrelation 0.95.

These last three result in transients of magnitude approximately 0.02, 0.05, and 0.1x the trend magnitude respectively.

From these options we selected combinations that allowed us to explore scenarios of interest, e.g. the effect of a trend break in the BI ratio. For each scenario we chose, we generated ten random replicates (excepting those that had no random transients).

There were many other scenarios we could have considered, e.g. the inclusion of seasonal components, or moving-average terms in the transients. It would also have been desirable to test more than ten replicates for each model. Unfortunately due to time limitations we had to restrict the scope of our analysis. Note that each combination of benchmark/forecast methods needs to be run many times on the same data set in order to track revisions over time; the results presented here required several days of computer run-time.

Complementing scoreboards with composite indicators: the new business cycle clock

3

Gian Luigi Mazzi ([1])

Abstract: Statistical agencies traditionally disseminate the information via so-called dashboards or scoreboards. They constitute a homogenous set of indicators, associated or not with a ranking structure that aims to describe a given phenomenon or situation. They provide a very complete picture of the phenomenon under investigation. However, they sometimes deliver quite contradictory information so that a synthetic message is not always easy to extract. On the other hand, composite indicators are intended to provide said synthetic message or to extract synthetic signals generally hidden when looking at data. This paper shows how dashboards and composite indicators can complement each other to provide a better and clearer picture of a given phenomenon. We then apply it to the case of short-term economic monitoring based on Principal European Economic Indicators (PEEIs). We also demonstrate how messages delivered by composite indicators in a friendly and easily understandable way by means of advanced graphical representations such as the new business cycle clock can be disseminated. Finally, we expose how the business cycle clock can be an ideal complement to the PEEIs dashboard by responding to questions that are clearly answered by the PEEIs themselves.

Keywords: composite indicators, data presentation and dissemination, business cycle analysis, real-time turning point detection, non-linear time series models.

JEL codes: C32, E01, E32.

([1]) Eurostat, Unit C1 'National accounts methodology. Sector accounts. Financial indicators'.

1. Introduction

Major macroeconomic data users such as policy makers, analysts, central bankers, media, etc., expect from statistical offices to assure a regular dissemination of a timely, reliable, comprehensive and clear picture of the economic situation. Fulfilling this expectation is not an easy task. It involves selecting the most appropriate set of indicators and the best way to display them in an easily understandable way.

Usually, statistical offices answer to user requests by creating dashboards or scoreboards based on official statistics. While both dashboards and scoreboards have the merit of providing a detailed and often exhaustive picture of the economic situation, they do not necessarily allow for a prompt and easy identification of the key macroeconomic signals. The large number of variables can sometimes lead to confusion, especially amongst non-expert users.

Another possible approach to address user needs consists in constructing composite indicators based on official statistics ([2]). Those indicators aim to emphasise the key underlying macroeconomic signals, making them easily understandable to non-experts. Unfortunately, those indicators are particularly sensitive to the selection of component series as well as to their construction method. Both are often based on subjective criteria. Nevertheless, there is indicators a growing interest in this kind of indicators, especially as recent studies have opened up new perspectives for reducing the degree of subjectivity in the construction of composite indicators by replacing them with a set of statistically sound, robust data selection and a compilation of techniques. One of the main questions raised in this paper is whether dashboards and scoreboards on the one hand, and composite indicators on the other can be seen as alternative ways to describe the economic situation or if they can complement each other in order to facilitate the understanding of the economic evolution.

The paper is structured as follows: section 2 describes the main characteristics of dashboards and scoreboards, highlighting their advantages and drawbacks; section 3 is devoted to methods for prioritising and summarising information, such as ranking and composite indicators techniques. More specially, it focusses on the peculiarities of those methods within the macroeconomic context. Section 4 introduces the main Eurostat macroeconomic dashboards and scoreboards, such as the PEEIs dashboard and the MIP scoreboard, whilst section 5 investigates the possibility and the usefulness of applying partially orderer set techniques (POSET) in the macroeconomic context. Section 6 presents an alternative way to use composite indicators for detecting turning points and cyclical phases, and it introduces the new Eurostat business cycle clock based on these indicators. Section 7 draws some conclusions.

([2]) In this respect there are many overlaps with the synthetic indicator approach, and this paper concentrates on composite indicators.

2. Dashboards and scoreboards

One of the main tasks of statistical institutions is to identify the most suitable sets of statistical data in order to monitor a given phenomenon and to provide these data to users. The identification of such sets of data can be a complex process that involves continuous interactions between stakeholders, policy makers and official statisticians. The outcome of such a process can be a decision to develop new statistical indicators or to enhance existing ones in order to meet the needs and requests of policy makers and stakeholders in the best of ways. The sets of indicators identified in this process can cover a large amount of data and variables from different areas of statistical production that are characterised by differences in their construction and classification. In order to make the relevant sets of indicators more useful and understandable — and thereby to ensure that they have an impact on the policy formulation cycle — it is essential to identify an attractive and friendly way of presenting them.

Over the last years, Eurostat — but also other statistical institutions — have invested many resources in developing advanced graphical and/or tabular ways to present and disseminate large sets of statistical indicators. Without attempting to be exhaustive, one could specifically mention here the Business Cycle Dashboard developed by Statistics Netherlands ([3]) and the Economic Data Dashboard developed by the UK Office for National Statistics (ONS) ([4]). Taking into account the specificities of the phenomenon to be described, two ways of presenting large datasets have been identified ([5]). The first one is represented by the so-called 'dashboards', which are graphical and tabular ways of presenting statistical indicators describing the development over time of a given social, economic or socio-economic phenomenon. These kinds of tools are particularly useful to monitor some phenomenon, even if no specific quantified political objectives have been defined. Alternatively, the second approach is constituted by the so-called 'scoreboards' where the statistical indicators are related to policy objectives and/or thresholds and are presented accordingly. A typical example of the dashboard developed by Eurostat is the one for Principal European Economic Indicators (PEEIs), whilst the Macroeconomic Imbalance Procedure (MIP) and Sustainable Development Indicator (SDI) sets of indicators are probably the most known examples of EU-level scoreboards.

Both dashboards and scoreboards aim to describe in a very detailed way the phenomenon to be monitored and for this reason they can include a relatively large number of indicators coming from various statistical domains. One important implication of the development of such sets of indicators has been the overall data quality improvement of the constituent statistics, with beneficial effects in terms of coverage, relevance, harmonisation, reliability and timeliness. It is also important to underline that, especially in the social or socio-economic context, both dashboards and scoreboards may contain cardinal as well as ordinal indicators, which have implied additional efforts in finding the best way of presenting non-quantitative variables.

([3]) Available at: http://www.cbs.nl/en-GB/menu/themas/dossiers/conjunctuur/publicaties/conjunctuurbericht/inhoud/conjunctuurklok/toelichtingen/conjunctuurdashboard.htm

([4]) Available at: http://data.gov.uk/apps/uk-economic-data-dashboard

([5]) This paper does not go into the background literature on indicators, but it is interesting to see that the distinction in that literature between 'descriptive indicators' and 'performance indicators' can be analogous to the common descriptions of dashboards and scoreboards respectively.

Lastly, it should be underlined that while dashboards and scoreboards provide very detailed, precise and often almost exhaustive pictures of a given phenomenon, they do not necessarily allow for a quick and easy identification of the key messages delivered by the constituent indicators. This is especially true when the number of indicators is relatively large. It is not infrequent that, in the same set of indicators, variables move in different directions, complicating the synthetic evaluation of the phenomenon. As an example, looking at the PEEIs set, it is not evident whether, after having exited a recessionary phase, the European economy is growing above or below the trend and how other economic indicators (prices, employment etc.) are relevant for this analysis.

3. Prioritising and summarising information

In order to overcome the drawbacks of dashboards and scoreboards described at the end of section 2 it is important to identify ways of summarising the information from multiple indicators requires the utilisation of tools that capture and highlight the main driving forces or key events characterising the statistical indicators included in a given dataset. Obviously, within this process, a lot of information related to sectorial behaviours or to other specificities is lost, privileging a synthetic picture instead of a detailed one that would be available from a dashboard or scoreboard. There is a wide variety of methods and tools that have been developed to summarise information. They stem from graphical and mathematical techniques, non-parametric and parametric statistical methods, and linear and non-linear approaches.

The most intuitive and popular way to summarise information is by constructing one or more composite indicators built up on the basis of a preselected number of statistical indicators from a given set. Nevertheless, this approach has been widely criticised, especially outside the macroeconomic area, for several reasons. They will be shortly described later in this section. Providing an overview of the available methods and approaches to summarise information, even if in a non-exhaustive way, is a challenging task. The best way to proceed is to start from the intrinsic characteristics of data involved and then to identify the most commonly used techniques. It is helpful to distinguish between social and socio-economic phenomena and macroeconomic ones, mainly because in the first case, ordinal variables are widely used whilst, in the second case, mostly quantitative indicators are present.

When considering social or socio-economic phenomena, such as wellbeing, quality of life, etc., the first important consideration is that the reference variable is usually not measurable in a cardinal scale and that a unique measure of such phenomena is not available, even if in an ordinal way. Furthermore, many indicators involved in the measurement of such phenomena are themselves measured in an ordinal scale. This is the case for example of material deprivation, poverty, etc. Despite the specificities of such sets of indicators, composite indicators have been widely used, mainly to provide an approximation of the non-measurable and latent reference variables. The construction of composite indicators is usually based on three main steps: variable selection, definition of the weighting scheme and aggregation. The main critics on the use of composite indicators in this area mainly involves the three steps described above as they resort to statistical techniques originally developed for dealing with quantitative data. This implies that an implicit cardinalisation of the indicators is used by means of rescaling methods which can alter the intrinsic nature of the indicators.

Furthermore, the attempt to reduce subjective intervention in the construction of composite indicators by means of statistical techniques, such as principal components, linear correlation, etc. arguably contradicts the nature of the component indicators and of the latent reference variable which, by its nature, reflects subjective preferences. Starting from such criticism, a variety of alternative methods based on the prioritisation and ranking of variables has been developed, avoiding the construction of compostite indicators. By means of ranking techniques, it is possible to identify the most significant statistical indicators amongst a large set in order to summarise information content through a smaller number of indicators. Among these methods, the most commonly used are propably the partially ordered sets (POSET) which derive directly from the mathematical sets theory. This method is widely used in the social and socio-economic fields, but also outside, for example in finances and in the evaluation of fiscal policy. In section 5 there is a description of the POSET method and its application to PEEIs.

In the macroeconomic area, the situation is very different because the reference variable is usually measurable and available. This is the case of Gross Domestic Product or, in a more complex way, of the weighted combination of statistical indicators proposed by the US Conference Board (see A. Ozyildirim (forthcoming)). Furthermore, the large majority of indicators are quantitative and, even for some qualitative indicators such as opinion surveys, well established. Widely accepted quantification techniques are available and regularly used. In such a situation, the use of composite indicators is much less subjected to criticism, even if there are still ongoing debates on the use of purely aggregation techniques based on a more or less subjective weighting scheme versus fully model based composite indicators. Since the reference variable is directly measured in the macro-economic field, the reason for compiling composite indicators is rather different. Composite indicators are not intended to approximate the reference variable but mainly to fill gaps in existing statistics and to highlight hidden phenomena. Composite indicators are usually constructed for:

1. Providing an estimation of the current evolution of the reference variable and/or anticipating it in the near future;
2. Estimating some unobserved components of the reference variable such as the trend and cycle, and providing an estimation of their current and future behaviour;
3. Providing an estimation of the occurrence of rare events such as the cyclical turning points for the current period and the near future.

Macroeconomic composite indicators can be further classified according to the following main criteria: timing, construction method and reference variable. A detailed classification of composite indicators is proposed in Carriero and Marcellino (2007); while a taxonomy is provided in Carriero et al. (forthcoming). Concerning timing, they are usually classified into leading (anticipating the near future), coincident (now-casting the present) and lagging (replicating the past) indicators. It is important to note that lagging indicators are mostly relevant to the producers of indicators, which use them for ex-post validation of coincident and leading ones.

Regarding construction, indicators can be distinguished into those based on aggregation techniques, on non-parametric techniques (e.g. partially square), on parametric techniques (e.g. dynamic factor models), on linear time-series techniques (e.g. VAR models) and on non-linear time-series techniques (e.g. Markov-switching models). According to the reference variable, indicators can be distinguished into those based on a well identified statistical indicator (GDP, Industrial Production Index (IPI), etc.), on a combination of statistical indicators (Conference Board approach), on a historical estimation of the trend and/or of the cycle of a given statistical indicator or a combination of them, and on a previously established historical sequence, of turning points based on a single statistical indicator or a combination of them.

Finally, it is interesting to notice that in the macroeconomic context, a special case is made by so-called sentiment or climate indicators. The specificity of those two indicators rely on the fact that they are constructed without an explicit identification of a reference variable, even if they are implicitly strongly related to some quantitative variables such as GDP or the IPI. In section 6, we present a system of coincident indicators for detecting turning points, together with a graphical tool designed for their dissemination in a user-friendly way.

4. Dashboards and scoreboards for macroeconomic analysis

In this section we describe how concepts and principles of dashboards and scoreboards, presented in section 2, have been applied by Eurostat to the macroeconomic field. The common objective of macroeconomic dashboards and scoreboards has been to provide policy makers and analysts with friendly and clear tools for monitoring specific macroeconomic aspects such as short-term evolution and the presence of structural imbalances. We focus here on the PEEIs dashboard for short-term macroeconomic monitoring for a detailed description of the MIP scoreboard for the detection of macroeconomic imbalances, we refer to Ruggeri et al. (2015).

4.1. The PEEIs dashboard

In October 2007, Eurostat released the so-called 'selected PEEIs page'. For the first time, this tool presented statistical indicators available at different time frequencies and coming from different areas of official statistics in a single web page and framework. Furthermore, this page provided information on data availability and characteristics such as the link to the last available press release and to the date of the next one, a short description for each statistical indicator in a harmonised form and a full access to metadata. The statistical coverage was constituted by all available PEEIs plus a small number of monetary, financial and balance of payment indicators, as well as the economic sentiment indicator provided by DG ECFIN of the European Commission. The 'selected PEEIs page' was available for the euro area and the European Union only.

Despite the relatively small number of indicators, the dashboard provided a good picture of the short-term economic situation. In the following years, the 'selected PEEIs page' has been one of the starting points for the discussion and implementation of wider dashboards, such as the Principal Global Indicators (PGIs) and the data template of the United Nations Statistical Division (UNSD), to which Eurostat has actively cooperated.

The relevance of wider dashboards for monitoring the short-term economic situation has of course considerably increased as a consequence of the global financial and economic crisis. The main objective of such wider dashboards is to ensure a regular and almost real-time monitoring of all aspects of the short-term economic situation.

In 2015, Eurostat decided to replace the 'selected PEEIs page' with the new 'PEEIs dashboard' which keeps all the features of the previous one, plus some new ones like the possibility of direct data downloading, and increases the statistical coverage of the dashboard with respect to indicators (both key and additional indicators) and to Member State data. The live PEEIs dashboard can be viewed from the main Eurostat website (http://ec.europa.eu/eurostat) by following the dedicated PEEIs link.

4.2. Some limitations of macroeconomic dashboards: the case of PEEIs

In section 2, we have already discussed the fact that, despite their global overview of the situation, the reading of dashboards is not always easy, as they sometimes display contradictory messages and/or some relevant signals are somewhat hidden. In this subsection we develop these aspects by means of some examples.

If we look at the PEEIs dashboard today, it is quite clear that the message delivered is mainly positive, at least at the euro area and European Union level. Looking into details, we have selected some relevant PEEIs grouped into three main categories: economic growth, price evolution and labour market conditions, which are presented in table 1.

Table 1: Latest evolution of some euro area PEEIs

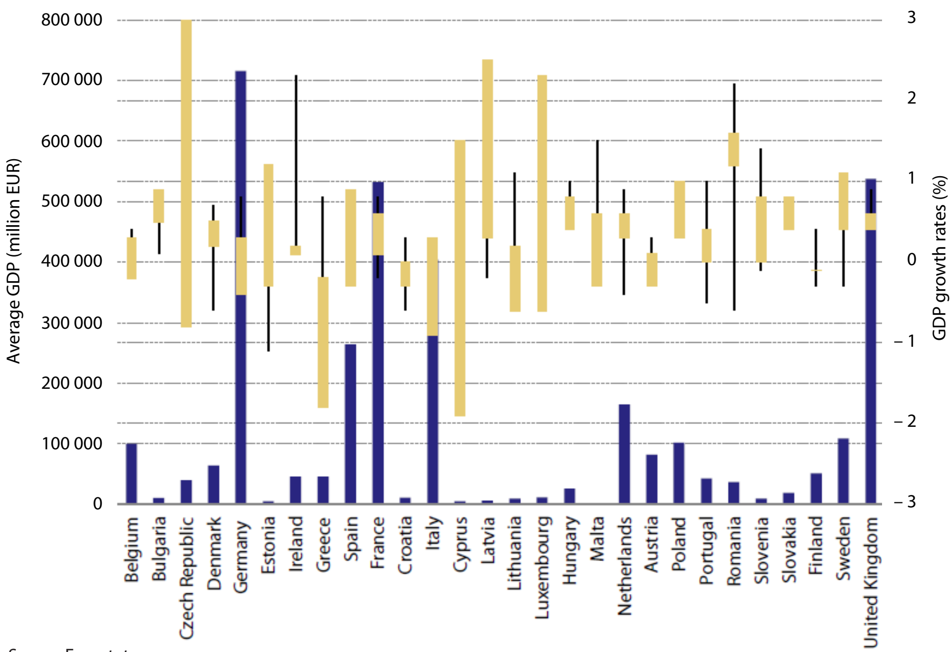
Euro area key short-term indicators						
Economic growth		**2014Q1**	**2014Q2**	**2014Q3**	**2014Q4**	**2015Q1**
	GDP growth rates (Q/Q – 1)	0.2	0.1	0.2	0.4	0.4
Consumer prices		**2015M01**	**2015M02**	**2015M03**	**2015M04**	**2015M05**
	Harmonised Index of Consumer Prices (HICP) (M/M – 1)	– 1.5	0.6	1.1	0.2	0.2
Business indicators		**2015M01**	**2015M02**	**2015M03**	**2015M04**	**2015M05**
	Industry producer prices (M/M – 1)	– 1.1	0.6	0.2	– 0.1	:
	Production in industry (M/M – 1)	0.0	1.0	– 0.4	0.1	:
	Retail trade deflated Turnover (M/M – 1)	0.3	0.1	– 0.6	0.7	:
Labour market		**2015M01**	**2015M02**	**2015M03**	**2015M04**	**2015M05**
	Unemployment rate (M/M – 1)	11.3	11.2	11.2	11.1	11.1
		2014Q1	**2014Q2**	**2014Q3**	**2014Q4**	**2015Q1**
	Employment rate (Q/Q – 1)	0.2	0.3	0.2	0.1	0.1

Source: Eurostat

We notice that GDP is characterised by a constantly positive evolution since the first quarter of 2014, which seems to accelerate from the end of 2014. However the industrial production index and retail trade deflated turnover show a less clear path, with some negative results also in 2015. This could raise some uncertainties amongst users concerning the consolidation of growth. Furthermore, looking at price evolution, represented by the HICP, it seems that the period of price decrease or stagnation has passed but, when comparing those results with the industrial producer price index, this trend appears less evident. The labour market data, represented by the unemployment rate and employment evolution, do not provide a clear insight on the impact of GDP growth.

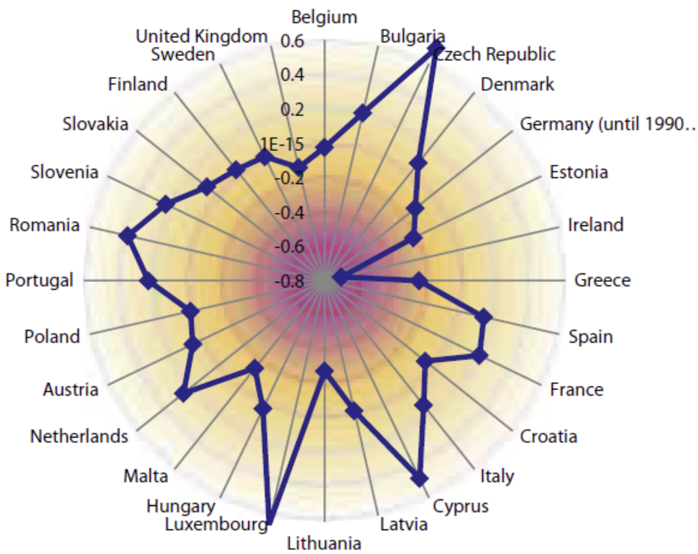
Another example is provided by the analysis of GDP growth at Member State level during the last five quarters. The evolution of GDP by Member States is shown in figures 1 and 2, which provide different visualisation approaches. The first one mainly focusing on the growth itself and the second one on the acceleration/deceleration of growth. The message from both approaches is that Member States are not growing in a homogenous way and it is not possible to evaluate which countries might be growing above or below the trend.

Figure 1: GDP growth and average by country from 2013Q1 to 2015Q1 (million EUR, %)



Source: Eurostat

Figure 2: GDP growth acceleration by country from 2014Q1 to 2015Q1
(average)



Source: Eurostat

It is thus important to highlight that, just by looking at the dashboard we are unable to answer relevant questions related to the current economic situation such as:

- Are the European economies growing below or above the trend?
- How synchronised are European economies?
- Which economies are still in a slowdown or in a recessionary phase in their economic cycle?

Answering the above questions is not necessarily easy and can sometimes imply sophisticated elaboration arguably going beyond the tasks of statistical agencies. Nevertheless, the use of some coincident indicators, such as those presented in section 6, could help in highlighting some hidden aspects of the economic situation. By combining the messages from dashboards and composite indicators, it becomes easier to answer some of the above questions.

5. Using POSET in the macroeconomic context

As already mentioned in section 3, the concept and definition of composite indicators can be subject to criticism, especially when dealing with phenomena which cannot be easily quantified, unless if we make strong and, in some cases, arbitrary assumptions. This is the case of several socio-economic phenomena such as material deprivation, poverty, quality of life, wellbeing, etc. In recent years, several studies aiming to identify alternative ways to replace at least partially the use of composite indicators have been conducted. Those studies have concentrated their attention on so-called ranking methods, which allow the creation of an order among groups, countries, etc. An excellent overview of the possibilities offered by POSET for ranking multi-indicators sets has been proposed by Brüggemann and Patil (2011), whilst Brüggemann et al. (2014) and Fattore et al. (2011) present very interesting applications of the POSET theory to poverty and material deprivation, respectively. By contrast, Badinger and Reuter (2014) apply the POSET approach to a very different domain, namely the evaluation of fiscal rules across countries.

5.1. The description of the POSET method

Within the set and ordering theories, a partially ordered set (POSET) formalises and generalises the intuitive concept of ordering, sequencing, or arranging of the elements of a set. A POSET consists of a set coupled to a binary relation indicating that, for certain pairs of set elements, one of them precede the other. This ordering is called 'partial' since it is necessarily possible to define a binary relation for all the elements of a given set so that one of them precede another or vice versa.

A (non-strict) partial order is a binary relation '$\leq$' over a set P which is reflexive, antisymmetric, and transitive, i.e., which satisfies for all x, y, and z in P (Davey and Priestley (2002), Neggers and Kim (1988), Schroeder (2003), Fattore et al. (2011)):

1. $x \leq x$ (reflexivity);
2. if $x \leq y$ and $y \leq x$ then $x = y$ (antisymmetry);
3. if $x \leq y$ and $y \leq z$ then $x \leq z$ (transitivity).

If $x \leq y$ or $y \leq x$, then x and y are called comparable. Otherwise they are said to be incomparable (written $x \parallel y$). A partial order P where any two elements are comparable is called a chain or a linear order. On the contrary, if any two elements of P are incomparable, then P is called an antichain. Thus, partial orders generalize the more familiar total orders, in which every pair is related. A finite POSET can be visualized through its Hasse diagram, which depicts the ordering relation.

In order to understand how the POSET theory can be applied to socio-economic phenomena, we consider, following Fattore et al. (2011), a set of k ordinal variables, v_1; ... ; v_k, associated to a given socio-economic phenomenon. Each possible sequence of ordinal scores on v_1; ...; v_k defines a different profile. Profiles can be (partially) ordered in a natural way, by the following dominance criterion:

Definition: Let *s* and *t* be two profiles over $v_1; \dots ; v_k$; we say that *t* dominates s if and only if $v_i\ (s)\ \leq v_i\ (t)\ \forall i = 1, \dots, k$ where $v_i\ (s)$ and $v_i\ (t)$ are the ordinal scores of *s* and *t* so on v_i.

Since not all the profiles can be linearly ordered based on the previous definition, they constitute a POSET.

5.2. Possible applications of POSET to the PEEIs

In the large majority of studies, POSET is used when ordinal variables are present to overcome some drawbacks of composite indicators (Fattore et al. (2011)). Nevertheless, there are no formal obstacles to the use of such ranking techniques, also in presence of cardinal variables. Obviously, the ordering rules have to be appropriately defined also taking into account the quantitative nature of the variables. Furthermore, the added value associated to the use of those techniques, especially in relation to composite indicators and dashboard/scoreboard, has to be carefully evaluated.

At this preliminary stage we have identified four possible applications of POSET to the PEEIs. The first one aims to detect the presence of cross-sectional effects in financial markets simultaneously to some growth/business cycle phases. By ranking stocks according to their book-to-market and capital size, Liew and Vassalou (2000) demonstrated that these cross-sectional factors contain significant information about future GDP growth. Their approach consists in building factors based on rankings, so they do not use rankings directly. The POSET would allow a direct approach. Close to what would be a POSET approach, a direct use of rankings to study cross-sectional effects has been proposed by Billio et al. (2011) and Billio et al. (2012). In the second application, given an order following which PEEIs can be classified (for example as 'good', 'medium' or 'bad'), we could combine them to explain the economic phases, like expansion, slowdown, recession, etc., each of them associated to a natural number. In the third one, the ranking of some countries' PEEIs could be used to explain the ranking of countries according to another variable (GDP growth rates, etc.).

Finally, the fourth application could consist in associating a ranking to each of the economic phases. Following Harding and Pagan (2006), a distance between these ranks could be used to measure the diffusion of a crisis. The country ranks could be combined to estimate the economic phase of individual countries or aggregate. Combining past rankings to explain the current economic phase would be interesting to study the synchronisation among countries. It is worth to note that the information provided by the cyclical indicators presented in 6.1 and displayed by the business cycle clock in 6.2 could constitute the ideal input for this application so that, for the first time, composite indicators and POSET can be used together in order to assess relevant cyclical phenomena, such as the synchronisation and the diffusion of turning points.

The results obtained until now are still preliminary and not very conclusive. What has emerged is that the third application is the easiest to implement, even if the results could be quite obvious, so that its added value will be relatively low. By contrast, the first, second and fourth applications appear to be more challenging due to the fact that some quite complex hypotheses have to be formulated but their informational content from analysts' point of view is expected to be relatively high.

6. Turning points detection and the new business cycle clock — an example of composite indicators

In this section we focus our attention on the construction of composite indicators aiming to detect economic turning points in a timely way. Since turning points are relatively rare phenomena, not occurring at regular intervals, and since they indicate discontinuities in the regular path of a time series, non-linear modelling techniques appear the most appropriate ways to deal with them. Since the publication of the seminal paper from Hamilton (1989), Markov-Switching (MS) models have been considered the most reliable tools for turning point detection and have been applied in several studies and research. Alternatively, some other researchers have concentrated their attention on binary regression models such as PROBIT and LOGIT, e.g. Chauvet and Potter (2005) and Harding and Pagan (2011).

Since 2007, Eurostat has been involved in the construction of turning points coincident indicators based on MS models (Anas et al. (2008)) which more recently have evolved into the use of multivariate MS models (MS-VAR) (Billio et al. (2015) and Anas et al. (forthcoming)).

The use of MS models at Eurostat has also been empirically validated in comparison to other non-linear models (i.e. Billio et al. (2013)). The most appealing feature of the MS model is constituted by the fact that they allow for a different dynamic according to the regime in which the phenomenon under evaluation is situated. In particular, by considering a two-regime representation where the regime could be assimilated to expansion and recession phases, if the economy is in recession at time T, at time $T+1$ it can either continue to stay in the same regime or to switch to expansion. The probabilities to stay in a phase or to switch phases can be estimated and they determine the expected duration of each phase. Given a threshold usually assumed equal to 0.5 (natural rule), when the recession probability is above/below 0.5, the MS model is expected to stay in the recessionary/expansionary regime for the time of the respective duration. Crossing the threshold in any of the two directions indicates the presence of a turning point.

In the subsection 6.1 we will present a step by step approach to the construction of the Eurostat cyclical composite indicators, while subsection 6.2 will be devoted to the description of a new graphical tool, called the business cycle clock, to disseminate the results of these indicators in an easy to read and intuitive way.

6.1. Step by step construction of cyclical composite indicators

Five steps may be identified, as follows.

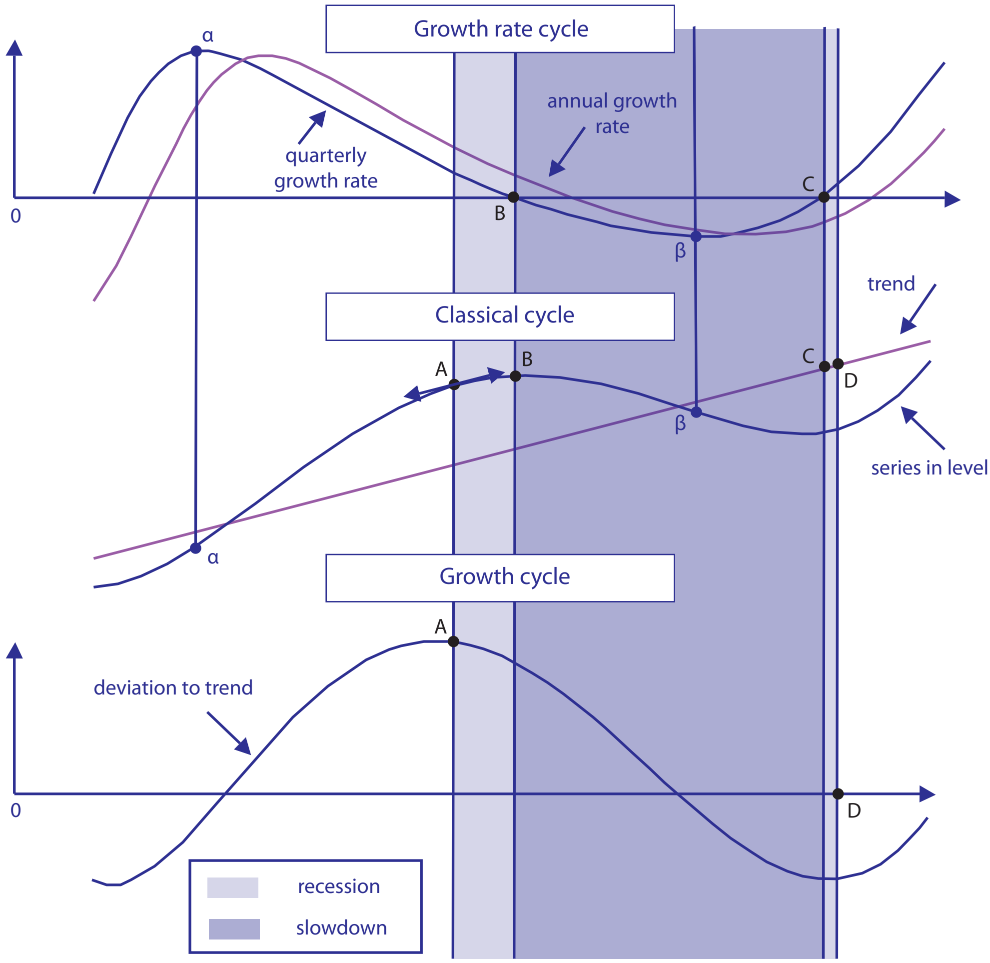
Step 1: Identification of the cycle to be monitored.

A. Classical Business cycle (Burns and Mitchell definition (1946)), which is very relevant for detecting recessions but not very informative during usually quite long expansion phases.

B. Growth cycle (Output gap), which is very relevant to understand the position with respect to the potential output (trend) and more informative also during the expansion phases of business cycle. It leads to the peaks and troughs of the business cycle by some months but it doesn't necessarily detect the start and the end of recessions.

C. Growth rate cycle (Acceleration cycle), which is characterised by the highest number of fluctuations and a high degree of volatility. It leads to the growth cycle peaks and business cycle troughs corresponding to the inflexion points of the classical business cycle. They determine the acceleration and deceleration phases of the economy.

Figure 3: Integrated framework for cyclical monitoring



Source: Anas and Ferrara (2004)

The approach retained by Eurostat consists in jointly monitoring cycles (Anas and Ferrara (2004)) within an integrated framework:

i. Growth cycle and Business cycle (*ABCD* sequence)

ii. Also including Acceleration cycle (*αABβCD* sequence)

The sequence of turning points is presented in figure 3.

Step 2: A historical dating chronology is computed for the classical cycle, the growth cycle and the acceleration cycle by means of a simple non-parametric dating rule (Harding and Pagan (2002)) applied to GDP, IPI and unemployment rate. The historical dating chronologies are constructed following the *αABβCD* approach (Anas et al. (2008) and Anas et al. (forthcoming)) and turning points are supposed to remain constant after a given number of years.

Step 3: Construction of a middle-sized dataset mainly containing original PEEIs and opinion surveys indicators together with their most appropriate data transformation highlighting cyclical movements.

Step 4: Variable selection based on the ability of timely and precisely detecting turning points within a real-time simulation exercise against the non-parametric historical turning point historical dating (Step 2).

Step 5: Selected variables are used to identify and estimate a number of autoregressive Markov-Switching models (MS-VAR):

MSIH(K) – VAR(L), where *H* indicates the presence of heteroskedasticity, (*K*) is the number of regimes and (*L*) the number of lags of the autoregressive part.

Remark: Dealing simultaneously with growth cycle and business cycle implies a number of regimes not smaller than 4, while the heteroskedastic part can or cannot be present depending on the degree of fluctuation asymmetry.

Step 6: From step 5, *N* best fitting models are identified, each of them producing a pair of coincident indicators for the growth cycle and the business cycle respectively, labelled as *MS-VAR GCCI* (multivariate growth cycle coincident indicator) and *MS-VAR BCCI* (multivariate business cycle coincident indicator): *MS-VAR GCCI (j)* and *MS-VAR BCCI (j); j=1 …N.*

Remark 1: Each composite indicator is defined between 0 and 1, and can be viewed as a composite probability of being in a recessionary phase for the *MS-VAR BCCI (j)* and in a slowdown phase for the *MS-VAR GCCI (j)*.The recession/slowdown regions are defined on the basis of a threshold, usually equal to 0.5.

Remark 2:

- *MS-VAR BCCI (j) > 0.5 = recession*
- *MS-VAR GCCI (j) > 0.5 = slowdown*

By construction, *MS-VAR BCCI (j) > 0.5* → *MS-VAR GCCI (j) > 0.5*, so that the *ABCD* sequence is always fulfilled.

Step 7: The indicator for the acceleration cycle, labelled as *ACCI* (acceleration cycle coincident indicator), cannot be modelled within the multivariate framework described in step 6 for the growth cycle and business cycle coincident indicators due to purely mathematical reasons. It is then independently computed by means of a univariate two regimes Markov-Switching model fit to the economic sentiment indicator.

Step 8: Within a real-time simulation exercise, the *N* pair of composite coincident indicators is compared with the non-parametric historical turning point dating.

Step 9: The identification of the best performing pair of coincident indicators is based on the outcome of step 7, using the following criteria:

- Maximisation of the Concordance Index
- Minimisation of the Brier's Score (QPS)
- Minimisation of type-2 errors: detection of false cycles
- Minimisation of type-1 errors: missing cycles

Remark: Due to the trade-off between type-2 and type-1 errors, the simultaneous minimisation of both is unachievable. A conservative approach suggests privileging the minimisation of type-2 errors, i.e. the detection of false cycles.

It may be seen from the steps described above that the compilation of composite indicators is technically demanding, and involves the use of multiple assumptions and modelling specifications.

6.2. The new business cycle clock

The outcome of the cyclical indicators described in 6.1 can be presented either in a graphical or in a tabular form. Figures 4 and 5 show, for the euro area, the evolution of the univariate acceleration cycle coincident indicator (*ACCI*) and the multivariate growth cycle and business cycle coincident indicators (*MS-VAR GCCI* and *MS-VAR BCCI*), respectively. They also show the results of corresponding historical dating chronologies.

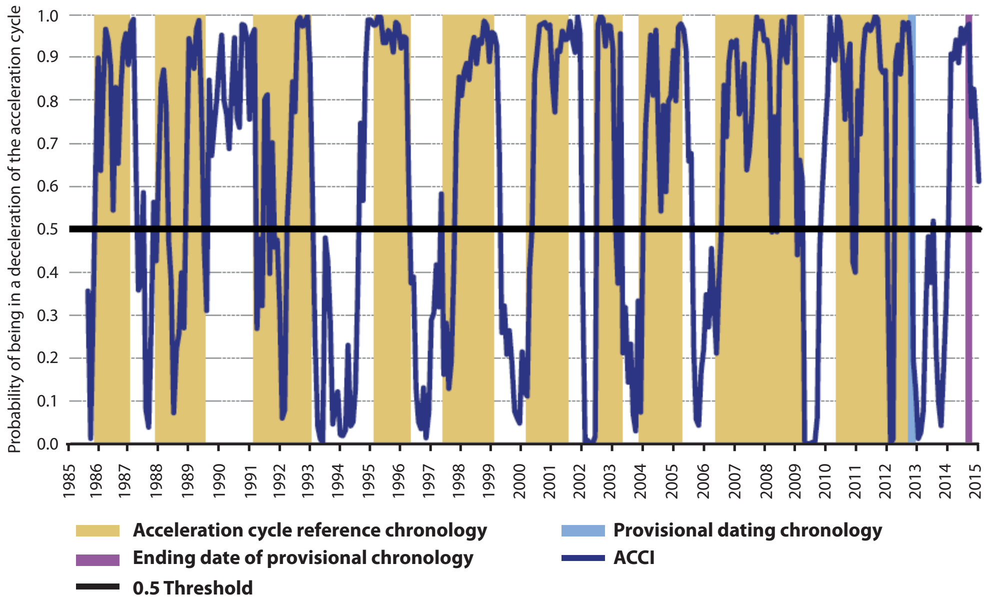
Both figures show in the vertical axis the recession probabilities of each indicator and in the horizontal one the time scale. The horizontal line at 0.5 indicates the chosen threshold. Figure 5 shows *MS-VAR BCCI* on the top panel and *MS-VAR GCCI* in the bottom one since they are simultaneously derived from a single multivariate model. The grey areas in both figures indicate the historical dating chronology.

Table 2: Latest Euro area peaks and troughs

		Peak	Trough	Peak	Trough	Peak	Trough	Peak
Growth cycle	Provisional dating	2008Q1	2009Q3	–	–	2011Q3	2013Q2	
	MS-VAR GCCI	Dec-07	Sep-09	–	–	May-11		
Business cycle	Provisional dating	2008Q1	2009Q2	–	–	2011Q3	2013Q1	
	MS-VAR BCCI	Apr-08	Sep-09	–	–	Jul-11	May-13	
Acceleration cycle	Provisional dating	2006Q2	2009Q1	2010Q2			2012Q4	
	ACCI	Jun-06	Mar-09	Dec-10	Dec-11	Mar-12	Oct-12	Dec-13

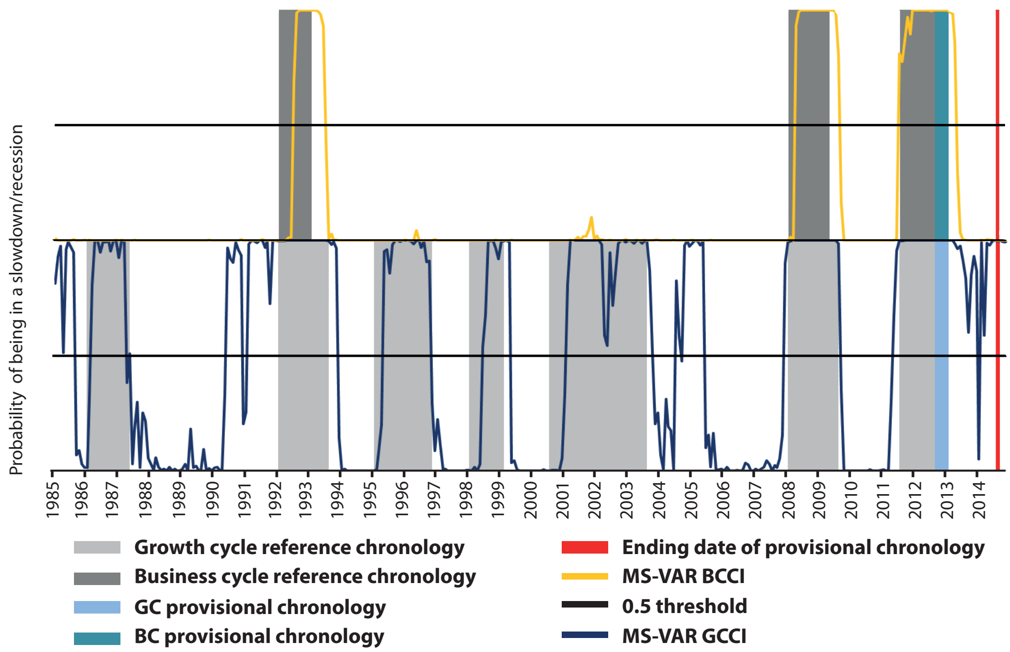
Source: Author's calculations

Figure 4: Euro area ACCI univariate



Source: Author's calculations

Figure 5: Euro Area MS-VAR BCCI and GCCI multivariate



Source: Author's calculations

Table 2 summarises the latest euro area peaks and troughs for the three cyclical indicators comparing them with the results of the historical dating chronologies (in grey).

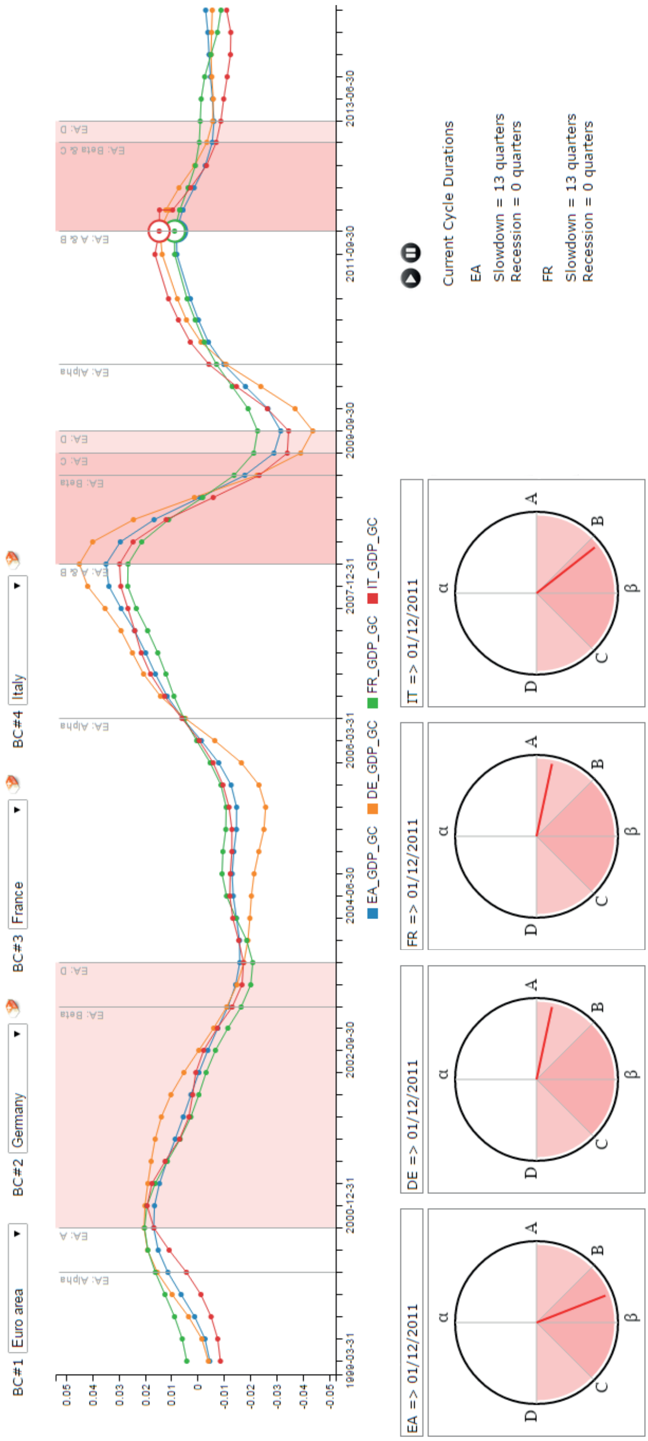
While the interpretation of the indicators' outcome is not particularly challenging for sophisticated users, it is quite clear that this is not a friendly way to present the data to a wider audience. Furthermore, since the indicators are presented individually, it is not easy to understand the relations between them so that the global assessment of the cyclical situation, which is the main added value of this system, is often hidden. In order to overcome this, a graphical tool is under development. It is intended to provide an intuitive, easy to read and user-friendly picture of the cyclical situation based on the outcome of indicators, which are not directly displayed but used as the data source to animate the tool.

The tool under development is mainly based in a clockwise representation of the cyclical movements. In the last year, several institutions including Eurostat have developed clock-based representations of the cyclical movements. This is the case for Statistics Netherlands with the business cycle tracer, of the business cycle monitor of the German Federal Statistical Office and the OECD business cycle clock. What is really innovative in the Eurostat proposal is that the representation of the cycles within the clock is given by the set of cyclical composite indicators presented in 6.1. The layout of the new business cycle clock we are proposing is presented in figure 6.

Figure 6 is divided in three main parts: on the top there is an historical graphical representation based on the evolution of GDP; on the lower left corner, one or more clocks are displayed; while on the lower right corner some statistics associated to the cycles are presented. The upper part contains a didactical representation consisting of the GDP deviations from the trend, where the peaks and troughs of the cycles are highlighted. The slowdown phases are represented in pink; the recession phases are represented in dark pink; each point of the $\alpha AB\beta CD$ cycle is represented by a vertical line. The graph is based on the data obtained by the historical dating described in step 2 of section 6.1. For this reason, it does not contain information for the latest time periods but it provides a historical overview of the cycles over a long time horizon.

It is worth noting that the clock and graph representation are dynamic. A play button sets time running. The current position in the graph representation is highlighted and the clock hand runs.

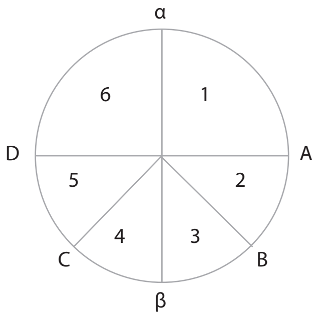
Figure 6: Structure of the new business cycle clock



Source: Author's calculations

The clock on the lower left part is structured according to the *αABβCD* approach, presented in step 1 of section 6.1 (see figure 7).

Figure 7: Clock structure



Noon is *α*, peak of the growth rate cycle; 3 pm is *A*, peak of the growth cycle; 4.30 pm is *B*, peak of the business cycle; 6 pm is *β*, trough of the growth rate cycle; 7.30 pm is *C*, trough of the business cycle; 9 pm is *D*, trough of the growth cycle. Those turning points delimitate six sectors in the clock which correspond to various phases of the business cycle. The location of the hand in the clock is based on the values of the three cyclical coincident indicators for the acceleration, growth and business cycles described in 6.1, as well as on their positioning with respect to the 0.5 threshold. Table 3 synthetically presents the characteristics and meaning of the various sectors.

Table 3: The clock sectors and the cyclical composite indicators

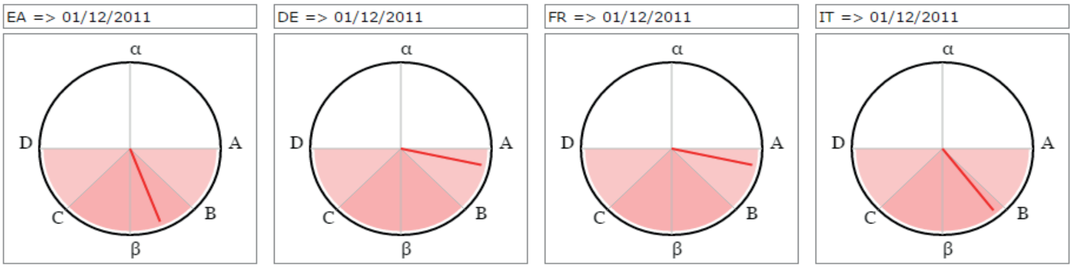
		ACCI			
		<0.5		>0.5	
		BCCI		**BCCI**	
		<0.5	>0.5	<0.5	>0.5
GCCI	<0.5	6 *Recovery*	/	1 *Deceleration*	/
	>0.5	5 *Expansion*	4 *Acceleration*	2 *Slowdown*	3 *Recession*

Source: Author's calculations

From table 3, we can say that in sector 1 the economy is growing above the trend but its growth is progressively decelerating. In sector 2, the still positive growth is below the trend while in section 3 it becomes negative. In sector 4, the negative growth starts to accelerate approaching zero. In sector 5, it becomes positive but still below the trend, while in sector 6 the economy is growing above the trend and accelerating.

The new business cycle clock aims to assess and compare the situations among different countries. For instance, in figure 8 we illustrate the comparison of Germany, France and Italy with the Euro Area. The new business cycle clock can display up to 4 countries simultaneously.

Figure 8: Cross-country comparison in December 2011



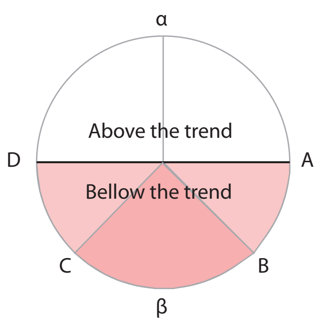
Source: Author's calculations

It is worth noting that the new business cycle clock tool will be accompanied by substantial documentation including standard metadata files for the tool itself and for the cyclical indicators as well as methodological notes. The tool and the cyclical indicators will be regularly monitored by Eurostat and a quality assessment will be disseminated annually together with the description of any improvements introduced.

6.3. Using the clock

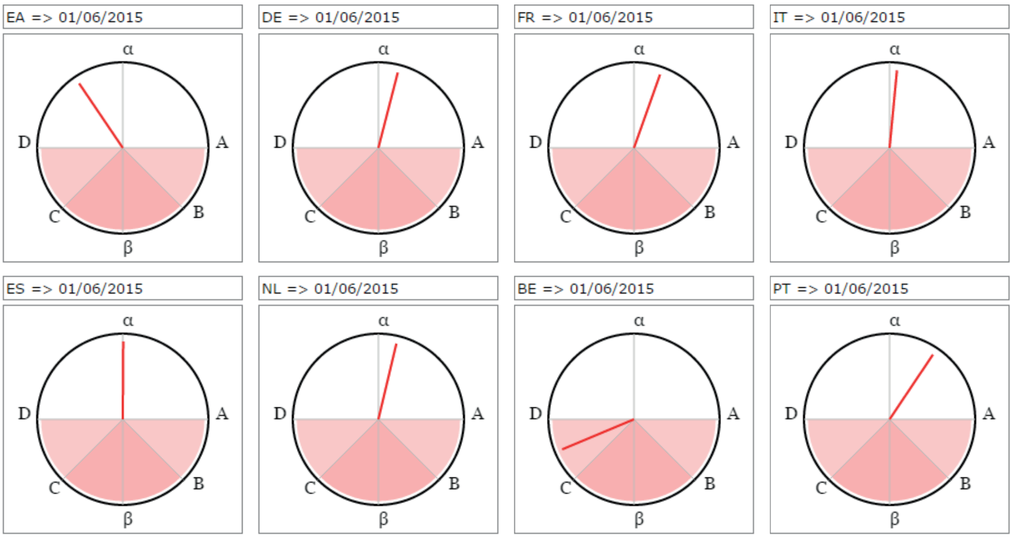
In this subsection we will try, by using the information delivered by the business cycle clock, to find an answer to some questions raised in the subsection 4.3, to which the PEEIs dashboard could not provide clear evidence. The first question is related to the identification of which economies are growing above the trend and which ones are still below. Since the growth cycle is defined as the deviation from the trend, so that $GC = Y_t - Y^p_t$ for $t = 1, 2, \ldots, T$; where GC is the growth cycle, Y is the actual growth and Y^p is the trend, it is possible to show that the growth cycle will cross the trend in A (in a descending phase) and in D (in an ascending phase) of the clock. By drawing a line between A and D we can conclude that, in the sectors of the clock above the line, the economy is growing above trend (sectors 6 and 1), while in the others the economy is growing either below trend or even decreasing. This is shown in figure 9.

Figure 9: The clock and the economic growth



The main difference between staying in sector 6 or in sector 1 is that, in the first case, the economy is growing above the trend and it is still accelerating, while in the sector 1 it has started a deceleration phase while still growing above the trend. By using those results we can analyse in a comparative way the growth of some Euro area member countries by referring to figure 10.

Figure 10: Cyclical situation of the Euro area and some member countries in June 2015

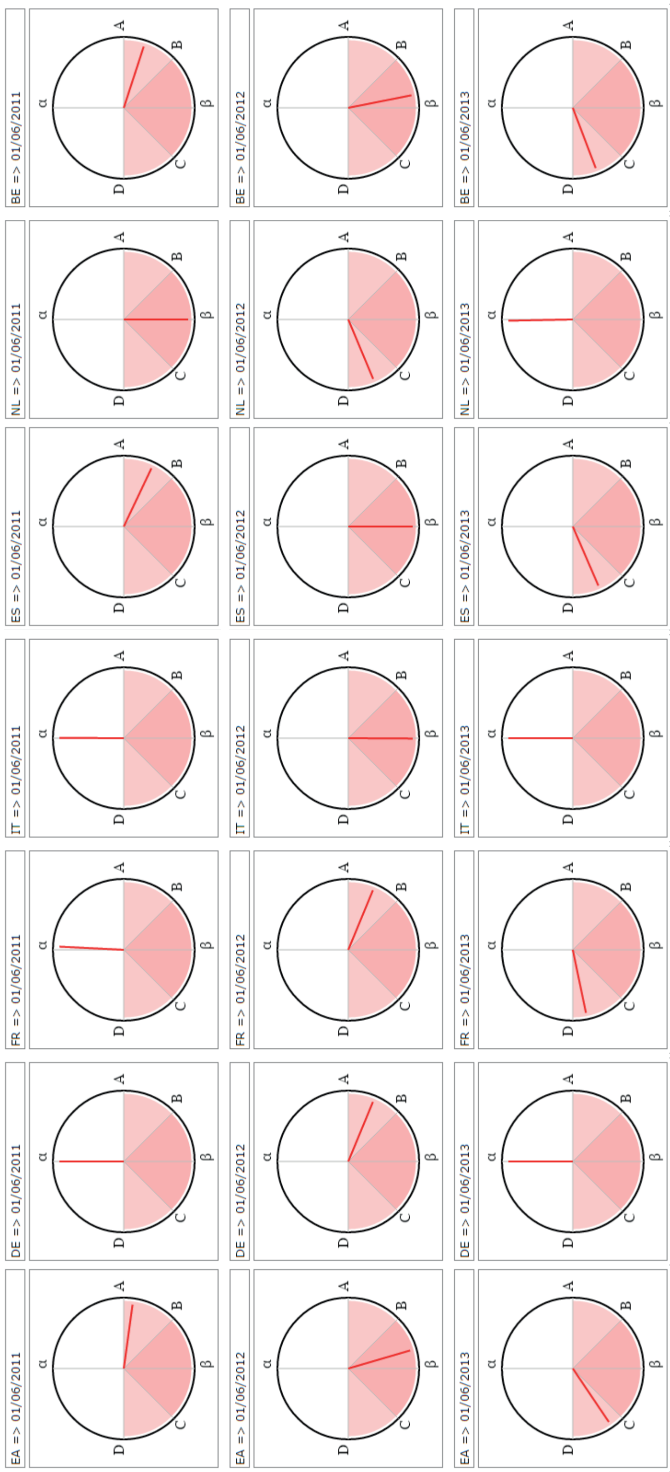


Source: Author' calculations

By looking in detail to the various clocks it emerges that almost all economies are growing above trend except Belgium which is still growing below. Furthermore, the Euro area is still in an acceleration phase, while Spain has achieved the peak of the acceleration cycle. For the remaining economies, we can observe that Italy is just at the beginning of the deceleration phase while France, Germany and the Netherlands, as well as Portugal, have a more consolidated deceleration phase. Since for all those economies, the hand is located in the first half of the sector 1, we can also conclude that the risk of reaching point A and therefore starting to grow below the trend, is very low. In this case, by combining the information delivered by the dashboard and the one delivered by the clock, it is possible not only to rank the countries according to the intensity of growth (above or below the trend) but also to obtain useful insight srelated to the acceleration or deceleration of their growth.

Another question raised in section 4.3 concerns the degree of cyclical synchronisation among Euro area countries. To answer the question, we analyse the evolution of the cyclical situation, represented by a series of clocks at different points in time.

Figure 11: Analysis of the 2012/2013 recession at Euro area and member countries level



Source: Author's calculations

In figure 11, we consider the Euro area plus six member countries (Germany, France, Italy, Spain, the Netherlands and Belgium). Their cyclical situation is assessed respectively in June 2011, 2012 and 2013. By looking at the countries and Euro area behaviour at the three different points in time, we can observe the following:

- June 2011 shows the immediate entering into slowdown of Spain, Belgium and the Euro area, while Germany, France and Italy are still in expansion; the Netherlands is the only country already in recession, achieving the trough of the acceleration cycle.
- June 2012 shows Spain, Belgium, Italy and Euro area in recession. Germany and France are in slowdown only. The Netherlands exited the recession but remained in slowdown.
- June 2013 shows the exit of Euro area, Spain and Belgium from recession, remaining in slowdown, while France continues to be in slowdown. Germany, Italy and the Netherlands are in expansion.

This analysis shows how, during 2011–2013, the cyclical movements in the Euro area economies were neither synchronised nor diffused. The lack of diffusion is clearly shown by the fact that some economies entered in recession while others just experienced a slowdown. The lack of synchronisation is shown by the fact that peaks and troughs are shifted among economies. This confirms the prevailing idiosyncratic behaviour characterising the Euro area economies, especially after the 2008/2009 financial and economic crisis.

The two cases analysed in this subsection show how, by combining the information contained in the PEEIs dashboard and in the business cycle clock, it is possible to obtain a much better picture of the economic situation. In this way, it has been possible to find answers to some relevant questions and also to obtain insight going beyond the questions themselves, such as the acceleration/deceleration of growth in the first case and the presence/lack of turning point diffusion in the second one.

7. Conclusions

This paper has discussed the characteristics of dashboards, scoreboards and composite indicators, with reference to examples from macroeconomic data published by Eurostat. In each case there are both identified advantages and disadvantages to their use. However it may also be seen that a combination of the approaches – with a close eye on communication to users — can provide a complementary economic picture.

With regard to composite indicators, the biggest challenges appear to be the technical complexity of model/assumption selection, and their appropriate presentation to users. The Business Cycle Clock is one example of a graphical approach to presentation of a complex model. Alternative methods might be explored — for example non-parametric approaches such as POSET. However they are more appropriate in some statistical domains than others.

Acknowledgements

The author is grateful to Walter Radermacher, Silke Stapel-Weber and John Verrinder for their precious comments and suggestions. He is also thankful to Rosa Ruggeri Cannata and Dario Buono for their technical advice and useful suggestions in some specific parts of the paper.

References

Anas, J., M. Billio, L. Carati, L. Ferrara and G.L. Mazzi (forthcoming), 'Detecting turning points within the ABCD framework', *Handbook on Cyclical Composite Indicators*, ed. by G.L. Mazzi and A. Ozyildirim.

Anas, J., M. Billio, L. Ferrara, G.L. Mazzi. (2008), 'A system for dating and detecting turning points in the euro area', *The Manchester School*, 76 (5), pp. 549-577.

Anas, J. and L. Ferrara (2004), 'Detecting Cyclical Turning Points: The ABCD approach and Two Probabilistic Indicators', *Journal of Business Cycle Measurement and Analysis*, Vol. 1, No 2, OECD.

Badinger, H. and W.H. Reuter (2014), 'Measurement of fiscal rules: Introducing the application of partially ordered set (POSET) theory', *Journal of Macroeconomics*, 43 pp. 108–123.

Billio, M., L. Calès and D. Guégan (2011), 'Portfolio Symmetry and Momentum', *European Journal of Operational Research*, 214(3), 759-767.

Billio, M., L. Calès and D. Guégan (2012), 'Cross-sectional Analysis through Rank-based Dynamic Portfolios', Working paper, 36, Centre d'Economie de la Sorbonne — ISSN: 1955-611X.

Billio, M., L. Ferrara L., D. Guégan and G.L. Mazzi (2013), 'Evaluation of regime switching models for real-time business cycle analysis of the euro area', *Journal of Forecasting*, vol. 32, n°7, pp. 577-586.

Billio, M., L. Ferrara, G.L. Mazzi (2015), 'An integrated system for euro area and member states turning points detection', short paper in proceedings of the NTTS conference, Brussels.

Brüggemann, R. and G.P. Patil (2011), 'Ranking and Prioritization for Multi-indicator Systems — Introduction to Partial Order Applications', Springer, New York.

Carriero, A. and M. Marcellino (2007). 'A comparison of methods for the construction of composite coincident and leading indexes for the UK', *International Journal of Forecasting*, 23, pp. 219–236.

Carriero, A., M. Marcellino, and G.L. Mazzi (forthcoming), 'Definitions and Taxonomy of Indicators forthcoming', *Handbook on Cyclical Composite Indicators*, ed. by G.L. Mazzi and A. Ozyildirim.

Chauvet M. and S. Potter (2005), 'Predicting Recessions Using the Yield Curve', *Journal of Forecasting*, 24 (2), pp. 77–103.

Davey, B. A. and B.H. Priestley (2002), 'Introduction to lattices and order', Cambridge: CUP.

Fattore, M., F. Maggino and F. Greselin, (2011), 'Socio-economic evaluation with ordinal variables: integrating counting and POSET approaches', *Statistica and Applicazioni*, 01/2011, Special Issue, pp. 31–42.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57, 357–384.

Harding, D. and A. Pagan, (2002), 'Dissecting the cycle: a methodological investigation', *Journal of Monetary Economics*, 49(2), 365–381.

Harding, D. and A. Pagan (2006), 'Synchronization of cycles', *Journal of Econometrics*, 132 (1), 59-79.

Harding, D. and A. Pagan (2011), 'An Econometric Analysis of Some Models for Constructed Binary Time Series', *Journal of Business and Economic Statistics*, 29 (1), 86–95.

Liew, J. and M. Vassalou (2000), 'Can book-to-market, size and momentum be risk factors that predict economic growth?', *Journal of Financial Economics*, 57 (2), August 2000, pp. 221–245.

Neggers, J., and H.S. Kim (1988), 'Basic posets', World Scientific, Singapore.

Ozyildirim, A. (forthcoming), 'Business Cycle Indicator Approach at the Conference Board', *Handbook on Cyclical Composite Indicators*' ed. by G.L. Mazzi and A. Ozyildirim.

Ruggeri Cannata, R., D. Buono and F. Biscosi (2015), 'The Macroeconomic Imbalances Procedure and the scoreboard: ensuring data coverage', *EURONA*, Issue 2/2015, p. 101.

Schroeder, B. (2003), 'Ordered sets', Birkhäuser, Boston.

The Macroeconomic Imbalances Procedure and the scoreboard: ensuring data coverage

4

Rosa Ruggeri Cannata [1], Dario Buono [2] and Ferdinando Biscosi [3]

Abstract: The Macroeconomic Imbalance Procedure (MIP) is a surveillance mechanism that aims to identify potential risks early on, prevent the emergence of harmful macroeconomic imbalances and correct the imbalances that are already in place. It is a system for monitoring economic policies and detecting potential harms to the proper functioning of the economy of a Member State, of the Economic and Monetary Union, and of the European Union as a whole. The MIP is supported by the analysis of a set of headline and auxiliary indicators, whose data coverage can reach twenty years due to data transformations. In order to ensure the necessary time series length for policy makers, statisticians can resort to statistical techniques such as backcalculation. This paper illustrates the MIP in the European Union policy context and some applications of backcalculation to two MIP indicators.

Keywords: economic governance, policy decision, time series.

JEL codes: C32, C82, E60.

(1) Eurostat, Task Force Macroeconomic Imbalances Procedure.

(2) Eurostat, unit B1 'Methodology and corporate architecture'.

(3) Eurostat, unit C2 'National and regional accounts production. Balance of payments'.

1. Introduction

The Macroeconomic Imbalances Procedure (MIP) is a system for monitoring macro-economic developments and policies and detecting potential threats to the proper functioning of the economy of the European Union (EU). It is part of a surveillance system for budgetary and economic policies, implemented via the European Semester, the EU's policy-making calendar.

An essential tool in the procedure is the MIP scoreboard — a set of fourteen headline indicators intended to screen internal and external macroeconomic imbalances, covering a time span of ten years for the twenty-eight EU Member States. It acts as a first filter in a broader process seeking to disentangle the existence and severity of macroeconomic imbalances in the EU Member States, starting every year in autumn with the Alert Mechanism Report identifying countries and issues for which a closer analysis is deemed necessary. In this process, the scoreboard is used by policy makers, together with a set of auxiliary indicators, for their economic reading.

Policy makers need an as complete as possible picture of the economy; it might be, then, necessary to apply statistical techniques to ensure optimal data coverage, in particular when official statistics are subject to events which could disrupt time series length, such as the adoption of new classifications or changes in the production process.

Considering the definitions of some of the MIP headline and auxiliary indicators (e.g. averages or percentage changes over several years), needed data coverage can reach twenty years. In this context, statisticians have to ensure the availability of the required length for time series needed for MIP indicators and, where necessary, to apply statistical techniques such as backcalculation for this purpose.

Section 1 of this paper introduces the main characteristics of the MIP, and its role within the context of the economic governance; section 2 briefly illustrates the methodological approach used for a backcalculation exercise in general; sections 3 and 4 introduce two applications, the first one focuses on backcalculation of the House Price Index (HPI), while the second one covers the unemployment rate and can be considered more as an exercise in temporal disaggregation exercise; section 5 concludes and introduces some research lines for future work.

2. The Macroeconomic Imbalances Procedure

2.1. Policy makers' needs and sets of indicators for policy making

Policy makers often require aggregated indicators which provide a synthetic and clear picture of the different areas of interest. Several sets of indicators have been designed and are currently being used in order to permit the planning and monitoring of European policies, such as the Principal European Economic Indicators (PEEIs) (for more information about principles of backcalculation of PEEIs see Mazzi et al. (2010)), a comprehensive set of infra-annual macro-economic statistics aiming to describe the economic and labour market situation as well as price developments in the euro area and the European Union, and the Sustainable Development Indicators (SDIs), a system developed to monitor progress in the implementation of the European Union's sustainable development strategy, aiming to improve continuously the quality of life through reconciling economic development, social cohesion and protection of the environment.

Another relevant policy in the economic context is the Europe 2020 strategy, which places emphasis on a new growth path that can lead to a smart, sustainable and inclusive economy; a path that aims to overcome the structural weaknesses in Europe's economy, improve its competitiveness and productivity and underpin a sustainable and social market economy. The Europe 2020 strategy is accompanied by a set of indicators designed to monitor progress towards targets related to its key objectives at the European Union level.

The financial and economic crises and the sovereign debt crisis that swept over Europe in 2008 and the following years lead to a number of new European Union initiatives in 2010. As a response to weaknesses in its economic governance system revealed by the crisis, the European Union has taken a wide range of measures to strengthen economic governance and to achieve sustained convergence, economic growth and employment. The Van Rompuy Task Force report on 'Strengthening economic governance in the European Union' ([4]) of 21 October 2010 and the European Commission proposal of 29 September 2010 on an 'Enhanced Economic Policy Coordination' ([5]) included several suggestions for improving the European Union economic surveillance, which were formalised in the legislative packages known as the six-pack and two-pack (see European Parliament and Council (2011a and 2011b)), entering into force at the end of 2011 and in 2013 respectively.

The legislation aims for a closer coordination of economic policies through a strengthening of budgetary surveillance under the Stability and Growth Pact (SGP), the introduction of a new procedure in the area of macroeconomic imbalances, the establishment of a framework for dealing with countries experiencing difficulties with financial stability, and the codification in legislation of integrated economic and budgetary surveillance in the form of the European Semester.

With the MIP, the surveillance of economic policies of the Member States was broadened beyond budgetary issues; the MIP has been introduced with the objective to detect, prevent, and correct problematic economic trends, such as internal and external imbalances, falling competitiveness, real estate bubbles or banking crises. It is part of a system for monitoring economic policies and detecting at an early stage potential threats to the proper functioning of the economy of a Member State, of the Economic and Monetary Union, and of the European Union as a whole.

In subsection 1.2 we focus on the surveillance system for budgetary and economic policies, implemented via the European Semester, the European Union's policy-making calendar, while subsection 1.3 will be dedicated to MIP headline and auxiliary indicators. Subsection 1.4 illustrates the evolution of the scoreboard over time while section 1.5 contains some considerations on the role of official statistics within the MIP context.

([4]) https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/117236.pdf.

([5]) http://ec.europa.eu/economy_finance/articles/eu_economic_situation/pdf/com2010_525en.pdf.

2.2. The MIP in the context of the European Semester

The European Semester, introduced in 2010, ensures that Member States discuss their budgetary and economic plans and structural reforms with their European Union partners at specific times throughout the year. This framework governs the:

- implementation of structural reforms to ensure progress towards the agreed goals of the Europe 2020 strategy;
- implementation of fiscal policies under the SGP to strengthen economic governance and ensure budgetary discipline;
- prevention and detection of excessive macroeconomic imbalances through the MIP.

The cycle starts in November each year with the publication of the Commission's Annual Growth Survey and the Alert Mechanism Report (AMR). The Annual Growth Survey sets out general economic priorities for the European Union and provides Member States with policy guidance for the following year. At the beginning of the following year, the Commission publishes country reports for each Member State, analysing the economic situation and policies of each Member State and assessing whether imbalances exist in those Member States where an in-depth review was carried out. The publication of country reports is followed by a series of bilateral meetings between the European Commission and the Member States. In spring, Member States present their national reform and stability or convergence programmes, containing reforms and measures envisaged to make progress towards smart, sustainable and inclusive growth and the country's plans for sound public finances. The Council concludes the European Semester in summer by agreeing on a set of country-specific recommendations, highlighting areas where European Union Member States need to take further actions to boost growth, job creation, training and education opportunities, research and innovation.

The AMR is the annual starting point of the MIP; it is based on a scoreboard of indicators and on their economic reading, where the MIP scoreboard acts as a filter to identify countries and issues for which a closer analysis (in-depth review) is deemed necessary. The outcome of these in-depth reviews forms the basis for further steps under the MIP whereby a graduated approach is followed reflecting the severity of imbalances.

The MIP includes a preventive and a corrective arm; the corrective arm is triggered by the excessive imbalance procedure, which can eventually lead to sanctions for euro area Member States if they repeatedly fail to meet their obligations.

2.3. MIP headline and auxiliary indicators

An essential tool in the MIP procedure is the MIP scoreboard — a set of fourteen annual headline indicators intended to screen internal and external macroeconomic imbalances, covering a time span of ten years for the twenty-eighth European Union Member States. It acts as a first filter in a broader process seeking to disentangle the existence and severity of macroeconomic imbalances in the European Union Member States. In this process, the scoreboard is used by policy makers at the initial stage, together with a set of auxiliary indicators and all available information, for their economic reading with a view to select the countries for which a more in-depth analytical assessment is required.

For the headline indicators, indicative thresholds based on historical data have been set at alert levels; such thresholds can have both an upper and lower alert level for some indicators, and can have different values for euro area and non euro area Member States when justified by specific features of the monetary union. The overall number of breaches of thresholds, the severity of individual breaches as well as the combination of breaches, potentially signalling broad based problems, are all taken into account in the Commission economic reading, together with any other relevant information such as the most recent forecast picture. All of this is done without any automaticity.

BOX 2.1: LIST OF SCOREBOARD HEADLINE INDICATORS

External macroeconomic imbalances and competitiveness

Current account balance as a percentage of GDP (3-year backward moving average)

Net international investment position as a percentage of GDP

Real effective exchange rate based on HICP/CPI deflator (3-year percentage change)

Export market share (share of world exports) as 5-year percentage change

Nominal unit labour cost (3-year percentage change)

Internal macroeconomic imbalances

HPI relative to the final consumption deflator (year-on-year change)

Private sector credit flow — consolidated — as a percentage of GDP

Private sector debt — consolidated — as a percentage of GDP

General government gross debt (EDP) as a percentage of GDP

Unemployment rate (3-year backward moving average)

Total financial sector liabilities (year-on-year change)

Labour market indicators

Activity rate as percentage of total population aged 15–64 (3-year change in percentage point)

Long-term unemployment rate as percentage of active population aged 15–74 (3-year change in percentage point)

Youth unemployment rate as percentage of active population aged 15–24 (3-year change in percentage point)

The choice of indicators focuses on the most relevant dimensions of macroeconomic imbalances and competitiveness developments. The scoreboard composition might evolve over time, if necessary, in order to adapt to the changing nature of macroeconomic imbalances due, inter alia, to evolving threats to macroeconomic stability or enhanced availability of relevant statistics. The indicators can be used to monitor external imbalances, competitiveness positions and internal imbalances. Internal imbalances include those that can arise from public and private indebtedness; financial and asset market developments, including housing; the evolution of private sector credit flow; and adjustment through the evolution of unemployment. External imbalances include those that can arise from the evolution of current account and the net investment positions of Member States, real effective exchange rates, share of world exports, and nominal unit labour cost.

Furthermore, in order to both monitor existing imbalances and give signal of emergent ones at an early stage, the scoreboard consists of a combination of stock and flow indicators which can capture both shorter-term rapid deteriorations as well as the longer term gradual accumulation of imbalances.

BOX 2.2: LIST OF AUXILIARY INDICATORS

Real GDP growth rate

Gross fixed capital formation as percentage of GDP

Gross domestic expenditure on R & D (GERD) as percentage of GDP

Net lending-borrowing/current plus capital account as percentage of GDP

Net external debt as percentage of GDP

Inward FDI flows as percentage of GDP

Inward FDI stocks as percentage of GDP

Net trade balance of energy products as percentage of GDP

Real effective exchange rate — euro area trading partners 3-year percentage change

Terms of trade (goods and services) 5-year percentage change

Share of OECD exports 5-years percentage change

Export market shares (good and services) in volumes growth rate

Labour productivity growth rate

Nominal unit labour cost index, 10-year percentage change

Unit labour costs performance relative to euro area, 10-year percentage change

Nominal HPI, 3-year percentage change

Residential construction as percentage of GDP

Private debt (non-consolidated) as percentage of GDP

Financial sector leverage (debt-to-equity)

Employment growth rate

Young people neither in employment nor in education and training (% total population)

People at risk of poverty or social exclusion rate (% total population)

At risk of poverty after social transfers rate (% total population)

Severely materially deprived people (% total population)

People living in households with very low work intensity (% total population)

Data transformations, such as the 5-year percentage change or the 3-year average, have been chosen in order to smooth the effect of a particular year on indicators development, to focus more on structural developments (for more details: European Commission (2011) and European Commission (2012)).

In order to facilitate users access to the set of MIP indicators, Eurostat offers a dedicated set of webpages to the MIP, covering both figures as published in the statistical annex and up-to-date values of MIP indicators, together with a large set of information and metadata on methodologies, legislation, relevant publications and a complete set of graphical presentations and download facilities: http://ec.europa.eu/eurostat/web/macroeconomic-imbalances-procedure/indicators.

2.4. Role of official statistics

Statisticians commit to supply policy makers with fit-for-purpose data of the highest possible quality, including data comparability across time and countries which is essential for policy decisions in particular at the European level. Eurostat and the European Statistical System (ESS) follow an encompassing quality management approach based on the European Statistics Code of Practice covering the statistical domains underlying MIP indicators. To assess country performance on a sound base, it is necessary to define not only the concepts to be measured, but also a harmonised framework in which indicators can be effectively compared. MIP indicators stem from several statistical areas, following specific quality assurance frameworks, including national accounts, balance of payments statistics, price statistics, Excessive Deficit Procedure (EDP) statistics or the labour market statistics. Moreover, some statistics are not produced by statistical offices but by central banks; it is then important that the ESS and the European System of Central Banks (ESCB) maintain a continuous effort to enhance the relevant quality frameworks of the statistics underlying the MIP indicators.

The role of official statistics in the MIP is essential in other aspects too. Eurostat is involved in the identification of the relevant statistics measuring as well as possible the economic phenomena under analysis. As illustrated in the preceding subsection, this is an dynamic process because new statistics can become available or new classification be adopted, thanks also to the entering into force of new legislation. As a result of these factors, ESS statisticians are constantly engaged in monitoring and enhancing data quality in an evolving scenario, to adapt to emerging users' needs and new data sources and to satisfy policy makers' needs. Scoreboard indicators are then regularly reviewed while the underlying statistical methodology and the statistical production processes are constantly improved. As a consequence, the length of the time series used to compute MIP indicators could be shortened, jeopardising the possibility for the policy maker of looking at a 10-year timespan for all Member States. It is in this context that Eurostat has started to work in cooperation with Member States to maximise indicators' availability to improve data coverage, an essential dimension of quality. The rest of this paper provides two examples that illustrate what kind of efforts can be performed in this direction, introducing first the methodological background of backcalculation techniques and then circumstantiating it to the context of production of official statistics, which need to be validated at country and European level.

3. Ensuring data coverage: backcalculation

When looking at quality in the context of policy making, one relevant aspect is data coverage; policies need to be assessed on a medium to long-term period and this imposes constraints on time series length. Considering the definitions of some of the MIP headline and auxiliary indicators (e.g. averages or percentage changes over several years), desirable data coverage can reach twenty years.

Moreover, the adoption of new classifications or regulations (as the new European System of Accounts (ESA) 2010 and the Balance of Payments Manual, 6th edition (BPM6), both in 2014), the availability of new primary data, or a change in the production process, might imply a disruption in the time series length and a lack of comparability over time. Experience has shown some lack of data concerning the MIP indicators scoreboard, in particular for indicators not entirely covered by a regulation, such as the HPI, where some past data are missing for several member states ([6]).

Policy makers need an as complete as possible picture of the economy. In order to satisfy thid need, statisticians have to ensure the availability of the required length for time series needed for the computation of MIP indicators; when this is not possible, they can try to fill data gaps, for example by applying statistical techniques such as backcalculation (for more details on methodological aspects see Caporin and Sartore (2006)).

Backcalculation (also referred as back-recalculation, back-casting, retropolation or reconstruction) of long time-series is the statistical process allowing to project back in time values of a given time-series by using all relevant available information. In other words, given a time-series for which values cover the time interval t to $t + k$, the backcalculation exercise aims to estimate the missing values $t - 1$, $t - 2$,…, $t - j$ for a certain j.

Backcalculation can be based on estimation techniques, such as retropolation, making use of all available information, using static or dynamic modelling.

3.1. Horizon of backcalculation

When performing a backcalculation exercise, one of the first decisions to take is the targeted horizon back in time for the time series in consideration; the horizon will be a function of several factors; a crucial point is which information is available, in fact it could not be feasible or statistically relevant to extrapolate back in time a series much longer than the available length of relevant indicators used in the retropolation model.

A second point to consider when fixing a backcalculation horizon is the historical context. As an example, it would not be meaningful to talk about Czech or Slovak data before 1993.

A third point very relevant in the context of MIP indicators, is the availability of data at national level; backcalculation should optimally be assessed by the national producers who have a better insight into the data and can have a clearer understanding of the time series evolution. Moreover, new data could become available also for the past and they should be then considered in the backcalculation exercise.

Finally, special attention has also to be given to the availability of primary data for past periods, in particular when the time series has shortened as a consequence of a major revision (see European Commission (2013) for a definition of different kinds of revisions) due to a classification updating, some new items considered in the updated classification could not have been recorded in the past so that any retropolation must carefully consider the meaningfulness of the extended series; as an example the relevance of IT expenses in the '70s could be difficult to model and justify, the same could hold for house price indexes.

([6]) The current regulation for HPI requires transmission of quarterly data starting from 2008Q1

3.2. Choice of the model and of the method

The choice of the model (see Gatto (2007) for a complete review) to consider could be based on several characteristics; first of all, for the sake of transparency, it is important for the model to be simple, robust, and parsimonious in the number of parameters.

When choosing indicators to be included in the model, the set of potential candidates should include data directly related to the target variable, for example series measuring similar quantities with minor differences in definitions, as can be the case of consolidated and not consolidated data in the financial accounts.

It is also important to check for the need of data transformations in order to assess stationarity in the mean and in the variance. In principle, the backcalculation could be performed on the levels, on the logs or on any difference of the series. The choice has to take into account the characteristics of the series, for example growth rates will not require the same treatment of levels. In the presence of alternative models, criteria for the choice of the optimal one have to be defined, for instance the Root Mean Square Error (RMSE) or other indexes could be used as indication of the model accuracy.

In order to evaluate the information content of other indicators, several tests could be considered aiming to assess the degree of co-movement and the type of relation of the candidate and the targeted variable; co-integration tests (common root for stationarity) and correlation ones should be always considered. The final aim is to reproduce in the past the same 'shape' of the available time series. A basic, but clear and robust, approach would be to project back in time the growth rate of the series under consideration.

MIP indicators being annual, when the same indicator is also available at quarterly frequency with a prolonged coverage, this higher frequency information should be considered and integrated by using benchmarking techniques (see Di Fonzo (2003) for a dynamic model approach). When low-frequency (e.g. annual) benchmarks and sufficiently long time series on related indicators are available, it is possible to develop a constrained regression-based retropolation methodology which extends the classical solution given by Chow and Lin (1971, 1976) to the problem of optimal disaggregation of a time series by means of related indicators. The idea behind this approach (see Buono and Kocak (2011), for more model-based-link detail) is to estimate missing past quarterly values for which only annual benchmarks are available by using the available, more recent quarterly series and one or more related series covering the quarters for which back-estimates are needed.

A possible choice for the model is a linear regression model (with or without constant), possibly extended with SARIMA terms in the residuals; estimation can be performed by maximum likelihood and standard tests can be used to evaluate the estimated coefficients and the residuals. The case of missing values in the related indicators has to be treated as it could lead to different results.

Some international institutions, such as the European Central Bank (ECB), the Bank of International Settlements (BIS) and OECD, are using the so-called ratio method, which is equivalent to the Ordinary Least Squared (OLS) with no intercept.

In our applications, detailed in the following two sections, the OLS method has been used. The model, which includes the intercepts, can be represented by the following expression:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

The model is not dynamic, i.e. there is no lag included. The contemporaneous value of the targeted series is regressed on the value of the indicator for the sample period. The level values should be tested for stationarity in the mean and in the variance and should undergo an opportune co-integration test. In order to deal with potential serial correlation, one possible solution is adopting an appropriate box-cox transformation (e.g. log-transformation) and operating on differentiated series.

The correlation index and the relative R^2 can be used as a measure of model fitness.

An additional alternative would be to use structural models with local trend or with seasonal component. A similar approach, based on a state-space formulation of the missing data problem and using the Kalman filter to retropolate the series of interest, can also be used. With regard to this procedure, the method proposed here is easy to formulate and flexible enough to capture simple dynamic relationships, usually excluded, or not conveniently captured, by the static-in-nature formulation of the Chow and Lin estimation procedure.

Some relevant definitions

The reconstruction of previous values of the time-series can be obtained by using either univariate or multivariate methods. Hereafter are some definitions to be used as reference:

Definition 1.1: Univariate backcalculation is the reconstruction of past values of a time-series based only on the information contained in the available part of the time-series itself. This applies, for instance, when the same time series is available at both annual and quarterly frequencies. Where only annual series are available, a quarterly linear trend can be used as a proxy. The latter would be the case of the temporal disaggregation.

Definition 1.2: Multivariate backcalculation is the reconstruction of past values of a time-series based on the information contained in the available part of the time-series itself and in one or more related indicators, which have at least an overlapping period with the series to be backcalculated and non-missing values for the period on which the backcalculation has to be performed. This applies, for instance, when the targeted time series is available at annual frequency and multiple series are available at quarterly frequency.

Finally, in the context of the production of official statistics, it is important to reach a high degree of automation in order to be able to run a backcalculation exercise whenever necessary and possibly with a very limited need for subjective judgement.

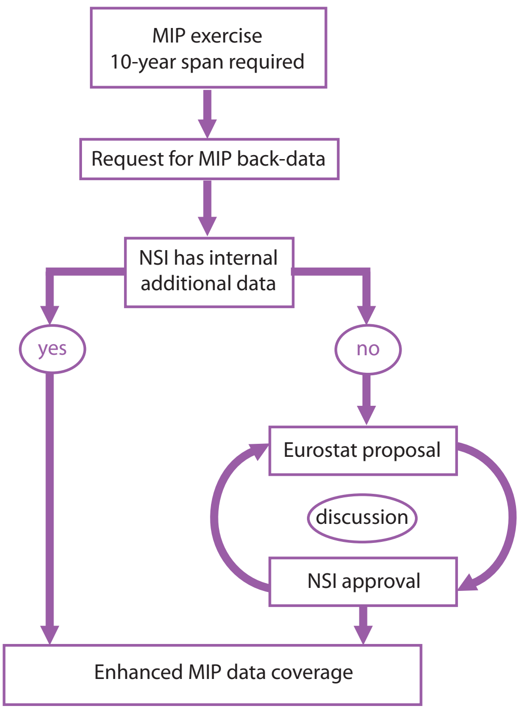
3.3. Validation of results

Once obtained a set of backcalculated values, criteria to assess and validate the results have to be defined; the backcalculation exercise could be repeated enlarging the set of indicators or modifying the chosen model until some criteria are met; the setting of clearly measurable criteria will contribute to make more automatic the modelling process and to limit subjective judgement.

Usually, a high value of R^2 between the original series and the backcalculated one should be reached. Moreover, it is important not to 'rewrite the history', so that the most recent values of the time series should be the officially available ones; as a consequence, a break could appear and has to be treated.

As stated above, when the backcalculation exercise is performed at European level, it is important to involve the national data producers in the validation process. For this reason, as shown in the figure below, when Eurostat backcalculates certain MIP indicators the relevant Member State will be informed via bilateral contacts in order to share all available information and to reach a common agreement on the prolonged time series.

Figure 3.1: Coverage enhancement workflow



3.4. Limits of the exercise

The backcalculation exercise described here does not imply a micro-level backcalculation, being mainly targeted to main aggregates with a limited breakdown. It's aim is not to change the past pattern of the series but, on the contrary, to keep it as far as possible unchanged, especially concerning turning points occurrence and the series' cyclical shape. The rationale of this approach is primarily to preserve the historical characteristics of the series in order to reflect the economic and monetary policy decisions which have been taken on the basis of the information available in the past.

This exercise is mainly based on the homogenisation of existing, partially overlapping segments of time series in order to eliminate breaks and inconsistencies. This is in line with users' expectations, in particular for those users involved in econometric modelling and policy making.

4. House Price Index (HPI) case

One of the headline indicators in the MIP is the HPI; since the beginning of 2013, with the entering into force of the reference legislation for this indicator, the monitoring of changes in house prices is based on data regularly compiled by Member States and transmitted to Eurostat. However, both because of primary data unavailability and as the regulation does not require annual values before 2008, HPIs time series for some Member States are shorter than what is needed for the MIP. Eurostat, together with the European Commission, the ECB, the OECD, and the BIS continues to work on backward data calculation aiming at increasing the length of back series by using all available data, with particular attention to the historical coverage required in the context of the MIP.

4.1. A first application: the HPI for Malta

The initial data set consists of the following series:

- Eurostat HPI quarterly series, covering the period 2004Q1 to 2012Q2;
- ECB HPI quarterly series, covering the period 2000Q1 to 2012Q2;
- ECB HPI annual series, covering the period 1980 to 2011.

The arithmetic average of the ECB HPI quarterly figures gives the respective annual value. This exercise aims at backcalculating the Eurostat (2010 = 100) series for the period 2000Q1 to 2003Q4 by linking it with the available proxy variable provided by the ECB. The correlation index for the time span in common (2004Q1 to 2012Q4) is found to be 0.32 on the raw indexes. Both indexes are then log transformed and first order differentiated. Delta logs can be interpreted as growth rates. The correlation in delta log form is 0.51. OLS regression is run on the delta logs for the period 2004Q2 to 2012Q2 (given the loss of one observation at the beginning of the series) on the common time span and found parameters are used to backcalculate the Eurostat series. The data and related chart are available in figure 1.1 and table 1.1 of the annex.

4.2. Further developments

Following the results of the 'Quarterly House Price Indices long time series Joint Residential Property Prices Indices Study Group', Eurostat started to analyze the possibility to enlarge the deflated HPI coverage for the MIP purpose.

The Study Group involved several international organisations and the following datasets were taken into account:

1. Eurostat HPI quarterly series;
2. ECB HPI quarterly series;
3. BIS HPI quarterly series;
4. National Central Bank quarterly series;
5. OECD quarterly series.

As for the Maltese example mentioned before, this exercise aimed at backcalculating the Eurostat (2010 = 100) annual series for the missing period by deriving it from proxy quarterly series available. The procedure is the same for the next four applications and can be summarised in the following steps. The first step is calculating the correlation index between the official series and the auxiliary ones over the overlapping period. The second step is taking as proxy the series that displays the highest correlation with the official one. Afterwards both series are log-transformed and first order differentiated. The following step is carrying out the OLS regression on transformed (delta logs) overlapping period and using the parameters to estimate the back series. The final step is to calculate the annual series as average of the four quarters.

The exercise was run for several Member States including Spain, Cyprus, Lithuania and Latvia.

The following chart gives a graphical representation of the results obtained for the Spanish HPI series. Table 3.1 contains an overview of the backcalculation exercise done by Eurostat. The data and its graph are available in the annex.

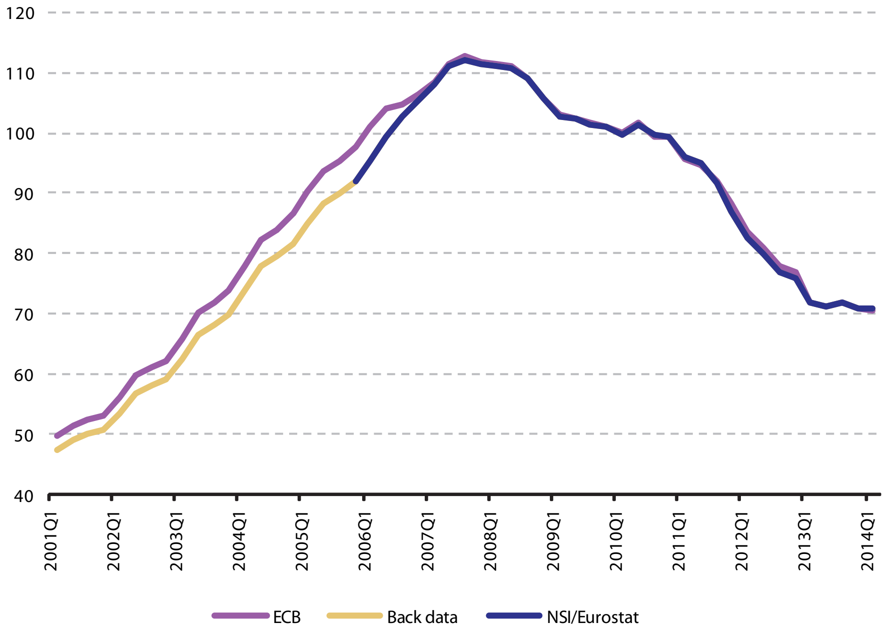
Table 4.1: Summary of backcalculation exercise

Country	Correlation index	Period covered	Proxy source dataset	Proxy series coverage
Cyprus	0.73	2002–2004	ECB (*)	2002Q1–2014Q2
Latvia	0.96	2000–2005	ECB (*)	2000Q1–2014Q1
Lithuania	0.97	2000–2005	ECB (*)	1998Q4–2014Q1
Malta	0.99	2000–2004	ECB (*)	2000Q1–2012Q2
Spain	0.51	2000–2005	ECB (*)	1987Q1–2014Q1

Source: Eurostat

(*) Residential property prices, new and existing dwellings in good and poor condition, whole country.

Figure 4.1: Quarterly HPI backcalculation, Spain
(2010 = 100)



Source: Eurostat

5. Croatian unemployment data

Croatia joined the EU on the 1st July 2013 and its data where then included in the MIP scoreboard. Annual unemployment data are derived from the monthly figures, which for Croatia are produced by Eurostat using the data from the quarterly Labour Force Survey (LFS) and the monthly number of unemployed persons registered with the public unemployment office. From 2000 until 2006, the Croatian LFS was conducted twice a year, while in 2007 a continuous quarterly survey was started; Eurostat launched an empirical exercise with the aim to produce comparable figures for the period back to 2000 on the base of the existing LFS data.

The exercise consisted of:

- outlining a pattern for the quarterly LFS, including seasonality;
- estimate the values of monthly series for periods not available from the LFS.

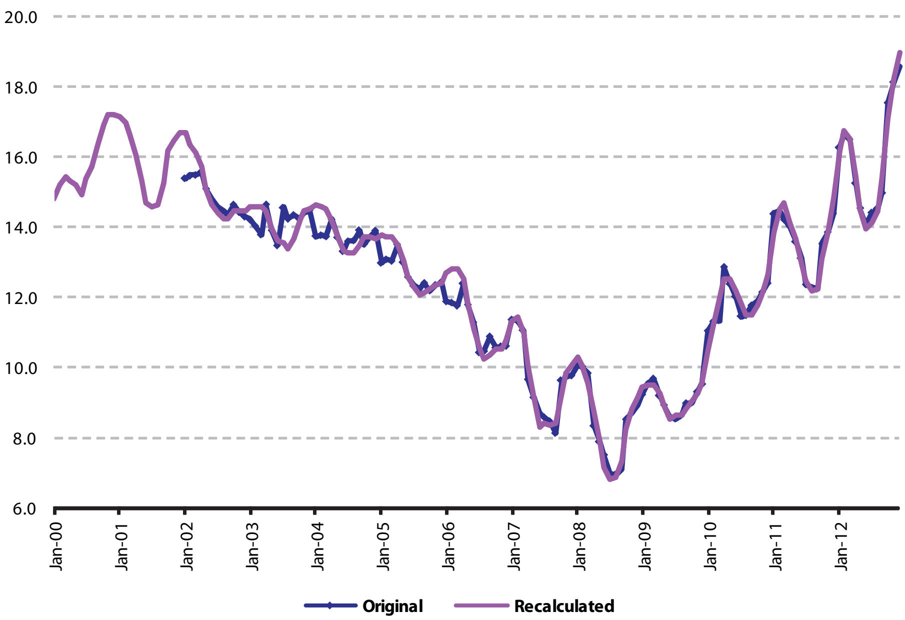
This exercise can be considered more as a disaggregation issue from quarterly to monthly data; the adopted methodological approach was the proportional Denton procedure, which ensures that the resulting monthly figures on average equal the corresponding quarterly data. This method was applied at a disaggregated level, that is on the eight primary series for employment and unemployment levels, each broken down by sex and the two age groups, young (15–24 years) and adults (25–74 years). Aggregates (e.g. total employment or total unemployment) were obtained by summing up the disaggregated series. The method was applied to the disaggregated series via the following steps:

1. For each year, half-yearly observations were allocated to the 2nd and 4th quarter, respectively;
2. Missing values for quarters 1 and 3 were estimated using a SARIMA model;
3. Quarterly series were temporally disaggregated into monthly ones using the proportional Denton method.

The following graph shows the recalculated monthly series for total unemployment rate together with the original one. It is clearly visible how the reconstructed series displays a consistent seasonal pattern before and after 2006.

Figure 5.1: Unemployment rate, total, monthly data before and after recalculation, not seasonally adjusted

(%)



Source: Eurostat

6. Conclusions and further work

This paper has introduced the MIP and the related scoreboard in the context of policy making, and has illustrated the main issues related to the backcalculation of MIP indicators; the aim of the exercise is to enlarge data coverage in order to fulfil the needs of policy makers and other stakeholders. The paper also includes two applications: the first one is a typical backcalculation exercise while the second one focuses on a disaggregation approach.

The backcalculation of time series for MIP indicators aims to provide required long series of data needed for the assessment of emerging or persistent macroeconomic imbalances in a country. It can become a systematic activity because statistics are subject to many events which could disrupt time series length. It is then particularly important for data providers to be able to manage this recurrent task; the adopted methodology for backcalculation uses a simple and robust approach, easily replicable and well documented.

Data used as input in this exercise are also subject to revisions, meaning that a re-estimation of the model could yield different results. Past data however, should not depend too much on present data, especially for far away periods. When input data are revised, we can distinguish between major and routine revisions. A major revision would require a through reconsideration of input datasets and of the models, with an assessment of the impact. When dealing with routine revisions, the stability of backdata should be favoured.

Based on these considerations, it will probably be necessary to continue to perform backcalculation exercises of MIP indicators in order to fill data gaps and satisfy policy makers' needs.

Acknowledgements

This paper was originally prepared for the Conference of European Statistics Stakeholders (CESS — Rome, November 2014) and the 7th Eurostat Conference on New Technologies and Techniques for Statistics (NTTS 2015 — Brussels, March 2015). The authors are grateful to participants who commented the paper at the conferences, in particular, Duncan Elliott and Emanuele Baldacci who gave constructive comments. Moreover, we would like to acknowledge the NTTS scientific committee for encouraging the paper submission, as well as Eurostat for being supportive in doing so.

Last but not least, a special thank also goes to the two blind peer reviewers for providing encouraging feedback and constructive comments as well as the Eurona team in charge of layout editing and proofreading the English.

References

Buono, D. and A. Kocak (2011), 'Backward recalculation of seasonal series affected by economic crisis: a Model-Based-Link method for the case of Turkish GDP', *Euroindicators working papers*, ISSN 1977–3331, EWP 2011/005.

Caporin, M. and D. Sartore (2006), 'Methodological aspects of time series back-calculation', University Ca' Foscari of Venice, Dept. of Economics Research Paper Series, No 56/06.

Chow G. and A.L. Lin (1971), 'Best linear unbiased interpolation, distribution and extrapolation of time Series by related series', *The Review of Economics and Statistics*, 53(4).

Chow, G. and A.L. Lin (1976), 'Best linear unbiased estimation of missing observations in an economic time series', *Journal of the American Statistical Association*, 71, 355, pp. 719–721.

Di Fonzo, T. (2003), 'Constrained retropolation of high-frequency data using related series: a simple dynamic model approach', *Statistical Methods & Applications*, 12, pp. 109–119.

Gatto, R. (2007), 'The SUH Approach in Time Series Back-Recalculation', 18th annual EC2 conference Recent Advances in Econometric Time Series Analysis.

Mazzi, G.L., F., Moauro and R., Ruggeri Cannata (2010), 'Back-calculation of European aggregates: some general considerations', European Conference on Quality in Official Statistics, Helsinki.

European Commission (2013), ESS guidelines on revision policy for PEEIs, Publications Office of the European Union, Luxembourg, http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-13-016.

European Parliament and Council (2011a), 'Regulation (EU) No 1174/2011 of 16 November 2011 on enforcement measures to correct excessive macroeconomic imbalances in the euro area', Official Journal of the European Union, L 306/8.

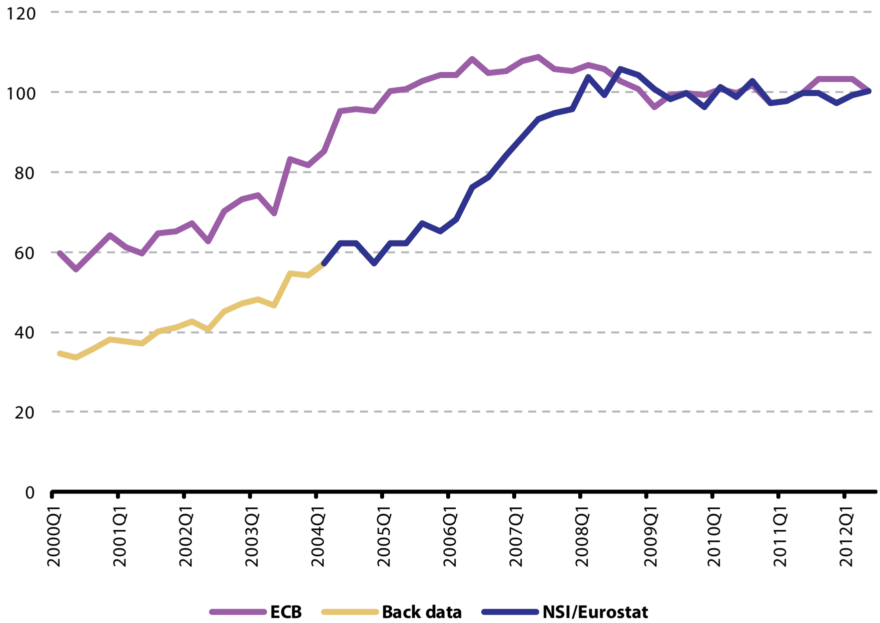
European Parliament and Council (2011b), 'Regulation (EU) No 1176/2011 of 16 November 2011 on the prevention and correction of macroeconomic imbalances', Official Journal of the European Union, L 306/25.

European Commission (2011), 'Scoreboard for the surveillance of Macroeconomic Imbalances: envisaged initial design', Commission staff working paper, SEC(2011) 1361.

European Commission (2012), 'Completing the Scoreboard for the MIP: Financial Sector Indicator', Commission staff working paper, SEC(2012) 389.

Annex — HPI quarterly data

Figure 1.1: Maltese quarterly HPI
(2010 = 100)



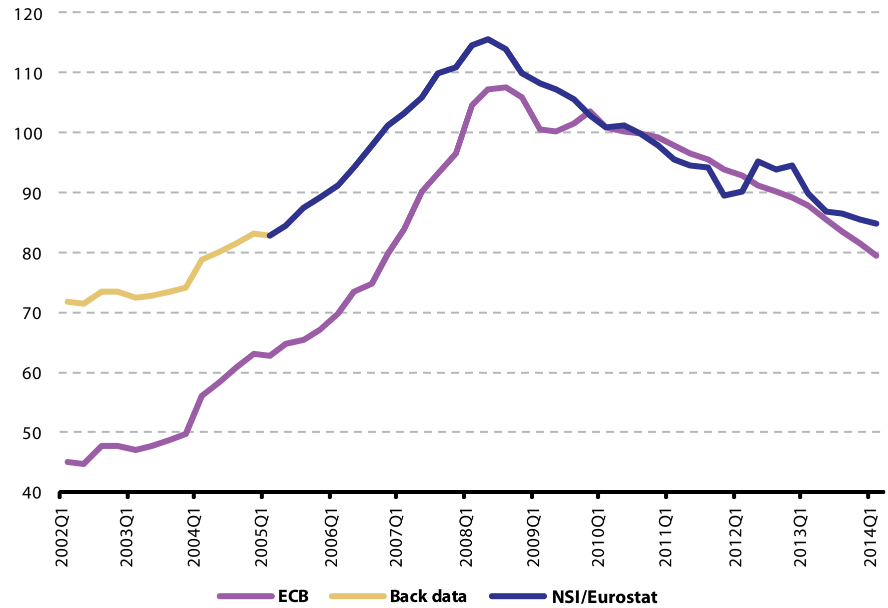
Source: Eurostat

Table 1.1: Maltese quarterly HPI
(2010 = 100)

	Backcasted	ECB		Eurostat	ECB		Eurostat	ECB
2000Q1	34.74	59.70	2004Q1	c	85.45	2008Q3	105.94	102.68
2000Q2	33.35	55.70	2004Q2	c	95.59	2008Q4	104.56	100.96
2000Q3	35.75	59.84	2004Q3	c	96.02	2009Q1	100.77	96.43
2000Q4	38.23	64.09	2004Q4	c	95.32	2009Q2	98.61	99.57
2001Q1	37.44	61.34	2005Q1	62.23	100.45	2009Q3	99.82	100.12
2001Q2	37.23	59.86	2005Q2	62.04	101.11	2009Q4	96.56	99.53
2001Q3	40.15	64.83	2005Q3	67.19	103.02	2010Q1	101.15	100.75
2001Q4	40.97	65.40	2005Q4	65.28	104.45	2010Q2	98.92	100.04
2002Q1	42.45	67.27	2006Q1	68.48	104.57	2010Q3	102.82	101.65
2002Q2	40.61	62.48	2006Q2	76.17	108.59	2010Q4	97.12	97.56
2002Q3	45.07	70.19	2006Q3	78.64	104.91	2011Q1	97.61	98.09
2002Q4	47.27	73.33	2006Q4	84.45	105.45	2011Q2	99.65	100.07
2003Q1	48.33	74.18	2007Q1	88.85	107.79	2011Q3	99.84	103.48
2003Q2	46.77	69.91	2007Q2	93.10	108.81	2011Q4	97.41	103.57
2003Q3	54.44	83.47	2007Q3	94.73	106.09	2012Q1	99.13	103.47
2003Q4	54.38	81.95	2007Q4	95.68	105.56	2012Q2	100.27	100.13
			2008Q1	103.70	106.99			
			2008Q2	99.45	105.92			

Source: Eurostat

Figure 1.2: Cypriot quarterly HPI
(2010 = 100)



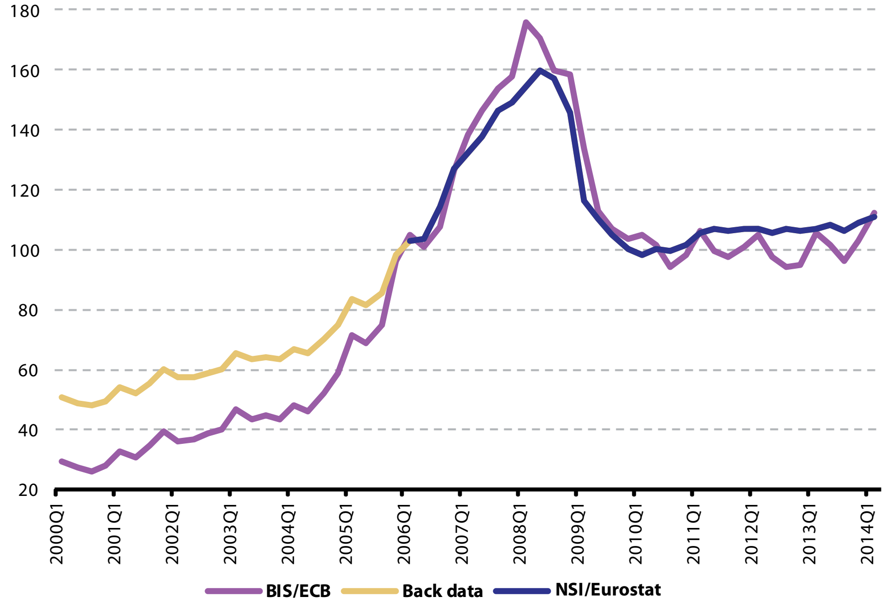
Source: Eurostat

Table 1.2: Cypriot quarterly HPI
(2010 = 100)

	Backcasted	ECB		Eurostat	ECB		Eurostat	ECB
2002Q1	71.64	44.9	2005Q1	82.74	62.9	2010Q4	97.74	99.2
2002Q2	71.28	44.7	2005Q2	84.48	64.7	2011Q1	95.51	97.8
2002Q3	73.41	47.6	2005Q3	87.50	65.5	2011Q2	94.47	96.6
2002Q4	73.37	47.8	2005Q4	89.14	67.1	2011Q3	94.10	95.5
2003Q1	72.60	47.1	2006Q1	91.27	69.8	2011Q4	89.58	93.8
2003Q2	72.81	47.6	2006Q2	94.28	73.4	2012Q1	90.32	92.9
2003Q3	73.58	48.8	2006Q3	98.00	74.9	2012Q2	95.20	91.1
2003Q4	74.09	49.7	2006Q4	101.12	79.8	2012Q3	93.98	90.2
2004Q1	78.79	56.1	2007Q1	103.11	83.7	2012Q4	94.46	89.3
2004Q2	80.20	58.3	2007Q2	106.03	90.1	2013Q1	89.93	87.9
2004Q3	81.64	60.7	2007Q3	109.88	93.2	2013Q2	86.83	85.6
2004Q4	83.04	63.0	2007Q4	110.82	96.6	2013Q3	86.46	83.6
			2008Q1	114.78	104.6	2013Q4	85.54	81.5
			2008Q2	115.52	107.2	2014Q1	84.77	79.4
			2008Q3	113.99	107.7	2014Q2		77.8
			2008Q4	109.80	105.8			
			2009Q1	108.42	100.5			
			2009Q2	107.33	100.2			
			2009Q3	105.57	101.6			
			2009Q4	102.99	103.5			
			2010Q1	100.94	100.8			
			2010Q2	101.34	100.2			
			2010Q3	99.98	99.9			

Source: Eurostat

Figure 1.3: Lithuanian quarterly HPI
(2010 = 100)



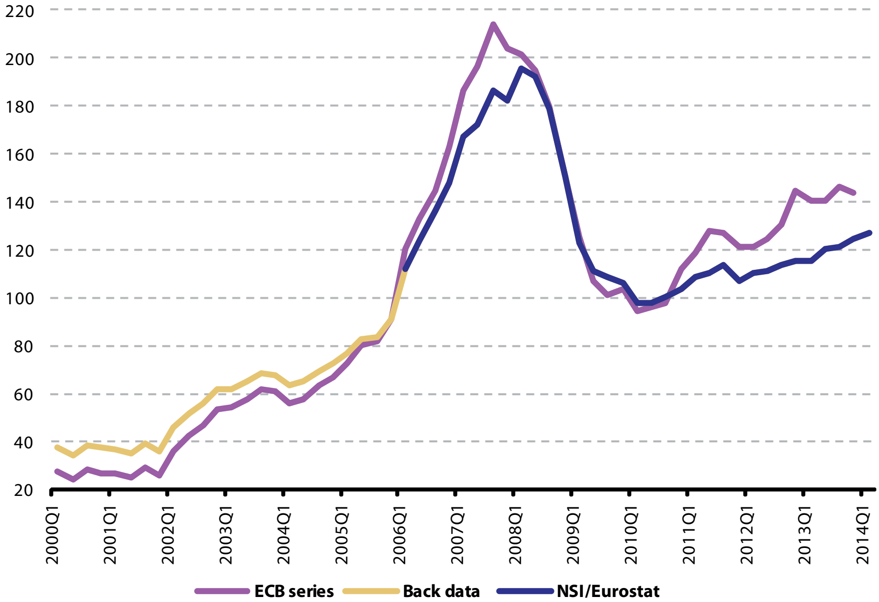
*Source:*Eurostat

Table 1.3: Lithuanian quarterly HPI
(2010 = 100)

	Backcasted	ECB		Eurostat	ECB		Eurostat	ECB
2000Q1	50.60	29.4	2006Q1	102.85	105.1	2012Q1	106.88	105.2
2000Q2	48.87	27.5	2006Q2	103.80	100.7	2012Q2	105.75	97.3
2000Q3	47.74	26.3	2006Q3	114.26	107.4	2012Q3	106.81	94.3
2000Q4	49.67	28.2	2006Q4	126.81	126.3	2012Q4	105.99	95.1
2001Q1	53.94	32.9	2007Q1	132.13	138.6	2013Q1	106.72	105.4
2001Q2	52.15	30.8	2007Q2	137.60	146.3	2013Q2	108.31	101.5
2001Q3	55.60	34.6	2007Q3	146.57	153.6	2013Q3	106.41	96.0
2001Q4	59.87	39.7	2007Q4	149.25	157.9	2013Q4	109.16	103.1
2002Q1	57.12	36.2	2008Q1	154.27	175.6	2014Q1	110.84	112.1
2002Q2	57.53	36.6	2008Q2	159.72	170.7			
2002Q3	59.03	38.4	2008Q3	156.87	160.1			
2002Q4	60.26	39.8	2008Q4	145.52	158.2			
2003Q1	65.56	46.5	2009Q1	116.35	133.5			
2003Q2	63.43	43.6	2009Q2	110.20	113.2			
2003Q3	64.24	44.5	2009Q3	105.13	107.1			
2003Q4	63.64	43.7	2009Q4	100.24	103.4			
2004Q1	67.11	48.1	2010Q1	98.28	105.1			
2004Q2	65.62	46.0	2010Q2	100.44	101.8			
2004Q3	70.07	51.9	2010Q3	99.68	94.6			
2004Q4	74.83	58.6	2010Q4	101.60	98.5			
2005Q1	83.31	71.5	2011Q1	105.93	106.6			
2005Q2	81.81	69.0	2011Q2	106.74	99.8			
2005Q3	85.56	74.8	2011Q3	106.49	97.6			
2005Q4	97.99	96.2	2011Q4	107.25	101.2			

Source: Eurostat

Figure 1.4: Latvian quarterly HPI
(2010 = 100)



Source: Eurostat

Table 1.4: Latvian quarterly HPI
(2010 = 100)

	Backcasted	ECB		Eurostat	ECB		Eurostat	ECB
2000Q1	37.71	29.4	2006Q1	111.95	120.3	2012Q1	109.93	121.0
2000Q2	34.06	27.5	2006Q2	123.48	133.0	2012Q2	111.49	124.2
2000Q3	38.26	26.3	2006Q3	136.32	144.3	2012Q3	113.84	130.6
2000Q4	37.17	28.2	2006Q4	148.34	163.0	2012Q4	115.25	144.7
2001Q1	36.37	32.9	2007Q1	167.51	186.8	2013Q1	115.38	140.3
2001Q2	35.40	30.8	2007Q2	172.47	196.9	2013Q2	120.13	140.8
2001Q3	38.98	34.6	2007Q3	186.09	213.9	2013Q3	121.55	146.6
2001Q4	35.46	39.7	2007Q4	182.67	203.7	2013Q4	124.72	144.1
2002Q1	45.87	36.2	2008Q1	195.45	201.5	2014Q1	127.40	112.1
2002Q2	51.96	36.6	2008Q2	191.90	194.5			
2002Q3	55.49	38.4	2008Q3	179.00	179.7			
2002Q4	61.63	39.8	2008Q4	150.23	150.1			
2003Q1	61.98	46.5	2009Q1	123.18	125.1			
2003Q2	64.81	43.6	2009Q2	110.82	106.9			
2003Q3	68.61	44.5	2009Q3	109.02	101.3			
2003Q4	67.83	43.7	2009Q4	106.21	103.8			
2004Q1	63.35	48.1	2010Q1	97.72	94.1			
2004Q2	65.13	46.0	2010Q2	98.07	96.4			
2004Q3	69.43	51.9	2010Q3	100.60	97.7			
2004Q4	72.51	58.6	2010Q4	103.61	111.9			
2005Q1	77.09	71.5	2011Q1	108.25	118.6			
2005Q2	83.00	69.0	2011Q2	110.15	127.9			
2005Q3	83.75	74.8	2011Q3	113.70	127.4			
2005Q4	91.16	96.2	2011Q4	107.30	121.6			

Source: Eurostat

Table 1.5: Spanish quarterly HPI
(2010 = 100)

	Backcasted	ECB		Eurostat	ECB		Eurostat	ECB
2000Q1	43.83	45.81	2005Q4	91.94	97.60	2011Q3	91.56	91.98
2000Q2	45.02	47.09	2006Q1	95.36	101.03	2011Q4	86.81	88.15
2000Q3	45.63	47.73	2006Q2	99.44	103.92	2012Q1	82.48	83.76
2000Q4	45.69	47.79	2006Q3	102.69	104.72	2012Q2	79.80	81.04
2001Q1	47.53	49.76	2006Q4	105.34	106.48	2012Q3	76.82	77.98
2001Q2	49.14	51.48	2007Q1	107.87	108.30	2012Q4	75.74	76.88
2001Q3	50.19	52.60	2007Q2	110.95	111.40	2013Q1	71.90	71.82
2001Q4	50.69	53.14	2007Q3	112.15	112.60	2013Q2	71.33	71.28
2002Q1	53.64	56.29	2007Q4	111.35	111.80	2013Q3	71.89	71.80
2002Q2	56.88	59.78	2008Q1	110.94	111.38	2013Q4	70.97	70.89
2002Q3	58.18	61.16	2008Q2	110.65	111.04	2014Q1	70.75	70.66
2002Q4	59.28	62.34	2008Q3	108.89	109.19			
2003Q1	62.51	65.82	2008Q4	105.53	105.76			
2003Q2	66.49	70.10	2009Q1	102.71	102.93			
2003Q3	68.22	71.97	2009Q2	102.25	102.48			
2003Q4	69.96	73.84	2009Q3	101.32	101.55			
2004Q1	73.73	77.91	2009Q4	100.91	101.17			
2004Q2	77.84	82.35	2010Q1	99.65	99.91			
2004Q3	79.43	84.07	2010Q2	101.27	101.52			
2004Q4	81.75	86.58	2010Q3	99.59	99.33			
2005Q1	85.07	90.16	2010Q4	99.49	99.23			
2005Q2	88.43	93.80	2011Q1	96.10	95.78			
2005Q3	89.87	95.36	2011Q2	94.95	94.65			

Source: Eurostat

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy: via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm); from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm); by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Priced subscriptions:

- via one of the sales agents of the Publications Office of the European Union (http://publications.europa.eu/others/agents/index_en.htm).

KS-GP-15-002-EN-N

Publications Office