How accurate are the National Bank of Belgium's macroeconomic projections ?

G. Minne
T. De Keyser
G. Langenus

Introduction

This article assesses the track record of the macroeconomic projections that are regularly produced by the National Bank of Belgium (NBB)[1]. While these projections also support the Bank's analysis and understanding of the Belgian economy, they are specifically tailored to inform monetary policy for the euro area in the context of a joint exercise with the European Central Bank (ECB) and the other national central banks of the Eurosystem.

According to its mandate, the objective of the ECB's monetary policy is to maintain price stability, which is operationalised as reaching an inflation rate close to but below 2 % in the medium term. Attainment of this target is monitored by means of a two-pillar strategy consisting of monetary and economic analysis. The monetary analysis specifically focuses on the money supply growth rate, while the latter takes a broader view and assesses whether the current and future macroeconomic developments are in line with the ECB's objective. In this connection, the regular macroeconomic projection exercises conducted by the Eurosystem institutions constitute a key input for the ECB Governing Council's economic analysis. The projections are independently prepared by the staff of the participating institutions and the Governing Council typically takes note of them in the monetary policy decision-making process.

In practice, two large-scale projection exercises are undertaken each year in a coordinated procedure by the ECB and the national central banks. They are finalised in June and in December. The results are published for the euro area and the individual euro area countries. Between those two exercises, smaller-scale intermediate updates are provided by the ECB – with some involvement from the national central banks – in March and in September. These updates are also published under the responsibility of the ECB but only cover developments at euro area level. In addition, some national central banks, such as the Banque de France and the Banco de España, also provide updates for the outlook of their national economies in March and September.

As regards the coordinated exercises in June and in December, the euro area outlook is derived in a bottom-up manner from the projections for the different individual countries made by the national central banks. However, joint ownership of the aggregate results is ensured by various coordination procedures that include peer reviews of the country results, common international and financial assumptions, as well as a trade consistency module. The latter ensures that intra-euro area trade flows are projected in a consistent manner. In addition, common projection guidelines are to be followed, notably as regards the extent to which announced fiscal policy measures can be taken into account.

(1) We would like to thank our ECB colleagues and especially A. Page, G. Kontogeorgos and K. Lambrias for the use of a large database of the Eurosystem projections.

The projection exercises give the outlook for a broad range of macro variables in the current and the next two to three[1] years. Apart from growth and inflation, projections are made for the demand components, the labour market, public finances, the current account, etc. This is important as an assessment of the inflation projections, for monetary policy purposes, requires a correct view on the drivers of inflation and, hence, the broader developments in the economy. In that connection, it should be stressed that not just the numbers of the projections are important, but also the economic story behind these numbers. That is why the publication of the projections – both by the ECB for the euro area results and by the national central banks for the individual country results – typically takes the form of a comprehensive article that describes the outlook in detail[2].

Turning to the specific projections for Belgium, the NBB uses both a set of econometric analytical models and experts' judgement to elaborate the macroeconomic outlook. As regards the former, the workhorse econometric model is Noname (Jeanfils & Burggraeve, 2005), a quarterly, medium-scale, neo-Keynesian model for the Belgian economy. This model is driven by intertemporal optimisation behaviour of representative agents and covers the most important macroeconomic aggregates. It favours a "story-telling approach" and delivers projections that are consistent with economic theory. However, other complementary tools, that provide a more detailed and granular approach, are used for the public finances projections, as well as for the inflation projections, in particular for the short term. In addition, the short-term estimates for GDP growth are anchored to two specific nowcasting models – BREL (Piette, 2016) and R2D2 (Basselier *et al.*, 2017) – that use a very broad range of indicators, including information from surveys and high-frequency hard data, such as on turnover and retail sales. Several groups of experts specialised in different fields are involved in the forecasting process and analyse thoroughly the details and figures derived from econometric models. Expert judgement, that may modify pure model forecasts, is part and parcel of our projection approach.

This article reviews the performance and the reliability of the macroeconomic projections for the Belgian economy since 2001, as produced by the NBB twice a year. It uses a database that includes projections that were released by national central banks and some major international institutions but also the different vintages of data published by the National Accounts Institute (NAI) for the macroeconomic variables on which these projections were based. The article specifically evaluates the forecast errors – defined as the difference between projections and actual values, notably in comparison with other institutions. Further statistical tests are also used to detect an over- or underestimation bias in the NBB projections and to evaluate their directional accuracy. We mostly focus on the projections for GDP and the demand components but provide results for inflation and employment as well. Given the specific rules regarding the public finances projections[3], we do not extend the analysis to the outlook for government deficit and debt.

The findings in this article should be interpreted with some caution : a number of caveats should be mentioned explicitly. First, the period considered is necessarily rather short, especially when sub-sets of the available data are considered, which may limit the statistical significance of the results. Second, the average forecast performance over the whole period is significantly affected by the sharp unexpected decline in economic activity during the great recession. This is not an idiosyncratic feature of the Eurosystem projections. As argued by Alessi *et al.* (2014), the forecast performance during the financial crisis was notably worse than before it but remained comparable to that of other central banks and forecasters. Third, methodological or operational statistical changes can influence the results to the extent that statistics on the (final) outcomes are not produced using the same methodology as the one that applied at the time of the projections. Despite those limitations and notwithstanding the fact that past performance is not necessarily indicative of future results, the findings in this article may be used to fine-tune forecasting approaches.

The remainder of this article is organised as follow. The next section focuses on the NBB projection errors for Belgian GDP growth while the second one compares the forecast accuracy of the NBB with those of other institutions. This is followed by a section that brings together different statistical tests to characterise the projection errors. The fourth section then analyses forecast accuracy in greater detail, namely by looking into the forecast errors for the different components of aggregate demand, as well as those that are caused by the common Eurosystem assumptions. The fifth section compares the forecast accuracy of the Belgian GDP projections with those for the euro area. The sixth and seventh sections then briefly analyse the projection errors for inflation and employment, respectively. The last section concludes and reiterates the key findings.

(1) In the December exercise, the year t+3 is added to the projection period. All the other projections cover the period up to year t+2.

(2) The first article in this Economic Review describes the December 2018 projection results for Belgium.

(3) In accordance with the projection guidelines, fiscal measures are only incorporated if they are already specified in sufficient detail and likely to pass the legislative process. As in most countries, the budget cycle still mainly has an annual frequency, which implies that the public finance projections, in particular for the last few years covered, do not always correspond to a 'most likely' scenario. They only describe the outlook in the absence of additional measures. In reality, such measures can be taken in subsequent budgets, which then modifies the outlook for the budget deficit and public debt.

1. Annual GDP projection errors for Belgium: concept and first assessment

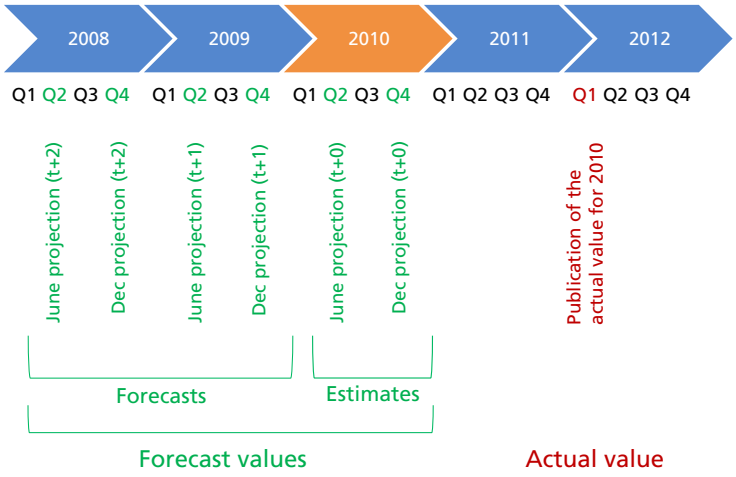
This article focuses on projection errors for Belgium over the years from 2001 to 2017, the period for which published data are available. The main yardstick for measuring projection errors is the *actual value* that the National Accounts Institute supplies for a given time period, less a *forecast value* for that time period. Hence, a negative value attributed to this measure suggests that the variable of interest has been overestimated.

As regards the *actual value*, selecting different vintages of published data can considerably influence the results since published national accounts data can be revised and corrected as more recent and accurate statistics become available. Choosing the final or current statistics as the benchmark typically increases the aforementioned risk that the comparison will be biased by methodological changes in the compilation of the national accounts. Against this background, the actual value for annual data is defined as those data published two years after the year has ended, and, more precisely, the vintage published in the first quarter[(1)]. This choice represents a trade-off between using the first estimates, which are based on relatively little hard data and, thus, potentially subject to significant revisions, and using the most recent releases for the period under review, which might be subject to methodological changes.

The *forecast value* is produced as part of the projection exercise. Projections whose forecast horizon is t+3 are disregarded as they were only introduced in December 2016 (for the requirements of the stress tests for financial institutions), which implies that the evaluation sample is too short. Moreover, it is explicitly acknowledged that projections that are that far in the future come with a very large degree of uncertainty and should be considered more as a technical scenario, rather than an actual forecast. Depending on the forecast horizon under consideration, the forecast consists of the estimated or projected value of a macroeconomic variable for the year for which the forecast was made, the year after or the year after that (i.e. the value up to two years in the future). If one disregards the year t+3, each projection exercise includes two annual forecasts (i.e. t+1 and t+2) and one annual estimation (i.e. t+0) for each variable and, consequently, the variable in each target year has been forecast or estimated six times before the first NAI statistics become available.

CHART 1 NBB PROJECTION EXERCISES FOR THE YEAR 2010

(June and December projections)



Source: NBB.

(1) For the most recent years and quarters, the data availability limits the construction of the error measure and for those cases, the realisation value used is extracted from the most recent available data, as published by the NAI.

A specific example can further clarify the approach. For annual GDP growth in 2010, for instance, the actual value is fixed as the GDP growth for that year published by the NAI in the first quarter of 2012. In order to estimate the maximum (two-year) forecast horizon for GDP growth, we use the growth rate for that year included in the June 2008 (June, t+2) NBB projections. Afterwards, five other projections have targeted the same variable, from the projections published in December 2008 to those in December 2010. As mentioned, forecasts for time periods beyond a horizon of 2 years are disregarded, given their very limited availability, but also because the predictive power of projecting macroeconomic variables that far into the future is low.

Turning to a first analysis of the forecasting accuracy, the mean of the forecast error on GDP growth comes to –0.84 pp for the maximum horizon, which corresponds to the June projection two years before the target year. So, on average, the NBB tends to overestimate GDP growth at this forecast horizon. However, a large part of this bias is caused by the fact that the substantial downturn in the great recession that was not predicted in the years building up to it, neither by the Bank, nor by other forecasters. Excluding the financial crisis (the observations related to 2008 and 2009) would cut the overestimation at longer horizons by about half. For the shorter horizons – i.e. the projections realised in the current year or in December the year before –, the mean error is much smaller and ranges between –0.1 and 0.1 of a percentage point. For the December exercise of the actual year, it falls to just 0.02 pp. This is to be expected, as projections for GDP released in December of the actual year include data, or at least first vintages, that has already been published on quarterly GDP growth for the first three quarters of the target year. Forecast uncertainty is significantly reduced under those circumstances, but not removed altogether due to the unknown fourth quarter figure and data revisions.

TABLE 1 AVERAGE AND ABSOLUTE ERRORS: GDP GROWTH 2001-2017
(in percentage points, annual data in volume)

	t+2		t+1		t+0	
	June	December	June	December	June	December
Average forecasting error						
Mean	**–0.84**	**–0.63**	**–0.45**	**–0.07**	**0.05**	**0.02**
Mean (2008-09 excl.)	*–0.46*	*–0.26*	*–0.10*	*0.17*	*0.05*	*0.02*
Median	–0.37	–0.12	–0.02	0.21	0.10	0.03
Median (2008-09 excl.)	*–0.14*	*0.10*	*0.12*	*0.26*	*0.10*	*0.03*
Absolute forecasting error						
Mean	**1.02**	**0.99**	**1.08**	**0.67**	**0.38**	**0.20**
Mean (2008-09 excl.)	*0.67*	*0.68*	*0.83*	*0.52*	*0.32*	*0.16*
Median	0.59	0.41	0.62	0.56	0.27	0.16
Median (2008-09 excl.)	*0.39*	*0.31*	*0.59*	*0.51*	*0.14*	*0.24*

Sources: NAI, NBB.

Large positive errors can be offset by large negative errors of the same size, resulting in a misleadingly low mean error. It is possible to get round this problem by taking the absolute value of the error and switching to the mean absolute error (MAE). Unsurprisingly, the absolute errors confirm that, as the projection horizon becomes larger, the uncertainty behind the projections increases. The MAE of the NBB projections released in June of the year before and earlier years is approximately 1 percentage point. As regards the shorter horizons, the absolute error comes down to 0.38 and 0.2 of a percentage point for the projections made, respectively, in June and December of the actual year. Again, disregarding the observations corresponding to the financial crisis results in a significantly lower MAE, particularly for longer horizons.

The projection for GDP growth in 2009 released in December 2007 gives an illustration of the importance of the errors caused by the great recession. The absolute error of this forecast worked out at 4.9 percentage points (the forecast

value for GDP growth was 2.3 % while the actual value was –2.6 %). Depending on the horizon, dropping the crisis years from the evaluation sample would reduce the MAE for the entire sample period by 15 % to 35 %. The MAE for the two-years-ahead December projections tripled during the crisis, compared to the pre-2008 figures. Strikingly, in the period 2008-2010, the MAE is first driven by large negative forecast errors due to the unexpected drop in economic activity, but during the recovery phase also by positive forecast errors due to the slightly stronger-than-expected pick-up in growth.

CHART 2 PROJECTION ERRORS: GDP GROWTH

(in %, data in volume)

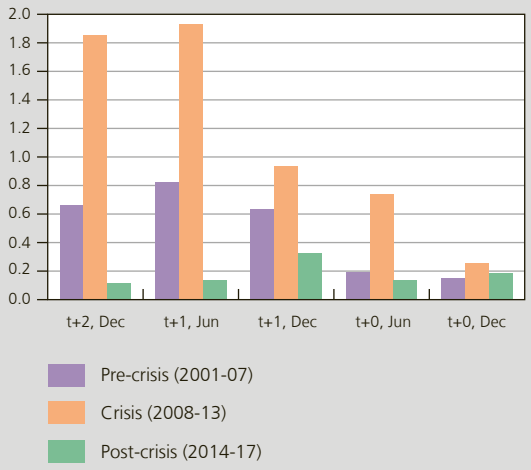
Sources: ECB, NBB.

Calculating the median error can suggest a "typical" forecast error that is not affected as much by outliers in the evaluation sample and, in this case, by the bias introduced by the great recession. When ranked by size, the median of the absolute forecast error falls in the middle of the dataset. Results confirm that, in general, as horizons lengthen, forecast uncertainty increases. However, for horizons beyond a year, the forecast error does not continue to get bigger (as is evident from chart 2). The gap between the median and the mean absolute error widens as the horizon lengthens due to the outliers related to the financial and sovereign debt crises.

All in all, the great recession had a profound effect on the average forecast performance. Against this background, it is worthwhile taking a look at how the projection accuracy has varied over time. To this end, we consider three sub-samples to assess the forecasts, even though this reduces the statistical significance of the evaluation due to the lower number of observations. The crisis-related sub-sample incorporates not only the great recession period itself, but also the consecutive recovery and the European debt crisis. The pre-crisis sub-sample covers the period from 2001 to 2006, while the post-crisis sub-sample spans the years from 2014 onwards. The MAE tends to be lower in the post-crisis period for projections made one or two years ahead and the fact that forecast errors tend to be smaller in recent years is also visible from chart 1. The interpretation of those figures, however, is not straightforward: it does not necessarily mean that the more recent projections are of any better quality. The post-crisis GDP growth rate is significantly lower and less volatile than the pre-crisis growth rate and this needs to be taken into account in the interpretation of the results when comparing the different periods[1]. For the shorter time horizons, the difference between the pre-crisis and post-crisis forecast error appears to be smaller.

CHART 3 IMPACT OF THE CRISIS ON PROJECTION ACCURACY: MEAN ABSOLUTE ERROR FOR THE GDP GROWTH RATE

(in percentage points, data in volume)



Sources: ECB, NBB.

(1) A measure of forecasting error scaled by the standard deviation could be considered as a valid tool in theory, but in this case the low number of observations in the post-crisis sub-sample (only four) makes it difficult to use the standard deviation as a benchmark.

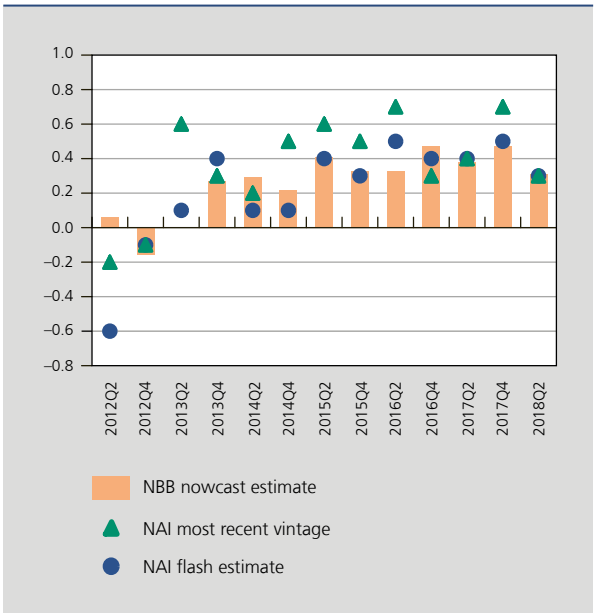
Accuracy of nowcasting models

In addition to the main macroeconometric model, specific tools are used to estimate GDP growth in the short run in the context of the Bank's twice-yearly macroeconomic projection exercises. This is mostly because a broader dataset of high-frequency indicators, consisting of both certain hard data and survey information, is available for nowcasting or short-term estimates. Duly taking these data into consideration is likely to improve forecast accuracy. In practice, the first quarter for which no NAI releases for GDP are available at the time of the projection (this is either the second, for the June projections, or the fourth quarter of the year, for the December projections) is estimated on the basis of short-term indicators and expert judgment. In 2015, the methodology for short-term estimates was revamped and since then, specific formal nowcasting models have been routinely used, even though expert judgment still plays a role, in particular when the estimates from the different models diverge. We specifically use a bridge model with predictor selection, called BREL (Piette, 2016), and a dynamic factor model (Basselier *et al.*, 2017), referred to as R2D2 in its current version, to support our short-term estimates of quarterly GDP growth.

In this box, we briefly look into the accuracy of these estimates using two evaluation benchmarks. The first one is the first quarterly (the so-called 'flash') estimate for GDP, published by the NAI about 30 days after the end of each quarter. The second is the latest vintage of GDP growth for that quarter, which incorporates all currently available information and hence can significantly deviate from the first estimate due to data revisions by the NAI. It should be noted that the current statistics used here are by no means final. GDP growth continues to be revised for figures dating back several years.

QUARTER-ON-QUARTER GDP GROWTH: NBB NOWCAST AND FIRST AND CURRENT NAI STATISTICS (2012-2017)

(seasonally and calendar adjusted volume data, in %)



Sources: NAI, NBB.

A comparison for the recent period shows that the NBB nowcasts tend to be quite close to the NAI flash estimate, while the deviation (i.e. the forecast error) with respect to the most recent NAI statistics can be somewhat more important (as the latter can be quite different from the first NAI statistics). The average forecast error with respect

to the flash estimate is about zero over the period under scrutiny. This implies that the nowcast of the NAI flash estimate of quarterly GDP growth is unbiased, as no consistent over- or underestimation is found. The accuracy of the estimates can be judged by considering the average absolute forecast error over the period. The nowcast generally turns out to be very accurate, with the flash estimate having an average absolute error below 0.13 of a percentage point. Looking at two different sub-periods, the accuracy of the estimates seems to have improved significantly from 2015 onwards, even when no account is taken of the outlier related to the flash estimate for the second quarter of 2012. This could suggest that bringing in the new models has improved the nowcasts, at least with respect to the NAI flash estimate.

When compared to the most recent data vintage, the nowcasts turn out to have a slight downward bias. On average, growth has been underestimated by 0.09 percentage point. The accuracy of the forecasts is also slightly worse, as the average absolute error rises to nearly 0.2 percentage point. Errors have again become smaller in recent years, although this may also be partly due the fact that these quarters have been less subject to statistical revisions.

Overall, the specific nowcasting procedures used in the context of the twice-yearly macroeconomic projections seem to produce satisfactory results, especially when compared with the NAI flash estimate. The clear improvement in forecast performance after 2015 suggests the nowcasting models that were introduced back then are fairly accurate.

2. Comparing GDP different institutions' growth projections for Belgium

This section compares the NBB's forecasting performance for annual Belgian GDP growth with other institutions' results, using the officially published projection figures. The maximum horizon considered in the comparison is the forecast released in the fourth quarter two years ahead of the target year[1]. Projection errors are constructed following the methodology described in the previous section and both the average error and the MAE are used as a gauge[2]. The reference data for the actual value of the variables under analysis are common for all institutions (corresponding to data published by the NAI).

While comparing the forecast performance, differences in the timing of each institution's publication should be acknowledged, as the available set of information at the time of the projection exercise may differ. Having a more up-to-date information set can be a considerable "advantage" for institutions releasing their results at a later moment in time. To minimise the impact of this, we restrict the dataset to projections published in the second and the fourth quarters, as most institutions conduct a projection exercise then[3]. Nevertheless, the comparison must still be interpreted with some caution, as there will still be some timing differences.

2.1 Comparison with the Federal Planning Bureau

Bearing in mind those caveats, the small sample size and the non-adjustment for any outliers during the sovereign and financial crises, an initial comparison can be made between the NBB's projections and those made by the Federal Planning Bureau (FPB). The best basis for comparison is probably the publication by the FPB of the Economic Budget in the second quarter (in place since 2013). As the horizon used by the FPB in this Economic Budget is shorter than in the NBB projections, the comparison is limited to the current and the following year. The comparison of forecast errors between the FPB and the NBB suggests that, over this short period of time, the signs and magnitudes of the forecast errors of the projections released by the Federal Planning Bureau are very close to, albeit slightly larger than, those for the NBB. Overall, the forecasting accuracy of both institutions has tended to be comparable in recent years.

(1) Neither the OECD nor the European Commission conduct projections in the second quarter two years ahead of the target year.
(2) Using the root-mean-square error (RMSE) does not change the conclusions and is less intuitive.
(3) May/November for the European Commission (spring and autumn forecasts), June/November for the OECD (Economic Outlook), April/October for the IMF (World Economic Outlook), June for the Federal Planning Bureau (Economic Budget) and June/December for the NBB (economic projections for Belgium)

TABLE 2 ANNUAL FORECAST ERROR: GDP GROWTH IN 2013-2017

(2nd quarter projection exercise, in percentage points, data in volume)

	NBB		Federal Planning Bureau	
	t+0	t+1	t+0	t+1
	Current year	One year ahead	Current year	One year ahead
Mean error	0.168	–0.163	0.183	–0.177
Mean absolute error	0.168	0.336	0.170	0.368

Sources: FPB, NBB.

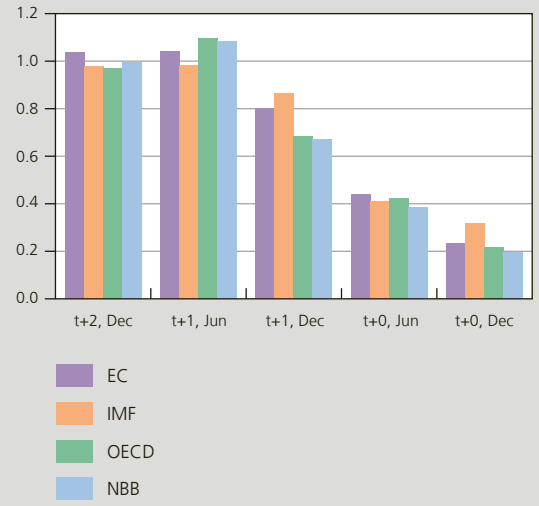
2.2 Comparison with international institutions

The comparison with the other institutions – the OECD, the European Commission and the IMF – can be made on the basis of a larger dataset, both in terms of the evaluation period (2001-17) and projection exercises performed in a comparable timeframe (2nd and 4th quarters).

The results indicate that, for the projections with the shortest horizons – between the 4th quarter of the preceding year and the 4th quarter of the current year –, the NBB's GDP growth projections have the lowest mean absolute error. That being said, differences remain limited and, for example, projection errors for GDP growth estimates produced in December of the actual year (the shortest possible horizon) by all institutions on average do not exceed 0.32 pp in absolute terms. For the longer horizons, the NBB's projections are not better than those of other institutions. The IMF tended to marginally outperform the other institutions for the forecasts produced in the second quarter of the previous year (t+1, June) while the OECD did so for the forecasts produced in the fourth quarter two years ago (t+2, Dec).

CHART 4 ABSOLUTE FORECAST ERROR: ANNUAL GDP GROWTH

(in percentage points, annual data in volume)



Sources: EC, IMF, OECD, NAI, NBB.

However, given the small sample size and the volatility of the underlying errors, the difference in forecasting errors between the institutions is mostly non-significantly different from zero. Overall, the forecast performance appears to be very similar across the institutions.

While comparing averages is informative, important caveats apply. This analysis does not take into consideration, for example, the distribution of the forecast error and therefore the dispersion within the sample. The simple comparison above can be backed up with a test developed by Diebold and Mariano (1995), which assesses the statistical significance of the differences between two forecast series. For GDP growth projections, the conclusion remains robust to the Diebold-Mariano methodology: the NBB forecasts tend to outperform those of other institutions for shorter horizons, while for longer horizons the gap between the institutions is not statistically significant. As regards the NBB's better performance for the shorter horizons, it should be borne in mind that the NBB projections typically come a bit later in the quarter than those by the international institutions and so may be based on a larger and more recent dataset.

3. Additional statistical tests

The forecast errors for GDP growth made by the NBB and the major international institutions can be formally analysed and compared in greater detail by applying standard statistical tests to the sample.

3.1 Unbiasedness

The test of unbiasedness is based on the simple mean forecast errors and aims to detect a systematic tendency to under- or overpredict the variable of interest. The forecast error is regressed on a constant term, as is common in the literature and in similar analyses[1]. The test is performed in a recurrent manner for the projections with a horizon of 0, 1 and

TABLE 3 ADDITIONAL TESTS OF FORECAST ERRORS FOR BELGIAN GDP GROWTH
(annual data in volume)

	Unbiasedness test[1]			Median test[2] (in %)		
	t+0	t+1	t+2	t+0	t+1	t+2
NBB	0.03	–0.25	–0.73***	66.7*	54.8	37.9
OECD	0.06	–0.37	–0.73***	57.6	45.2	33.3
IMF	0.13	–0.33	–0.70***	69.7**	54.8	48.3
EC	0.08	–0.30	–0.64	60.6	51.6	46.7
	Directional accuracy test[3] (in %)			Comparison with naive projections[4]		
	t+0	t+1	t+2	t+0	t+1	t+2
NBB	87.9***	48.4	55.2	0.97***	0.41**	0.27*
OECD	87.9***	45.2	40.0	0.94***	0.40**	0.29
IMF	84.9***	35.5	58.6	0.91***	0.36*	0.28
EC	78.8***	45.2	53.3	0.92**	0.36*	0.22

Sources: EC, ECB, IMF, OECD, NBB.
The significance level: *** indicates a p-value under 1 %, ** a p-value under 5 % and * a p-value under 1 %
(1) Based on coefficients from regressions over a constant term and on the use of robust standard errors.
(2) Based on the percentage of forecasting errors>0 and the exact binomial test.
(3) Based on the success rate – the share of observations for which the projected direction of change matches that seen in the actual value – and the statistical test is a 2-sided Fisher's exact test.
(4) Based on the difference in pp. between the MAE of the naïve projection (previous year growth fixed by the NAI at the time of the forecast) and the institution's projection and on the Diebold-Mariano test.

(1) The regressions are based on robust standard errors and the null hypothesis is rejected if this constant is significantly different from zero, indicating that there may be a bias in the forecast.

2 years ahead of the target year, i.e. combining the spring and autumn projections. The sign of the coefficient specifies the direction of the bias (negative values indicating overestimation).

For real GDP growth, overall, the constant in the regressions tends to become more negative for longer forecast horizons. This confirms the earlier finding that there was a tendency to overestimate GDP growth between 2001 and 2017, which is heavily influenced by the presence of outliers in the sample, caused by the great recession. On the shorter end, there is no strong bias in the projections. The fact that additional short-term indicators such as business or consumer confidence indicators and the first quarterly figures for GDP growth are integrated into those projections obviously reduces the average bias.

On average, all institutions have over-predicted GDP growth for the two-years-ahead projections, as confirmed by the test of unbiasedness. For shorter horizons, the bias was not significantly different from zero and the NBB's projections had the lowest bias among the four institutions.

3.2 Median test

The median test checks whether the median of the forecast errors is significantly different from zero[(1)]. The key objective is to determine whether the forecast errors have been positive and negative in equal measure while disregarding the size of the error. A neutral forecast exercise is not expected to be systematically in favour of either an overvaluation or undervaluation (i.e. 50 % of the errors are positive).

Overall, the median test shows that the median forecast error remains close to zero and that the signs of the errors are, more or less, equally likely to be positive or negative. However, looking at specific horizons, the forecasts made in the current year often show a positive error. So, these projections have generally tended to be too pessimistic. However, the imbalance is related to a specific period and seems to stem mainly from a persistent under-estimation of the recent economic recovery (2013-2017).

This does not relate to specific errors in the NBB projections as the other institutions exhibit a comparable profile across the forecast horizons under analysis. While more than half of the projections targeting the current year tend to be associated with a positive forecast error, more than half of the projections made two years ahead tend to have a negative forecast error. To some extent, the different institutions are leaning towards an underestimation for the short-term horizons but an overestimation in the two-years-ahead projections.

3.3 Directional accuracy

The test for the directional accuracy does not focus on the value of the forecast but rather on the *change* in the projected variable. In other words, it aims to check whether the predicted direction taken by the variable of interest (increase or decrease of the variable) is the right one rather than measuring the size of the error. The test result is expressed as a "success rate", which is defined as the share of observations for which the projected direction of change in the variable in question matches that seen in the actual figures. For the NBB, the preceding year projections or the two-years-ahead projections for GDP growth turn out not to be significantly different from 50 %, meaning that the projections do not fare any better at predicting accelerations or decelerations of GDP than a simple flip of the coin. The success rate for the projections for the actual year's GDP growth is much higher: in 87.9 % of the cases, the predicted direction was confirmed by the direction of the actual figure. Reaching strong results for short horizons but a poorer performance for longer horizons is relatively common in similar studies on forecasting performance.

The results of the other institutions show a similar picture, although their success rate seems to be slightly lower, on average, than that of the NBB.

3.4 Performance against a naive projection

Forecast performance can also be tested against a simple benchmark model: a naive projection, such as, for example, the GDP growth figure of the previous year at the time of the projection. The question then becomes: "is the projection

(1) The 2-sided p-values are calculated using the exact binomial test, with the null hypothesis being a perfectly balanced sample between positive and negative errors.

released by the institution more informative of the future than simply taking the growth rate of the previous year?". Even though this type of test is relatively common, the results should be interpreted with caution as naive projections are not conditioned on external assumptions and do not require the baseline projection to follow a consistent economic scenario or to comply with some technical assumptions.

The test results indicate that the projections made by all institutions produce better results than the naive ones. The mean absolute forecasting error of the naive projection appears to be higher than NBB projections for each horizon considered. However, the difference in MAE shrinks with the forecast horizon because the MAE associated with the naive projections is quite stable, whereas the MAE of the NBB is rising. The simple comparison between averages can be backed up with the Diebold-Mariano test, which checks whether the expected difference between the NBB forecast and the naive projection is relatively small. The test results show that the two are not equal and that the mean absolute errors associated with the naive projections tend to be significantly higher than those associated with the NBB projections for every horizon. This conclusion also holds for other institutions for the horizon t+0 and t+1, but NBB projections are the only ones to beat the naive forecast by any significant margin at the t+2 horizon.

4. Aggregate demand components and importance of common Eurosystem assumptions

4.1 Aggregate demand components

As already indicated, a typical projection exercise consists of a complete and consistent economic scenario incorporating notably variables characterising the real economy, public finances, the labour market and price dynamics. More specifically, projection exercises not only include forecasts for GDP but also for the demand components, such as private consumption, investment, etc.

Interestingly, on average, projections for GDP growth tend to be more accurate than projections for major demand components, implying that positive errors made in forecasting certain demand components tend to be offset by negative errors made in forecasting other components. The findings indicate that NBB projections were generally too pessimistic for business investment growth for the shorter forecast horizon but too optimistic about developments in private and public consumption.

As errors for more volatile series (e.g. business investment or exports) are compared to time series that are more stable (e.g. private consumption or GDP), it may be worthwhile controlling for this volatility by scaling the absolute error using the standard deviation of the underlying variable and therefore by constructing the scaled MAE. The more volatile series are marked by a relatively high standard deviation, so using the standard deviation as a scaling factor could give a more comparable indicator of forecast accuracy.

Two conclusions emerge. On the one hand, based on the scaled MAE, both private and public consumption are singled out as variables with a relatively large forecast error. Regardless of the forecast horizon, there is a tendency to overestimate both private and public consumption growth. This is consistent with and may be caused by the errors on real disposable income, which is a measure of households' purchasing power and a key determinant of private consumption. Another variable which tends to be overestimated is public investment.

On the other hand, projections for business investment seem to be among the most accurate in relative terms as the scaled MAE is somewhat lower, especially in the short run. However, the results for the simple average error indicate a tendency to underestimate the variable for shorter forecast horizons. This is partly related to a number of large specific transactions made recently, notably important purchases of investment goods (mostly vessels and immaterial assets) abroad. These specific transactions are difficult to foresee and have significantly boosted investment growth but have not affected GDP growth (as they were offset by imports).

TABLE 4 NBB PROJECTION ERRORS: BELGIAN GDP COMPONENTS

(2nd quarter and 4th quarter projections exercises are aggregated, data in volume)

	Mean Error			Scaled Mean Absolute Error[1]		
	t+0	t+1	t+2	t+0	t+1	t+2
GDP	0.03	–0.25	–0.73	0.22	0.65	0.73
Private consumption	–0.03	–0.27	–0.72	0.69	0.80	1.06
Public consumption	–0.29	–0.58	–0.75	0.77	0.88	1.04
Gross fixed capital formation	0.35	0.13	–0.52	0.41	0.68	0.71
Business investment	0.41	0.42	–0.21	0.44	0.72	0.78
Public investment	–0.49	–0.67	–1.11	0.69	0.75	0.92
Housing investment	0.33	0.06	–0.55	0.67	0.76	0.71
p.m. Real disposable income	*–0.05*	*–0.51*	*–0.84*	*0.70*	*0.75*	*0.92*
Net exports[2]	–0.02	–0.20	–0.16	0.68	1.00	0.83

Source: NBB.

(1) Rescaled based on the standard deviations of the underlying variables.

(2) In percentage points.

4.2 Influence of external assumptions

In line with the Eurosystem projection guidelines and procedures, the NBB's projections are based on set of common external and financial assumptions (with a view to enhancing the consistency of the forecasts across the countries of the euro area). These assumptions are defined commonly by and for all participating institutions.

The contribution to the forecast errors coming from errors in the common assumptions can be approximated using the basic features of the Noname macroeconometric model used for the projections. The projection can in particular be reproduced using the actual values of the variables covered by the common assumptions (oil prices, external demand, interest rates, etc.). It should be stressed that this is only a rough mechanical approximation of the contribution of errors in the common assumptions to the forecast errors, as it assumes, among other things, that the expert judgment added to the model outcomes would have been identical for an alternative set of common assumptions, which in actual practice might not have been the case. In addition, the set of common assumptions has been expanded and refined throughout the years, something that should be borne in mind when comparing their contribution to the forecast errors over time: in principle, a larger part of the error can be traced back to the common assumptions in more recent years, simply because there are common assumptions for more variables now.

The results indicate that if one use conventional Noname model elasticities, forecast errors for Belgian GDP growth are to a large extent driven by the common Eurosystem assumptions for the one-year-ahead and two-years-ahead projections[1]. The part of the error that is not explained by the external assumptions – the residual error – does not constitute the main part of the total error on average. As regards the projections published in June of the preceding year, for instance, the common assumptions, on average, account for about two-thirds of the total forecast error. The longer the projection horizon, the bigger the contribution of the assumptions is.

In the one-year-ahead and two-years-ahead projections, the foreign demand assumptions, in particular, have significantly contributed to the forecast errors. In this connection, one should keep in mind that the forecasts on foreign demand are not only targeting global growth but also its trade intensity and this makes it especially difficult to accurately predict the foreign demand indicator. This assumption seems to have an important impact on the forecast error especially around major turning points in the business cycle (see for instance its contribution to forecast error in 2009). Remarkably, foreign

(1) The current-year projections are not taken into account as the common assumptions are less relevant for the shorter time horizons, that also rely, for instance, on specific nowcasting models that are less directly connected to these assumptions.

demand has been almost persistently overestimated in the last few years, which has led to an overly optimistic view on GDP growth and therefore a negative forecast error between 2012 and 2016. As a consequence, the common assumptions on foreign demand and, more specifically for the Belgian projections, the forecast growth of Belgian export markets have been gradually revised downwards.

CHART 5 EXTENT TO WHICH THE GDP FORECAST ERRORS CAN BE ATTRIBUTED TO THE COMMON EUROSYSTEM ASSUMPTIONS

(Estimated using basic Noname model elasticities)

Sources : ECB, NBB.

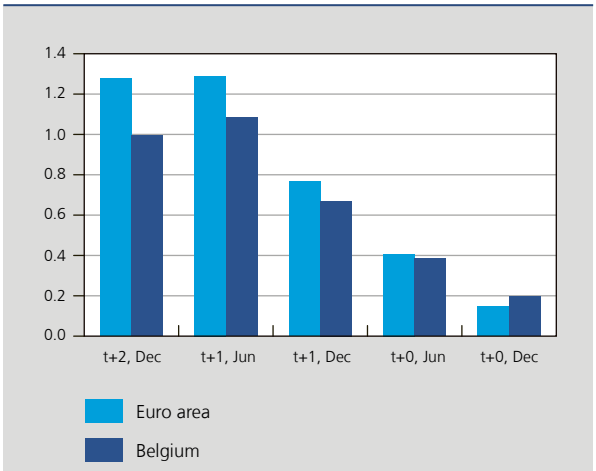
However, despite a relatively large contribution of the foreign demand hypothesis to forecast errors in absolute terms, the most recent years are characterised by low projection errors. Part of this can be attributed to the fact that the latter effect was offset either by the residual or by other variables that turned out to foster GDP growth more than initially expected (e.g. low oil prices). In 2015, for example, compared to what was taken into account in the June forecast of 2014 (t+1, June), oil prices had fallen more than expected and the euro had weakened by more than anticipated against the US dollar. This was unexpected and had a positive effect or contribution on the overall forecast error, indicating that this could have led to an underestimation of GDP growth. However, because actual foreign demand was generally weaker than expected, the overall error was low. While the offsetting contributions of errors in different individual assumptions may simply be down to 'good luck', the offsetting contribution of the residual error, that has typically helped keep the forecasting error low in the one-year-ahead projections for recent years for instance, may suggest that expert judgment was at times used to attenuate the impact of the assumptions (because of perceived risks related to either those assumptions or the mechanically estimated impact of them in the recent period).

5. Projections for the euro area

As is the case for other national central banks' projections, the NBB's macroeconomic projections are integrated into an overall euro area forecast. Formal and informal interactions among the participating institutions ensure that the country aggregate reflects the area-wide scenario and is jointly owned by all participating central banks[1]. Note that, due to the euro area enlargement, more countries are involved in the euro area projection exercises at the end of the period under analysis than at the beginning.

CHART 6 MEAN ABSOLUTE ERROR: ANNUAL GDP GROWTH

(in percentage points, available Belgian obs. & corresponding EA obs., balanced sample)



Sources: ECB, NBB.

The forecast accuracy for Belgian GDP can be compared with that for the euro area GDP. To have an idea of where the Belgian forecast accuracy ranks alongside other member countries, the mean absolute error for euro area GDP growth over the period 2001-2017 can be compared to the corresponding error for Belgian GDP growth predicted by the NBB[2]. The euro area GDP growth forecasts seem to have slightly fewer errors than their Belgian counterparts at the shortest horizon but exhibit a clearly larger forecasting error for longer horizons.

(1) For further details about the Eurosystem/ECB staff macroeconomic projection exercises, see the ECB website (https://www.ecb.europa.eu/pub/projections/html/index.en.html)

(2) One should remain cautious about the comparison between countries when considering an indicator scaled with the standard deviation of the underlying variable as it remains heavily influenced by outliers and in this case by the crisis episodes.

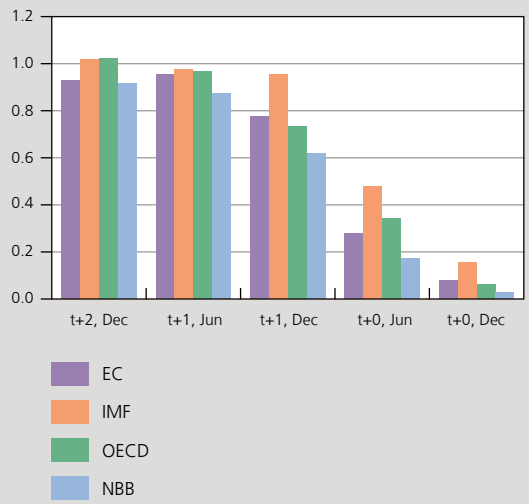
When comparing the forecast accuracy for Belgium and the euro area, it should be kept in mind that forecast errors for the latter can be reduced by the fact that, especially given the decentralised nature of the actual projection procedures, overestimations for certain countries can be partially offset by underestimations for other countries or vice versa. This may in particular be an issue for the shorter horizons as the longer-term projections depend to a greater extent on common assumptions. At the same time, it should be acknowledged that, since the great recession, the volatility in euro area GDP has been somewhat larger than for Belgian GDP. This may imply that euro area growth is more difficult to predict than Belgian growth.

6. HICP inflation

As price stability is the ECB's objective, the accuracy of inflation projections is particularly important. Forecast accuracy for inflation should in principle be better than for GDP and not be compared to it as inflation statistics are available on a monthly basis, which is not the case for GDP. This also implies that small differences in the timing of publication can significantly affect forecast accuracy, in particular for the shorter time horizon, as more recent projections may be based on more monthly statistics.

Bearing this caveat in mind, it should be stressed that the MAE of the NBB inflation projections is typically clearly lower than that in the international organisations' projections, when comparing again the projections produced and published in the second and fourth quarter of each year. This applies to all projection horizons, but the difference is larger for the shorter horizons. For the projection for the current year published in June, the NBB's MAE is only about 40 % of the one associated with the three other institutions and for the December exercise, this even drops to about 20 %.

CHART 7 MEAN ABSOLUTE ERROR: ANNUAL HICP RATE
(in percentage points)



Sources: EC, FPS Economy, IMF, OECD, NBB.

The median test and the unbiasedness test both show that the NBB projections are tilted towards positive forecast errors when the forecasting horizon widens (i.e. a tendency to underestimate inflation, unlike the projections for euro area inflation) and towards zero for shorter horizons. The directional test indicates a relatively high success rate for the nowcasting exercises and confirms the ability to predict turning points to a certain extent. All in all, the inflation projections published by the NBB are quite accurate, particularly for the shorter horizons, both in absolute terms and relative to other institutions.

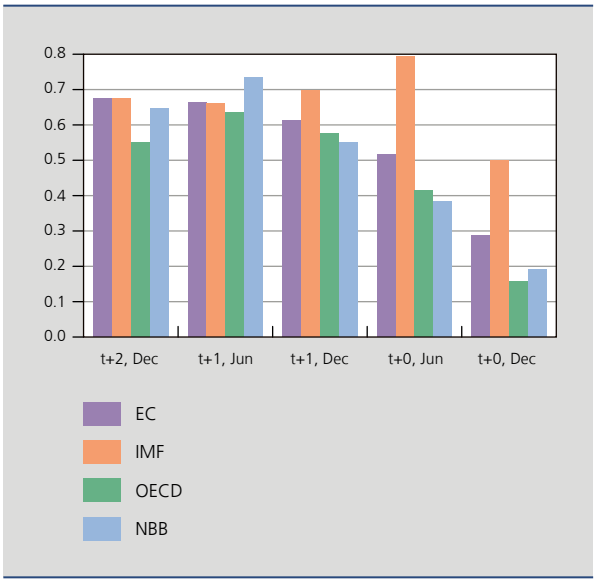
7. Employment growth

Employment is obviously also a key variable in the macroeconomic projections. As before, issues related to the timing of publication play a key role in explaining the differences in forecast accuracy between the institutions.

For longer horizons (t+2, Dec and t+1, Jun), the OECD projections have the lowest MAE. For shorter horizons, the NBB and OECD forecasts tend to be more accurate and are relatively close. From a general perspective, the MAE of employment growth improves more gradually as horizons shorten than those for GDP and errors on employment growth exhibit a somewhat "flatter profile". For the NBB, for example, the MAE for the December projections for the actual year is about one-third of the MAE associated with December projections two years ahead. For GDP growth, the MAE falls to one-fifth.

CHART 8 MEAN ABSOLUTE ERROR: ANNUAL EMPLOYMENT GROWTH

(in percentage points)



Sources: EC, IMF, NAI, OECD, NBB.

The additional tests suggest that the NBB projections for employment growth tend to be positively biased for the current-year and one-year-ahead horizons. This suggests a tendency to underestimate employment growth. Remarkably, the forecast error on employment growth between 2015 and 2017 has been positively biased, while a similar positive bias has not been observed in the same magnitude for GDP forecast at the same time. This suggests that we have recently clearly underestimated the employment intensity of growth. This may be related to a larger-than-expected impact of the labour cost moderation policies, as well as the structural reforms on the labour market. Finally, the directional accuracy seems to be significantly higher for the projections on employment growth than for other macro variables.

Conclusion

This article has assessed the accuracy of the NBB's twice-yearly macro forecasts for Belgium in the context of the coordinated Eurosystem projection exercises that form a key input for monetary policy decision-making in the ECB's Governing Council. A number of general conclusions can be drawn. Absolute yardsticks for projection errors are difficult to establish: how large is an 'acceptable' forecast error? Hence, projection accuracy is best assessed in relative terms, i.e. for different projection horizons, in terms of change over time or in comparison to other forecasts.

The GDP growth projections are obviously more accurate for the shorter horizons than for longer time periods. For the evaluation period considered, the forecast accuracy was significantly reduced by the great recession episode in particular. Disregarding the forecasts for 2008 and 2009 substantially reduces the average and mean absolute errors. The most recent projections, dating from the period after the great recession, tend to be marginally more accurate than the pre-crisis ones but this may simply because GDP growth has recently been less volatile and, hence, easier to predict. Regarding the very short run, GDP estimates for the first quarter of the projection horizon are estimated on the basis of short-term indicators and expert judgment and seem to produce satisfactory results, especially after the introduction of specific nowcasting models. When comparing forecasts that are published around the same time, the NBB projections for Belgian growth tend to be at least as accurate as those in the June Economic Budget prepared by the Federal Planning Bureau. In addition, the accuracy of the NBB forecasts is broadly in line with that of the projections of the major international organisations if a longer forecast horizon is considered, but somewhat better in the case of a shorter-term estimates. The differences in the timing of publication, however, may affect the comparability between institutions and some caution is warranted when interpreting the results.

With respect to the demand components of GDP, the forecast error associated with the aggregate measure of GDP growth tends to be smaller than for the individual components. In other words, any overestimation of certain components is offset by an underestimation of others. Remarkably, private consumption has been regularly overestimated and business investment underestimated. After taking into account the volatility of the underlying variables, private consumption stands out as the variable for which forecasting accuracy should be further improved, in particular for the shorter forecast horizons.

In accordance with the Eurosystem projection guidelines, the NBB projections need to be anchored to a set of common assumptions (regarding the oil price, external demand, interest rates, etc.) determined at the level of the Eurosystem. We show that the forecast errors in the evaluation period can be mostly traced back to errors in these common assumptions, notably related to Belgian export markets. The recent improvement in accuracy of the GDP projections can also be interpreted in this way : the overestimation of foreign demand has been offset more by either other exogenous variables that turned out to be more favourable for growth than assumed (such as the oil price in the 2015-2016 period) or by the residual error, which may be related to expert judgment.

Turning to inflation, the NBB projections clearly outperform those of the international institutions at every projection horizon, but even more so for the shorter term. Finally, while the labour market projections are no less accurate than those made by other international organisations, employment growth has recently been underestimated in the NBB projections. This may be related to a stronger-than-expected impact of the structural reforms and the wage cost moderation policies.

Bibliography

Alessi L., E. Ghysels, L. Onorante, R. Peach and S. Potter (2014), "Central Bank Macroeconomic Forecasting During the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences", *Journal of Business & Economic Statistics*, 32(4), October.

Antonio Liedo, D. (2014), *Nowcasting Belgium*, NBB, Working Paper Research 256.

Bank of England (2015), *Evaluating forecast performance*, Independent Evaluation Office. Bank of England, November.

Basselier R., D. de Antonio Liedo and G. Langenus (2018), "Nowcasting real economic activity in the euro area: Assessing the impact of qualitative survey", *Journal of Business Cycle Research*, 145(1), 1-46, April.

Diebold F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-63.

EC (2016), *European Commission's Forecasts Accuracy Revisited: Statistical Properties and Possible Causes of Forecast Errors*, Discussion Paper 27, March.

ECB (2016), *A guide to the Eurosystem/ECB staff macroeconomic projection exercises*, July.

Jeanfils P. and K. Burggraeve (2005), *Noname – A new quarterly model for Belgium*, NBB, Working Paper Research 68, May.

Piette Ch. (2016), *Predicting Belgium's GDP using targeted bridge models*, NBB, Working Paper Research 290, January.