WORKING PAPER 3-12

Track record of the FPB's short-term forecasts

An update

February 2012

Ludovic Dobbelaere, ldo@plan.be

Igor Lebrun, il@plan.be

Avenue des Arts 47-49 – Kunstlaan 47-49
1000 Brussels

E-mail: contact@plan.be
http://www.plan.be

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency.

The FPB performs research on economic, social-economic and environmental policy issues. For that purpose, the FPB gathers and analyses data, examines plausible future scenarios, identifies alternatives, assesses the impact of policy measures and formulates proposals.

The government, the parliament, the social partners and national and international institutions appeal to the FPB's scientific expertise. The FPB provides a large diffusion of its activities. The community is informed on the results of its research activities, which contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic Enterprise label (three stars) for its environmental policy

url: http://www.plan.be
e-mail: contact@plan.be

Publications

Recurrent publications:
Forecasts
Short Term Update

Planning Papers (last publication):
The aim of the Planning Papers is to diffuse the FPB's analysis and research activities.

109 Satellietrekeningen Transport in 2005 / Comptes satellites des transports en 2005
Dries Goffin, Estelle Nayes - April 2011

Working Papers (last publication):

2-12 La loi du 26 juillet 1996 relative à la promotion de l'emploi et à la sauvegarde préventive de la compétitivité - Quelques éléments de réflexion sur la « norme salariale »
Luc Masure - January 2012

Responsible publisher: Henri Bogaert
Legal Deposit: D/2012/7433/3

Federal Planning Bureau
Avenue des Arts 47-49 – Kunstlaan 47-49, 1000 Brussels
phone: +32-2-5077311
fax: +32-2-5077373
e-mail: contact@plan.be
http://www.plan.be

Track record of the FPB's short-term forecasts

An update

February 2012

Ludovic Dobbelaere, ldo@plan.be
Igor Lebrun, il@plan.be

Abstract – The Federal Planning Bureau is responsible, within the National Accounts Institute, for producing the macroeconomic forecasts that are used to set up the federal government budget. This working paper presents an update of the ex post assessment of the quality of these forecasts. Compared to the previous working papers devoted to this topic, the analysis is extended in several ways. Firstly, the number of variables examined is markedly increased, as is the number of statistical tests. Secondly, an evaluation of the quality of the quarterly forecasts is presented for the first time. In addition, this information is used to calculate the probability distribution of these forecasts and to construct a so-called "fan chart".

Jel Classification – C53, E6

Keywords – Economic budget, Forecasts, Post-mortem assessment

Table of contents

List of tables

List of figures

Executive summary

The Federal Planning Bureau is responsible, within the National Accounts Institute, for producing the macroeconomic forecasts that are used to set up the federal government budget. This working paper presents an update of the ex post assessment of the quality of the forecasts published within this framework. Compared to the previous working papers devoted to this topic, the analysis is extended in several ways. Firstly, the analysis on an annual basis now covers the period 1994-2010, while in the preceding study the sample ended in 2005. Secondly, the number of variables examined is markedly increased, as the number of statistical tests checking whether the forecasts fulfil a series of criteria. Thirdly, an evaluation of the quality of the quarterly forecasts for GDP growth and inflation is presented for the first time. In addition, this information is used to calculate the probability distribution of these forecasts and to construct a so-called "fan chart", aimed at illustrating graphically the uncertainty surrounding the central scenario.

Although several hundreds of variables are forecast in an economic budget, only the most relevant ones to set up the government budget were selected for the evaluation on an annual basis. Real growth in GDP and expenditure components were chosen as the most important indicators of economic activity, while growth in the consumer price index and the GDP deflator were singled out as the most representative price variables. Finally, growth in employment and unemployment were considered the most relevant indicators for the labour market. As in previous post-mortem analyses, it was decided to define the outcomes used to assess the quality of the forecasts as the first available estimates of the national accounts. Forecast errors are defined as the difference between forecasts and realisations.

The quality of the annual forecasts is evaluated for three successive forecasting rounds, namely the one-year-ahead forecasts in September and the current-year forecasts in February and September, with the emphasis being put on the first two rounds. The statistical tests performed give the following results. Forecast errors for all variables decline as the horizon shortens, but one-year-ahead forecasts exhibit sizeable mean absolute errors and contain little information on the variation around the sample mean. Errors are particularly large in the case of severe economic downturns or sudden changes in inflation. The positive bias observed in first-round GDP forecasts is nonetheless not statistically significant and disappears completely in February. The vast majority of forecast errors do not feed into the next round. The February forecasts undeniably outperform naïve forecasts and predict growth accelerations and decelerations correctly in most cases. This shows the utility of the budgetary control in order to bring the economic parameters used in the budget closer to their outcome. Lastly, forecast revisions tend to be too smooth although this may be considered a prudent strategy in the context of the use of these forecasts within the elaboration of the federal budget.

The track record of the quarterly forecasts is gauged through an indicator of real economic activity (quarter-on-quarter GDP growth) and a price indicator (year-on-year growth of the CPI). The analysis shows that forecast errors for both indicators gradually rise as the forecast horizon is extended and that large errors are more common at longer forecast horizons. Up to three quarters ahead, neither GDP growth nor inflation forecasts suffer from a statistically significant bias, but the mean forecast error tends to be slightly positive for the former and rather negative for the latter.

The observed distribution of the forecast errors is then used to calculate probability intervals around the forecasts for GDP growth and inflation. Considering past forecast errors as a good measure of uncertainty is a disputable assumption, but it has the merit to provide a transparent and objective method of quantifying. After correction for outliers, these confidence intervals are subsequently used to construct, around the central scenario provided by the economic budget, a so-called fan chart, representing confidence bands painted in different colours.

1. Introduction

Since 1994, the National Accounts Institute (NAI) has been responsible for producing the macroeconomic forecasts that are used to set up the Belgian government budget and to perform budgetary control exercises. Within the NAI, the Federal Planning Bureau (FPB) is in charge of producing these forecasts, which are usually referred to as the "economic budget". The FPB has always been as transparent as possible on its forecasts, implying the availability of a clear description of the hypotheses behind the forecasts, but also of the tools and methods used to produce the forecasts[1]. To enhance transparency further, one also needs to produce ex post assessments of the quality of its forecasts (generally called "post-mortem analyses") at regular time intervals. This need for transparency is explicitly referred to in the new directive on requirements for budgetary frameworks of the Member States included in the so-called "Six-Pack" adopted recently by the European Parliament and the Council. The Directive states in particular that "The macroeconomic and budgetary forecasts for fiscal planning shall be subject to regular, unbiased and comprehensive evaluation based on objective criteria, including ex post evaluation. The result of that evaluation shall be made public and taken into account appropriately in future macroeconomic and budgetary forecasts".

The FPB published post-mortem analyses of the economic budget in 1998, 2004 and 2006[2]. This working paper updates these analyses and extends them in two ways: the spectrum of examined variables is broadened and the number of statistical tests that allow checking whether our forecasts satisfy a series of desirable properties is increased. Moreover, as the FPB started to use a quarterly model[3] to produce the economic budget in 2002, an evaluation of the quality of the quarterly forecasts for GDP growth and inflation is presented here for the first time.

It should be noted that an ex post evaluation of forecast errors should not be considered as the ultimate quality check of the FPB's macroeconomic forecasts. In fact, a post-mortem analysis only examines properties such as the size of the forecast errors regardless of other features of the forecasting exercises such as completeness and consistency, which are just as desirable in the framework of a government budget elaboration. Nonetheless, the aim of this post-mortem analysis is to give external users a broad idea of the uncertainty surrounding the forecasts and also to detect possible methodological weaknesses that need to be addressed internally.

This working paper is structured as follows: the next chapter provides a description of the (annual) data that were used, discussing the choice of the forecasted variables and outcomes. The third chapter describes the results of the tests carried out on the annual forecasts to assess their quality. The fourth chapter presents a more limited post-mortem analysis of the quarterly forecasts for GDP growth and inflation. This information is used in the final chapter to calculate the probability distribution of these forecasts and to present them on a so-called "fan chart", as popularised by the Bank of England.

1 See Dobbelaere et al. (2003) and FPB (2006).
2 See respectively FPB (1998), Dobbelaere and Hertveldt (2004) and Bogaert et al. (2006).
3 See Hertveldt and Lebrun (2003).

2. Data description

The first section of this chapter discusses the variables that will be evaluated. The second section provides useful information on the economic budget, while the third one deals with the properties of the outcomes. Forecast errors are calculated as the difference between forecasts and realisations, implying that positive errors are associated with overestimations and negative errors to underestimations.

2.1. Variables

Although several hundreds of variables are forecasted in an economic budget, only the most relevant ones to set up the government budget will be considered in this evaluation:

- *Real growth in GDP and expenditure components* were selected as the most important indicators of economic activity. Even if not all expenditure components are equally crucial for the government budget, they are all examined because of their importance to determine GDP in the FPB's demand-driven forecasting models.
- *Growth in the consumer price index (CPI) and the GDP deflator* are the price variables that will be covered in this analysis. The CPI plays a key role in the Belgian economy that is characterised by a generalised automatic wage indexation[4], while the GDP deflator is almost as important as real GDP growth to estimate tax revenue.
- *Growth in employment and unemployment* are the indicators that will be used to assess the quality of labour market forecasts. These variables play a central role in estimating social benefits, social contributions and personal income tax revenue.

2.2. Outcomes

Determining the outcomes for the CPI and the number of unemployed persons is straightforward as they are quickly available and not subject to revisions. For all other variables the outcomes have to be taken from the national accounts that can be revised substantially over time. Consequently, the choice of the outcomes is not without importance for the results of the post-mortem analysis as it can influence the size and the properties of the forecast errors. Although one could argue that economic fluctuations are better captured in revised versions of the national accounts, it was decided to compare each forecast with its earliest estimate in the national accounts for the following reasons:

- Revisions of the national accounts are often related to methodological adjustments, which are impossible to forecast.
- The considered variables in this analysis have been forecast with the FPB's quarterly macroeconometric model since 2002, thereby integrating all available data from the quarterly

[4] Although wages and social benefits are indexed on the basis of the health index (CPI corrected for the price evolution of motor fuels, alcoholic beverages and tobacco products), the CPI is considered here to ensure international comparability.

national accounts. This implies that the economic budget can be considered as a forecast of the first estimates of the national accounts.

Since the start of the quarterly national accounts in 1998, the first estimate for the year t has been published in April of the year t+1. In earlier years, the first estimate was available in June of the year t+1.

2.3. Forecasts

The FPB publishes two economic budgets per year: one in September of the year t-1 that is used to set up the budget for the year t and one in February of the year t that serves as an input for the budgetary control of the year t. The economic budget has been finalised in September since 2002, while it was available by the beginning of July before. Since 2002, the FPB has continued to produce a preliminary forecasting exercise in June of the year t-1 in order to allow other institutions to start their budgetary exercise for the year t, enabling them to prepare the production of the final figures in September. As these preliminary exercises are not published, they will not be considered in this post-mortem analysis. In what follows, “September forecasts” should be understood as a mix of forecasts that were finalised in July (up to 2001) and September (from 2002 on).

Table 1 Publication data of forecasts and annual national accounts

month \ year	t-1	t	t+1
September	forecast t-1 (3rd round) forecast t (1st round)		
February		estimate t-1 (4th round) forecast t (2nd round)	
April		outcome t-1	
September		forecast t (3rd round) forecast t+1 (1st round)	
February			estimate t (4th round) forecast t+1
April			outcome t

Table 1 shows that four forecasts/estimates are published in subsequent economic budgets before the first outcome is available. As forecast errors of the fourth round estimate are generally very small, only the September forecasts for the year ahead (first round) and the February and September forecasts for the current year (second and third round respectively) will be discussed. As forecasts are available from 1994 onwards and observations up to 2010, only 17 data points per round of forecasts can be calculated. Due to this limited sample, statistical test results should be interpreted with caution.

2.4. Graphical inspection of forecasts and outcomes

Before applying formal tests to the forecast errors in chapter 3, this section presents first and second round forecasts and outcomes of all variables in a number of graphs. While the tests generally translate the properties of the forecast errors into a single figure, this section provides a time perspective of forecast errors that allows us for example to identify outliers.

Graph 1 Forecasts and outcomes of GDP and expenditure components
Real growth rates in percent

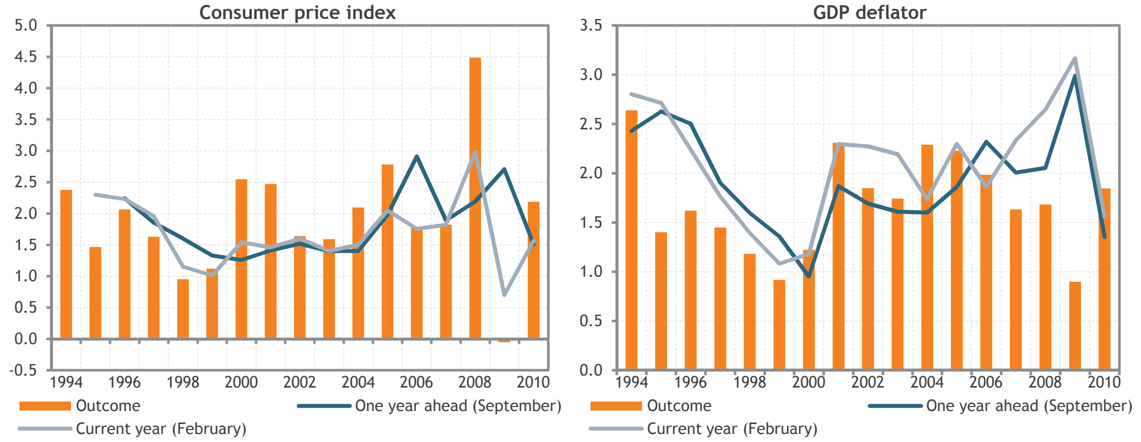
Source: NAI, FPB

Graph 1 allows identifying some properties of the forecasts for GDP and the main expenditure components. Firstly, current year forecasts in February are generally much closer to the outcome than the year-ahead forecasts produced in September. This should not come as a surprise as business cycle related information is far from abundant one year ahead. Moreover, it shows the utility of the budgetary control in order to bring the economic parameters used in the budget closer to their outcome. Secondly, forecasts are typically less volatile than outcomes, implying the appearance of negative forecast errors during growth accelerations and positive ones during growth decelerations.

Consequently, forecast errors become much larger when business cycle fluctuations are more pronounced. This was notably the case during the recessions of 2001-2002 and 2008-2009. Thirdly, forecast errors tend to be significantly higher for volatile variables. While forecast errors for private consumption are only rarely bigger than 1 percentage point, those for exports and imports can easily amount to 5 percentage points in years characterised by a pronounced upturn or downturn. Finally, there are no clear signs of consistent over- or under-predictions, except for exports and imports where over-predictions prevail, especially during the last decade. Fortunately these forecast errors cancel each other out when GDP growth is calculated.

Variables related to public expenditures should be interpreted with caution as their forecasts are traditionally based on the unchanged policy assumption. This is motivated by the fact that the economic budget is the basis for the calculations that should provide policy makers an image of public finances in the absence of additional measures. Even if this approach does not improve forecast accuracy, it is a desirable property of the economic budget.

Graph 2 Forecasts and outcomes of price variables
Growth rates in percent



Source: NAI, FPS Economy, FPB

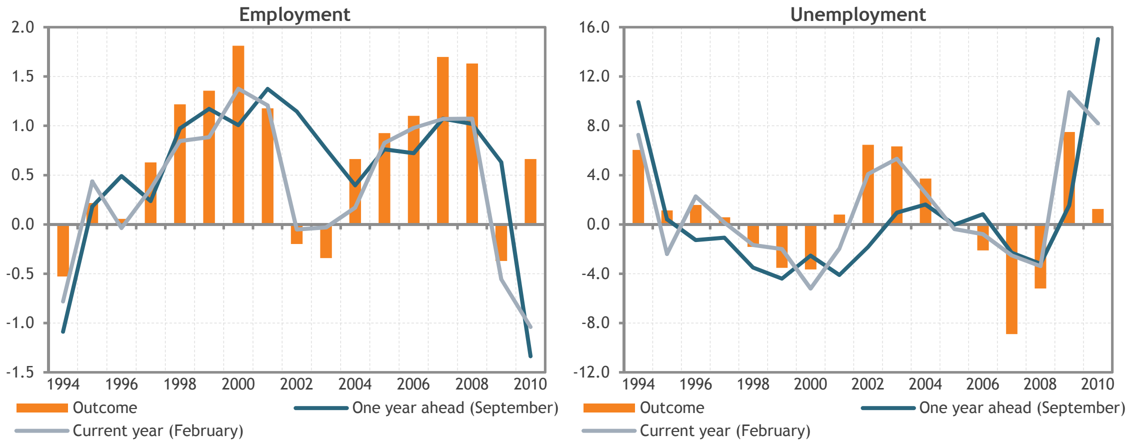
Although the CPI and the GDP deflator generally exhibit the same growth profile, the CPI is forecasted relatively well (especially in the February exercise), while forecasts for the GDP deflator are less satisfying, mainly between 2007 and 2009. During that period, the GDP deflator was consistently over-estimated and instead of reducing the forecast error in February as compared to September, a larger discrepancy between forecasts and outcomes was seen. During the second half of the nineties, over-predictions prevailed for the CPI as well as the GDP deflator. The large forecast errors for the CPI in 2008 and 2009 are related to extreme energy price developments.[5]

When assessing forecasts for the number of employed and unemployed persons, it should be noted that the magnitude of the growth rates of both variables cannot be compared as 1% of employment represented 44 900 persons in 2010, while 1% of unemployment only amounted to 6 500 persons. Consequently, forecast errors expressed in percentage points will be larger for unemployment than for employment.

5 The Brent oil price expressed in USD increased by 34% in 2008 and declined by 36% in 2009. In euro, price developments were somewhat less pronounced with a rise of 25% in 2008 and a fall of 33% in 2009.

Overall, over-predictions of employment went hand in hand with under-predictions of unemployment indicating that forecast errors for the labour force were generally smaller than those for employment. As was also the case for GDP, forecast errors for (un)employment were particularly big during the recessions at the beginning and the end of the last decade. However, as employment typically reacts with a lag to developments in GDP, large forecast errors generally show up one year later for employment than for GDP. Another striking fact is that the upturn in employment in 2010, unlike that in GDP, was completely missed in both the September and the February forecast, indicating an erroneous assessment of productivity after the initial phase of the financial crisis.

Graph 3 Forecasts and outcomes of labour market variables
Growth rates in percent



Source: NAI, RVA/ONEM, FPB

3. Quality of forecasts

The quality of the FPB forecasts will be evaluated on the basis of a number of criteria. The formulas that are used to calculate the indicators as well as the formalisation of the statistical tests applied in this section can be found in Annex 1. Note that the methods presented below are those traditionally proposed in the literature[6] and used by national institutions like the CPB in the Netherlands[7] or international organisations such as the OECD, the EC and the IMF[8] to assess the accuracy of their own forecasts.

3.1. Size of forecast errors

Two aspects of the size of the forecast errors will be evaluated. They should be small and they should decline as the forecast horizon shortens. The most intuitive indicator to measure forecast errors is the mean absolute error (MAE), which provides the average (positive or negative) deviation of the forecast from the outcome in percentage points. As could be expected, this measure is the highest for the most volatile variables such as exports, imports and gross fixed capital formation. The MAE declines for all considered variables as more information can be integrated in the forecasts. The decline in the error between the September (round 1) and the February (round 2) forecast is most pronounced for broad aggregates such as GDP and employment, which does not come as a surprise as these variables are generally less subject to revisions in the quarterly national accounts. The decline of the MAE going from round 2 to round 3 is most obvious for variables that are available on a monthly basis (i.e. unemployment and CPI) as almost three quarters of the year are observed in September.

Table 2 Size of forecast errors

	Mean absolute error			Root mean square error			Corrected RMSE		
	round 1	round 2	round 3	round 1	round 2	round 3	round 1	round 2	round 3
GDP	1.29	0.76	0.40	1.60	0.86	0.53	1.06	0.57	0.35
Pm: EC GDP forecast	*1.38*	*1.12*	*0.68*	*1.75*	*1.34*	*0.81*	*1.16*	*0.89*	*0.53*
Private consumption	0.89	0.77	0.33	1.09	0.86	0.44	0.96	0.76	0.39
Public consumption	0.76	0.69	0.58	0.92	0.85	0.78	1.37	1.27	1.16
Gross fixed capital formation	2.67	2.15	1.53	3.29	2.61	1.95	0.95	0.75	0.56
Exports	4.04	2.95	1.88	5.60	3.73	2.27	1.08	0.72	0.44
Imports	3.43	2.61	1.51	5.23	3.65	1.99	1.05	0.73	0.40
CPI	0.82	0.51	0.12	1.13	0.67	0.16	1.23	0.72	0.17
Pm: EC CPI forecast	*0.92*	*0.81*	*0.41*	*1.16*	*1.08*	*0.54*	*1.25*	*1.16*	*0.58*
GDP deflator	0.55	0.51	0.36	0.72	0.75	0.45	1.50	1.57	0.94
Employment	0.61	0.38	0.27	0.79	0.53	0.36	1.05	0.71	0.48
Unemployment	3.81	2.15	0.86	5.08	2.86	1.03	1.16	0.65	0.24

Note: Statistics computed based on the Economic Forecasts of the European Commission for GDP and CPI are mentioned in italic.

6 See for example Carnot et. al. (2005).
7 See Kranendonck et al (2009).
8 See respectively Vogel (2007), Melander et al. (2007) and Timmerman (2006).

Another frequently used indicator is the root mean square error (RMSE) that penalises large errors more than the MAE. In fact, the larger the standard deviation of the forecast errors, the more the RMSE will be above the MAE; it will only be equal to the MAE if all forecast errors in the sample have the same magnitude. The RMSE generally declines more than the MAE between the first and the second forecasting round, indicating not only that forecast errors are smaller in February than in September, but also that errors significantly above the MAE are less common in February.

To correct for the fact that volatile series are generally accompanied by large forecast errors, a corrected RMSE was calculated by dividing the RMSE by the standard deviation of the series. This brings the forecast errors much closer to each other in all forecasting rounds. Only forecasts for public consumption and the GDP deflator seem to perform much worse than the other ones. Another property of this indicator is that it will be bigger than one if the sample average of the variable would have produced forecasts that are more accurate on average than the FPB forecasts[9]. The corrected RMSE is close to or above one for all variables in the first forecasting round, indicating that these forecasts do not provide more accurate information than a forecast based on average growth rates. This will be discussed more extensively in paragraph 3.5. From the second forecasting round onwards, FPB forecasts perform much better than average forecasts, except for public consumption and the GDP deflator.

To put the size of the forecast errors of the economic budget into perspective, the statistics in Table 2 were also calculated for Belgian GDP and CPI growth as forecasted by the European Commission. The results for round 1 cover the forecasts for year t published in spring of the year t-1, those for round 2 the forecasts for year t released in autumn of the year t-1 and round 3 relates to the spring forecasts for year t issued that same year. This timing implies that each European Commission forecasting round is released a few months ahead of the corresponding round of the economic budget. The statistics show that the economic budget of September improves slightly the quality of the Commission spring forecasts for GDP growth and inflation and that the economic budget of February upgrades the autumn forecasts markedly. This demonstrates the utility of the economic budget, keeping in mind that in addition it provides a more detailed forecast for the Belgian economy.

3.2. Absence of bias

Irrespective of the size of the errors, forecasts should be unbiased. In other words, they should not be systematically too optimistic or too pessimistic, implying that positive and negative forecast errors should offset out each other on average. The presence of a bias suggests that forecast accuracy can be improved using this information.

A first indication can be provided by the percentage of positive forecast errors, which should be close to 50% in the case of unbiased forecasts. Table 3 shows that this is the case for most first and second round forecasts. However, under-predictions seem to prevail for public consumption and employment, while over-predictions are more common in second round forecasts for the GDP deflator.

---

[9] Note that in practice it is impossible to use the sample average as a forecast because it is not known at the time the forecast is produced.

Table 3 Unbiasedness of forecasts

	Percentage of positive forecast errors			Mean error		
	round 1	round 2	round 3	round 1	round 2	round 3
GDP	53	47	29	0.32	-0.05	-0.12
Private consumption	59	53	35	0.12	-0.05	-0.04
Public consumption	35	35	41	-0.34	-0.42*	-0.27
Gross fixed capital formation	47	53	71	0.52	0.11	0.17
Exports	65	65	53	1.40	0.69	-0.76*
Imports	59	59	35	1.09	0.65	-0.74*
CPI	47	31	50	-0.13	-0.22	0.00
GDP deflator	53	71	76	0.23	0.39**	0.20**
Employment	29	24	12	-0.13	-0.29**	-0.12
Unemployment	35	53	59	-0.25	0.59	0.17

Note: * and ** denote that the mean error is significantly different from zero at the 10% and the 5% level respectively. Significance levels were calculated on the basis of standard errors robust to autocorrelation.

To test whether the bias is statistically significant, forecast errors were regressed on a constant. The estimated constant of this equation is equal to the mean error (ME) that is positive for GDP and most expenditure components in the year-ahead forecasts, although not significantly different from zero. It should be noted that the positive ME for GDP is mainly related to the positive outlier in the forecast errors in 2009 that was only partially compensated for by the negative error in 2010. If the last two years are left out of the sample, the ME is reduced for all expenditure categories (except for public consumption) and falls to 0.19 for GDP growth. The influence of this outlier disappears in the second and third forecasting round, bringing the ME generally closer to zero and even slightly into negative territory for GDP and most expenditure components. Among the expenditure categories, only public consumption, exports and imports suffer from a statistically significant bias in one forecasting round. For the two last ones, the significance of the bias mainly stems from the surprisingly vigorous upturn in world trade after the initial phase of the financial crisis.

CPI forecast errors register a small negative bias that is not statistically significant, but the GDP deflator suffers from a substantial positive bias that is significantly different from zero in the second and third forecasting round.

The ME for both employment and unemployment first round forecasts is negative, but not statistically significant, pointing to a small under-prediction of the labour force on average. Current-year forecasts for employment also register a negative bias that is even significantly different from zero in the second forecasting round. This is mirrored in second and third round forecasts for unemployment that were somewhat too high on average.

3.3. No persistence in forecast errors

Next to the significance of the forecast bias, the equations estimated in paragraph 3.2 also allow us to check for serial correlation in forecast errors. The presence of serial correlation implies that past forecast errors could be used to improve the forecast in year t, which is in contradiction with the weak efficiency assumption (see next paragraph). First and second-order correlation coefficients are reported in Table 4.

Table 4 Serial correlation of forecast errors

	1st order serial correlation			2nd order serial correlation		
	round 1	round 2	round 3	round 1	round 2	round 3
GDP	-0.17	-0.24	-0.09	-0.20	-0.23	-0.27
Private consumption	0.02	-0.09	0.41*	-0.21	-0.31	-0.10
Public consumption	0.51**	0.17	0.19	0.17*	0.03	0.05
Gross fixed capital formation	0.18	0.46**	0.35	-0.13	0.07	0.03
Exports	-0.28	-0.24	-0.05	-0.13	-0.11	-0.43
Imports	-0.29	-0.30	-0.04	-0.13	-0.10	-0.40
CPI	-0.38	-0.08	-0.09	-0.14	0.07	-0.07
GDP deflator	0.11	0.08	-0.11	0.11	0.03	-0.23
Employment	-0.04	0.08	-0.07	-0.06	0.09	-0.05
Unemployment	0.07	0.39*	-0.13	0.04	0.27*	0.07

Note: * and ** denote that the null hypothesis of no serial correlation is rejected at the 10% or the 5% level respectively. The hypothesis is tested on the basis of the Ljung-Box Q-statistic.

Only few forecast errors suffer from serial correlation. During the first two forecasting rounds, statistically significant positive first order autocorrelation is only seen in the forecast errors of investment, public consumption and unemployment, the last two also suffering from positive second order serial correlation. The serial correlation found in third round forecast errors for private consumption should not necessarily be a source of concern as it concerns relatively small errors.

3.4. Efficiency tests

Forecasts are said to be efficient if all information available at the time of the forecast was used optimally. Because the annual estimation in round 3 partly relies on data available for the first half of the year and is thus not entirely a forecast we did not include it in this analysis.

3.4.1. Weak efficiency

Weak efficiency assumes that the forecasts are unbiased (see paragraph 3.2) and the forecast errors uncorrelated with past values of the forecasts or with the errors themselves (see paragraph 3.3). One test suggested in the literature consists in running a regression with the outcome as dependent variable on a constant and the forecast. Weak efficiency requires the constant to be equal to zero, the coefficient of the forecast to be equal to one and the residuals to be white-noise. In this specification, uninformative forecasts have values for their coefficient close to zero while negative values point even to a misleading projection. A non-zero value for the constant should be incorporated directly into the forecast, as well as the non-random pattern of the residuals in case of serial correlation.

The values of the two coefficients are given in Table 5. Formally, two statistical tests need to be performed to examine the assumption of weak efficiency, see Annex 1 for the details.

Table 5 Results of the weak efficiency test
(estimated constant ; estimated coefficient)

	Round 1	Round 2
GDP	(1.15 ; 0.31)	(-0.21 ; 1.15)
Private consumption	(0.36 ; 0.72)	(-0.01 ; 1.04)
Public consumption	(1.32 ; 0.29)**	(1.28 ; 0.34)**
Gross fixed capital formation	(0.33 ; 0.72)	(-0.01 ; 0.96)
Exports	(5.81 ; -0.49)**	(-2.74 ; 1.50)
Imports	(3.67 ; -0.02)*	(-3.60 ; 1.70)
CPI	(2.27 ; -0.18)**	(-0.40 ; 1.37)
GDP deflator	(1.63 ; 0.04)**	(1.25 ; 0.21)**
Employment	(0.43 ; 0.46)*	(0.35 ; 0.86)*
Unemployment	(0.47 ; 0.36)**	(-0.35 ; 0.80)*

Note: * and ** denote that the null hypothesis of weak efficiency (constant = 0, coefficient = 1 and white-noise residuals) is rejected at the 10% or the 5% level respectively. Significance levels were calculated on the basis of standard errors robust to autocorrelation.

In round 1 only GDP[10], private consumption and GFCF appear to be efficient[11]. This is not surprising as we saw in Table 2 that only for private consumption and GFCF the forecast outperforms the mean of realisations in round 1. Note that for public consumption, employment and unemployment the coefficient of the forecast is significantly different from zero, meaning that though not efficient, the forecasts still provide some useful information. The picture for round 2 is rosier, only for the variables which were previously shown to be biased (public consumption, GDP deflator, employment) or exhibiting serial correlation (unemployment), the property of weak efficiency is rejected.

3.4.2. Informational efficiency

In principle, the forecast should include all publicly available information at the time it is produced. If this is true, forecasts errors cannot be predicted using this information. Given the potentially open-ended information set, we will restrict our analysis to two crucial variables. As a small open economy, forecasts for Belgium depend heavily on assumptions regarding the evolution of world trade and world prices. These assumptions rely mainly on forecasts produced twice a year by the European Commission and on developments in consensus forecasts and market expectations ("futures"). These forecasts may be summarised in two key exogenous variables for Belgium, namely growth of potential export markets and growth of oil prices in euro. To test whether this information was completely taken into account in our forecast we run a regression with the forecast errors of the GDP or CPI growth rate depending on the forecasts of these two international variables. If all the information contained in these two indicators was incorporated, we expect both coefficients not to be significantly different from zero. Table 6 provides the value of these two coefficients for round 1 and 2.

[10] Although the estimated constant and coefficient for GDP are clearly different from respectively zero and one, statistically they are not at the 10% level.

[11] As a matter of comparison, the one-year-ahead GDP forecasts for the G7 by the OECD fail to pass the weak efficiency test (see Vogel, op. cit.).

Table 6 Results of the informational efficiency test
(coefficient potential export markets ; coefficient oil prices in euro)

	Round 1	Round 2
GDP	(0.30 ; -0.00)	(0.01 ; -0.02)
CPI	(0.03 ; -0.02)	(-0.01 ; -0.05)

Note: * and ** denote that the null hypothesis of informational efficiency (both coefficients = 0) is rejected at the 10% or the 5% level respectively.

No coefficient appears individually or jointly significantly different from zero, meaning that the forecasts for GDP and CPI in the economic budget encompass all the information contained in the world trade and oil price indicators.

3.5. Comparison with naïve forecasts

Another way to evaluate the forecasts is to compare them with those produced by alternative methods. In particular one wants to make sure that model-based forecasts using all the information available, like those contained in the economic budget, outperform those formed with naïve methods. As we already discussed in paragraph 3.1, using the sample mean of realisations as a forecasting rule is not a viable option because the sample mean is not known at the time of the forecast. At best, the forecaster could use the mean of the outcomes up to the day of the forecast. Therefore we use the average growth of the fifteen years prior to the forecasting exercise as a naïve alternative for extrapolation. Another rule-of-thumb method consists in simply taking the growth rate of the observation for the last available year[12]. Comparison with naïve forecasts usually relies on the Theil coefficient which is computed as the ratio between the RMSE of the FPB forecast and the RMSE based on the naïve approach. A Theil coefficient below unity is expected.

Table 7 provides the results based on the last year observed (Theil 1) and on the fifteen-year moving average (Theil 2).

Table 7 Theil coefficients

	Theil 1			Theil 2		
	round 1	round 2	round 3	round 1	round 2	round 3
GDP	0.74	0.40	0.24	1.00	0.54	0.33
Private consumption	0.65	0.51	0.28	0.94	0.74	0.38
Public consumption	0.78	0.73	0.95	1.24	1.16	1.08
Gross fixed capital formation	0.63	0.50	0.45	0.85	0.68	0.52
Exports	0.81	0.54	0.27	1.03	0.69	0.40
Imports	0.79	0.55	0.24	1.02	0.71	0.37
CPI	0.85	0.50	0.10	0.79	0.43	0.11
GDP deflator	0.98	1.03	0.81	0.54	0.57	0.39
Employment	0.72	0.49	0.44	1.01	0.68	0.48
Unemployment	0.70	0.40	0.20	0.99	0.56	0.21

The forecasts are more accurate than the ones using the latest observed growth rate in all cases except for the GDP-deflator in round 2. The Theil coefficient is also declining as the forecast horizon shortens

[12] This means that for round 1, the realization of the year t-1 is used to predict t+1, for round 2 the same outcome is employed while for round 3, the outcome of the previous year can be utilized.

for all variables excepting the GDP-deflator in round 2 and public consumption in round 3. Using a moving average as a benchmark provides less convincing results in round 1 where a majority of the coefficients are close to unity. As already mentioned, relatively few information related to the business cycle is available in September for the upcoming year and under those circumstances it is extremely difficult to produce a better forecast than the expected trend growth. The Theil coefficient decreases markedly during round 2, once more business cycle indicators become accessible. Logically, it is further reduced in round 3. Only in the case of public consumption does it remain above unity in all three rounds.

3.6. Forecast revisions

One further important requirement is that forecast revisions should not be predictable. Indeed, if these revisions exhibit a particular pattern, this information should be taken on board in the preceding forecasting round. Moreover, these revisions should be independent from each other. To test the first property the revisions are regressed onto a constant which should not be significantly different from zero. In our dataset two revisions are available for a given year: the revision of the forecast in February (round 2 – round 1) and the revision in September (round 3 – round 2). To examine the second property, the September revisions are regressed onto those made in February and the coefficient is expected not to be statistically different from zero. The values of the regression coefficient for both tests are provided in Table 8.

Table 8 Unbiasedness and independence of forecast revisions

	Unbiasedness of revisions		Independence
	round 2	round 3	
GDP	-0.37*	-0.07	0.39**
Private consumption	-0.17	0.01	0.11
Public consumption	-0.08	0.15**	-0.07
Gross fixed capital formation	-0.41	0.06	0.13
Exports	-0.71	-1.45**	0.99**
Imports	-0.44	-1.39**	1.10**
CPI	-0.17	0.22	0.60**
GDP deflator	0.17	-0.19	0.38
Employment	-0.16	0.17*	0.43**
Unemployment	0.83	-0.42	0.20

Note: * and ** denote that the null hypothesis of unbiasedness or independence is rejected at the 10% or the 5% level respectively. Significance levels were calculated on the basis of standard errors robust to autocorrelation.

Revisions in round 2 exhibit a downward trend (except for unemployment) but it is only significantly different from zero in the case of GDP. The latter is not surprising because, as already shown in Table 3, the positive mean forecast error in round 1 disappears in the next round. The revisions in round 3 are significantly systematic for public consumption, employment and external trade. But we also know from Table 3 that the negative bias for public consumption and employment is reduced in round 3 while on the contrary a negative bias shows up for exports and imports. The last column tells us that with positive coefficients (with the exception of public consumption) revisions are smoothed across the three rounds. These coefficients are significantly different from zero in half of the cases confirming the findings made on a large sample of countries that real GDP growth forecasts are generally revised in a very smooth way (see Loungani and Rodriguez, 2008).

3.7. Directional accuracy

Next to the size and the properties of the errors, the capacity of forecasts to predict growth pick-ups and slowdowns is also a desirable feature. In Table 9 we report the percentage of times that an acceleration or deceleration in growth was predicted correctly. Note that for round 1 and 2 the growth rate for the preceding year is still a forecast while for round 3 the previous year is the outcome as defined in paragraph 2.2.

In round 1, the acceleration or deceleration in GDP growth was correctly predicted in 75% of the cases. The other variables obtain a comparable score with exports on the lower end and employment on the high side. As expected this percentage increases in round 2 except for public consumption and employment. In round 3 the percentages climb well above 80% (they even reach 100% for GDP) with the exception of public consumption and the GDP deflator.

Table 9 Proportion of correctly predicted growth accelerations and decelerations
In % of total forecasts

	Round 1	Round 2	Round 3
GDP	75	94	100
Private consumption	69	71	82
Public consumption	69	65	76
Gross fixed capital formation	75	88	88
Exports	63	88	82
Imports	81	82	82
CPI	73	88	94
GDP deflator	69	71	76
Employment	88	82	88
Unemployment	75	88	88

3.8. Conclusions

To sum up, the forecasts of the variables considered in this analysis display a number of desired properties. The forecast errors for all variables decline as the horizon shortens. The positive bias observed in the forecasts for GDP and most expenditure categories in the first round is not statistically significant and disappears in the following round. CPI forecasts exhibit in both rounds a small negative bias that is not statistically significant. The vast majority of the forecast errors are exempt from serial correlation. GDP and CPI growth forecasts encompass the information contained in world trade and oil price indicators. Second-round forecasts undeniably outperform naïve forecasts and predict growth accelerations and decelerations correctly in most cases.

Our analysis also shows some weaknesses. Although very much comparable to those of international institutions, one-year-ahead forecasts exhibit sizeable mean absolute errors and contain little information on the variation around the sample mean. Errors are particularly large in the case of severe economic downturns or sudden changes in inflation. Forecasts revisions tend to be too smooth although this may be considered a prudent strategy in the context of the use of these forecasts within the elaboration of the federal budget. In terms of desirable properties, the forecasts for public consumption and the GDP deflator clearly fall behind. What the exact reasons are behind these two poor individual performances should be the subject of further research.

It is beyond the scope of this working paper to make an exhaustive comparison of the performance of the forecasts contained in the economic budget with those produced for Belgium by, for instance, international organizations[13]. However, an examination of the size of the European Commission forecast errors for Belgian GDP growth and inflation reveals much resemblance with the size of the forecast errors reported for the economic budget.

13 Note that such an investigation is presented in a previous post-mortem analysis (see Dobbelaere and Hertveldt, 2004).

4. Track record of quarterly forecasts

The National Accounts Institute has published national accounts on a quarterly basis for the Belgian economy since 1998. The FPB developed a quarterly macroeconomic forecasting model that has been used since 2002 to set up the economic budget. Producing quarterly forecasts offers the advantage that all available monthly and quarterly information can be explicitly taken into account. Moreover, business cycle developments, which are key in short-term forecasting, are better captured with quarterly than with annual data. In practice, forecasts for the annual aggregates considered in this paper are set up on a quarterly basis.

Even if quarterly forecasts are indirectly assessed in the analysis presented in the previous chapters, they were never evaluated explicitly. This chapter presents a tentative assessment of quarterly GDP growth and inflation forecasts.[14] In the first part of this chapter, the publication calendar of the quarterly national accounts is briefly presented to determine the outcomes. The magnitude of the errors and the presence of a bias are assessed in the second part. The post mortem analysis will not be detailed further as the number of data points per forecasting horizon is still fairly limited, which could lead to a lack of robustness of the statistical test results.

4.1. Forecasts and outcomes

As the FPB's quarterly forecasts are released together with the annual forecast, their publication calendar can be found in paragraph 2.2. The calendar of the quarterly national accounts releases has varied over time. Up to 2003, the flash estimate for GDP growth was published two months after the end of the quarter and the first complete estimate of the components of GDP (value added, expenditures and income categories) was released three and a half months after the end of the quarter. From 2004 onwards, the release of both the flash estimate and the first complete estimate has been accelerated to respectively 30 and 70 days after the end of the quarter. The timing is generally postponed by a few weeks for the publication of the fourth quarter, allowing for enhanced robustness of the first estimate of the annual figure.

As for the annual series, the first available figures will be considered as the outcome. A strict interpretation of this rule would mean that the flash estimate has to be used as the outturn for GDP growth. Although the flash estimate generally gives a good indication of evolution of economic activity, it is based on incomplete information and it does not provide the composition of GDP. For these reasons, it was decided to consider the first complete estimate as the outcome for GDP growth. This also offers the advantage that a consistent set of outcomes can be used if the quarterly post-mortem analysis is extended to other variables (e.g. expenditure components of GDP) in the future. Defining the outcomes for the CPI is straightforward as they are quickly available and not subject to revisions. In this evaluation, outcomes up to 2011Q3 were taken into account.

[14] GDP growth is measured on the basis of quarter-on-quarter growth rates, while inflation is calculated as the year-on-year growth rate of the national index of consumer prices.

As a rule of thumb, the quarter in which the economic budget is published is considered as the first quarter to be forecasted.[15] This implies that six and four quarters have to be forecasted in the September and the February exercises respectively. Nineteen economic budgets are considered in the evaluation of the quarterly forecasts (from February 2002 to February 2011).

Table 10 Basic characteristics of forecast errors

	GDP growth		Inflation	
	Number of observations	Standard deviation	Number of observations	Standard deviation
current quarter	19	0.34	19	0.16
1 quarter ahead	19	0.43	19	0.57
2 quarters ahead	19	0.53	19	0.94
3 quarters ahead	18	0.59	18	1.09
4 quarters ahead	9	0.27	9	1.63
5 quarters ahead	8	0.72	8	1.35

Table 10 provides an overview of the sample size and the standard deviation of the forecast errors[16] per forecast horizon. The sample size is rather limited at all forecasting horizons, but extremely small for the 4- and 5-quarters-ahead forecasts as these are only available in the September economic budgets. Consequently test results at these forecast horizons risk lacking robustness. This can already be seen in the evolution of the standard deviation of the forecast errors: it creeps gradually up as the forecast horizon is extended up to three quarters, but it exhibits some strange fluctuations for the longest horizons. It should also be noted that forecast errors are strongly influenced by the recession related to the financial crisis that led to the deepest plunge in economic activity since World War II. The preceding economic budgets failed to forecast this evolution. If outliers are removed from the sample, the standard deviation of GDP growth forecast errors declines by more than 30% on average. Finally, it should not come as a surprise that the standard deviation of inflation forecast errors is generally bigger than that of GDP growth errors as it matches differences in standard deviation of the series of observations, which is related to the fact that inflation is measured here as year-on-year growth rates while GDP growth is expressed in terms of quarter-on-quarter growth rates.

4.2. Quality of forecasts

Due to the limited sample and the small number of business cycles that is covered by the FPB's quarterly forecasts, only the size of the forecast errors and the absence of a bias in the forecasts will be evaluated. The mathematical formulas can be found in Annex 1.

4.2.1. Size of forecast errors

Forecast errors for both GDP growth and inflation gradually rise as the forecast horizon is extended. Again, the RMSE increases faster than the MAE, indicating not only that forecast errors become bigger, but also that large errors are more common at longer forecast horizons. Errors for current quarter

15 Although this hypothesis does not hold for the February economic budgets up to 2004, in which the previous quarter (i.e. the fourth quarter of the previous year) also had to be forecasted, mixing previous with current quarter estimates was avoided in this analysis.

16 Note that the standard deviation of the forecast error will differ from the root mean square error if the mean error is different from zero (see Annex 1 for the respective definitions).

estimates are generally smaller for inflation than for GDP growth, which is due to the fact that partial information (1 or 2 months of the current quarter) is generally available for inflation at the time of finalising the forecasting exercise, while this is not the case for GDP. Forecast errors for future quarters are higher for (year-on-year) inflation than for (quarter-on-quarter) GDP growth, which is related to the differences in variability of the outcomes. This can be seen from the RMSE corrected for the standard deviation of the outcomes that is even lower for inflation than for GDP growth up to 3-quarters-ahead forecasts. The corrected RMSE for GDP growth is already close to unity at a forecast horizon of two quarters, indicating that the FPB's quarterly forecasts do not perform significantly better than the sample average or put differently, that forecasts fail to capture much of the variability in the outcomes.

Table 11 Size of quarterly forecast errors

	GDP growth			Inflation		
	MAE	RMSE	Corrected RMSE	MAE	RMSE	Corrected RMSE
current quarter	0.24	0.34	0.61	0.11	0.17	0.13
1 quarter ahead	0.28	0.44	0.79	0.44	0.61	0.47
2 quarters ahead	0.36	0.54	0.97	0.78	0.98	0.76
3 quarters ahead	0.37	0.61	1.10	0.94	1.16	0.89
4 quarters ahead	0.22	0.29	0.52	1.39	1.70	1.31
5 quarters ahead	0.35	0.78	1.40	1.19	1.38	1.07

As was already mentioned, the summary statistics for GDP growth are strongly influenced by a few outliers related to the financial crisis. When correcting for these outliers, the MAE for GDP growth declines by 0.1 percentage point on average. The size of the forecast errors also starts to fluctuate quite heavily at a forecast horizon of 4 or 5 quarters. Considering the small sample size, this should not be too much of a concern.

4.2.2. Absence of bias

GDP growth as well as inflation forecasts do not suffer from a statistically significant bias. Only at the 4-quarter forecasting horizon, a statistically significant bias is measured. Again, this should not be considered as an important conclusion as its robustness is not assured due to the very limited sample size.

Table 12 Unbiasedness of quarterly forecasts

	GDP growth		Inflation	
	Percentage of positive forecast errors	Mean error	Percentage of positive forecast errors	Mean error
current quarter	37	-0.03	39	-0.06
1 quarter ahead	42	0.05	37	-0.20
2 quarters ahead	47	0.11	26	-0.27
3 quarters ahead	44	0.14	28	-0.40
4 quarters ahead	44	0.10*	33	-0.48*
5 quarters ahead	63	0.28	38	-0.28

Note: * denotes that the mean error is significantly different from zero at the 10% level. Significance levels were calculated on the basis of standard errors robust to autocorrelation.

In line with the conclusions for the annual forecast errors, the bias measured on the basis of the mean error – even if it is not significantly different from zero – tends to be slightly positive for GDP growth forecasts and rather negative for inflation forecasts. The combination of a percentage of positive forecast errors below 50 and a positive mean error for GDP forecasts reveals that the positive mean error is mainly the result of the effect of some positive outliers. This “contradiction” does not show up for inflation forecasts although the magnitude of the mean error is also influenced by some large negative outliers.

5. Probability distribution of quarterly forecasts

The Bank of England (BoE) has a long-standing tradition of reporting its inflation forecasts as a probability distribution. To enhance the awareness with respect to the uncertainty surrounding economic forecasts, presenting forecasts in the form of fan charts has become increasingly popular during the last decade.[17] In this chapter we will calculate a probability distribution for our quarterly GDP and inflation forecasts up to three quarters ahead.

The first part of this chapter provides an overview of the existing methodologies to determine the probability distribution and reports the distributions for the FPB forecast on the basis of the selected methodology. The second part presents the fan charts, i.e. the graphical representation of the calculated distributions, for our GDP and inflation forecasts.

5.1. Methodology

To quantify the risk around a central forecast, the probability distribution of the coefficients and the residuals of the estimated equations in the macro-econometric model or the statistical properties of the forecast errors can be used. The former option is more restrictive as it only integrates the fraction of the forecast errors related to the model specification and not the part that is caused by an incorrect evaluation of the exogenous variables or other external information[18] underlying the forecast. In line with the methodologies of other national institutions like the Office for Budget Responsibility (OBR) in the UK or INSEE in France[19], the observed distribution of the forecast errors will be used to calculate probability intervals around the FPB forecasts for GDP growth and inflation.

Assuming that past forecast errors provide a good measure of uncertainty surrounding any given forecast may be considered a disputable assumption, but it has the merit to provide a transparent and objective method of quantifying. Other methods, like the one proposed by the BoE or the IMF[20], offer the possibility to adapt the historical parameters to take on board the subjective view of the forecasters on the asymmetry in the balance of risks. However, such an approach does not fit the philosophy of the economic budget that supposes that the risks around the central scenario are balanced.

In order to obtain a probability distribution that is as robust as possible, all available forecast errors will be exploited in this chapter. This implies that, in contrast to the error tests described in previous chapters, forecast errors from the preliminary exercises prepared in June (cf. paragraph 2.3) will be added to the data set. This raises the number of data points from 18 or 19 to between 26 and 28 per forecasting horizon.

[17] For instance, the euro-area growth forecasts by the European Commission and the IMF global growth projections have been accompanied in recent years by such a chart.

[18] Forecasts, even if they are model-based, rely also on external information introduced into the model through add-factors. See Dobbelaere et al. (2003) for a description of the elaboration of the economic budget.

[19] See Office for Budget Responsibility (2010) and INSEE (2008) for a description of the methodologies of these institutions.

[20] See respectively Britton et al. (1998) and IMF (2009) for more details.

We first have to determine the shape of the probability distribution function. Two approaches can be found in the literature. On the one hand, the OBR uses a so-called "two-piece normal distribution" that integrates a degree of skew in the normal distribution, in other words a distribution in which large errors are more plausible on the downside than on the upside or vice versa. On the other hand, INSEE assumes that forecast errors fit a normal distribution. The Jarque-Bera test was used to check whether the quarterly forecast errors discussed in Chapter 4 are normally distributed around their mean. In many cases, the null hypothesis of normally distributed forecast errors is rejected. However, a closer look at the data[21] reveals that this is mainly due to some outliers combined with the limited power of this test in case of a small sample size. The rejection of the null hypothesis is avoided by correcting for outliers, except in the case of current quarter inflation forecasts. Consequently, it is assumed that forecast errors fit a normal distribution.

In this case, the mean and the standard deviation of the forecast errors determine the full probability distribution of the quarterly forecasts. The goal of the FPB has always been to produce unbiased forecasts, an objective that seems to be met for the quarterly forecasts considered here up to three quarters ahead (see paragraph 4.2.2). Consequently, it makes sense to assume that the mean of the probability distribution of the forecast errors is equal to zero, implying that the FPB forecast are considered as the most probable ones. The standard deviations used to construct the probability distribution were corrected for outliers in the same way as in Annex 2. Having only a limited set of data points per forecasting horizon at our disposal, the effect of one or two outliers on confidence intervals is assessed to be excessive. The corrected standard deviations at different forecasting horizons and their associated confidence intervals are summarised in Table 13.

Table 13 Standard deviations (corrected for outliers) and confidence intervals
Percentage points

	Current quarter	One quarter ahead	Two quarters ahead	Three quarters ahead
GDP growth				
Standard deviation	0.24	0.30	0.35	0.37
20% confidence interval	-0.06/+0.06	-0.08/+0.08	-0.09/+0.09	-0.09/+0.09
40% confidence interval	-0.12/+0.12	-0.16/+0.16	-0.19/+0.19	-0.19/+0.19
60% confidence interval	-0.20/+0.20	-0.26/+0.26	-0.30/+0.30	-0.31/+0.31
80% confidence interval	-0.30/+0.30	-0.39/+0.39	-0.45/+0.45	-0.47/+0.47
Inflation				
Standard deviation	0.10	0.40	0.77	1.00
20% confidence interval	-0.03/+0.03	-0.10/+0.10	-0.20/+0.20	-0.25/+0.25
40% confidence interval	-0.05/+0.05	-0.21/+0.21	-0.41/+0.41	-0.52/+0.52
60% confidence interval	-0.08/+0.08	-0.34/+0.34	-0.65/+0.65	-0.84/+0.84
80% confidence interval	-0.13/+0.13	-0.52/+0.52	-0.99/+0.99	-1.28/+1.28

5.2. Fan charts

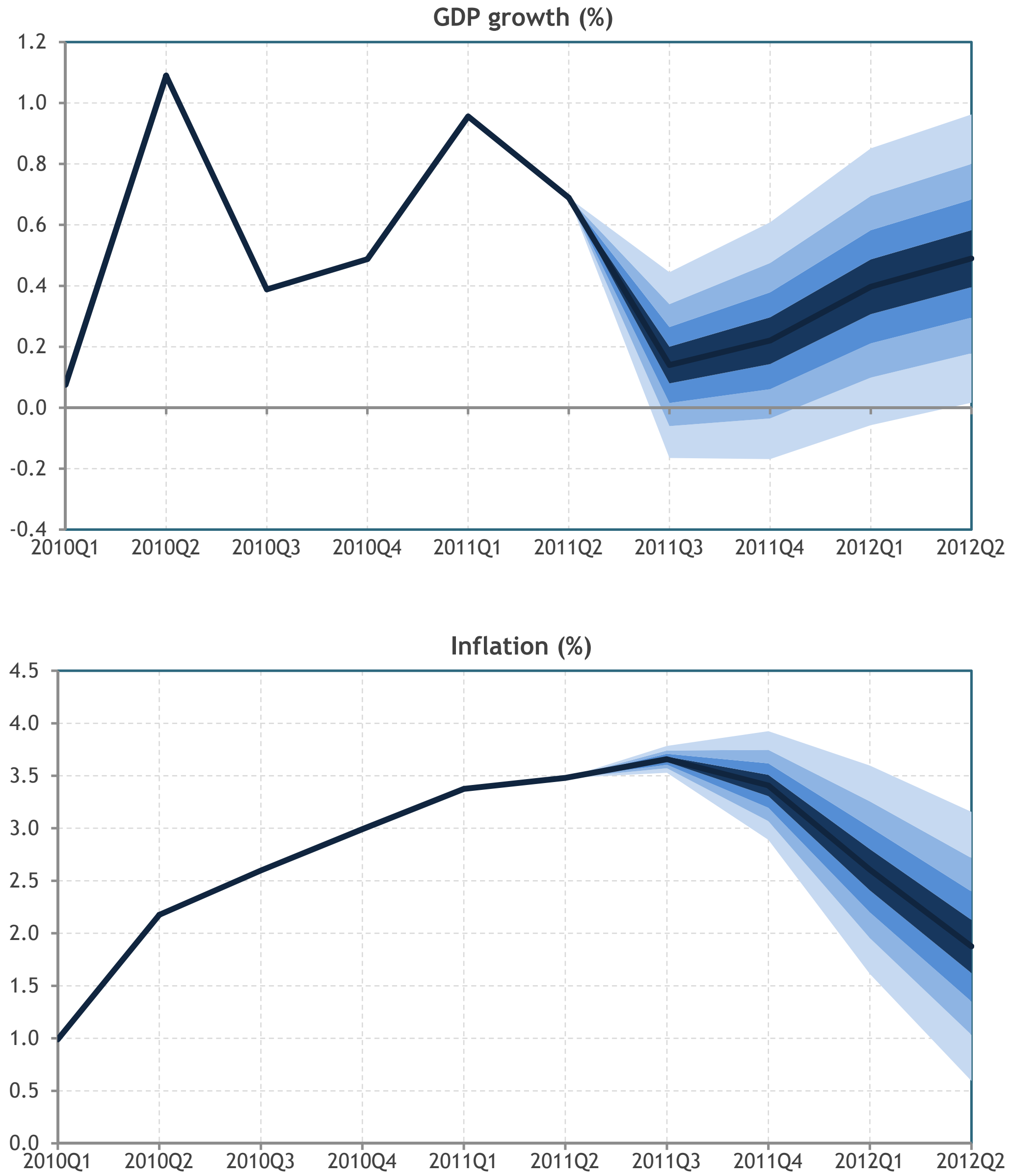
Having calculated the confidence intervals around the forecasts, one can summarise this information in a so-called fan chart. Graph 4 provides the fan charts for GDP growth and inflation that can be

[21] A detailed description of the histograms for GDP growth and inflation forecast errors as well as the Jarque-Bera tests is provided in Annex 2.

constructed on the basis of the economic budget published in September 2011. As the probability distribution is only calculated up to three quarters ahead, the fan chart reports forecasts from 2011Q3 up to 2012Q2.

The central projection that is published and commented in the economic budget is depicted as a line in the graph. During the forecasting period, this line is surrounded by a "fan", representing the probability distribution of the forecasts. The further we move away from the central forecasts, the higher the probability of the confidence interval and the lighter the colour in the fan: the darkest area around the central forecast represents the 20% confidence band, followed by the 40% confidence band in a somewhat lighter colour, up to the 80% confidence band in the lightest colour.

Graph 4 Fan charts for GDP growth and inflation



These charts can be interpreted as follows: at the time the economic budget of September was finalised qoq GDP growth in, for example, 2012Q1 would be within the +0.31/+0.49% - interval with a probability of 20%, within the +0.21/+0.58% - interval with a probability of 40%, within the +0.10/+0.70% - interval with a probability of 60% and within the -0.06/+0.85% - interval with a probability of 80%. Note that the fan charts only report the area of possible outcomes up to the 80% confidence interval, implying that there is still a probability of 20% to obtain an outcome that falls out of this area. This indicates for instance that the probability of negative economic growth in 2012Q2 was evaluated to be less than 10%.

A similar reading can be done for inflation forecasts. For example, at the end of the forecasting horizon the probability that inflation would remain above the threshold of 2% was estimated to be around 45%. Observe also that, as the probability distributions have a zero mean and skew, the central forecast can be calculated as the average of the upper and the lower bound of a confidence interval.

References

Bogaert H., Dobbelaere L. Hertveldt B., and Lebrun I. (2006),”Fiscal councils, independent forecasts and the budgetary process: lessons from the Belgian case”, Federal Planning Bureau, Working Paper 04-06, June.

Britton E., Fisher P. and Whitley J. (1998), “The Inflation Report projections: understanding the fan chart”, Bank of England, Quarterly Bulletin, February, pp. 30-37.

Carnot N., Koen V. and Tissot B. (2005), Economic Forecasting, Palgrave McMillan.

Dobbelaere L., Hertveldt B., Hespel E. and Lebrun I. (2003), “De opmaak van de economische begroting: een handleiding”, “Tout savoir sur la confection du budget économique”, Federal Planning Bureau, Working Paper 17-03, October.

Dobbelaere L. and Hertveldt B. (2004), “10 jaar Economische Begroting: Een terugblik op de kwaliteit van de vooruitzichten”, Federal Planning Bureau, Working Paper 13-04, April.

Federal Planning Bureau (1998), “The accuracy of the FPB short-term economic forecasts since 1994”, Special Topic, Short Term Update 4-98, November, pp. 3-4.

Federal Planning Bureau (2006), “Tools and methods used at the Federal Planning Bureau”, Working Paper 7-06, September.

Hertveldt B. and Lebrun I. (2003), “MODTRIM II: A quarterly model for the Belgian economy”, Federal Planning Bureau, Working Paper 6-03, May.

IMF (2009), “Fan Chart for Global Growth”, World Economic Outlook, April, pp. 55-58.

INSEE (2008), “Éclairage - Appréhender l’incertitude autour de la prévision”, Note de conjoncture, June, pp. 15-16.

Kranendonk H., de Jong J. and Verbruggen J. (2009),” Trefzekerheid CPB-prognoses 1971-2007”, CPB Document No 178, February.

Loungani P. and Rodriguez J. (2008), “Economic Forecasts: Too smooth by far?”, World Economics, Vol. 9, No. 2, April-June.

Melander A., Sismanidis G. and Grenouilleau D. (2007), “The track record of the Commission's forecasts - an update”, European Commission, European Economy - Economic Paper, No. 291, October.

Office for Budget Responsibility (2010), “Construction of fan charts”, Annex A, Pre-Budget forecast, June, pp. 65-72.

Timmerman A. (2006), “An Evaluation of the World Economic Outlook Forecasts”, IMF Working Papers, 06/59, March.

Vogel L. (2007), “How do the OECD growth projections for the G7 economies perform? A post-mortem”, OECD, Economics department working paper, No. 573, September.

Annex 1. Summary statistics and statistical tests

Summary statistics

The standard deviation of the outcome X is a measure of the variable volatility:

$$STD = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(X_t - \bar{X}\right)^2}$$

The forecast error E is defined as the forecast F minus the outcome:

$$E_t = F_t - X_t$$

The mean absolute error provides a measure of the size of the error:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|E_t\right|$$

The root mean square error also provides a measure but giving more weight to the largest errors:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}E_t^2}$$

The corrected RMSE ensures comparability between series with different volatilities:

$$RMSE_{corr} = \frac{RMSE}{STD}$$

The mean error measures the bias of the forecast:

$$ME = \frac{1}{N}\sum_{t=1}^{N}E_t$$

The Theil coefficient compares the ratio between the RMSE of the forecast and the RMSE of an alternative forecasting method:

$$THEIL = \frac{RMSE}{RMSE_{alt}}$$

Statistical tests

Bias

Testing for the statistical significance of the bias is done by regressing the forecast error on a constant:

$$E_t = \alpha + \varepsilon_t$$

The absence of bias requires α = 0. The restriction is tested with the t-statistic that is always corrected for the (possible) presence of autocorrelation and/or heteroskedasticity in the residuals.

Persistence of forecast errors

The persistence of forecast errors is tested by checking the residuals from the above-mentioned equation for serial correlation. This test is carried out on the basis of Ljung-Box Q-statistics that are asymptotically X^2 distributed with the degrees of freedom equal to the number of autocorrelations that is tested. Note that, for example, the test for second order serial correlation tests the joint significance of first and second order serial correlation.

Efficiency

Testing for weak efficiency is done by regressing the outcome on a constant and the forecast:

$$X_t = \alpha + \beta F_t + \varepsilon_t$$

Weak efficiency requires α = 0, β = 1 and ε to be white-noise residuals. The joint restriction on the coefficients is tested with the F-statistic, while the white-noise properties of the residuals are examined with the Ljung-Box Q-statistics.

Testing for informational efficiency is done by regressing the forecast error on the forecast of the international indicators:

$$E_t = \alpha + \beta_1 F1_t + \beta_2 F2_t + \varepsilon_t$$

Informational efficiency requires $\beta_1 = 0$ and $\beta_2 = 0$. The joint restriction is tested with the F-statistic.

Revisions

Testing for the absence of bias in the revisions is done by regressing the revisions on a constant:

$$\Delta F_t = \alpha + \varepsilon_t$$

The absence of bias requires α = 0. The restriction is tested with the t-statistic.

Testing for independent revisions is done by regressing revisions in round 3 onto revisions in round 2:

$$\Delta F_t^3 = \alpha + \beta \Delta F_t^2 \varepsilon_t$$

Independent revisions requires β = 0. The restriction is tested with the t-statistic.

Annex 2. Histograms of forecast errors

The graphs and tables below provide the information that was used to determine whether GDP growth and inflation forecast errors at different forecasting horizons can be considered as normally distributed. To perform the Jarque-Bera test, a statistic is calculated in which the skew and the kurtosis of a distribution are compared to the parameters of a normal distribution (respectively 0 and 3). Under the null hypothesis of a normal distribution, this statistic is X^2-distributed with two degrees of freedom.

Graph 5 Histograms of quarterly GDP growth forecast errors
Number of forecast errors within the specified ranges

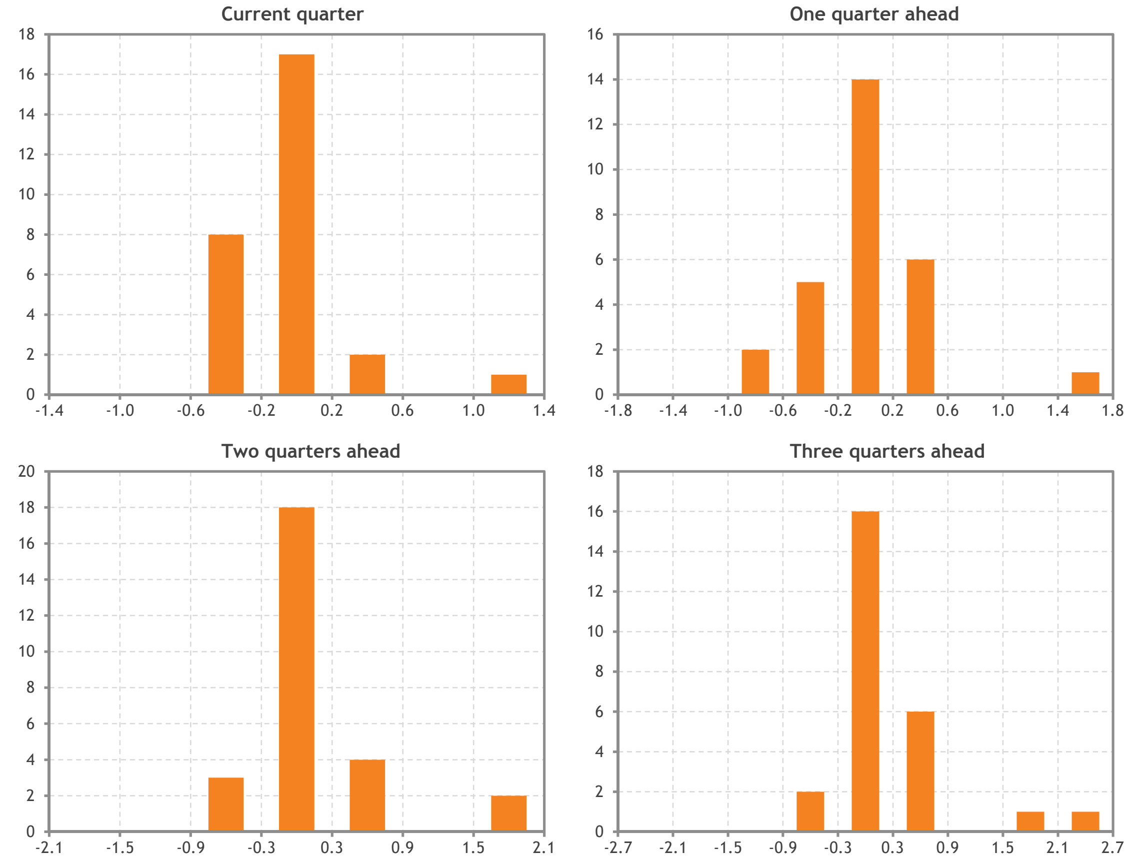


Table 14 Statistics for the histogram of quarterly GDP growth forecast errors

	Current quarter	One quarter ahead	Two quarters ahead	Three quarters ahead
Full sample				
Skewness	1.4	1.4	1.6	1.8
Kurtosis	6.3	8.2	5.7	6.2
Jarque-Bera	21.6**	40.7**	20.2**	25.4**
After outlier correction				
Skewness	0.3	-0.7	0.4	0.3
Kurtosis	3.1	4.1	3.1	3.2
Jarque-Bera	0.3	3.6	0.7	0.4

Note: ** denotes that the null hypothesis of normally distributed data points is rejected at the 5% level.

The Jarque-Bera test was carried out on the full sample of forecast errors and on a sample excluding the biggest outlier(s). In the case of GDP growth, these (positive) outliers were at each forecast horizon seen during the recession related to the financial crisis of 2008. The biggest (negative) outliers in inflation forecast errors were all situated in the first quarters of 2008, a period in which Belgian consumer price inflation exploded due to strong increases in commodity prices and exogenous factors raising prices.

Graph 6 Histograms of quarterly inflation forecast errors

Number of forecast errors within the specified ranges

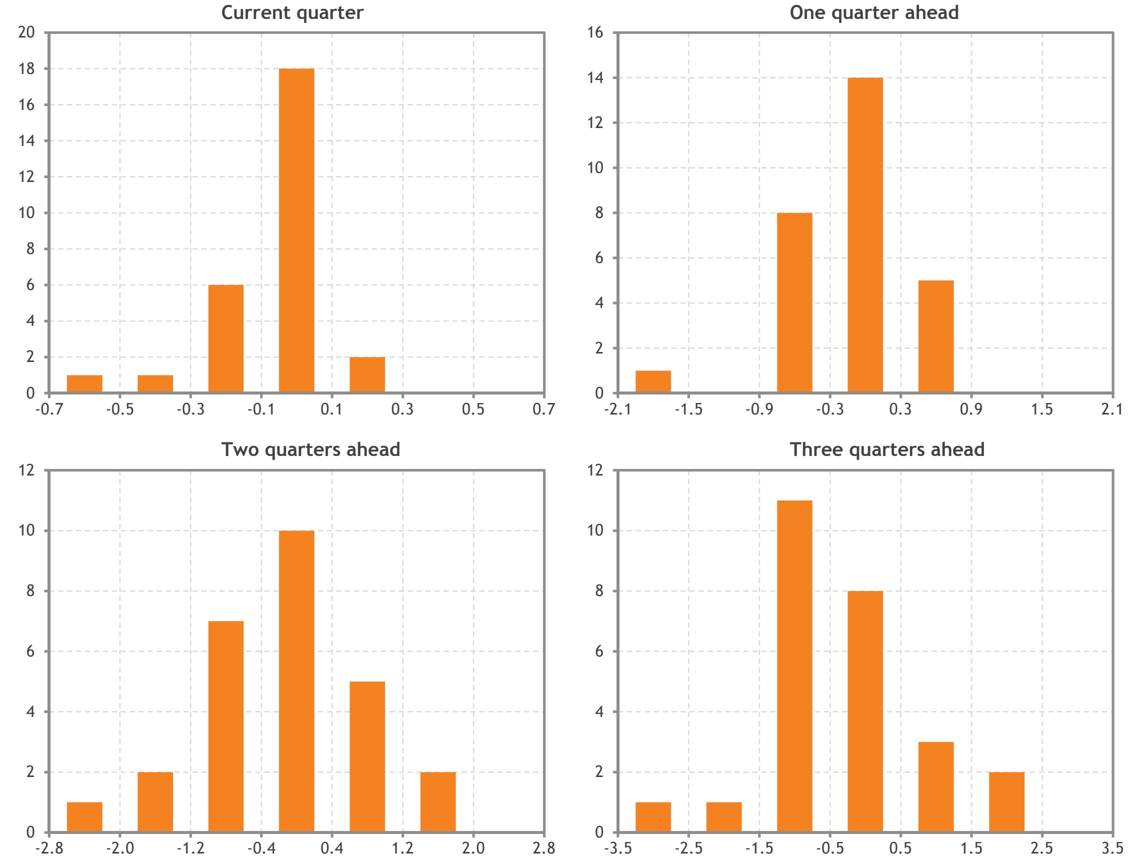


Table 15 Statistics for the histogram of quarterly inflation forecast errors

	Current quarter	One quarter ahead	Two quarters ahead	Three quarters ahead
Full sample				
Skewness	-1.9	-1.1	-0.1	0.2
Kurtosis	7.7	5.4	3.1	3.1
Jarque-Bera	43.2**	12.6**	0.1	0.3
After outlier correction				
Skewness	-0.9	0.1	0.5	0.7
Kurtosis	4.0	2.3	2.4	2.8
Jarque-Bera	4.9*	0.6	1.3	2.2

Note: * and ** denote that the null hypothesis of normally distributed data points is rejected at the 10% or the 5% level respectively.