Evaluating the forecast performance of GDP

2.1 Introduction*

"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. If the evaluation detects a significant bias affecting macroeconomic forecasts over a period of at least four consecutive years, the Member State concerned shall take the necessary action and make it public."

Council Directive 2011/85/EU – Article 4(6)

Macroeconomic forecasts have an important role in framing government policies, particularly the budget process. Given their pivotal role, regular assessment and evaluation of forecast performance are key to improving forecast quality and accuracy. Reliable economic forecasts build economic certainty and confidence and allow economic agents to make more efficient decisions. On the contrary, inaccurate forecasts beyond certain margins, whether they overpredict or underpredict, have consequences in misleading decisions and increasing costs.

This analysis evaluates the forecasting performance of the macroeconomic projections of the Ministry for Finance and Employment (MFE) for the 2004 – 2021 period, focusing on forecasts for real and nominal GDP growth. The Economic Policy Department within the MFE is responsible for producing macroeconomic projections for the Government of Malta through its Short-term Quarterly Forecasting Econometric Model for Malta (STEMM).[10] Macroeconomic forecasts produced by the MFE serve as important inputs in several key documents and policymaking decisions of the Government of Malta, including:

(a) the annual Update of the Stability Programme, presented by the Government of Malta to the European Commission every April, highlighting its

* This thematic chapter presents selected conclusions from an advanced working paper drafted by the Chief Economist Mr Gilmour Camilleri and Senior Economist Mr Kurt Davison, forthcoming as a publication in the MFAC Working Paper series.

[10] STEMM is an expenditure-driven model developed in collaboration with Cambridge Econometrics in 2002. For a detailed technical report on this macroeconomic model, please refer here.

macroeconomic and fiscal projections for years t up to t+3 in accordance with European Union Council regulations[11],

(b) annual Medium-Term Fiscal Strategy for Malta covering years t up to t+3 in accordance with the requirements of Article 15(8) of the Fiscal Responsibility Act,

(c) the Annual and Half-Yearly reports published by the MFE, and

(d) the Draft Budgetary Plan presented by the Government of Malta to the European Commission every year, including macroeconomic and budgetary projections covering years t and t+1.

In all of the above, the Malta Fiscal Advisory Council (MFAC) is mandated by law to:

(a) endorse, as it considers appropriate, the macroeconomic and fiscal forecasts prepared by the Ministry for Finance and provide an assessment of the official forecasts;

(b) analyse and assess whether the Government's Medium Term Fiscal Policy Statement and Medium-Term Fiscal Policy Strategy are compliant with the provisions of the Act, issue an opinion and any appropriate recommendations;

(c) in relation to each National Medium Term Fiscal Plan, Stability Programme, Annual Draft Budget and Annual Budget, provide an assessment of whether the fiscal stance for the year or years concerned is, in the opinion of the Fiscal Council, conducive to prudent economic and budgetary management, and in conformity with the provisions of this Act, including by reference to the provisions of the Stability and Growth Pact;

(d) assess the Government's budgetary performance against the fiscal targets and policies specified in the fiscal strategy and its compliance with the provisions of this Act;

(e) analyse and issue an opinion and any recommendations pursuant to the Government's publication of the half-yearly and the annual report on the execution of the budget.

Fiscal Responsibility Act (Cap. 534 of the Laws of Malta) – Article 13(3), *ad verbatim*

[11] European Union Council Regulations – Council Regulation (EC) 1466/97 on the strengthening of the surveillance of budgetary positions and the surveillance and coordination of economic policies, as amended by Council regulation (EC) 1055/2005 of 27 June 2005 and Regulation (EU) No 1175/2011 – the preventive arm of the Stability and Growth Pact.

Using forecast accuracy measures, this chapter assesses the forecasting performance of the MFE, for real and nominal GDP growth, and compares the results with those of the European Commission, the Central Bank of Malta, and other benchmark models. Forecasts by the MFE are published bi-annually: in spring (Update of the Stability Programme) and in autumn (Draft Budgetary Plan). This analysis distinguishes between the two.

This analysis builds and updates a similar exercise by Camilleri and Vella (2015), who test for forecast accuracy and biasedness, and present the uncertainty surrounding the macroeconomic projections using fan charts for the period 2004-2013. The main findings from this study indicate that the forecast performance by the Ministry compares favourably to other small open economies and even across other independent institutions' forecasts. They also do not find any systematic bias in the spring forecasts for nominal and real GDP but do find bias in the separate expenditure components.

In this chapter, the authors update the sample period to include post-2013 macroeconomic projections and extend the analysis to include both the forecasts published in the Update of Stability Programme and the Draft Budgetary Plan. The analysis also adopts a broader rigorous assessment to evaluate forecast performance. In fact, forecast performance is assessed on three pillars: accuracy, unbiasedness, and benchmarking. As in the 2015 study, this evaluation is conducted for two macroeconomic variables: real and nominal GDP, but the analysis presented in this chapter focuses on these main aggregates and does not include expenditure components.[12]

The objectives of this research are twofold; first, it is being carried out in the context of Council Directive 2011/85 of the European Union on the requirements for budgetary frameworks of the Member States on the evaluation of forecast biasedness. Second, the MFAC believes that assessments of the forecast performance of the projections produced by the MFE are key to identifying areas of improvement and issuing recommendations and advice in that regard.

[12] A forthcoming publication by the MFAC will also assess forecasting efficiency and will also include an analysis of the forecast performance of the expenditure components.

2.2 Methodology

This section describes how the MFE's forecasting performance was evaluated. Specifically, the assessment involved the following two pillars:

- How close the predictions are to the actual outcome (accuracy) and whether the forecasts produced by the MFE have been more or less accurate than the forecasts produced by other institutions, and forecasts from simple models (benchmarking exercise).
- Whether forecasts have been consistently optimistic or conservative (unbiasedness).

We acknowledge that these measures of performance are interrelated in the sense that if forecast accuracy is high, there is less scope for forecast biasedness. Nonetheless, each measure provides diverse ways to assess forecast performance that is worth evaluating.

2.2.1 Forecast accuracy and benchmarking

To measure forecast accuracy, we calculated the mean error (ME), the mean absolute error (MAE), the root mean squared error (RMSE), Theil's U statistic (U) and the mean relative absolute error (MRAE). These are explained below:

- The mean error (ME) is the average of forecast errors. More formally,

$$\text{ME} = \frac{1}{T}\sum_{i=1}^{T} \hat{y}_t - y_t = \frac{1}{T}\sum_{i=1}^{T} e_t \qquad (1)$$

  where, the forecast of variable y for period t is denoted by $\hat{y}_t$ and the actual value by y_t, and e_t is a forecast error, defined as forecast–outturn.

  The ME needs to be interpreted with caution because a small result is not necessarily indicative of good forecast accuracy, due to the fact that negative forecast errors offset positive forecast errors. Furthermore, it is not meant for comparing and evaluating a method's forecast accuracy across multiple data of different magnitude.

- The mean absolute error (MAE) is the average of the absolute error, which is the deviation of forecasts from actual points, disregarding the sign of the error. Formally,

$$\text{MAE} = \frac{1}{T}\sum_{i=1}^{T}|\hat{y}_t - y_t| = \frac{1}{T}\sum_{i=1}^{T}|e_t| \qquad (2)$$

  where again, the forecast of variable y for period t is denoted by $\hat{y}_t$ and the actual value by y_t, and e_t is a forecast error, defined as forecast-outturn.

- The root mean squared error (RMSE) is a common forecast accuracy measure calculated as the standard deviation of the forecast errors. This measure disproportionately penalises forecast accuracy according to the magnitude of the forecast errors i.e., RMSE accounts for the fact that large forecast errors are considered more problematic than small ones. More formally,

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(\hat{y}_t - y_t)^2} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(e_t)^2} \qquad (3)$$

  The sensitivity of the RMSE to data outliers is the most common reason for using this scale-dependent measure. As a result, this forecast accuracy indicator is considered superior to the measures previously outlined.

- Theil's U statistic (U) is a relative accuracy measure that compares the predictions with a naïve forecast.[13] This forecast accuracy measure can be interpreted as the ratio of the RMSE to the standard deviation of the forecast errors from the naïve model. Similar to the RMSE, this measure also gives more weight to large errors by squaring the deviations. If Theil's U statistic exceeds one, it means that the forecast from the model is no more accurate than a naïve forecast. Theil's U statistic is calculated by using the following formula:

$$\text{U} = \sqrt{\frac{\frac{1}{T}\sum_{i=1}^{T}(\hat{y}_t - y_t)^2}{\frac{1}{T}\sum_{i=1}^{T}(y_t - y_{t-1})^2}} \qquad (4)$$

[13] Naïve Forecast is a forecasting technique in which the forecast for the current period (y_t) is set to the actual value from the previous period (y_{t-1}).

- The mean relative absolute error (MRAE) is an alternative to the mean absolute error (MAE) as a scale-dependent measure. The MRAE implies taking an average of the absolute value of the relative share of errors i.e., the forecast error based on the forecasts published by the institutions of interest as a share of the forecast error obtained from the benchmark method. Usually, the benchmark method is the random walk without drift model where f_t^* is equal to the last observation. The MRAE is calculated by using this formula:

$$\text{MRAE} = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{\hat{y}_t - y_t}{f_t^* - y_t}\right| \qquad (5)$$

  A deficiency of this measure is that if the forecasting error obtained from the benchmark method is zero, the use of the random walk without drift model as a benchmark method would no longer be possible because it would involve dividing by zero.

A simple comparison of the different forecast accuracy measures was also conducted across institutions (MFE, European Commission and Central Bank of Malta), and also with forecasts generated through simple statistical models (naïve forecast, moving average (MA) of the past two years and moving average of the past three years).

2.2.2 Unbiasedness

In the history of the EU's Stability and Growth Pact, some governments have justified fiscal expansions or postponed fiscal adjustments by being more optimistic when predicting medium-term growth (Larch et al., 2021; Frankel, 2011). In this context, we also assessed the real and nominal GDP growth forecasts produced by the Ministry for Finance and Employment for any potential upward or downward bias. To carry out this evaluation, we employed the Least Squares methodology and regressed the forecast errors on a constant with a null hypothesis that the constant was zero. In case of biasedness, the constant would take a non-zero value. Formally, we estimated the following regression:

$$e_t = \beta_0 + \varepsilon_t \qquad (6)$$

where e_t is the forecast error of variable y for period t and where ε_t is a zero-mean error term. Under the null hypothesis of unbiasedness, $\beta_0 = 0$. If $\beta_0 < (>) 0$, forecast has been

systematically too low (high). In line with the methodology employed by the Bank of England (2015) when evaluating for the presence of bias in macroeconomic forecasts, we estimated the regression using OLS with Heteroscedasticity and Autocorrelation (HAC) standard errors.[14]

2.2.3 Sample and data sources

Time series data for nominal and real GDP were collected for the 2004 – 2021 period from the autumn and spring forecast rounds (Update of Stability/Convergence Programme) and for the 2013 – 2021 period from the autumn forecast round (Draft Budgetary Plan).[15] For the benchmarking exercise, data was obtained from the Quarterly Reviews of the Central Bank of Malta, and the spring and autumn forecasts of the European Commission.

From 2004 to 2009, the Stability-Convergence Programmes used to be published in November, while from 2011 onwards it was published in April.[16] As a result, for comparability purposes, data for the European Commission for this period were collected from the autumn forecast round issue (published in November) while the rest of the sample was collected from the spring issue (published in May).[17] In 2010, the Stability Programme was not published by the Ministry for Finance, however, forecast data for this year were still available internally, and were used in this study in order to have a complete time series.[18]

With regards to the Central Bank of Malta's projections, these are available from 2008. Again, to maintain data comparability, data for 2008 and 2009 were collected from the fourth Quarterly Review issue (published in December), while the rest of the sample was collected from the second Quarterly review issue (published in May). On the other hand, the forecasts from the Draft Budgetary Plan are compared to the European Commission's autumn forecasts (published in November) and the fourth Quarterly Review by the Central Bank of Malta (published in December).

[14] The results are based on a HAC adjustment using Andrew's Automatic bandwidth method.
[15] The first publication of Malta's Draft Budgetary Plan was on 15 October 2013. This following regulation (EU) No 473/2013 of the European Parliament and of the Council of 21 May 2013 on common provisions for monitoring and assessing draft budgetary plans and ensuring the correction of excessive deficit of the Member States in the euro area.
[16] The change in publication date reflects developments at EU-level in relation to the European Semester.
[17] For ease of reference the forecasts published in the Stability/Convergence Programme will be referred to as Spring.
[18] This was a result of the shift to the European Semester.

Since the Central Bank of Malta published neither nominal GDP nor GDP deflator forecasts, the comparison is only carried out for real GDP. In the case of the European Commission, the nominal GDP forecast is derived using the published GDP deflator. In addition, while forecast data for the Ministry for Finance and Employment in the Stability/Convergence Programmes are available up to year t+3, the European Commission and the Central Bank of Malta only report forecasts for years t and t+1. In the autumn period, forecast data for all institutions is available for years t and t+1.

2.2.4 Limitations

At this stage, some limitations of the study are worth mentioning. In particular, forecast error evaluations generally cover at least 20 years of data, whereas our sample size, especially for the autumn forecast period is notably smaller, with only nine data points. Additionally, the data limitations and different cut-off points across institutions highlighted earlier, constrain the cross-institution benchmarking and its results should be interpreted with caution.

It is important to note that forecast errors can also be affected by statistical errors in national accounts data. Forecast errors are influenced by the vintage of input data used in the forecasting model, and the vintage of data used as a benchmark to estimate the forecast errors. While the most up-to-date statistical data gives a more accurate and reliable estimate of forecast errors, it may underestimate the efficiency of a forecasting model by diluting statistical errors with the pure forecast errors of a given economic model. Using the latest national accounts release as a benchmark on which to compute forecast errors and the relatively small sample size, will add a further degree of uncertainty to the analysis in that any further revisions to national accounts data will have an impact on the accuracy of forecast projections both ex-ante and ex-post.[19] The ex-ante effect is mainly through the trajectories of the forecast projections, while ex-post forecast errors may emerge since the base on which the forecast projections were estimated in the first place would have changed.[20]

[19] The actual data used as a benchmark to compare with the forecast vintages is NSO News Release 218/2022 available here.
[20] This implies that the ex-ante statistical revisions may change the current trajectory of your forecasts, while ex-post the trajectory of forecasts may be less accurate given a change in the actual figures used to estimate those forecasts. More detail on the effects of statistical revisions is provided in section 2.4.3.

2.3 Forecast errors of GDP growth projections

This section describes the forecast errors observed for nominal GDP and real GDP growth projections by MFE. Forecast errors in this study are defined as the forecast at time t minus the actual data at time t+1. More formally,

$$iptcre_{t,t} = y_{t,t} - y_t \text{ for the current year; and}$$
$$e_{t+1,t} = y_{t+1,t} - y_{t+1} \text{ for the following year,}$$

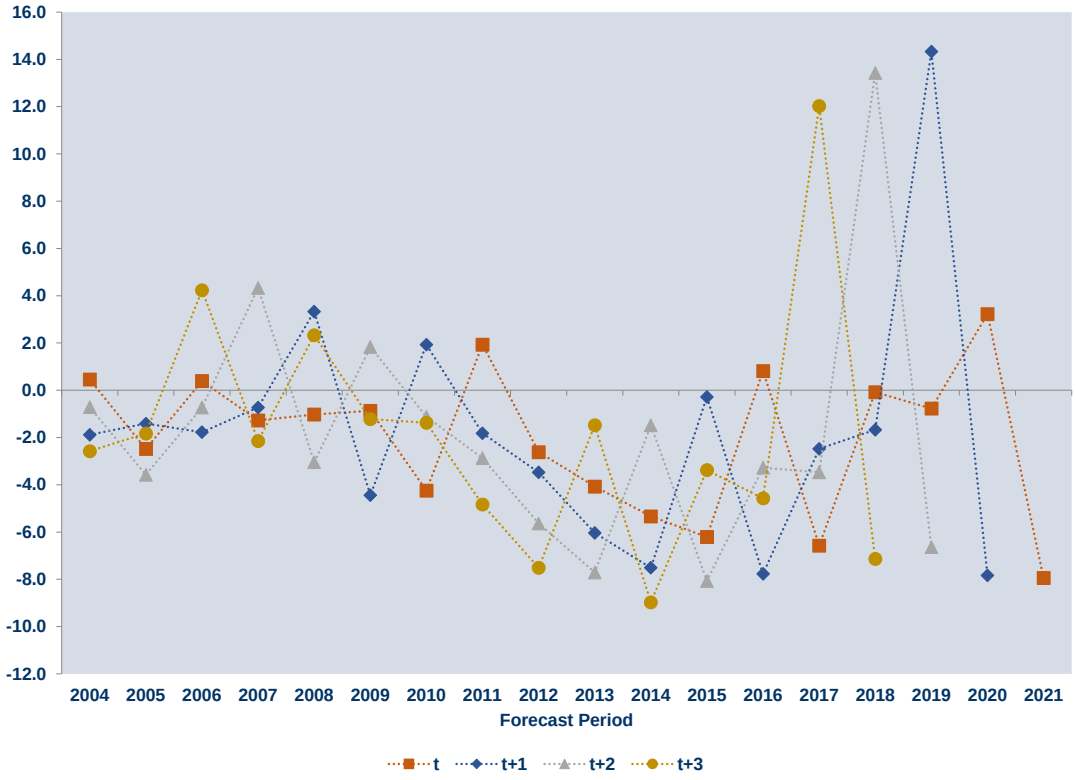
where $y_{t,t}$ and $y_{t+1,t}$ are the projections made at time t and t+1 respectively, y_t is the actual data of variable y for year t, and y_{t+1} is actual data for variable y for year t+1. Therefore, a positive forecast error for real GDP growth implies an overestimation of the rate of growth while a negative value implies an underestimation of the rate of growth.

Forecast errors for real GDP in the Update of the Stability Programme categorised by forecasting vintage since 2004 for current, one-year, two-year and three-year ahead forecasts are presented in Chart 2.1. The horizontal axis represents the year in which the forecast is undertaken. For instance, for 2008, the one-year ahead forecast error represents the difference between the forecast for 2009 made in 2008 and the actual data for 2009. Similarly, the two-year ahead forecast error represents the difference between the forecast produced in 2008 for 2010 and the actual for 2010 whilst the three-year ahead forecast error reported in the 2008 vintage represents the forecast error for 2011.

Forecast errors for real GDP growth range between +14.2 and –9.0 percentage points. However, the overestimation of growth is largely due to unexpected economic shocks. Indeed, when removing the financial crisis years, 2009 and 2010 and the COVID-19 crisis years, 2020 and 2021, the forecast errors for real GDP range between +2.3 and -9.0 percentage points. Overall, there seems to be more tendency to underestimate growth. This could be due to structural changes in the economy not incorporated in the model or statistical revisions. We also note that the tendency to underestimate real GDP growth has increased during the last decade, with forecast errors excluding the crisis years for the period 2004 to 2010 ranging from +2.3 to -3.6 percentage points,

while the forecast errors for the period 2011 to 2019 ranged from +1.9 to -9.0 percentage points.[21]

Chart 2.1: Forecast error: real GDP (forecast – actual, pps.) (USP)



Similarly, Chart 2.2 shows forecast errors for real GDP in the Draft Budgetary Plan categorised by forecasting vintage since 2013 for current and one-year ahead forecasts. Forecast errors for real GDP growth range between +12.9 and -7.4 percentage points. When removing the COVID-19 crisis years, the forecast errors for real GDP range between +0.5 and -7.4 percentage points. Thus, as in the case of the Stability Programme, the overestimation is largely due to the unexpected economic shock. The tendency to underestimate real GDP growth is also present in the autumn forecast round, albeit to a somewhat lower extent.

A similar pattern is evident for nominal GDP growth although forecast errors are marginally higher than for real GDP (See Chart 2.3 and Chart 2.4). Indeed, in the Update of the Stability Programme, forecast errors for nominal GDP growth range from

[21] The time periods analysed here are based on an eyeball view of the data.

+15.5 to -10.0 percentage points while in the Draft Budgetary Plan these range between +13.7 and -9.4 percentage points.

Chart 2.2: Forecast error: real GDP (forecast - actual, pps.) (DBP)

With the exclusion of the crisis years, forecast errors for nominal GDP growth range from +3.4 to -10.0 percentage points in the Stability Programme, and from +0.8 and -9.4 percentage points in the autumn round of forecasts published in the Draft Budgetary Plan. This suggests that similar to the case of real GDP growth, there is a tendency for the Ministry for Finance and Employment to underestimate nominal GDP growth, which increased in the more recent period between 2011 and 2019 in the case of the Stability Programme.

Indeed, the forecast errors for nominal GDP growth from 2004 to 2010 were in the region of +3.4 to -3.3 percentage points, while from 2011 to 2019, the forecast errors ranged between +3.1 and -10.0 percentage points.

Chart 2.3: Forecast error: nominal GDP (forecast - actual, pps.) (USP)

Chart 2.4: Forecast error: nominal GDP (forecast - actual, pps.) (DBP)

2.4 Analysis and empirical results

The forecast accuracy is evaluated using five summary statistics over the period 2004 to 2021 in the case of the Update of Stability/Convergence Programme and over the period 2013 to 2021 in the case of the Draft Budgetary Plan for both real GDP and nominal GDP (See Section 2.4.1). In the Annex to this chapter, Table 2.1 and Table 2.2 provide the results obtained from these summary statistics. As indicated in the methodology, the Root Mean Squared Error (RMSE) is considered superior to the other summary statistics in measuring forecast accuracy and is given more weight in the analysis. Notwithstanding this, the Mean Error, Mean Absolute error, Theil's U statistic and Relative Mean Absolute error are also presented and discussed.

These results for MFE's forecasts are also benchmarked to the forecasting performance of the European Commission and the Central Bank of Malta for the current and one-year ahead forecasts. In addition, a comparison of the different forecast accuracy measures is also conducted in relation to simple statistical models, that is, a naïve forecast and two other forecasts based on the moving average of the past two years and the moving average of the past three years. Also, we undertake a simple analysis utilising the RMSE to determine whether forecast accuracy has improved across the forecast horizon in the Update of Stability/Convergence Programme forecast round. To this end, we compare the sample analysed in the working paper published by Camilleri, G. and Vella, K. in 2015, from 2004 to 2013, to the additional sample included in this analysis, from 2014 to 2019, excluding only the crisis periods. We then test unbiasedness on MFE forecasts for both the Update of Stability/Convergence Programme and Draft Budgetary Plan forecast rounds. Biasedness is also assessed through the different sample periods, from 2004 to 2013, 2014 to 2021 and the full sample to determine whether there were any changes in this regard. The unbiasedness is also tested with the crisis period excluded from the sample to eliminate the impact and difficulties of forecasting in uncertain periods.

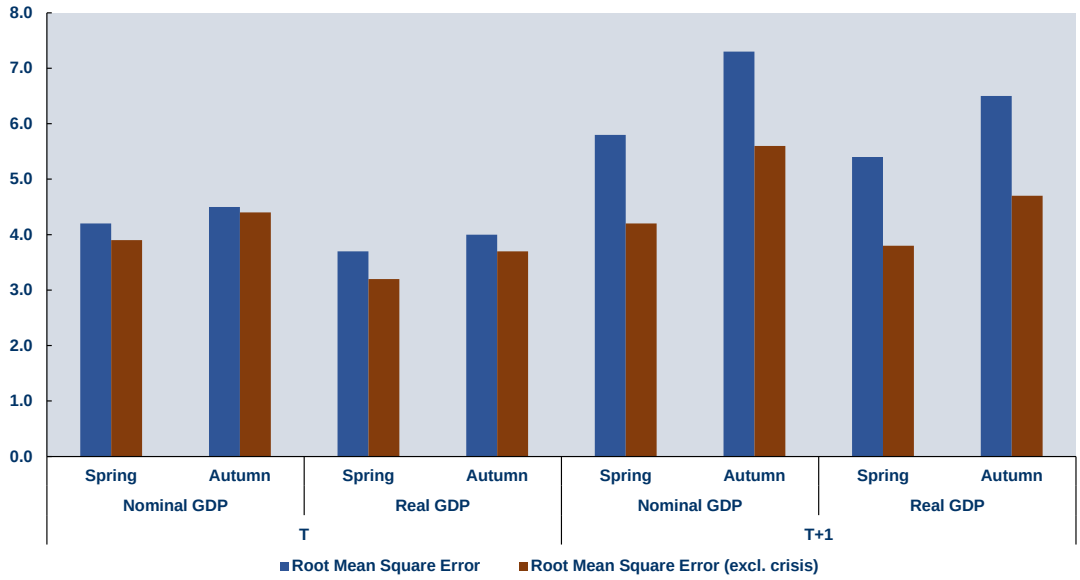
An important caveat is that the cause of the forecast errors, inaccuracy and biases reported in this analysis are not strictly all explained by errors made by the forecaster at the time of the forecasting exercise but are also partly attributable to statistical revisions in the data. Indeed, given that the forecast errors were compared to the most recent available data, the revisions carried out by NSO throughout the years may have increased or decreased these errors substantially. The statistical revisions carried out by the NSO are analysed in Section 2.4.3.

2.4.1 Accuracy

Although the mean error is generally considered to be a weak measure of forecast accuracy, it is the only measure which provides a sense of the general direction of the forecast errors. The results obtained suggest that in general, all institutions tend to be more cautious when estimating both real GDP growth and nominal GDP growth, thus the outturn is generally higher than initially projected for that year. This applies to both the Update of Stability/Convergence Programme and the Draft Budgetary Plan forecast rounds.

Focusing on the RMSE statistic, this is generally lower in the MFE's Update of Stability/Convergence Programme than the Draft Budgetary Plan forecast round for both years t and t+1 (See Chart 2.5). Also, the forecast error in the one-year ahead forecast tends to be slightly higher. This is expected as assumptions considered in the forecast are more susceptible to revisions and information is more limited the further away the forecast.

Chart 2.5: Accuracy of MFE for year t and t+1 forecasts, full sample and excluding crisis (Spring (USP) and Autumn (DBP)) – RMSE



When comparing the RMSE of the MFE forecast with that of the other institutions, we observe similar patterns and magnitudes of inaccuracy (See Chart 2.6 and Chart 2.7). In all instances and across institutions, forecast accuracy tended to improve when eliminating the crisis years (both the financial and COVID-19 crises) from the sample.

This implies that, as generally expected, the accuracy of forecasts tends to lessen at times of crisis.

Chart 2.6: Accuracy of COM for year t and t+1 forecasts, full sample and excluding crisis (Spring and Autumn) - RMSE

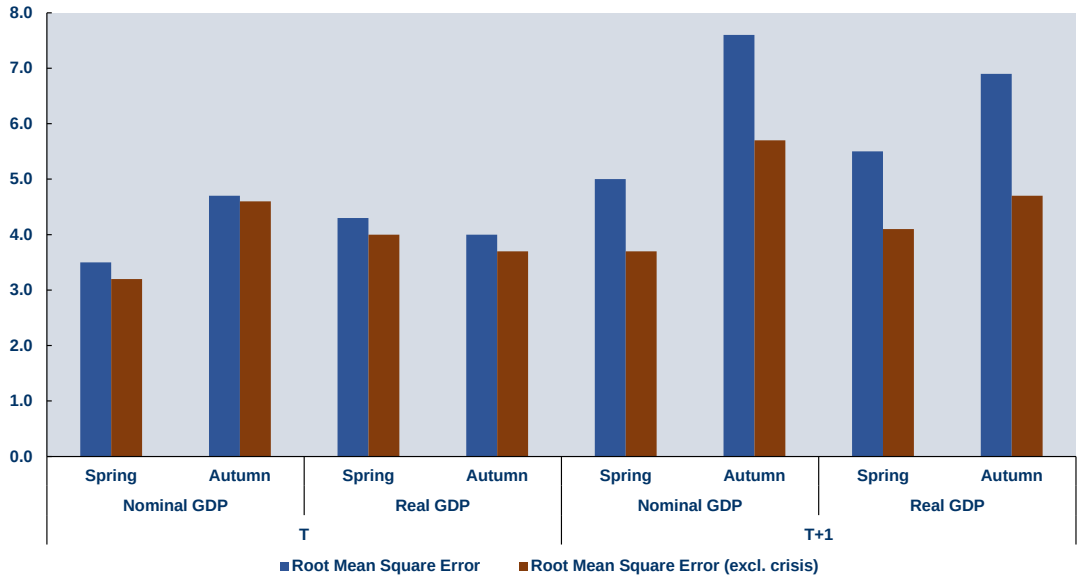
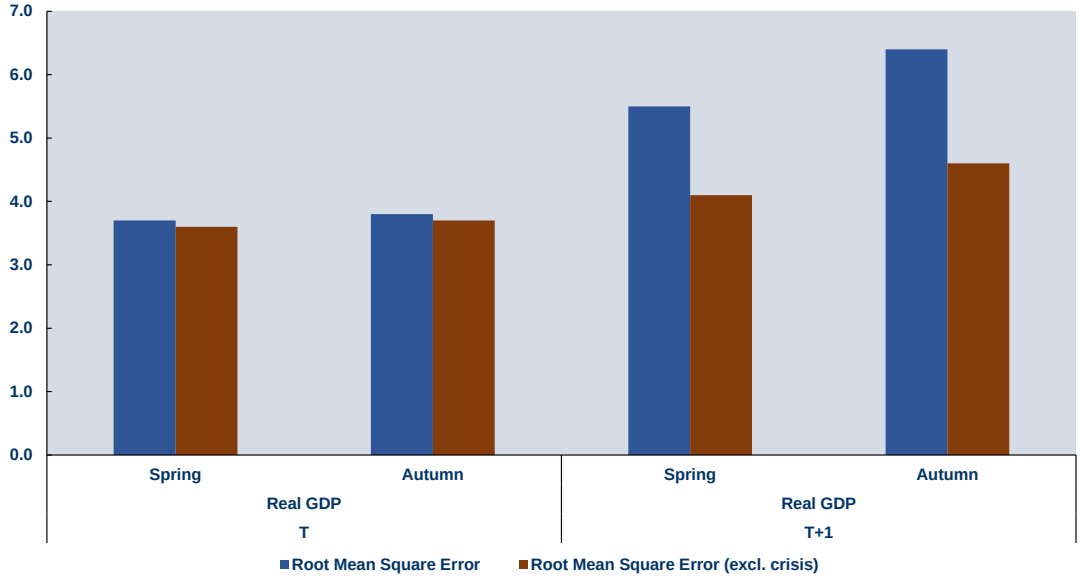


Chart 2.7: Accuracy of CBM for year t and t+1 forecasts, full sample and excluding crisis (Spring and Autumn) - RMSE



Analysing the results of Theil's U statistic, we note that the estimations carried out by the various institutions across both forecast rounds are in general more accurate than a simple naïve forecast, especially when forecasting at times of crisis (See Chart 2.8

and Chart 2.9). Indeed, Theil's U statistic is much closer to one and in some instances also exceeds one in the sample which does not include the crisis years, meaning it is less accurate than a simple forecasting model. It is also interesting to note that, as expected, the forecast by all institutions of the one-year ahead forecasts for both real GDP and nominal GDP is generally less accurate than the current year forecast when compared to a naïve forecast.

Chart 2.8: Theil's U statistic of real GDP, comparison across institutions

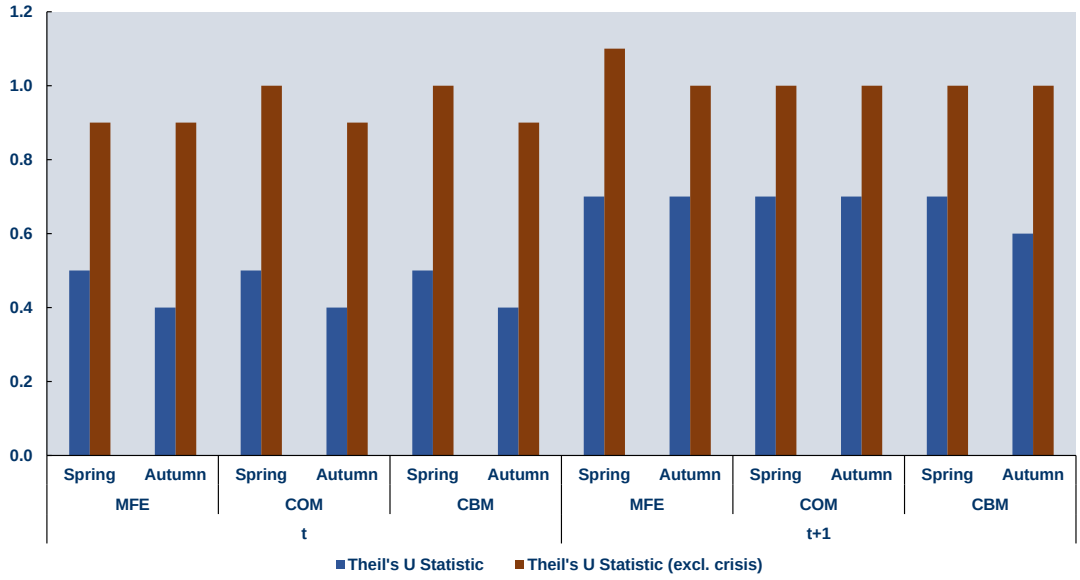
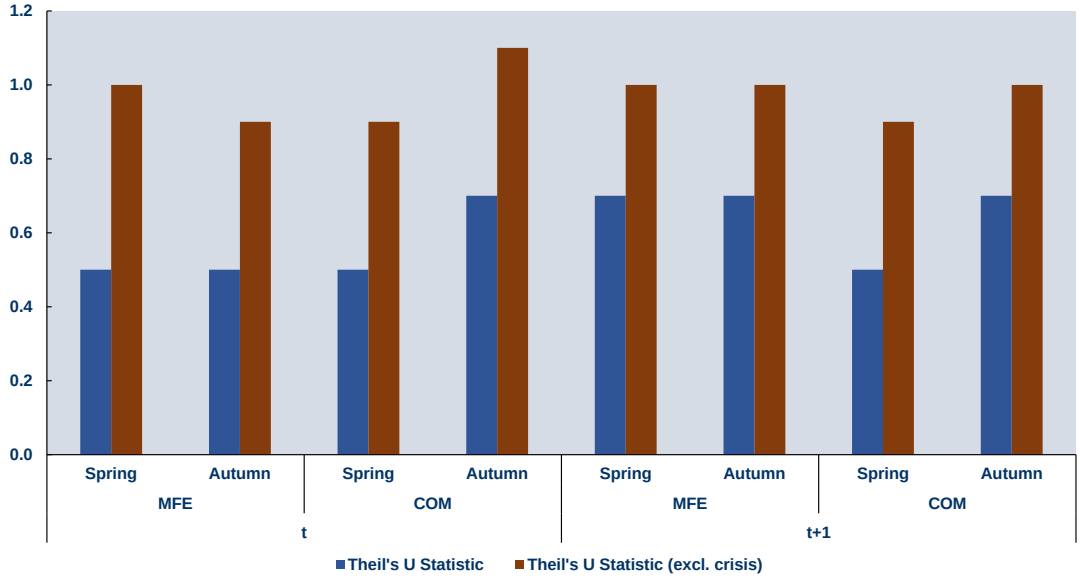


Chart 2.9: Theil's U statistic of nominal GDP, comparison with COM



Indeed, the RMSE resulting from MFE forecasts when compared to that of a naïve, MA(2) and MA(3) forecast also show that the forecast by MFE for both nominal and real GDP are generally better at estimating current year forecasts, denoted by a lower RMSE in year t (See Chart 2.10 and Chart 2.11). However, the RMSE of MFE from its one-year ahead forecast is higher than the RMSE of the simple models estimated, especially the naïve and the MA(2) forecast.

Chart 2.10: RMSE of MFE benchmarked with simple statistical models (USP)

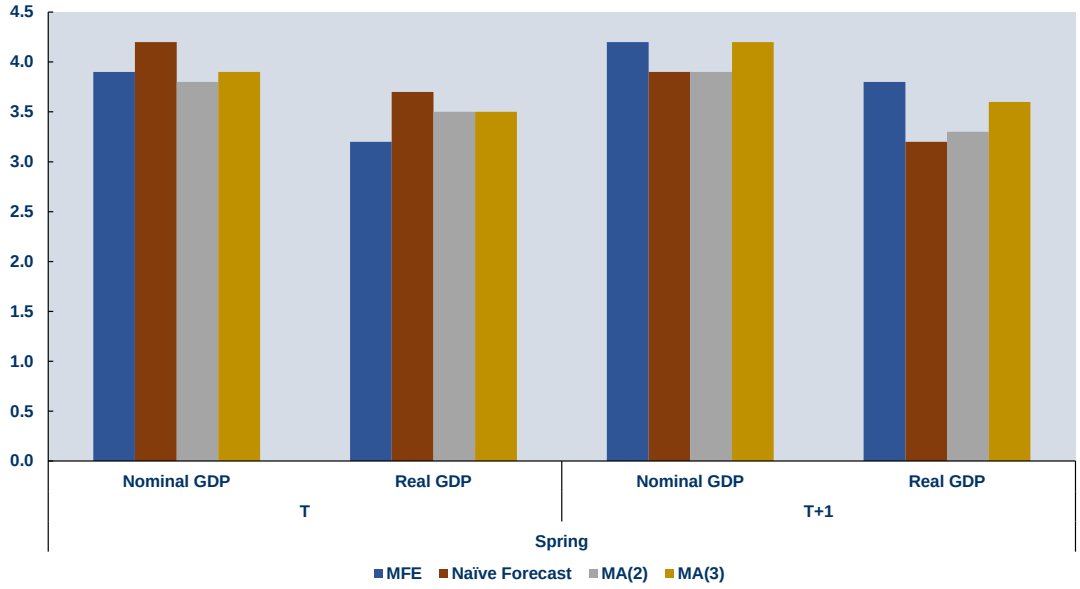
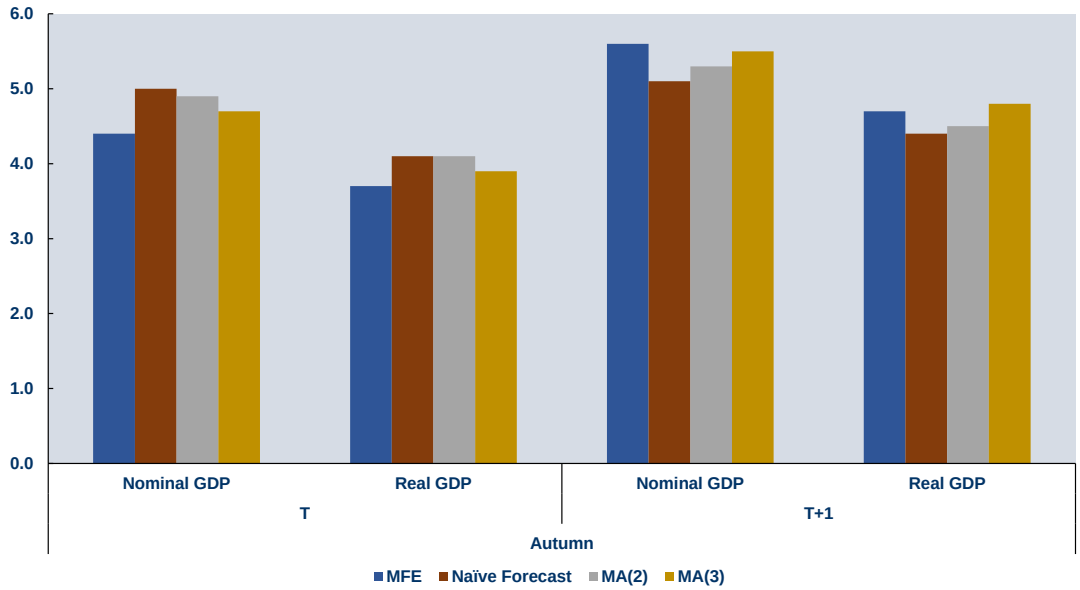
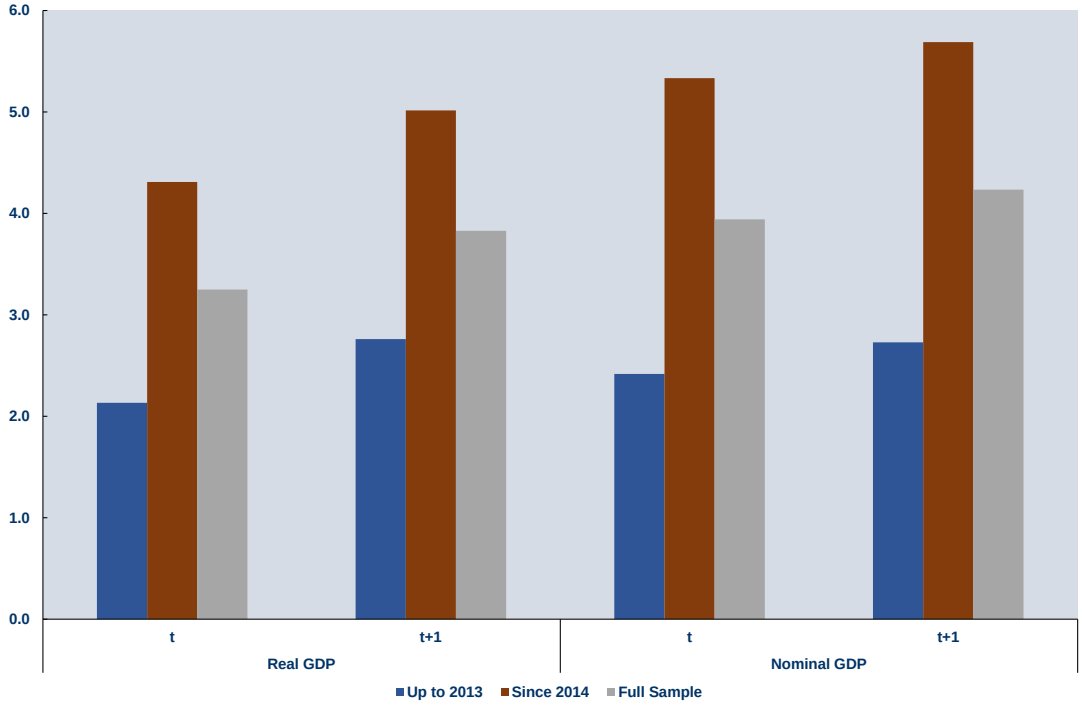


Chart 2.11: RMSE of MFE benchmarked with simple statistical models (DBP)



The accuracy of MFE forecasts is also evaluated by analysing how the RMSE changed over different sample periods. Chart 2.12 shows the RMSE for both real GDP and nominal GDP for years t and t+1 over different sample periods. Results from this simple analysis show that the RMSE has almost doubled in the sample from 2014 onwards, suggesting that forecasts by MFE have become less accurate over the more recent period.

Chart 2.12: Accuracy of MFE current year and one-year ahead forecasts, comparison over time (Stability Programme) - RMSE



2.4.2 Test for forecast unbiasedness

Results for forecast unbiasedness are presented in Table 2.3. This table presents regression results for MFE for both the Update of Stability/Convergence Programme and the Draft Budgetary Plan forecast round across different samples periods (2004 – 2013, 2014 – 2021 and 2004 – 2021 for the Update of Stability/Convergence Programme forecast round and 2013 – 2021 for the Draft Budgetary Plan forecast round) and includes results with crisis and without crisis years. Looking at the results for both real and nominal GDP, significant instances of bias (based on a confidence interval at the 90% level) can be detected. Indeed, the implication is that overall forecasts appear to have been systematically too low.

When analysing the unbiasedness in the Update of Stability/Convergence Programme forecast round, we find that when considering the full sample without excluding the crisis periods, the forecast for year t tended to be biased downwards for both real and nominal GDP, whereas the forecasts for year t+1 to year t+3 did not feature any bias. On the other hand, considering the sample 2004 – 2013, real GDP had been biased downwards in both years t and t+1, whereas the more recent sample shows that both nominal and real GDP tended to be underestimated in year t. Results for the Draft Budgetary Plan forecast period show a similar result whereby over the whole sample both real and nominal GDP were underestimated in year t, with no bias detected in year t+1.

Table 2.3 Test for Unbiasedness

USP	MFE				MFE (excl. crisis)			
Forecast period	**t**	**t+1**	**t+2**	**t+3**	**t**	**t+1**	**t+2**	**t+3**
Real GDP								
Full Sample	-2.03***	-1.74	-1.80	-1.90	-1.92**	-2.69*	-3.08**	-3.23**
	(0.00)	(0.13)	(0.15)	(0.14)	(0.02)	(0.07)	(0.01)	(0.01)
2004/13	-1.38**	-1.63*	-1.92	-1.64	-1.09*	-1.90*	-2.56*	-2.31*
	(0.01)	(0.08)	(0.16)	(0.20)	(0.09)	(0.06)	(0.08)	(0.08)
2014/21	-2.86**	-1.90	-1.60	-2.42	-3.03**	-3.96**	-4.10**	-5.67*
	(0.04)	(0.45)	(0.54)	(0.45)	(0.03)	(0.01)	(0.02)	(0.06)
Nominal GDP								
Full Sample	-1.98**	-1.68	-1.88	-2.17	-1.86	-2.63**	-3.15**	-3.68**
	(0.03)	(0.19)	(0.17)	(0.13)	(0.10)	(0.03)	(0.02)	(0.01)
2004/13	-1.04	-1.41	-2.12	-2.16	-0.64	-1.53	-2.69	-2.95*
	(0.12)	(0.14)	(0.19)	(0.15)	(0.42)	(0.19)	(0.14)	(0.07)
2014/21	-3.15*	-2.06	-1.47	-2.18	-3.48*	-4.40**	4.08**	-5.63*
	(0.07)	(0.46)	(0.58)	(0.51)	(0.06)	(0.02)	(0.03)	(0.05)
DBP	**MFE**				**MFE (excl. crisis)**			
Forecast period	**t**	**t+1**	**t+2**	**t+3**	**t**	**t+1**	**t+2**	**t+3**
Real GDP								
Full Sample	-2.99**	-1.91	na	na	-3.03**	-3.58**	na	na
(2013/21)	(0.01)	(0.38)			(0.01)	(0.01)		
Nominal GDP								
Full Sample	-3.39**	-2.35	na	na	-3.51**	-4.18**	na	na
(2013/21)	(0.01)	(0.34)			(0.01)	(0.02)		

Note: p-value in parenthesis; a p-value of less than 0.01 indicates the presence of bias at the 99% confidence interval (***), a p-value of less than 0.05 indicates the presence of bias at the 95% confidence interval (**), while a p-value of less than 0.1 indicates the presence of bias at the 90% confidence interval (*).

It is interesting to note that the results change completely when excluding the crisis period years from our sample. Indeed, in the Update of the Stability/Convergence Programme forecast period, real GDP has been biased downwards across all samples

and forecast years from t to t+3. On the other hand, in the case of nominal GDP over the full sample, we find no bias in year t, but year t+1 to year t+3 forecasts appear to have been biased downwards. Also, this bias is mostly stemming from the most recent period (2014 – 2021). The Draft Budgetary Plan forecast period shows comparable results, whereby both year t and t+1 forecasts had been biased downwards.

The increased biasedness noted in the forecasts by MFE when excluding the crisis years is to a certain extent understandable. Indeed, a reason for this result could be that since in general the MFE tend to underestimate their forecasts for nominal and real GDP, the inclusion of years where the actual data would have been lower due to a crisis year could mean that the MFE forecast would have been closer to actual data and thus the element of biasedness would decrease.

2.4.3 Statistical data revisions

It is important to note that statistical revisions can have a considerable influence on both forecast accuracy and assessing forecast unbiasedness. Indeed, both real and nominal GDP have undergone several revisions across the years as can be visually identified from Chart 2.13 and Chart 2.14, whereby the bars indicate the latest data of real and nominal GDP growth while the points indicate the growth rates of that same year which had been indicated in previous releases.

Chart 2.13: Real GDP revisions across NSO releases (% growth)

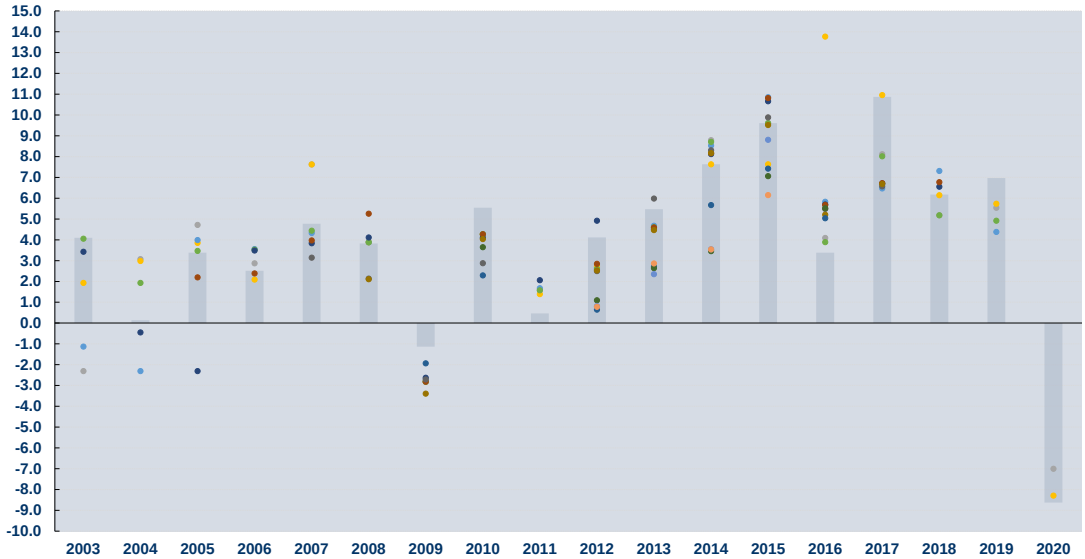
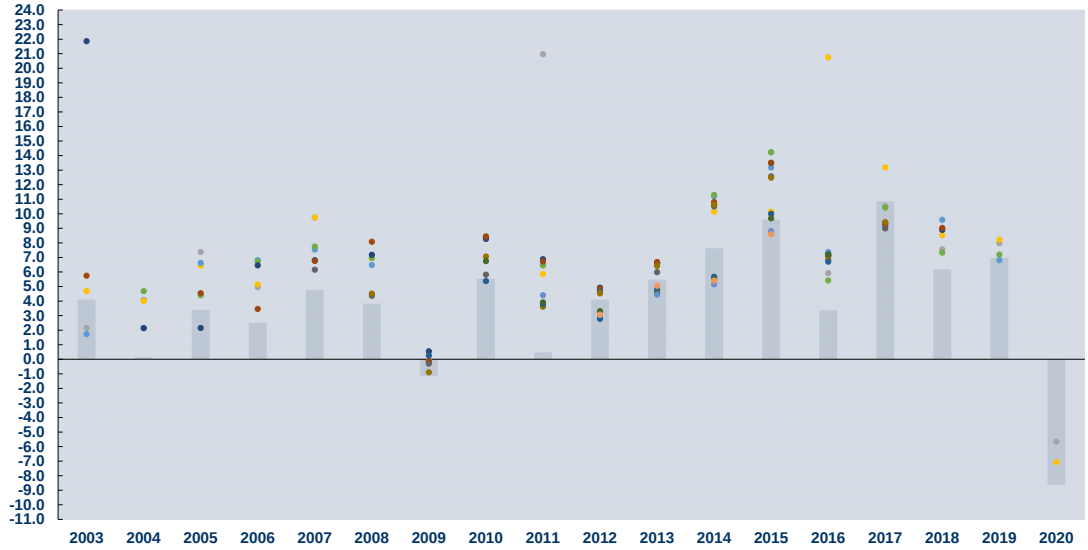


Chart 2.14: Nominal GDP revisions across NSO releases (% growth)

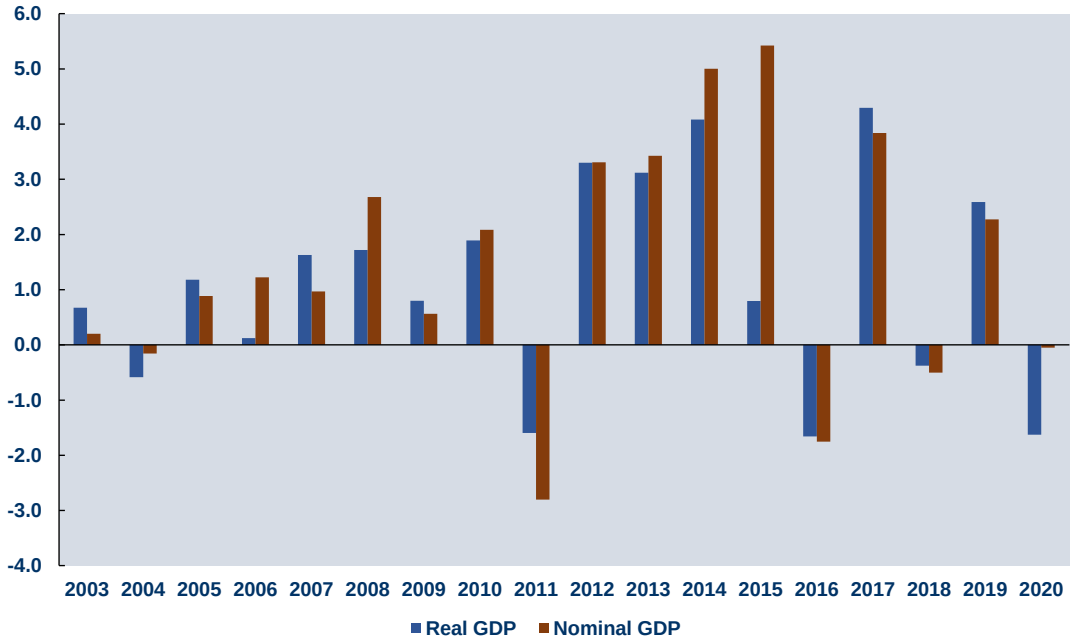


Comparing the real growth rate for t-n as reported in year t and the actual growth rate for that same year as reported in the latest national accounts news release, we note that most revisions have tended to be upwards (See Chart 2.15). In fact, we find evidence of systematic bias at the 99% confidence interval.[22] This means that in most cases the latest data release shows a higher growth than what was available at the time the forecasting exercise was carried out. As shown in Chart 2.15, this difference is quite notable, especially in a number of years in the latter half of the period under review.

For instance, growth in 2015 for nominal GDP was estimated at 8.8% in the March 2016 GDP release. However, the latest news release reports that growth in 2015 was 14.2%, an upward revision of 5.4 percentage points (pp). Indeed, in this case, forecasts carried out in 2016 were based on a weaker base of growth in 2015. On the other hand, although much less frequent, there are also instances where actual growth was revised downwards. For instance, real GDP growth in 2020 was recorded at -7.0% in the March 2021 GDP release. Whereas the more recent actual data shows that growth in real GDP actually fell by 8.6% in 2020, a downward revision of 1.6pp.

[22] This was calculated by applying the unbiasedness test on the statistical revisions.

Chart 2.15: Latest NSO data release less the NSO release available at the time of the USP (real GDP and nominal GDP growth statistical revisions) - pp



Since statistical revisions are sizable and tend to be inclined on the upside this may be a contributing factor in the bias and accuracy results previously identified. It is also interesting to note that the current year forecast errors of both real and nominal GDP for year t are also found to be inversely related to the statistical revision of real and nominal GDP growth for year t-1, with a negative coefficient of 0.55 and 0.53, respectively (See Chart 2.16 and Chart 2.17). This relatively high inverse correlation indicates that statistical revisions are indeed affecting the accuracy of the forecasts. The extent to which forecast errors have been affected by statistical revisions will be explored further in the forthcoming publication by the MFAC in its working paper series.

Chart 2.16: Real GDP – statistical revisions and forecast accuracy

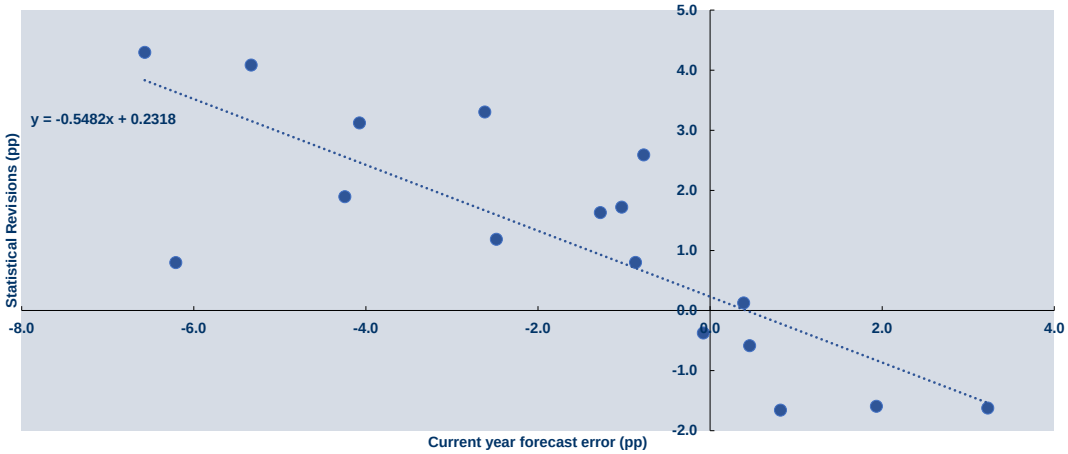
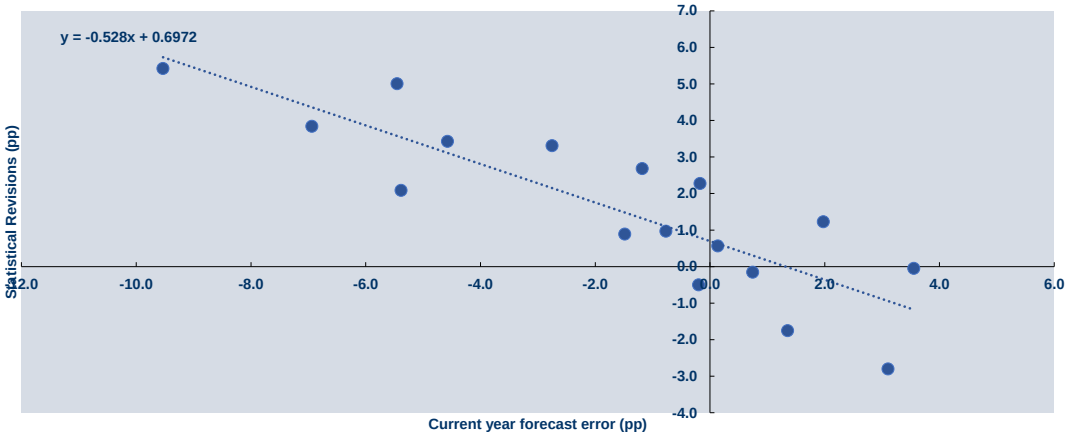


Chart 2.17: Nominal GDP – statistical revisions and forecast accuracy



2.5 Conclusion

The results presented in this chapter indicate that overall, real and nominal GDP growth forecasts produced by the MFE are generally underestimated. Indeed, although the accuracy of its forecasts is generally similar to those of other institutions, the MFE has over time experienced a slight reduction in forecast accuracy. At the same time, we find that as expected in times of crisis, the accuracy of forecasts tends to decrease. Similarly, results on the unbiasedness of forecasts show that over time, the forecasts by the MFE, especially when excluding the crisis years, appear to have been systematically too low, that is the actual data has consistently been higher than what initially was predicted by MFE both with regards to nominal and real GDP. Again, in

relation to forecast biasedness, we find that over the most recent sample, there was some increase in the downward biasedness of forecasts produced by the MFE.

Notwithstanding these results, statistical revisions which were carried out over time by the NSO, need to be taken into consideration. Indeed, a simple regression analysis shows that around half of the current year forecast errors are explained by these statistical revisions. In addition, revisions appear to have been larger across the most recent sample, which may have led to increased downward biasedness and lower accuracy of MFE forecast over time.

Such an analysis provides valuable insights for the MFAC. Indeed, regular assessments of the forecast performance of the projections produced by the MFE are key to identifying areas of improvement both in macroeconomic and fiscal projections and issuing recommendations and advice in that regard. To this end, a more detailed analysis is currently being undertaken, which will assess the different components of real GDP to identify and explore in more detail, the macroeconomic forecast performance of the Ministry for Finance and Employment.

Bibliography

[1] Bank of England. (2015). "Evaluating forecast performance". Independent Evaluation Office.

[2] Buturac, G. (2022). Measurement of Economic Forecast Accuracy: A Systematic Overview of the Empirical Literature. Journal of Risk and Financial Management 15: 1.

[3] Camilleri, G., & Vella, K. J. (2015). Interpolating forecast errors for assessing uncertainty in macroeconomic forecasts: An analysis for Malta.

[4] Celasun, O., Lee, J., Mrkaic, M, and Timmermann, A. (2021). An Evaluation of World Economic Outlook Growth Forecasts – 2004-17. IMF Working Paper, WP/21/216.

[5] Council Directive 2011/85/EU – Article 4(6) (2011). On requirements for budgetary framework of the Member States. Official Journal of the European Union.

[6] Eicher, T. S., Kuenzel, D. J., Papageorgiou, C, and Christofides, C. (2018). Forecasts in Times of Crises. IMF Working Paper, WP/18/48.

[7] Fiscal Responsibility Act (2014), Malta. Government Press. Retrieved February 19, 2015.

[8] Frankel, J. (2011). "Over-optimism in Forecasts by Official Budget Agencies and its implications". Oxford Review of Economic Policy 27(4): 536 – 562.

[9] Gupta, M, and Haris Minai, M. (2019). "An Empirical Analysis of Forecast Performance of the GDP Growth in India". Global Business Review 20(2): 368 – 386.

[10] Holden, K., and Peel, D. A. (1989). "Unbiased, efficiency and the combination of economic forecasts". Journal of Forecasting, 8(3): 175 – 188.

[11] Larch, M., Orseau, E., and Van der Wielen, W. (2021). "Do EU fiscal rules support or hinder counter-cyclical fiscal policy?". Journal of International Money and Finance 112.

[12] Sinclair, T. M., Joutz, F., Stekler, H. O. (2010). "Can the Fed Predict the State of the Economy?". Economic Letters, 108(1): 28 – 32.

Annex A

Table 2.1: Comparison of forecast errors for real GDP by institution

SPRING (USP)	MFE				EC		CBM		Naïve Forecast		MA(2)		MA(3)	
forecast period	**t**	**t+1**	**t+2**	**t+3**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**
Mean Error	-2.0	-1.7	-1.8	-1.9	-2.4	-1.9	-2.1	-1.6	-1.5	-1.1	-1.4	-1.7	-1.1	-1.7
Mean Error (excl. crisis)	*-1.9*	*-2.7*	*-3.1*	*-3.2*	*-2.2*	*-2.8*	*-2.3*	*-3.0*	*-1.2*	*-2.2*	*-1.8*	*-2.4*	*-1.7*	*-2.7*
Mean Absolute Error	2.8	4.0	4.2	4.4	3.5	4.2	3.1	4.4	4.7	3.6	4.2	4.0	4.2	4.0
Mean Absolute Error (excl. crisis)	*2.4*	*3.0*	*3.4*	*3.6*	*3.2*	*3.2*	*3.0*	*3.4*	*3.1*	*2.4*	*2.8*	*2.4*	*3.0*	*2.7*
Root Mean Square Error	3.7	5.4	5.4	5.4	4.3	5.5	3.7	5.5	6.6	5.2	5.7	5.8	5.4	5.6
Root Mean Square Error (excl. crisis)	*3.2*	*3.8*	*4.2*	*4.4*	*4.0*	*4.1*	*3.6*	*4.1*	*3.7*	*3.2*	*3.5*	*3.3*	*3.5*	*3.6*
Theil's U statistic	0.5	0.7	0.7	0.7	0.5	0.7	0.5	0.7	0.9	0.7	0.8	0.8	0.8	0.8
Theil's U statistic (excl. crisis)	*0.9*	*1.1*	*1.2*	*1.2*	*1.0*	*1.0*	*1.0*	*1.0*	*1.1*	*0.9*	*1.0*	*1.0*	*1.0*	*1.0*
Mean Relative Absolute Error	0.5	0.8	0.8	0.8	0.6	0.7	0.5	0.8	0.9	0.7	0.8	0.8	0.8	0.8
Mean Relative Absolute Error (excl. crisis)	*0.7*	*1.0*	*1.1*	*1.2*	*0.9*	*0.9*	*0.9*	*1.0*	*0.9*	*0.8*	*0.8*	*0.8*	*0.9*	*0.9*

AUTUMN (DBP)	MFE				EC		CBM		Naïve Forecast		MA(2)		MA(3)	
forecast period	**t**	**t+1**	**t+2**	**t+3**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**
Mean Error	-3.0	-1.9	na	na	-2.9	-2.3	-2.8	-1.9	-2.7	-1.3	-2.2	-2.3	-1.6	-2.1
Mean Error (excl. crisis)	*-3.0*	*-3.6*	na	na	*-3.0*	*-3.8*	*-2.9*	*-3.6*	*-2.5*	*-3.2*	*-2.9*	*-3.3*	*-2.6*	*-3.5*
Mean Absolute Error	3.4	5.2	na	na	3.4	5.6	3.2	5.1	6.4	5.2	5.5	6.0	5.3	5.9
Mean Absolute Error (excl. crisis)	*3.2*	*3.6*	na	na	*3.2*	*3.9*	*3.2*	*3.6*	*3.5*	*3.2*	*3.3*	*3.5*	*3.2*	*3.5*
Root Mean Square Error	4.0	6.5	na	na	4.0	6.9	3.8	6.4	8.8	7.1	7.2	7.9	7.0	7.7
Root Mean Square Error (excl. crisis)	*3.7*	*4.7*	na	na	*3.7*	*4.7*	*3.7*	*4.6*	*4.1*	*4.4*	*4.1*	*4.5*	*3.9*	*4.8*
Theil's U statistic	0.4	0.7	na	na	0.4	0.7	0.4	0.6	1.0	0.7	0.8	0.8	0.7	0.8
Theil's U statistic (excl. crisis)	*0.9*	*1.0*	na	na	*0.9*	*1.0*	*0.9*	*1.0*	*1.0*	*1.0*	*1.0*	*1.0*	*0.9*	*1*
Mean Relative Absolute Error	0.5	0.7	na	na	0.5	0.8	0.4	0.7	0.9	0.7	0.7	0.8	0.7	0.8
Mean Relative Absolute Error (excl. crisis)	*0.8*	*0.9*	na	na	*0.8*	*1.0*	*0.8*	*0.9*	*0.9*	*0.8*	*0.8*	*0.9*	*0.8*	*0.9*

Table 2.2: Comparison of forecast errors for nominal GDP by institution

SPRING (USP)	MFE				EC		CBM		Naïve Forecast		MA(2)		MA(3)	
forecast period	**t**	**t+1**	**t+2**	**t+3**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**
Mean Error	-2.0	-1.7	-1.9	-2.2	-2.2	-1.8	na	na	-1.9	-1.5	-1.7	-1.9	-1.4	-1.9
Mean Error (excl. crisis)	*-1.8*	*-2.6*	*-3.1*	*-3.7*	*-2.1*	*-2.8*	na	na	*-1.7*	*-2.8*	*-2.2*	*-2.7*	*-2.2*	*-3.1*
Mean Absolute Error	3.2	4.3	4.6	5.0	2.8	3.9	na	na	5.1	4.1	4.4	4.5	4.4	4.6
Mean Absolute Error (excl. crisis)	*2.9*	*3.1*	*3.7*	*4.2*	*2.5*	*3.0*	na	na	*3.3*	*2.8*	*2.8*	*2.8*	*3.1*	*3.2*
Root Mean Square Error	4.2	5.8	5.8	6.0	3.5	5.0	na	na	7.1	5.7	6.1	6.3	5.9	6.2
Root Mean Square Error (excl. crisis)	*3.9*	*4.2*	*4.7*	*5.0*	*3.2*	*3.7*	na	na	*4.2*	*3.9*	*3.8*	*3.9*	*3.9*	*4.2*
Theil's U statistic	0.5	0.7	0.7	0.7	0.5	0.7	na	na	0.9	0.7	0.8	0.8	0.8	0.8
Theil's U statistic (excl. crisis)	*1.0*	*1.0*	*1.1*	*1.1*	*0.9*	*1.1*	na	na	*1.0*	*1.0*	*0.9*	*1.0*	*0.9*	*1.0*
Mean Relative Absolute Error	0.5	0.7	0.8	0.8	0.6	0.7	na	na	0.9	0.7	0.8	0.8	0.8	0.8
Mean Relative Absolute Error (excl. crisis)	*0.8*	*0.9*	*1.1*	*1.1*	*0.7*	*1.0*	na	na	*0.9*	*0.8*	*0.8*	*0.8*	*0.9*	*0.9*

AUTUMN (DBP)	MFE				EC		CBM		Naïve Forecast		MA(2)		MA(3)	
forecast period	**t**	**t+1**	**t+2**	**t+3**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**	**t**	**t+1**
Mean Error	-3.4	-2.4	na	na	-3.6	-2.8	na	na	-3.1	-1.6	-2.5	-2.7	-1.8	-2.5
Mean Error (excl. crisis)	*-3.5*	*-4.2*	na	na	*-3.7*	*-4.4*	na	na	*-3.0*	*-3.7*	*-3.4*	*-3.9*	*-3.0*	*-4.1*
Mean Absolute Error	3.6	6.0	na	na	3.8	6.3	na	na	7.2	5.6	6.0	6.6	5.8	6.4
Mean Absolute Error (excl. crisis)	*3.6*	*4.4*	na	na	*3.8*	*4.6*	na	na	*4.2*	*3.7*	*3.8*	*4.0*	*3.8*	*4.1*
Root Mean Square Error	4.5	7.3	na	na	4.7	7.6	na	na	9.5	7.6	7.8	8.6	7.5	8.3
Root Mean Square Error (excl. crisis)	*4.4*	*5.6*	na	na	*4.6*	*5.7*	na	na	*5.0*	*5.1*	*4.9*	*5.3*	*4.7*	*5.5*
Theil's U statistic	0.5	0.7	na	na	0.5	0.7	na	na	1.0	0.7	0.8	0.8	0.8	0.8
Theil's U statistic (excl. crisis)	*0.9*	*1.0*	na	na	*0.9*	*1.0*	na	na	*1.0*	*0.9*	*1.0*	*1.0*	*0.9*	*1.0*
Mean Relative Absolute Error	0.4	0.7	na	na	0.5	0.8	na	na	0.9	0.7	0.7	0.8	0.7	0.8
Mean Relative Absolute Error (excl. crisis)	*0.8*	*0.9*	na	na	*0.8*	*1.0*	na	na	*0.9*	*0.8*	*0.8*	*0.8*	*0.8*	*0.9*