BANK ĊENTRALI TA' MALTA
EUROSISTEMA
CENTRAL BANK OF MALTA

CENTRAL BANK OF MALTA
WORKING PAPER

BANK ĊENTRALI TA' MALTA
EUROSISTEMA
CENTRAL BANK OF MALTA

An in-depth assessment of the Central Bank of Malta's projections[1]

Maria Christine Saliba*

WP/05/2026

[1] Maria Christine Saliba is a Senior Economist within the Economic Projections and Conjunctural Analysis Office within the Economic Analysis Department of the Central Bank of Malta

*The author would like to thank Governor Alexander Demarco, Deputy Governor Rita Schembri, Aaron G. Grech, Gilmour Camilleri, Noel Rapa, Ian Borg and Massimo Giovannini for their useful advice and insightful discussions. The views expressed are those of the author and do not necessarily reflect the views of the Central Bank of Malta or any other institutions. Any errors are the author's own.

Any queries or requests should be sent here.

ISSN 3134-707X (online)

Abstract

This paper presents an in-depth analysis of the performance of the macroeconomic projections produced by Central Bank of Malta (CBM) staff. A real-time macroeconomic database published by the CBM is utilized, allowing for the computation of forecast errors based on the first official data release. This provides a more just assessment of the forecast errors produced, than alternative measures of errors based on later releases, particularly in a context of substantial data revisions to Maltese data. A series of statistical tests are performed, evaluating the accuracy and optimality of the CBM projections. The main findings indicate that generally labour market and inflation data display the strongest accuracy results. These results vary across horizons, with the outer years marked by a deterioration in performance. The forecasts for headline inflation generally demonstrate informational efficiency, with forecast errors not being systematically predictable from past data or errors. Real GDP forecasts show slight inefficiency at the 10% significance level, primarily driven by distortions during the COVID-19 period, which is not detected when this period is excluded. Overall, the projections effectively incorporate available information. However, the uncertainty brought about by the COVID-19 pandemic had a marked impact on the accuracy and efficiency of the projections produced by the CBM. This was most evident for real GDP forecasts.

JEL Classification: C53, E37, E58

Keywords: CBM projections, forecast evaluation, forecast errors

1. Introduction

This paper presents an evaluation of the performance of the macroeconomic projections produced by the Central Bank of Malta (CBM). The analysis focuses on projection rounds published between 2007-2023. During the period 2007 to 2017 the CBM produced two projection updates annually, before transitioning to a quarterly schedule from 2018 onwards.[2] The projections considered are annual growth rates of a full year of key macroeconomic variables. Notably, CBM focuses on annual rather than quarterly estimates as the volatility and revisions in Maltese national accounts data limit the reliability of quarterly figures. The analysis covers annual growth rates for the current year and up to two years ahead.[3]

The CBM's forecasting process aims to generate coherent, clear, and credible forecasts that serve as inputs for the Eurosystem's Broad Macroeconomic Projection Exercise (BMPE), the Narrow Inflation Projection Exercise (NIPE), and the Bank's routine publications. The macroeconomic projections produced by the CBM are conditional forecasts as they are built on a set of technical assumptions provided by the European Central Bank (ECB), which are shared with all national central banks in the euro area as part of the BMPE and NIPE processes.

The work presented in this paper is an important step within the forecasting process as it enables an in-depth assessment of the forecasting performance of the CBM's projections in terms of their accuracy and reliability. Indeed, the importance of assessing forecast accuracy has been prominent in the literature since the early 1970s, including among others the work done by Mincer & Zarnowitz (1969) who put forward a set of criteria and methods for the assessment of economic forecasts. Among the literature on forecast evaluations, generally, focus is put on evaluating forecast errors via a set of statistical properties. In this regard, this paper follows a similar approach and is inspired in particular by the recent work done by Kontogeorgos & Lambrias (2019) who present an evaluation of the Eurosystem macroeconomic projections, focusing on real GDP growth and HICP inflation. Other studies, such as Bank of England (2015) and Davison, Camilleri, & Spiteri (2024) also served as an initial input to the work done in this paper.

In conducting this forecast evaluation, I compute forecast errors at each horizon over the forecast period and assess them using a number of statistical tests. In total, I collect the forecasts and data vintages of 16 variables relating to the real economic activity, labour market and prices. The statistical tests performed assess the accuracy and optimality of the forecasts produced by the CBM. The latter requires forecasts to be weakly efficient and unbiased. On one hand weak efficiency implies that all information available at the time of the forecast concerning the variable being predicted is fully incorporated into the forecast. On the other hand, projection errors should average zero overtime, reflecting the concept of unbiasedness.

A fair analysis of the CBM forecasting performance needs to account for the sizable revisions to Maltese National Accounts data. Indeed, Grech (2018), finds that such revisions are typical

[2] This does not apply to prices forecasts as these have been generated every quarter under the Narrow Inflation Projection Exercise.

[3] Notably, over the period considered throughout this forecast evaluation, changes have been made to the models used to produce forecasts and therefore this analysis considers projections generated by the real-time models available.

and are significant in both magnitude and volatility when compared to the initial release. These findings were confirmed by Debono & Mock (2024), who extended this study until 2024Q2. In addition to confirming the earlier results, they document a marked increase in revisions in aggregate real GDP, with revisions to exports and imports being the most prominent contributors.

In order to limit the impact that large revisions in Maltese national accounts data might have on forecast evaluation results, the exercise documented in this paper is based on real-time data. In particular, I compute forecast errors as the difference between the projected and actual outcomes as published in the Q1 data vintage released in the following year. In this respect, I use the real-time macroeconomic database published by the CBM and that originated from Grech (2018).

This approach differs from a recent study by Davison, Camilleri, & Spiteri (2024), which computes forecast errors based on the latest data release available at the time the study was conducted. The approach undertaken throughout this evaluation ensures a fairer and more realistic assessment by reflecting the information available to policymakers at the time the forecast was made and avoids hindsight bias.

This paper is structured as follows. Section 2 gives a brief overview of the forecasting process of the CBM, followed by a detailed description of the data and methodology used throughout this forecast evaluation. Section 4 provides an initial descriptive analysis of the forecast errors and how CBM's projections compare to the data outcomes. Sections 5 and 6 present the main results from the statistical tests performed as well as a decomposition of projection errors into the contributions from the main sub-components of real GDP growth and HICP. Section 7 compares the CBM's forecast errors to those produced by a benchmark model and to the forecast errors of other institutions. This is then followed by an analysis of the impact of technical assumptions and data revisions on CBM's projection errors in sections 8 and 9 respectively, while section 10 concludes.

2. CBM's Forecasting Process

The CBM utilises a structured macroeconomic forecasting process divided into three stages: preparation, projection, and evaluation. This process, as outlined by Borg, Farrugia, & Ellul (2016), aims to produce reliable forecasts that serve as input to the BMPE, NIPE and the CBM's publications.

The preparation stage lays the groundwork for the forecasting process by managing and updating a comprehensive database of variables for model-based and off-model projections. This stage involves evaluating the sources of recent forecast errors to enhance accuracy and assessing one-off factors, or "news," that deviate from typical business cycles. Such factors, often driven by policy or large-scale projects, are explicitly included in forecasting models. For example, energy sector projects in 2014 and 2015 significantly influenced private investment and were factored into the forecasts. Expert analysis is also crucial in this stage, leveraging economists' specialised knowledge of specific areas.

At the core of the projection stage is STREAM, the CBM's macro-econometric model, which integrates data from the preparation stage to produce detailed forecasts. The latest version of the model is STREAM version 3.2 as outlined in Borg, Cumbo, & Rapa (2024). Based on the neo-classical synthesis framework, STREAM captures short-term dynamics and long-term equilibria using error-correction equations. Additionally, satellite models complement STREAM, in areas such as inflation, fiscal performance, credit, and house prices. For instance, short-term inflation projections are prepared under the Eurosystem's Narrrow Inflation Projections Exercise (NIPE) using a dedicated set of equations outside STREAM. Similarly, other models such as a dynamic factor model as outlined in Ellul & Ruisi (2022) are used to provide short-term estimates for real GDP growth rates that also serve as input in the projections process.

Throughout the forecasting process, expert analysis complements model-based projections. CBM economists monitor specific economic variables and provide conjunctural analyses, ensuring their projections reflect both empirical data and expert judgment.

Moreover, as outlined previously, macroeconomic projections are conditioned on a set of technical assumptions provided by the ECB, which align the CBM's forecasts with the Eurosystem's BMPE and NIPE, ensuring consistency across euro area central banks. These assumptions are treated as exogenous variables and include among others the future path of foreign demand for Malta, competitors' prices for Malta on the import and export side, exchange rates, international oil prices, interest rates and food commodities.

3. Data and Methodology

3.1 Data

To construct forecast errors, I collect historical data on real GDP and its components - private consumption, government consumption, investment, exports and imports. together with labour market data, including data on the unemployment rate. With regards to prices, I collect historical data for the overall HICP index and its main subcomponents: unprocessed and processed food, non-energy industrial goods, services and energy.

Moreover, for the variables pertaining to the real economic activity with the exception of the unemployment rate, I collect actual data for each data vintage from a real-time GDP database, which is a collection of data vintages constructed by the CBM. For the purpose of this forecast evaluation, the data is transformed into year-on-year growth rates at annual frequency. Meanwhile, for the unemployment rate and inflation, the latest available data vintage is used across the entire period, as revisions to these variables are typically minimal.

Additionally, I collect annual projections for each variable and from each forecast round covering the period 2007-2023. In particular, for each round conducted in any given period t, and for each variable under consideration, I collect three years of annual forecasts, i.e. t, t+1 and t+2.

For measures of real economic activity, namely GDP and its components, as well as the unemployment rate, the projections frequency varies over time. Between 2007 and 2017, projections for these variables are available for two rounds per year, corresponding to publications in Q2 and Q4, reflecting the CBM's biannual forecasting cycle during that period. From 2018 onwards, the frequency increases to four projection rounds per year, with one publication in each quarter. Short-term inflation projections follow a different pattern, as they have been consistently available on a quarterly basis throughout the entire sample under the NIPE framework with the exception of a few missing data points. With regards specifically to GDP and its components, each round incorporates the most recently available GDP data release at that point in time. This alignment is consistent across all years from 2007 to 2023, ensuring that projections are conditioned on the latest observed economic information. A detailed mapping of each projection round to the corresponding GDP release is provided in Table 14 in the Annex.

3.2 Methodology

Given the nature of revisions in Maltese national accounts data as outlined in Grech (2018) and Debono & Mock (2024) and the potential impact these might have on the forecasting performance, this study defines forecast errors as the difference between the forecast and the first annual outturn within the Q1 data vintage released in the following year – that is the vintage which includes the first annual release for the previous year. Similar definitions are used by other studies in the literature which make use of real time datasets when performing forecast evaluations such as Kontogeorgos & Lambrias (2019). Nevertheless, revisions to historical data in between vintages could still have a marked impact on the estimated forecast error, and future work will attempt to control for this.

More formally, the h horizon projection error at annual frequency is defined as the outturn minus the forecasted value as shown in the following notation:

$$e_{th} = \left[y_t^{(t_{Q1}+1)}\right] - f_{th} \tag{1}$$

where t is the time period being forecasted, e_{th} is the h horizon forecast error for the relevant variable, $\left[y_t^{(t_{Q1}+1)}\right]$ represents the actual value of that variable for time t, using the vintage released in Q1 of the following year and f_{th} is the forecast of that variable for period t produced in the forecast rounds in year $(t - h)$. In this context, h denotes the forecast horizon, specifying how many years ahead the prediction is made. Thus, h = 0 refers to the same-year forecast, h = 1 to a one-year-ahead forecast, and h = 2 represents a two-year-ahead forecast. Given that one actual data vintage is used to calculate forecast errors across the different forecast rounds within any given year, the forecast errors presented throughout this paper represent fixed-event type of forecasts. Under this setup, the number of observations for forecast errors range from 37-61 across the three forecast horizons, with the same year horizon having the most data points.[4]

Following the computation of the projection errors, I perform a total of 6 statistical tests. These tests are performed both on the full sample as well as on an adjusted sample which excludes the years pertaining to the COVID-19 pandemic (i.e. 2020 and 2021).[5] Moreover, these statistical tests are divided into 3 groups being tests for: accuracy, unbiasedness and efficiency, the latter two being necessary conditions for optimality.

i. Accuracy Tests

The accuracy of the forecasts is assessed by measuring the standard deviations of the errors, the Root Mean Squared Error (RMSE), the mean error (ME) and mean absolute error (MAE). The three latter tests are computed as follows:

$$RMSE_h = \sqrt{1/_{Th} \sum_{t=1}^{Th} e_{th}^2} \quad (2)$$

$$ME_h = \frac{1}{Th} \sum_{t=1}^{Th} e_{th} \quad (3)$$

$$MAE_h = \frac{1}{Th} \sum_{t=1}^{Th} |e_{th}| \quad (4)$$

The ME and MAE measure the average of forecast errors and the average of the absolute errors respectively, identifying any tendency to overestimate/underestimate the projected variable as well as assessing the magnitude of the errors.

ii. Optimality

The optimality of the projections is obtained by minimising a given loss function, in this case a quadratic loss function as shown in equation 5:

$$\textit{Mean Squared Error loss:}\ L(e_{th}) = ae_{th}^2 \quad (5)$$

[4] The number of observations differs between macroeconomic data and prices data. The latter holds the most observations as forecasts are available across all quarters since 2007Q3, with the exception of 5 missing data points. Meanwhile, for macroeconomic variables, two forecasts per year are available from 2007-2017, and four forecasts per year are available from 2018-2023. For macroeconomic data the number of observations stands at 45, 41 and 37 for h=0, h=1 and h=2 respectively. Meanwhile, the number of observations for inflation data stands at 61, 48 and 42 for h=0, h=1 and h=2 respectively.

[5] The COVID-19 pandemic period was characterised by considerable revisions. Revisions were also evident during the 2008-2009 financial crisis period, however the pandemic period was unprecedented and resulted in uncertainty and volatility that far exceeded those observed during the financial crisis period. For this reason, 2020 and 2021 were excluded while the financial crisis period was not excluded from the sample.

where e represents the forecast error. Based on this quadratic loss function, projections reach optimality when they display three properties being unbiasedness, weak efficiency and the forecast error's variance is a non-decreasing function of the forecast horizon as outlined by Elliott & Timmermann (2016).

1. Unbiasedness

Unbiasedness suggests that on average, forecast errors are expected to be zero both conditionally and unconditionally:

$$E[e_{th}|\Omega_t] = 0 \quad (6)$$

$$E[e_{th}] = 0E[e_{th}] = 0 \quad (7)$$

where Ω is the information set available to the forecaster when the prediction is made. In other words, the forecasts should neither systematically overpredict nor underpredict the actual outcome. In order to test for unbiasedness, I follow the work undertaken by Ager et al. (2007), who propose a framework whereby forecasts for each variable are pooled together across horizons. This approach builds on the work of Clements et al. (2007) and is particularly suited for this study given that the projection errors presented here represent a fixed target type of forecasts. The latter causes errors to be serially correlated, and this approach allows me to correct the variance covariance matrix for the serial correlation of the projection errors. In implementing this approach, I test whether forecast errors are collectively unbiased across all horizons, starting off by estimating a common bias α:

$$e = iTH\alpha + v \quad (8)$$

Where e is the stacked vector of forecast errors, α represents the common bias parameter, iTH is a TH x 1 vector of ones that enables the model to estimate a single common intercept α across all horizons and time periods and v is the disturbance term. Subsequently the following null hypothesis is tested:

$$H0: \alpha = 0 \quad (9)$$

In addition, horizon-specific biases are jointly tested using a Wald test as follows:

$$H_0: a_1 = a_2 = \cdots a_H = 0 \text{ , for } 1 \leq h \leq 12 \quad (10)$$

Failing to reject the null hypothesis would indicate that forecasts are unbiased.

2. Weak Efficiency

Weak efficiency reflects a scenario where forecasts fully account for all available historical information, such as past forecast errors and previously observed data. Therefore, as outlined in Gavin and Mandal (2003), forecasts are considered informationally efficient when they take into account all recent data and previous forecast errors and when, therefore, past forecast errors and data have no explanatory power on the dependent variable, i.e. the forecast error. This implies that any patterns, trends, or anomalies in previous forecasts or actual values should not provide additional predictive power for future errors. Therefore, weak efficiency ensures that forecasts are not biased by overlooked historical information and are thus as accurate as possible given the available data.

To test for weak efficiency, forecast errors at the same-year horizon (h=0) are regressed on lagged forecast errors and lagged data represented by x_t in equation 12. The forecast errors in equation 12 reflect an average of same-year horizon forecast errors across the forecast rounds conducted by the CBM within one year. The lagged data is taken from the first quarter data release which represents the first release of annual observed data. Subsequently, the following null hypothesis is tested using HAC standard errors:

$$H0 : \gamma = 0 \tag{11}$$

The null hypothesis is tested through the following equation:

$$e_{t0} = a + \gamma x_t + u_t \tag{12}$$

3. Non-decreasing variance

This property is linked with the accuracy tests and highlights the fact that as the forecast horizon increases, the accuracy of the projections typically worsens. This property may be assessed through the RMSE and standard deviation of forecast errors as outlined above.

iii. Efficiency

To assess broader forecast efficiency beyond weak efficiency, two regressions were employed based on Blanchard and Leigh (2013) and Kanngiesser and Willems (2025). By going beyond weak efficiency, I am able to assess whether forecasts incorporated available information about key economic drivers in a way that is consistent with rational expectations. These drivers represent key underlying components or related indicators that shape the evolution of the forecasted variable, making them crucial for testing whether forecast errors stem from systematically neglecting relevant economic information. First, forecast errors were regressed on contemporaneous forecasts of key driver variables using robust linear regression

$$e_t = a + \beta' X_t^f + \varepsilon_t \tag{13}$$

where*:*

- e_t is the forecast error at time t,
- X_t^f is a vector of forecasts for relevant driver variables made at time t,
- β′ captures the contribution of each forecasted driver to the forecast error.

This specification tests whether the forecast errors can be systematically explained by information available at the time the forecast was made, which would imply inefficient use of available data. Robust regression is applied to reduce sensitivity to outliers, such as those occurring during the COVID-19 pandemic. A limitation of this approach is that it uses contemporaneous forecasts of the driver variables, whereas Kanngiesser and Willems (2025) rely on 2-quarter ahead forecasts of X. This may introduce horizon inconsistency, as forecasts can incorporate information arriving within the year, implying that estimated relationships may partly reflect differences in information sets rather than purely systematic inefficiencies. However, to maintain a consistent information set, in this paper I rely on contemporaneous forecasts which are taken from the first available forecast round of each year only, ensuring they remain fully ex-ante and avoiding contamination from within-year information updates. This approach is in line with the spirit of Olivier Blanchard and Daniel Leigh (2013) and Johannes Kanngiesser and Tim Willems (2025).

Second, Mincer-Zarnowitz (MZ) regressions were estimated for each driver variable to assess their forecast efficiency:

$$x_t = a + \delta x_t^f + \varepsilon_t \tag{14}$$

where:

- x_t is the actual realization of a driver variable at time t,
- x_t^f is its forecast made at time t,
- δ indicates the efficiency of the forecast.

A significant deviation of δ from 1 suggests biased or inefficient forecasts. A Wald ratio as outlined in equation 15 is also computed and its standard error is obtained using the delta method to assess the statistical significance of the ratio itself. The Wald ratio corrects the efficiency measure β with any bias implied by a δ ≠ 1 and provides a measure of whether forecasters correctly assess the pass-through from a given driver variable to the target variable, after correcting for any bias in the forecasts of that driver. A positive ω indicates that higher forecasts of the driver variable are associated with higher-than-expected outturns of the dependent variable, suggesting that the forecasting model underestimates the strength of the relationship. Conversely, a negative ω implies that higher forecasts of the driver are followed by lower-than-expected outturns, indicating an overestimation of the pass-through. A value of zero is consistent with efficient forecasts, where the impact of the driver variable is correctly incorporated.

$$\omega = \beta' / \delta \tag{15}$$

4. Assessing Forecast Errors

This section presents the forecast errors as well as a comparison of the forecasts with the outturn. At each data vintage, which ranges from 2007 to 2023, the figures below display the forecasts and the outturn for that particular year. The outturn is the data which became available in the first quarter of t + 1, where t is the time-period being forecasted. Moreover, the bars represent the range of forecasts for that year and include the median across the range of forecasts relating to horizon 0. In the earlier forecast rounds, projections were published twice a year, in June and December. Therefore, in the period 2008-2017 (for real economic activity and labour market variables) and 2007-2010 (for inflation variables), the range showed in the charts includes only 2 forecast errors instead of 4.

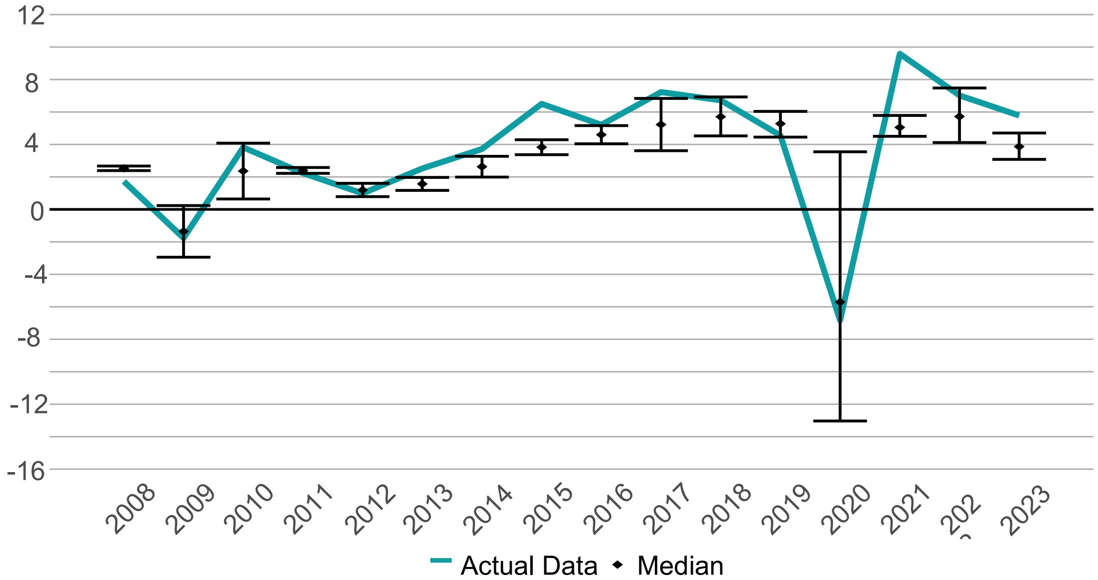
Similarly, the plot of the forecast errors ranges from 2007 to 2023, and the figures display the error bars for each particular year. The error bars represent the range in the forecast errors for that year and include the mid-point across the range of errors. Just like the forecasts, the forecast errors may vary from 2 to 4 data points depending on the number of forecast rounds produced each year.

Figures 1 till 3 show forecasts, forecast errors and actual data for real GDP, unemployment and HICP inflation respectively. The upper panel for each of the three figures compares the range of forecasts for each given year to the actual outturn of the variable under consideration while the second panel shows the distribution of forecast errors across time.

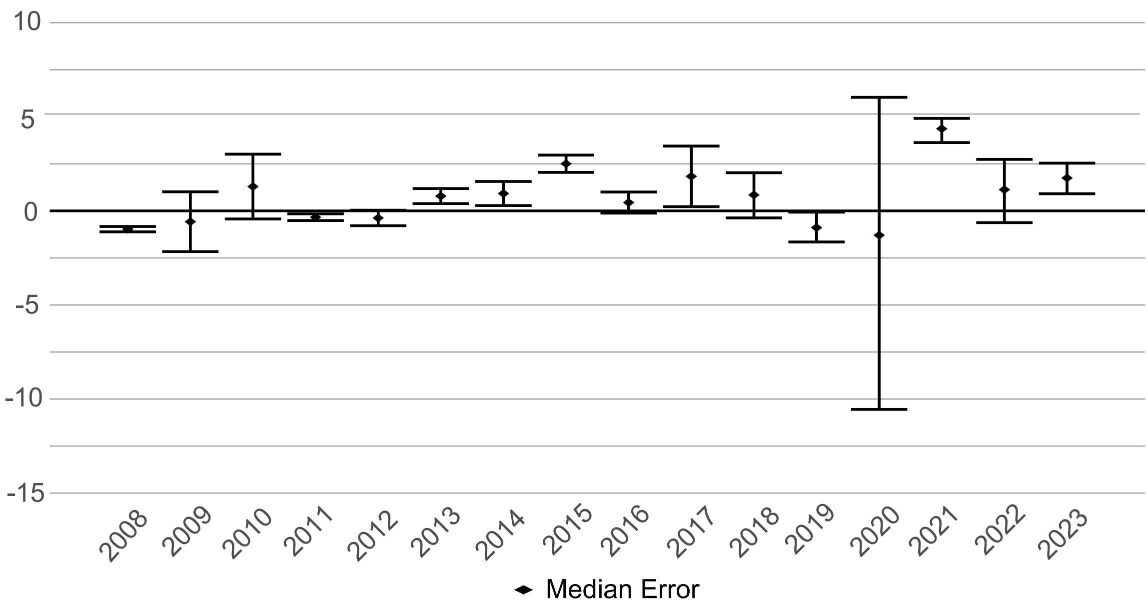
Starting from GDP results, Figure 1 shows that until 2014, the variability in the forecasts was relatively low and not too distant from the actual outturns registered during that period. Conversely, since 2017, the range of forecasts produced during each year was relatively wide, suggesting significant re-assessments from one forecast vintage to the next. In part this might reflect positive surprises in macroeconomic outcomes which led to very strong economic growth registered during this period. It is also interesting to note that since 2013 - except for the first pandemic year in 2020 - the actual data outturns often exceeded even the most optimistic projection, indicating that GDP growth projections have mostly underestimated the actual outturns.

Figure 1 Real GDP forecasts, forecast errors and actual data for same-year forecasts

a) Distribution of forecasts for GDP
(annual percentage change)



b) Distribution of forecast errors for GDP
(percentage point)



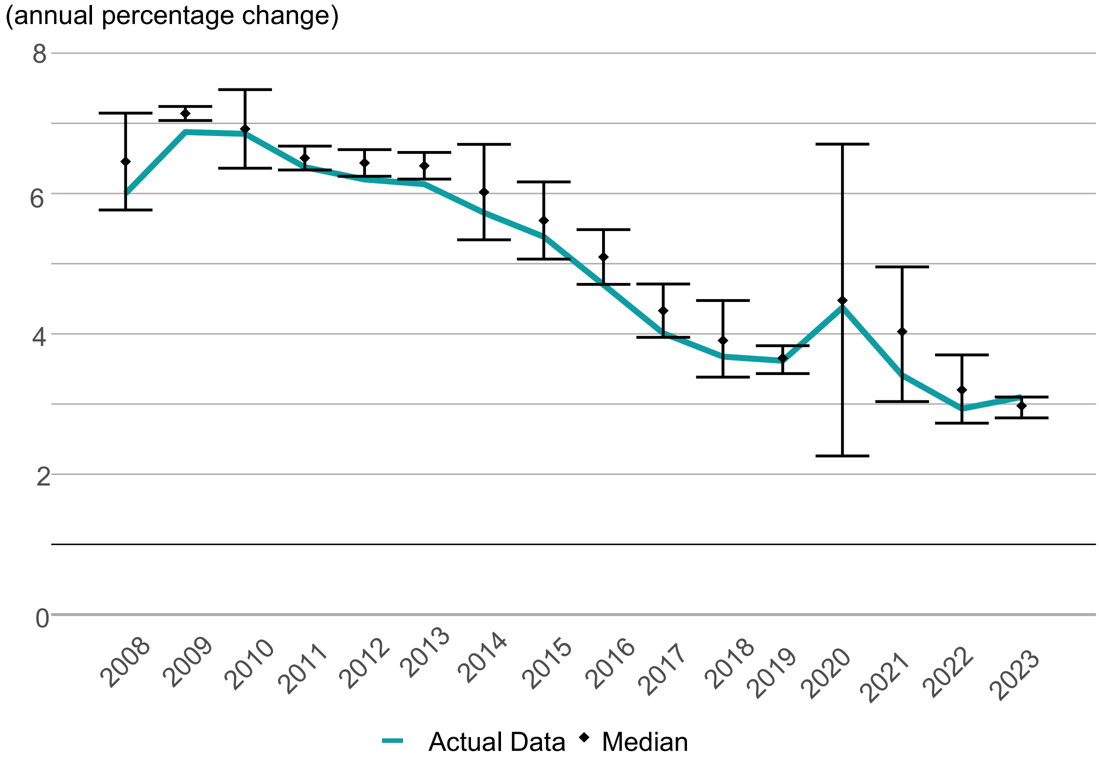
Source: Panel a): CBM projections, CBM real-time GDP database. Panel b): Author's calculations

Notes: Panel a): The bars represent the range in the forecasts conducted in that particular year. From 2008-2017, the range of forecasts includes 2 data points, from 2018-2023, the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t+1), where t is the time-period being forecasted. Panel b): The error bars represent the range in the forecast errors computed for that particular year. From 2008-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.

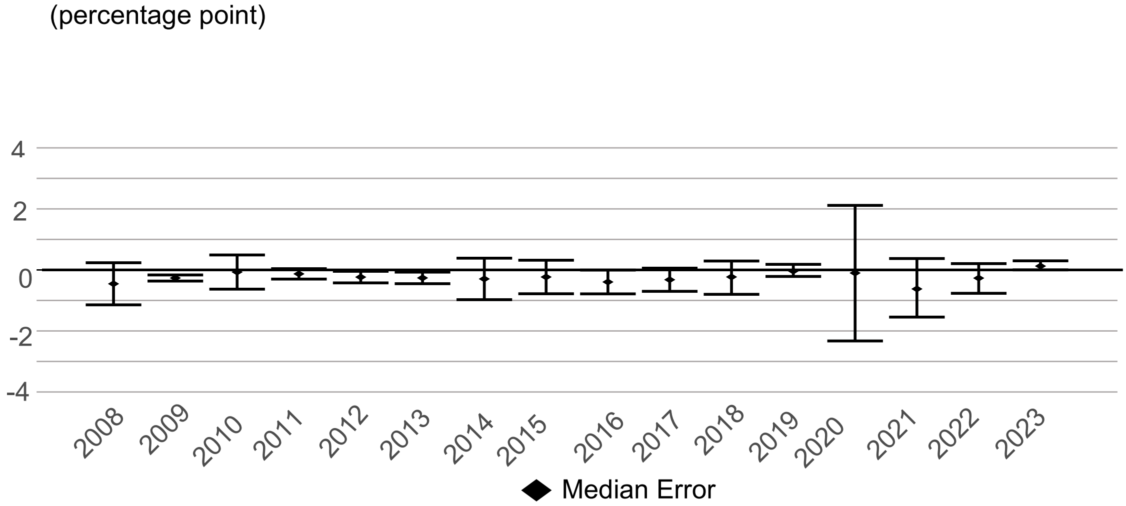
Turning to results for the unemployment rate, figure 2 shows that errors are generally smaller than those for GDP growth, indeed this variable is not revised as frequently as GDP and has less intra-year volatility. Nonetheless, the unemployment rate has been overestimated throughout the whole horizon according to the median point forecast. This is consistent with the general tendency shown in figure 1 to underestimate economic activity. However, the overestimation of the unemployment rate occurred also prior to 2014, despite generally low projection errors in GDP.

Figure 2 Unemployment rate forecasts, forecast errors and actual data for same-year forecasts

a) Distribution of forecasts for unemployment rate



b) Distribution of forecast errors for the unemployment rate

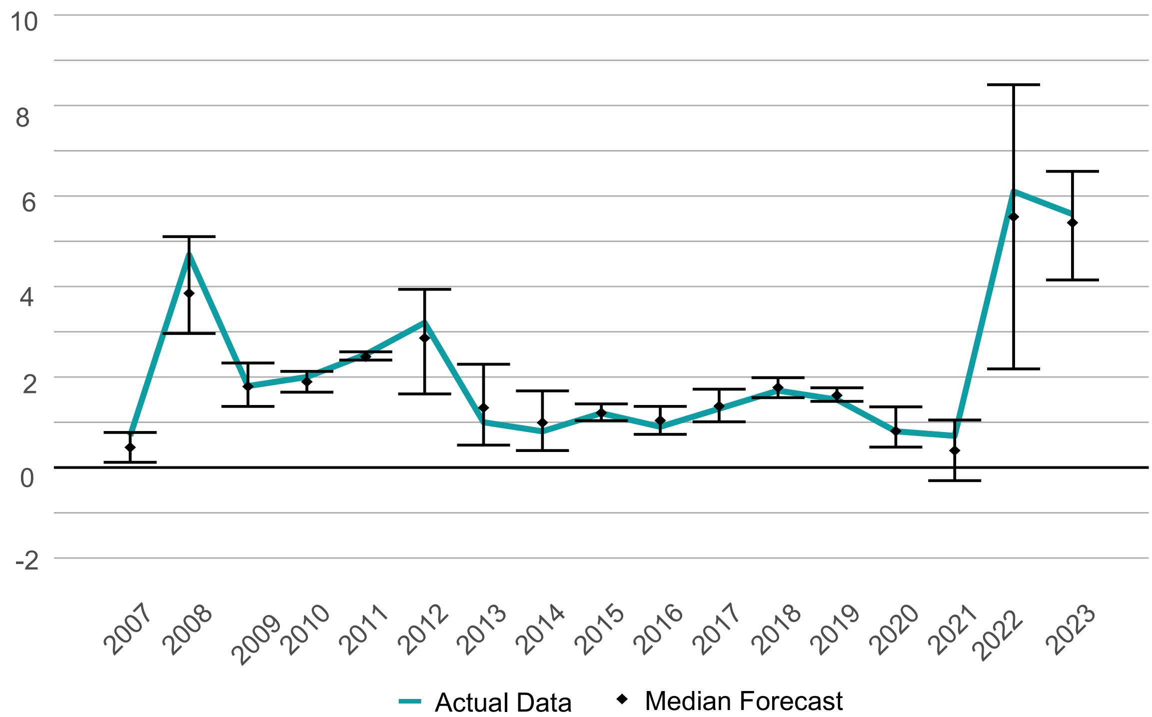


Source: Panel a): CBM projections, NSO. Panel b): Author's calculations. Notes: Panel a): The bars represent the range in the forecasts conducted in that particular year. From 2008-2017, the range of forecasts includes 2 data points, from 2018-2023, the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t+1), where t is the time-period being forecasted. Panel b): The error bars represent the range in the forecast errors computed for that particular year. From 2008-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.

Figure 3 HICP forecasts, forecast errors and actual data for same-year forecasts

a) Distribution of forecasts for HICP

(annual percentage change)



b) Distribution of forecast errors for HICP

(percentage point)

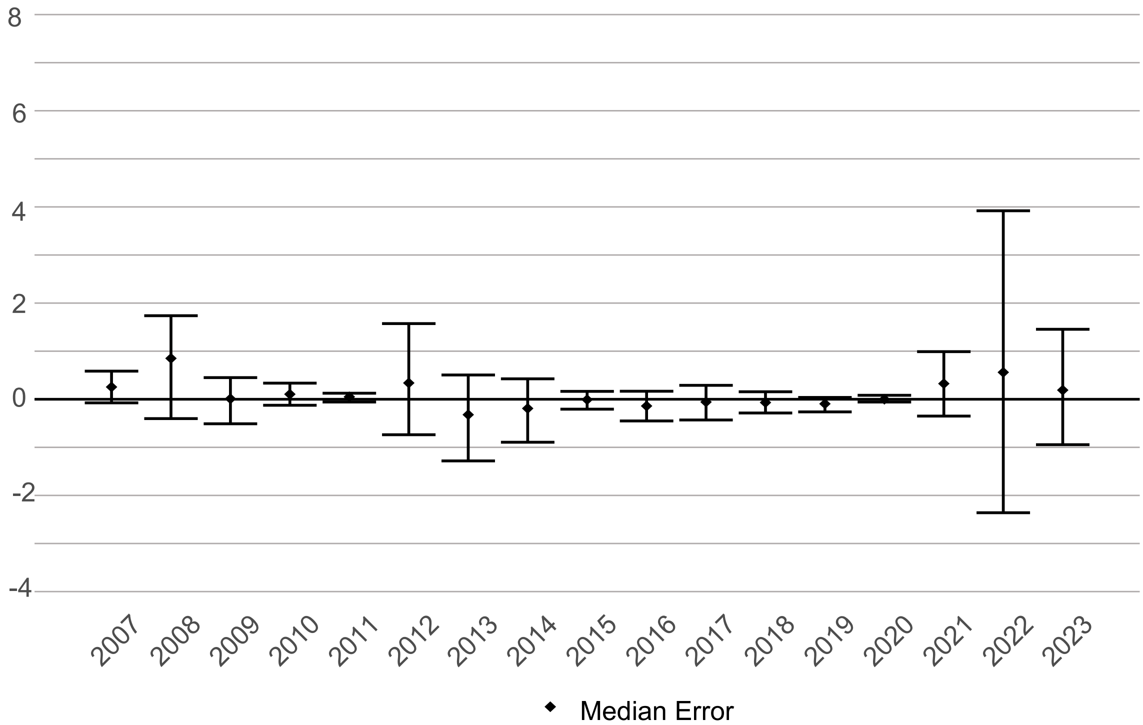


Figure 3 shows that despite some volatility in the actual data over the time-period analysed, the projections for inflation were typically quite accurate.[6] Notably, in 2022 the range in the forecast produced increased significantly when compared to the full time-period, reflecting the beginning of a period characterised by relatively high inflation and uncertainty in the context of the war in Ukraine. While forecasts for HICP are less accurate than those for the unemployment rate, they are slightly more accurate than GDP growth projections. Overall, HICP forecasts do not appear to present any systematic bias. However, one can note an underestimation of headline inflation in three of the years which presented the highest errors i.e. 2008, 2012 and 2022. Meanwhile, 2013 displayed a tendency to overpredict HICP inflation.

5 Main Results[7]

In what follows I discuss the main results for the formal statistical tests performed on projection errors for real GDP, the unemployment rate and HICP inflation. I also show results for a number of sub-components of GDP and HICP. The discussion is grouped into three sub-sections focusing on accuracy, bias and efficiency of the projections.

5.1 Accuracy

Table 1 outlines results for the accuracy tests performed on the forecasts for real GDP, the unemployment rate and HICP for the whole sample and for the sample excluding the Covid period in 2020 and 2021. As expected, the standard deviation and RMSE across variables increase with the forecast horizon. The latter is commonly observed in the forecasting literature and is explained in Diebold (2007), who highlights that as the forecast horizon increases, the information used to make the forecast becomes less reliable, available or relevant, consequently accuracy deteriorates. Meanwhile, when excluding the forecasts for 2020 and 2021, the standard deviation and RMSE decline significantly for GDP forecasts while for HICP and the unemployment rate, these remain broadly unchanged. This is corroborated by the mean absolute error which provides a measure of the average magnitude of errors. Similar findings were observed by Davison, Camilleri, & Spiteri (2024), who highlight that accuracy is strengthened when excluding crisis periods, while higher RMSEs are evident at longer horizons.

Meanwhile, the mean error indicates a tendency for underestimating growth for GDP and HICP with the exception of GDP at the 1-year ahead horizon in the full sample. On the other hand, the mean error shows a strong tendency to overestimate projections for the unemployment rate as this variable displays a negative mean error across all horizons in both the full and adjusted sample. At the same time, the unemployment rate displays the strongest forecast accuracy out of these three variables with a RMSE ranging from 0.4 to 1.0 percentage points in the full sample. The small magnitude of these errors is confirmed by the values of the mean absolute errors which are contained under 0.8 percentage points across the three horizons in

[6] The distribution of forecasts and forecast errors for HICP excluding energy is similar to that shown in figure 3.
[7] All statistical tests were also performed on a sub sample of forecast errors, starting from 2013 onwards. Results are similar to the ones presented in this section and are presented in the appendix section. Given that results are broadly unchanged, the full sample was chosen for the main analysis in order to make best use of the limited data we have available.

the full sample. Meanwhile GDP exhibits the weakest forecast accuracy with a RMSE ranging from 1.2 to 4.7 across both samples.

Table 1: Forecast performance metrics by horizon and dataset for the main macroeconomic variables

Variable	Horizon	Error Metrics (full sample/adjusted sample)			
		St. deviation	RMSE	Mean error	Mean absolute error
GDP	h = 0	2.31 / 1.11	2.36 / 1.24	0.57 / 0.59	1.52 / 1.03
	h = 1	4.21 / 1.88	4.16 / 1.99	-0.1 / 0.74	2.80 / 1.62
	h = 2	4.72 / 2.02	4.68 / 2.42	0.51 / 1.38	3.39 / 1.97
Unemployment Rate	h = 0	0.37 / 0.27	0.42 / 0.32	-0.21 / -0.19	0.32 / 0.25
	h = 1	0.57 / 0.45	0.71 / 0.69	-0.44 / -0.53	0.61 / 0.58
	h = 2	0.68 / 0.64	0.96 / 1.07	-0.69 / -0.86	0.79 / 0.92
HICP	h = 0	0.55 / 0.58	0.55 / 0.58	0.10 / 0.09	0.30 / 0.31
	h = 1	1.58 / 1.66	1.59 / 1.69	0.25 / 0.40	1.03 / 1.10
	h = 2	1.91 / 1.99	1.90 / 2.04	0.22 / 0.55	1.35 / 1.40

Source: Author's calculations

Table 2 presents the accuracy results for the main sub-components of real GDP. Across the main GDP components, growth in investment expenditure displays the lowest forecast accuracy with a RMSE ranging from 14.2 to 16.3 percentage points across all horizons and over the full sample. Additionally, as opposed to other variables, the RMSE of the projections for investment growth worsen further in the adjusted sample which excludes the pandemic. Similarly, government consumption, exports and imports exhibit weak forecast accuracy as evident by a RMSE which exceeds 4 percentage points starting from the same year horizon. These results are confirmed by the forecast range and the distribution of forecast errors for the latter variables in figures 14, 15, 18, 19, 20 and 21 in the appendix, Furthermore, a general tendency to underestimate growth is observed across most sub-components with the exception of private consumption and exports at the 1-year ahead horizon and investment and imports at the 1-year and 2-year horizons. The latter variables display a negative mean error at the specified horizons, suggesting a tendency to overestimate growth. Notably, the difference in performance of real GDP forecasts and the forecasts of its components reflects compensating effects. These will be discussed in more detail in section 6.

Table 2: Forecast performance metrics by horizon and dataset for the sub-components of real GDP

Variable	Horizon	Error Metrics (full sample/adjusted sample)			
		St. deviation	RMSE	Mean error	Mean absolute error
	h = 0	2.91 / 2.10	2.92 / 2.32	0.52 / 1.06	2.27 / 2.00
Consumption	h = 1	4.30 / 2.04	4.25 / 2.32	-0.16 / 1.15	2.81 / 1.97
	h = 2	4.81 / 2.79	4.76 / 3.25	0.41 / 1.75	3.65 / 2.77
	h = 0	4.06 / 3.97	4.02 / 3.93	0.13 / -0.35	3.21 / 3.24
Government consumption	h = 1	4.79 / 4.42	5.36 / 4.57	2.52 / 1.40	4.58 / 3.96
	h = 2	4.47 / 4.08	4.89 / 4.18	2.11 / 1.19	3.82 / 3.36
	h = 0	14.39 / 15.51	14.23 / 15.32	0.28 / -0.69	10.41 / 11.15
Investment	h = 1	14.89 / 15.81	15.48 / 16.37	-4.83 / -5.06	12.94 / 13.84
	h = 2	16.55 / 17.16	16.34 / 16.91	-0.52 / -1.24	14.07 / 14.19
	h = 0	4.97 / 4.87	4.95 / 4.81	0.55 / 0.19	3.52 / 3.30
Exports	h = 1	6.43 / 6.16	6.38 / 6.07	-0.61 / 0.19	4.33 / 3.86
	h = 2	6.35 / 5.54	6.26 / 5.51	0.10 / 0.80	4.82 / 3.92
	h = 0	4.60 / 4.41	4.57 / 4.35	0.43 / -0.05	3.37 / 3.23
Imports	h = 1	5.43 / 5.56	5.56 / 5.58	-1.46 / -1.06	4.34 / 4.28
	h = 2	5.74 / 5.59	5.67 / 5.50	-0.34 / -0.11	4.58 / 4.15

Source: Author's calculations

Table 3 below presents the forecast accuracy for the sub-components of HICP. While all variables considered in this section display the lowest standard deviation and RMSE at horizon 0, these two statistical measures do not generally improve when considering the adjusted sample that excludes 2020 and 2021. Nonetheless, with the exception of unprocessed food and energy, all other variables exhibit relatively strong forecast accuracy with a RMSE ranging from 0.5 to 1.0 percentage points across both samples at the first forecast horizon. As expected, this is then reflected in relatively small mean absolute errors. The variables with the strongest forecast accuracy include non-energy industrial goods inflation and services, which have a RMSE of 0.5, 0.6 and 0.7 percentage points respectively at the 0 horizon, in both samples. These components also have a superior performance at the 1-year and 2-year ahead horizon.

Table 3: Forecast performance metrics by horizon and dataset for the sub-components of HICP

Variable	Horizon	Error Metrics (full sample/adjusted sample)			
		St. deviation	RMSE	Mean error	Mean absolute error
Unprocessed food	h = 0	2.00 / 1.94	1.99 / 1.92	0.19 / -0.01	1.34 / 1.30
	h = 1	3.99 / 4.29	4.00 / 4.26	0.64 / 0.47	3.01 / 3.27
	h = 2	4.00 / 4.43	4.00 / 4.39	0.62 / 0.51	2.95 / 3.31
Processed food	h = 0	0.96 / 1.00	0.97 / 1.03	0.19 / 0.26	0.54 / 0.56
	h = 1	2.45 / 2.54	2.62 / 2.84	1.01 / 1.33	1.52 / 1.68
	h = 2	3.01 / 3.17	3.14 / 3.48	1.02 / 1.55	1.79 / 2.01
NEIG	h = 0	0.50 / 0.49	0.50 / 0.49	0.08 / 0.05	0.31 / 0.30
	h = 1	1.56 / 1.56	1.58 / 1.59	0.31 / 0.41	1.04 / 0.96
	h = 2	1.86 / 1.91	1.87 / 1.99	0.35 / 0.67	1.14 / 1.18
Services	h = 0	0.65 / 0.69	0.65 / 0.68	0.02 / 0.01	0.39 / 0.42
	h = 1	1.70 / 1.80	1.68 / 1.79	-0.01 / 0.14	1.27 / 1.34
	h = 2	1.80 / 1.87	1.77 / 1.87	-0.02 / 0.32	1.40 / 1.42
Energy	h = 0	1.94 / 2.06	1.97 / 2.09	0.43 / 0.48	0.77 / 0.86
	h = 1	3.10 / 3.33	3.08 / 3.30	-0.29 / -0.28	2.15 / 2.37
	h = 2	4.38 / 4.88	4.35 / 4.82	-0.42 / -0.34	2.48 / 2.88

Source: Author's calculations

Unprocessed food and energy inflation exhibit the lowest forecast accuracy with a RMSE starting around 2 percentage points at horizon 0 throughout both samples. At the same time, the mean error outlines a tendency to overestimate services and energy inflation at the first and second forecast horizons although for services this tendency is reversed when excluding the effect of the COVID-19 pandemic. On the other hand, all other HICP components typically tend to be underestimated with a positive mean error across all horizons in the full sample.

5.2 Unbiasedness

Unbiasedness is a necessary condition for optimal and rational forecasts, highlighting that on average, forecasts should not present any systematic over or under prediction. The forecast errors in this paper are based on a fixed-event type of forecasts, which generate overlapping forecast targets that induce serial correlation in the forecast errors. In order to account for this, I test for unbiasedness following the approach suggested in Clements et al. (2007) and Ager et al. (2007)

Table 4: Unbiasedness test - Pooled approach by dataset for the main macroeconomic variables

Variable	(full sample/adjusted sample)	
	Common bias	Wald χ^2
GDP	0.61* / 0.39 (1.81 / 0.90)	9.79 / 17.38
Unemployment Rate	-0.14** / -0.16** (-2.21 /-2.12)	22.14** / 17.00
HICP	0.05 / -0.002 (0.47 / -0.04)	4.20 / 14.36

Notes: Common bias refers to the value of α in equation 8. T-statistics are shown in parentheses*. The Wald statistic refers to the joint null hypothesis of horizon-specific bias equal to zero. *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

Table 4 outlines the overall estimated average bias and the Wald joint test results for real GDP growth, the unemployment rate and HICP.[8] While the common bias tests for a single overall systematic bias, the Wald test, is a joint test across all horizons and it jointly tests whether any of the horizons is biased. Consequently, results from these two tests may indeed vary. Strong evidence of bias is found for the unemployment rate, with outcomes typically below projections in both samples. The Wald statistic in the full sample is also found to be statistically significant, confirming the evidence of bias in at least one of the horizons. While significant only at the 10%, real GDP growth forecasts are found to be biased on average, across all horizons in the full sample. This differs somewhat from the results found in Davison, Camilleri, & Spiteri (2024), as they find mixed evidence for a downward bias in real GDP for CBM projections however the bias holds also when excluding crisis periods. Having said this, the reader is invited to exercise caution when comparing these results due to the pooling framework

[8] Horizon-specific bias tests were also performed and results are presented in tables 18-20 in the Appendix section.

undertaken in this analysis which differs from Davison, Camilleri, & Spiteri (2024). Meanwhile, no systematic bias is evident for HICP irrespective of the sample. The results shown in this section are similar to other studies such as Kontogeorgos and Lambrias (2019), who find no systematic bias for inflation projections, while for GDP, their results indicate a degree of biasedness particularly at the outer horizons.

Table 5 presents biasedness results for the main sub-components of real GDP. When looking at the overall average bias, results suggest a systematic positive bias in government consumption in both the full and adjusted sample. Meanwhile a strong negative systematic bias is evident for investment and imports in the adjusted sample, indicating that on average these forecasts were too optimistic. On the other hand, when testing for bias through the Wald joint test, the joint null hypothesis is rejected for private consumption in the adjusted sample, suggesting that at least one horizon has significant bias.

Table 5: Unbiasedness test - Pooled approach by dataset for the sub-components of real GDP

Variable	(full sample/adjusted sample)	
	Common bias	Wald χ^2
Consumption	0.45 / 0.40 (1.14 / 0.96)	5.92 / 20.78**
Government consumption	1.17** / 1.47** (2.42 / 2.22)	17.40 / 15.67
Investment	-1.35 / -4.00*** (-0.89 / -3.07)	6.09 / 17.57
Exports	0.35 / -0.92 (0.60 / -1.19)	3.17 / 4.09
Imports	-0.13 / -1.80*** (-0.25 / -2.79)	8.39 / 10.69

Notes: Common bias refers to the value of α in equation 8. T-statistics are shown in parentheses. *The Wald statistic refers to the joint null hypothesis of horizon-specific bias equal to zero. *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

Results indicate no statistically significant bias in exports and imports forecasts within the full sample. These results differ from those presented in Davison, Camilleri, & Spiteri (2024), who outline that CBM projections for both these two variables display a downward bias which holds also when excluding crisis periods. This is then also reflected in a downward bias for real GDP. In our case, as mentioned above an overall common bias is evident for government consumption and investment which indirectly is also affecting imports. For the latter two, the overall common bias is only introduced in the adjusted sample which excludes the crisis period. Such differences in results may be attributable in part to the different methodologies used in conducting these two forecast evaluations, in particular, the use of real-time data for the computation of forecast errors for this study.

Table 6 outlines the results for the main HICP sub-components. In the full sample, I find no evidence of systematic bias for the main sub-components of headline inflation both overall and across any of the horizons. However, in the sample excluding the COVID-19 pandemic, unprocessed food shows a tendency to be overpredicted on average across all horizons. At the same time, some evidence of bias in at least one of the horizons is confirmed for services inflation through the Wald joint test, although the joint null hypothesis in this case is only rejected at the 10% significance level. Meanwhile, strong evidence against biasedness is confirmed for all other sub-components.

Table 6: Unbiasedness test - Pooled approach by dataset for the sub-components of HICP

Variable	**(full sample/adjusted sample)**	
	Common bias	Wald χ^2
Unprocessed food	0.24 / -0.52** (0.83 / -2.10)	5.20 / 8.79
Processed food	0.09 / -0.01 (0.45 / -0.19)	13.23 / 8.30
NEIG	0.03 / -0.02 (0.29 / -0.42)	3.62 / 11.81
Services	0.03 / -0.13 (0.27 / -1.34)	4.78 / 20.54*
Energy	0.27 / 0.41 (1.04 / 1.21)	13.27 / 11.94

Notes: Common bias refers to the value of α in equation 8. T-statistics are shown in parentheses. *The Wald statistic refers to the joint null hypothesis of horizon-specific bias equal to zero. *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

5.3 Efficiency

The final set of tests performed in this section relate to the efficiency of forecasts. The latter is another necessary condition for forecasts optimality. We test for weak efficiency by regressing current forecast errors on past errors and past data outcomes of the variable of interest. In the case where the coefficients are not found to be significantly different from zero, forecasts are deemed to be informationally efficient.

Table 7 presents the results for real GDP, the unemployment rate and HICP. For inflation, the coefficient γ is not significantly different from zero and therefore one may conclude that forecasts for headline inflation are informationally efficient as both past errors and past observed data do not predict the current forecast error. At the same time, the unemployment rate is found to be informationally efficient only in the full sample. The test uncovers that over the full sample, GDP forecasts failed to take in consideration information contained in its lagged outcomes. Moreover, unemployment rate forecasts done on the adjusted sample seem to be informationally inefficient with regards to lagged errors. The Ljung-Box test is also performed as a robustness check in order to test for autocorrelation across the forecast errors. Failing to reject the Ljung–Box null hypothesis indicates that forecast errors do not display systematic autocorrelation and hence do not contain exploitable information from past forecast errors. Results from this test suggest that real GDP, unemployment rate and HICP forecasts are all weakly efficient.[9]

Table 7: Efficiency test by dataset for the main macroeconomic variables

Variable	Horizon	(full sample/adjusted sample) Lagged error P-values	(full sample/adjusted sample) Lagged outcome P-values
GDP error	h = 0	0.57 / 0.60	0.07* / 0.64
Unemployment rate error	h = 0	0.74 / 0.02**	0.46 / 0.99
HICP error	h = 0	0.25 / 0.68	0.92 / 0.78

*Notes: Values reported in the table represent the p-values for the null hypothesis $\gamma = 0$, estimated using HAC standard errors. *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively. Source: Author's calculations*

Tables 8 and 9 below present the results for the main sub-components of real GDP and headline inflation. Results are mixed across subcomponents and the two samples. Notably, government consumption, investment, NEIG inflation and services inflation present sound weak efficiency as the null hypothesis fails to be rejected with respect to both the lagged forecast error and lagged observed data across both samples. Meanwhile, energy inflation and exports present the strongest case for inefficiency among all the subcomponents as for

[9] Results from the Ljung-Box test for the main macroeconomic variables are presented in table 21 in the appendix section.

the latter two, both the lagged error and past observed data are found to have predictive power over the forecast error. These results also hold in the adjusted sample.

Table 8: Efficiency test by dataset for the sub-components of real GDP

Variable	Horizon	(full sample/adjusted sample)	
		Lagged error P-values	**Lagged outcome P-values**
Consumption	h = 0	0.48 / 0.03**	0.91 / 0.16
Government consumption	h = 0	0.11 / 0.24	0.33 / 0.53
Investment	h = 0	0.53 / 0.77	0.33 / 0.85
Exports	h = 0	0.03** / 0.07*	0.01** / 0.00***
Imports	h = 0	0.96 / 0.06*	0.73 / 0.00***

Notes: Values reported in the table represent the p-values for the null hypothesis $\gamma = 0$*, estimated using HAC standard errors. *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

Table 9: Efficiency test by dataset for the sub-components of HICP

Variable	Horizon	(full sample/adjusted sample) Lagged error P-values	Lagged outcome P-values
Unprocessed food	h = 0	0.00*** / 0.32	0.75 / 0.70
Processed food	h = 0	0.69 / 0.01**	0.93 / 0.01**
NEIG	h = 0	0.14 / 0.16	0.76 / 0.35
Services	h = 0	0.69 / 0.30	0.35 / 0.50
Energy	h = 0	0.00*** / 0.01**	0.00*** / 0.00***

Notes: Values reported in the table represent the p-values for the null hypothesis $\gamma = 0$, *estimated using HAC standard errors. *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations.

Beyond testing for weak efficiency, tests were also computed to analyse overall efficiency of the projections. Table 10 below presents results for Real GDP, whereby forecast errors were regressed on contemporaneous forecasts of GDP subcomponents using robust linear regression. This approach evaluates whether forecast errors can be systematically explained by information that was available at the time the forecasts were made. Wald ratios (β/δ) were computed for each driver variable, together with delta-method standard errors, to assess whether any of the subcomponent forecasts exhibit a statistically meaningful structural relationship with the corresponding forecast errors.

While the estimated coefficients point to potential asymmetries in how information from the different macroeconomic drivers was incorporated in the forecasts, the effects are not statistically significant. As such, the results do not provide strong evidence of systematic inefficiencies, though the direction of the estimates may indicate areas where forecasts could be improved.

Additionally, Mincer-Zarnowitz regressions as outlined in equation 14 were estimated for each subcomponent. The estimated Wald ratios are generally not statistically significant, providing limited evidence of systematic misperception. Positive estimates imply that higher forecasts of the driver variable are associated with higher-than-expected GDP outturns, suggesting an underestimation of their impact, while negative estimates indicate the opposite. Across all components, the 95% confidence intervals straddle the 0, indicating no strong evidence against forecast efficiency for real GDP.

Table 10: Efficiency test for real GDP based on Blanchard and Leigh (2013)

Variable	Horizon	Full sample		
		ß	δ	Wald ratio
Consumption	h = 0	0.39	1.41	0.28 (0.16)
Government consumption	h = 0	-0.06	0.58	-0.11 (0.15)
Investment	h = 0	-0.24	0.60	-0.39 (0.35)
Exports	h = 0	-1.42	0.78	-1.81 (1.17)
Imports	h = 0	1.88	0.90	2.08 (1.17)

Notes: Standard errors in parentheses.
Source: Author's calculations

For the unemployment rate, only one driver variable was used: the forecast of real GDP. The estimated Wald ratio of 0.05, is small and not statistically significant, indicating limited evidence that GDP forecasts systematically explain unemployment forecast errors. Additionally, the positive estimate implies that higher GDP forecasts are associated with higher-than-expected unemployment outturns, consistent with an underestimation of the impact of GDP on the unemployment rate. However, the 95% confidence interval again straddles 0 implying the estimate is statistically insignificant. As such, there is no evidence that unemployment forecast errors could be predicted using available GDP forecasts at the time. These results are presented in table 11 below.

Table 11: Efficiency test for the unemployment rate based on Blanchard and Leigh (2013)

Variable	Horizon	Full sample		
		ß	δ	Wald ratio
Real GDP	h = 0	0.06	1.34	0.05 (0.02)

Notes: Standard errors in parentheses.
Source: Author's calculations

Finally, for headline inflation, the forecast efficiency analysis was conducted using contemporaneous forecasts of its main subcomponents. The estimated Wald ratios are generally small and not statistically significant, as all confidence intervals include zero, indicating limited evidence of systematic misperception in the pass-through to HICP inflation. The positive estimates for unprocessed food, NEIG, energy and services suggest that higher forecasts of these components are associated with higher-than-expected inflation outturns, consistent with an underestimation of their inflationary impact. Conversely, the slightly negative estimate for processed goods points to a marginal overestimation of its pass-through. However, given the wide confidence intervals and associated uncertainty, these effects are not statistically distinguishable from zero.

Table 12: Efficiency test for HICP based on Blanchard and Leigh (2013)

Variable	Horizon	Full sample		
		ß	δ	Wald ratio
Unprocessed food	h = 0	0.04	0.73	0.05 (0.22)
Processed food	h = 0	-0.01	1.48	-0.01 (0.22)
NEIG	h = 0	0.31	1.07	0.29 (0.30)
Services	h = 0	0.02	1.21	0.01 (0.38)
Energy	h = 0	0.03	1.06	0.02 (0.05)

Notes: Standard errors in parentheses.
Source: Author's calculations

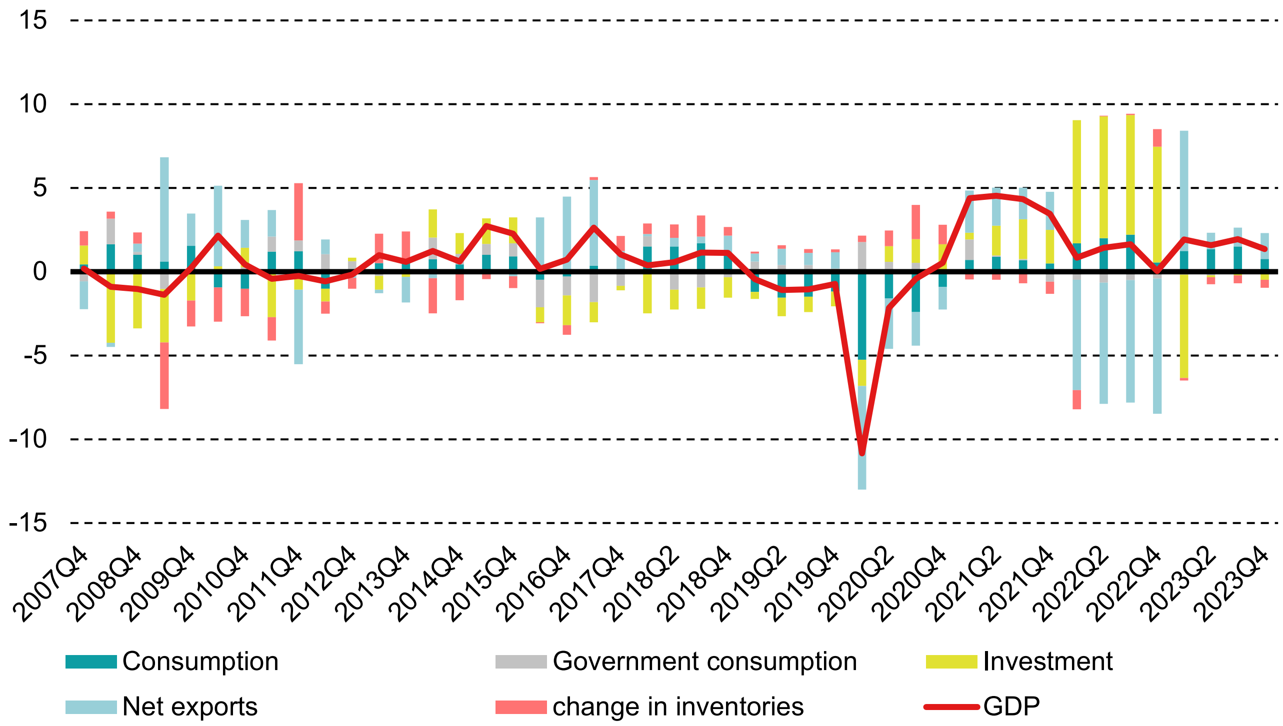
6. Analysing the sources of projection errors across sub-components

This section assesses the projection errors of real GDP and HICP through a decomposition based on the contributions of the projection errors of their main components. This analysis based on the same-year horizon forecast error over the entire sample of forecast rounds presented throughout this work.

Figure 4 presents the contributions from the forecast errors in the expenditure components to the overall forecast error for real GDP growth. Notably, over the sample period analysed, net exports and investment have been the main contributors to real GDP forecast errors. Marked contributions from these components can be observed during the 2008 financial crisis period and the more recent COVID-19 pandemic and post pandemic period. In particular, during 2022, net exports turned out to be weaker than expected (reflecting higher than expected import growth), while the opposite may be observed for investment. The underestimation in investment was partially driven by extraordinary outlays relating to the aviation sector which occurred during 2021 and the expected base effect for 2022. At the same time, the overestimation in net exports was mainly driven by an underprediction of imports which at the time were expected to decelerate due to the same base effect from extraordinary outlays in the aviation sector Central Bank of Malta (2022). While less striking, investment and net exports were also key contributors to real GDP forecast errors in 2023. However, throughout 2023, net exports were underestimated while investment turned out to be weaker than expected. The latter reflected the base effects owing to extraordinary outlays in investment equipment which occurred in 2022. Meanwhile, net exports growth turned out stronger than forecasted in 2023, as forecast errors in both exports and imports were positive. In particular, the anticipated normalisation in export growth driven by weaker foreign demand and post-pandemic recovery did not materialise.

While less pronounced, projection errors in private consumption also had a significant contribution, particularly in 2020, 2022 and 2023. Private consumption was overestimated in 2020 while it was underestimated in 2022 and 2023. In the initial projection round of 2020, prior to the onset of the pandemic, growth in private consumption was expected to be significant, driven by a positive labour market outlook. Subsequently, in the following rounds, private consumption was indeed expected to contract given pandemic-related measures, however the contraction proved to be larger than anticipated, reflecting the difficulty and uncertainty presented by these unprecedented times. On the other hand, throughout 2022, while robust private consumption growth was anticipated, reflecting the reopening of the economy, this was still below the outturn for 2022 as national accounts data generally surprised on the upside. At the same time, in 2023, a deceleration in private consumption was foreseen, driven by slower growth in real disposable income as well as the normalisation of consumer demand following a strong post-pandemic recovery. While a slowdown in private consumption did indeed materialise, this was less pronounced as private consumption remained robust. Meanwhile, projection errors in government consumption had a minimal contribution to the forecast error in GDP, mainly reflecting the relatively small share of this component in overall GDP.

Figure 4 Decomposition of same-year real GDP forecast errors
(percentage point)

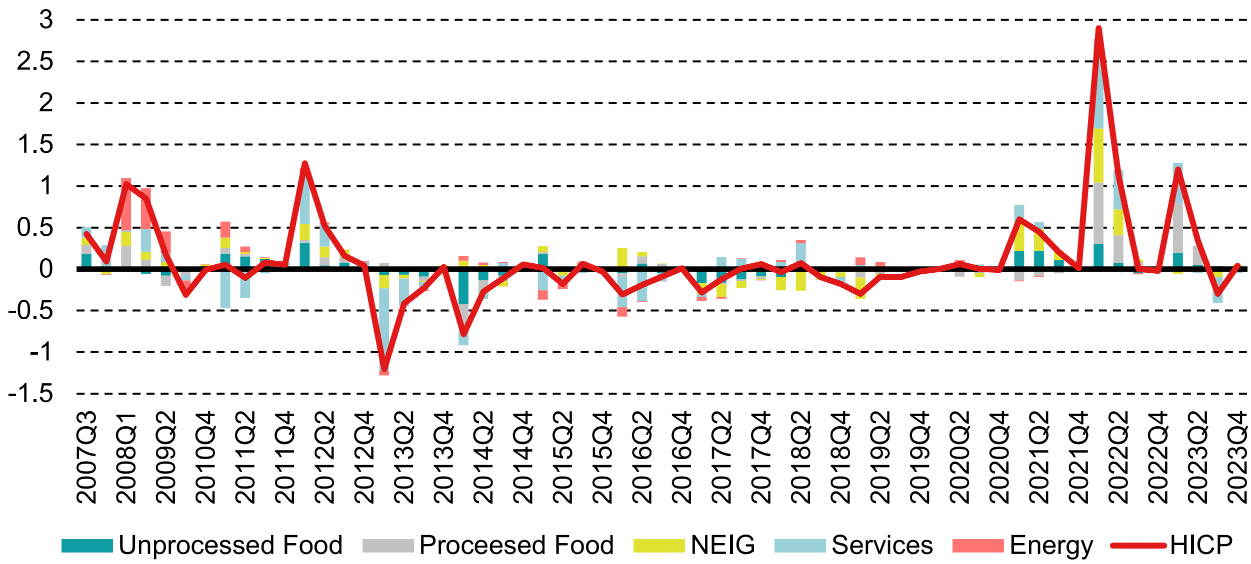


Notes: Contributions from the foreast errors in the main sub components to the overall forecast error are computed by multiplying the weight of the sub component with its forecast error. The forecast error is defined as the outturn minus the forecasted value as outlined in equation 1, for h = 0. A positive (negative) contribution indicates an underprediction (overprediction) of the component. The x-axis represents the projection rounds.
Source: Author's calculations, NSO

Figure 5 presents the contributions from the forecast errors in the HICP sub-components to the overall projection error for HICP. Overall, services prove to be the main contributor to HICP projection errors, followed by food and NEIG components. Errors in the latter were especially large during the first two rounds of 2022 as well as the initial round of 2023 for the food component. During these periods, inflation in these components turned out to be stronger than anticipated. Indeed, this was a time of persistently high inflation. Throughout both 2022 and 2023, services inflation was expected to be the main contributor to overall HICP inflation and while a surge in services inflation was envisaged, mainly reflecting spillovers from other sub-components and from increased labour costs, the anticipated surge was weaker than the outturn for both years.

At the same time, in 2022, the strong contributions from food and NEIG reflected marked increases in import price pressures, particularly increased prices for commodities, intermediate inputs and transport costs. Nonetheless, the outturns were more pronounced than the anticipated increases. Meanwhile, in 2023, contributions to projection errors from NEIG fell significantly, while those from food and in particular processed food given the higher weight, remained high. The latter in part reflected an overestimation of the effects of falling commodity prices and transportation costs at the time. Therefore, while inflation in processed and unprocessed food did indeed decline as a result of these falls, this decline was weaker than anticipated.

Figure 5 Decomposition of same-year HICP forecast errors
(percentage point)



Notes: Contributions from the foreast errors in the main sub components to the overall forecast error are computed by multiplying the weight of the sub component with its forecast error. The forecast error is defined as the outturn minus the forecasted value as outlined in equation 1, for h = 0. A positive (negative) contribution indicates an underprediction (overprediction) of the component. The x-axis represents the projection rounds.
Source: Author's calculations, NSO

7. Comparison of forecast errors

7.1 Performance against a benchmark model

This section presents results outlining the performance of forecasts relative to a simple benchmark model. For this purpose, the benchmark model used is a Random Walk (RW) with drift. A similar approach is undertaken in forecast evaluations such as those presented by Bank of England (2015) and Kontogeorgos & Lambrias (2019) who use a simple RW model for benchmarking the forecasting performance. The model used here extends the basic

Random Walk by including a constant term; *the drift*, which allows for a consistent trend in the data over time. In this model, future values are predicted as the last observed value plus the drift. It assumes that changes are unpredictable but may follow a systematic long-term direction dictated by the drift. The following equation reflects the forecasts based on this model:

$$\hat{y}_{t+k} = y_t + kd \tag{16}$$

Whereby, the k-step-ahead forecast is calculated as the last observed value plus k times the estimated drift, which represents the trend per period.

In order to evaluate the relative performance of our forecasts, we perform two statistical tests being the scaled RMSE and the Diebold-Mariano (DM, 1995) test. The latter is used to compare the predictive accuracy of two competing forecast models. Specifically, it evaluates whether the forecast errors from the two models are significantly different. Notably, the forecasts generated by the benchmark model differ from those produced by the CBM, given that the latter are conditioned on a set of assumptions while the benchmark model produces unconditional forecasts. Therefore, while such a comparison is useful for assessing the added value from a conditional forecast relative to a naïve forecast, the results presented in this section should be interpreted with caution.

In conducting the DM test, a quadratic loss function is assumed as outlined in equation 5. The test is based on the null hypothesis that the two models have equal predictive accuracy, meaning that the expected difference in their forecast errors is zero:

$$(H_o): E[d_t] = 0 \tag{17}$$

$$(H_1): E[d_t] \neq 0$$

The DM test statistic is computed by comparing the mean differences in errors, adjusted for autocorrelation, and is tested using a two-sided approach.

Additionally, the scaled RMSE is performed in order to compare the accuracy of projections to the simple benchmark model. This is done by dividing the RMSE of the forecasts shown earlier by the RMSE of projections obtained by the benchmark model. A scaled RMSE less than 1 indicates that the staff projections outperform those obtained from the benchmark model, and vice versa. The closer the scaled RMSE is to 1, the smaller the difference in performance between the staff projections and the RW model.

Table 13 displays the results for both the scaled RMSE and the DM test for real GDP, the unemployment rate and headline inflation for the full sample and the sample excluding the COVID-19 pandemic period. The scaled RMSE indicates that for all 3 variables, the Bank's projections outperform the simple RW benchmark model. For real GDP and headline inflation, the Bank's projections are significantly better than the benchmark model at horizon 0. However, the scaled RMSE increases at longer horizons suggesting that at such horizons, the difference in performance is lower. Meanwhile, for the unemployment rate, while the Bank's projections still outperform the benchmark model, the scaled RMSE is very close to 1 in the same year and for the 1 year ahead horizon. Moreover, in the outer horizon, it is equal to 1 suggesting both models perform equally well. These results are reiterated by the DM test which suggests a statistically significant difference in predictive accuracy between the Bank's projections and the RW model for real GDP at horizons 0 and 1 and HICP across all forecast horizons. In addition, the predictive accuracy of the Bank's projection of the unemployment rate, particularly in the outer years, is not found to be statistically significant with a p-value greater than 0.05.

Table 13: Relative forecasting performance by horizon and dataset for the main macroeconomic variables

Variable	Horizon	RMSE ratio (full sample/adjusted sample)	DM constant (full sample/adjusted sample)	DM p-value (full sample/adjusted sample)
GDP	h = 0	0.34 / 0.44	-3.18 / -3.74	0.00*** / 0.00***
	h = 1	0.56 / 0.31	-2.25 / -2.03	0.03** / 0.05**
	h = 2	0.65 / 0.40	-1.49 / -1.47	0.14 / 0.15
Unemployment Rate	h = 0	0.73 / 0.68	-2.19 / -1.65	0.03** / 0.11
	h = 1	0.87 / 0.85	-0.90 / -1.08	0.37 / 0.29
	h = 2	1.02 / 1.10	0.10 / 0.60	0.92 / 0.55
HICP	h = 0	0.29 / 0.29	-2.88 / -3.13	0.01** / 0.00***
	h = 1	0.65 / 0.64	-2.45 / -2.74	0.02** / 0.01**
	h = 2	0.81 / 0.79	-2.00 / -2.70	0.05** / 0.01**

*Notes: DM constant refers to the Diebold-Mariano test statistic. P-values are estimated using HAC standard errors. *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

7.2 Performance against other institutions

In this section, the forecast accuracy of the CBM projections are compared to those produced by the Ministry for Finance and Employment (MFE) and the European Commission (EC). The results of the CBM forecast performance are compared with both institutions whereby the same methodology applied throughout this evaluation is applied to the spring and autumn forecasts produced by the MFE and to the spring and autumn forecasts as well as the interim forecasts produced by the European Commission[10], in order to compute the forecast errors.[11]

In interpreting these results, a few caveats need to be kept in mind. Particularly, the sample size differs across institutions. Additionally, given that forecasts are not published simultaneously by the different institutions, this may reflect different information available to the institution and therefore this will impact the performance of the different forecasts.

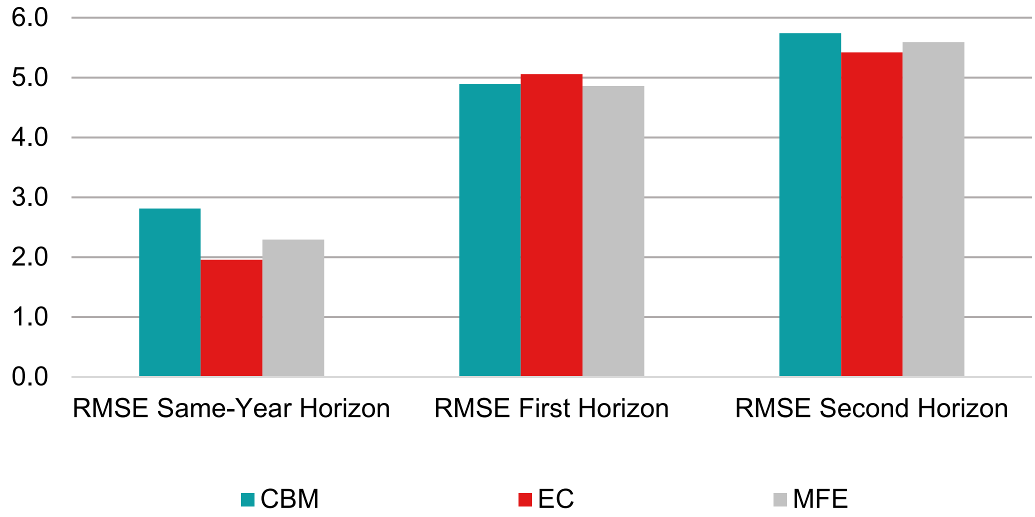
Figures 6, 7 and 8 below display the results of the RMSEs across the CBM, MFE and the EC forecast errors for real GDP, HICP and the unemployment rate.[12] Accuracy results are broadly similar across institutions. EC GDP and HICP forecasts at the same-year horizon perform better, than those produced by the CBM and MFE. Meanwhile, 1-year ahead CBM forecasts tend to perform marginally better than those of the other two institutions. At the same time, accuracy of forecasts for the unemployment rate are broadly the same across all three institutions.

[10] This is with the exception of the unemployment rate which is not available for the interim rounds for forecasts produced by the EC.

[11] The CBM and EC publish four rounds of forecasts per year whereas the MFE publishes two forecast rounds per year.

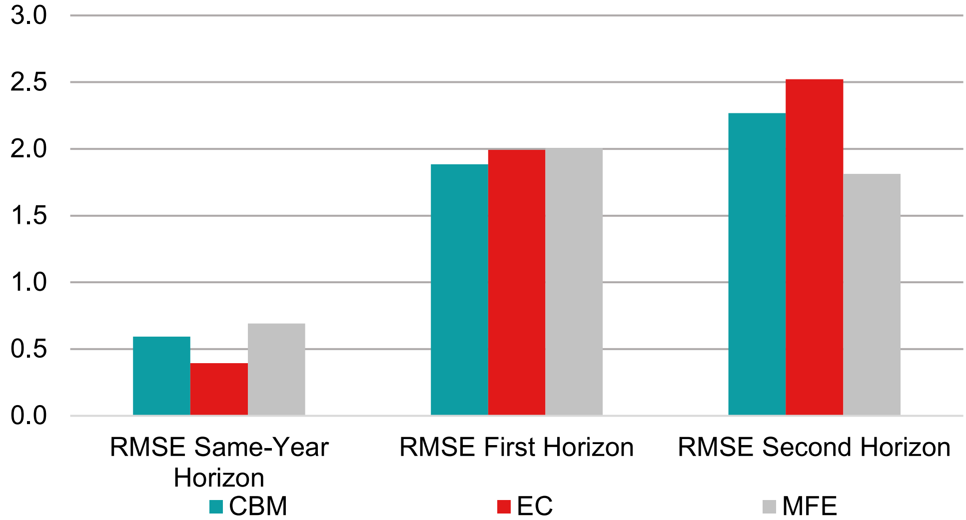
[12] While CBM and MFE forecast data are available from 2007 onwards, forecast data by the EC is available from 2015. Therefore, to ensure results are more comparable, results are computed on data from 2015 onwards across all three institutions. For this reason, the results for CBM portrayed in this section differ from those shown in Table 1.

Figure 6 RMSE across institutions for real GDP
(percentage point)



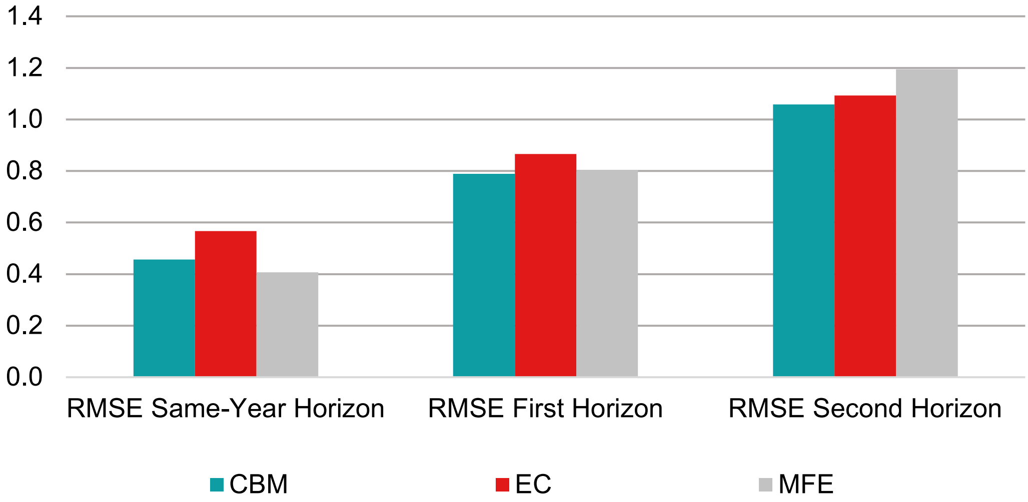
Note: The bars represent the RMSEs for real GDP growth projection errors at the same-year, first (1-year ahead) and second (2-years ahead) horizons based on the same methodology across all three institutions, where real-time data vintages were used to compute errors.
Source: Author's calculations

Figure 7 RMSE across institutions for HICP
(percentage point)



Note: The bars represent the RMSEs for HICP projection errors at the same-year, first (1-year ahead) and second (2-years ahead) horizons based on the same methodology across all three institutions. Given that HICP data is not affected by data revisions, the latest data vintage is used to compute forecast errors.
Source: Author's calculations

Figure 8 RMSE across institutions for the unemployment rate
(percentage point)



Note: The bars represent the RMSEs for the unemployment rate projection errors at the same-year, first (1-year ahead) and second (2-years ahead) horizons based on the same methodology across all three institutions. Due to data limitation, the latest data vintage is used to compute forecast errors.
Source: Author's calculations

8. Evaluating the impact of technical assumptions on forecast errors

8.1 Assessing impact via Basic Model Elasticities

Macroeconomic projections produced by the CBM are conditional forecasts as they are conditioned on a set of technical assumptions which are provided by the European Central Bank (ECB). Specifically, technical assumptions used in the CBM's models include among others, the future path of foreign demand for Malta, exchange rates, competitors' prices, the international oil price, international food prices, long-term yields and banks' interest rates Borg, Farrugia, & Ellul (2016).

Consequently, the forecasts produced are impacted by revisions in the technical assumptions. Therefore, this section aims to analyse the extent to which forecast errors are impacted by assumption revisions and this is done through the use of Basic Model Elasticities (BMEs) as estimated in Borg, Cumbo and Rapa (2024). Before delving further into the methodology used to evaluate this impact, it is important to note a few caveats, primarily it is assumed that BMEs are a good proxy of the main forecasting model used by CBM staff. In practice, when producing macroeconomic forecasts, additional tools are used particularly the inclusion of expert judgement. Therefore, it is possible that the impact of assumption revisions is underestimated when assessed within the BMEs framework. Indeed, this is highlighted by Kontogeorgos & Lambrias (2019) in their analysis of the Eurosystem projections as they note that biasedness results for GDP errors adjusted for assumption revisions actually deteriorate.

Essentially, BMEs provide a simplified, mechanical approach to assessing the economic impact of changes in key assumptions, offering a "rule of thumb" methodology. BMEs are used both to gauge the potential effects of hypothetical changes, such as a 10% rise in oil prices on real GDP growth and inflation, and to analyse the impact of shifts in assumptions between projection rounds. Their primary advantage lies in their ability to quickly and systematically calculate the effects of assumption changes across a broad range of variables. However, the tool's linear nature limits its precision to small deviations from baseline assumptions. Additionally, BMEs treat assumption changes as independent, enabling clear attribution of revisions but oversimplifying scenarios where economic interdependencies do exist.

For the purpose of this evaluation, the impact of technical assumptions on forecast errors is assessed for two key macroeconomic variables being real GDP and HICP at the same year horizon. In the process of evaluating this impact, assumption revisions were calculated for six technical assumptions which include foreign demand (WDR), competitors export prices (CXD), oil price (POU), US exchange rate against the euro (EXR), gas price (GAS) and food prices (HIF). Data on assumptions was first transformed into annual year-on-year growth rates and subsequently assumption revisions were computed as the difference in assumption growth rates between projection rounds. The impact is estimated using the respective BMEs for HICP and GDP, taking into account revisions for the forecast year and the preceding three years due to potential lagged effects of past revisions. [13]

Figure 9 portrays the impact of each of the six technical assumption revisions on real GDP forecast errors at the same-year horizon. Over the entire sample analysed, it is evident that revisions in foreign demand had the largest contribution to forecast errors, followed by competitors export prices. Additionally, the contribution of WDR in particular, was relatively large in periods of high uncertainty such as during the sovereign debt crisis, at the start of the COVID-19 pandemic and as geopolitical tension rose following the Russian war in Ukraine in 2022. Meanwhile, in normal periods, the effect of assumption revisions on forecast errors is much more muted.

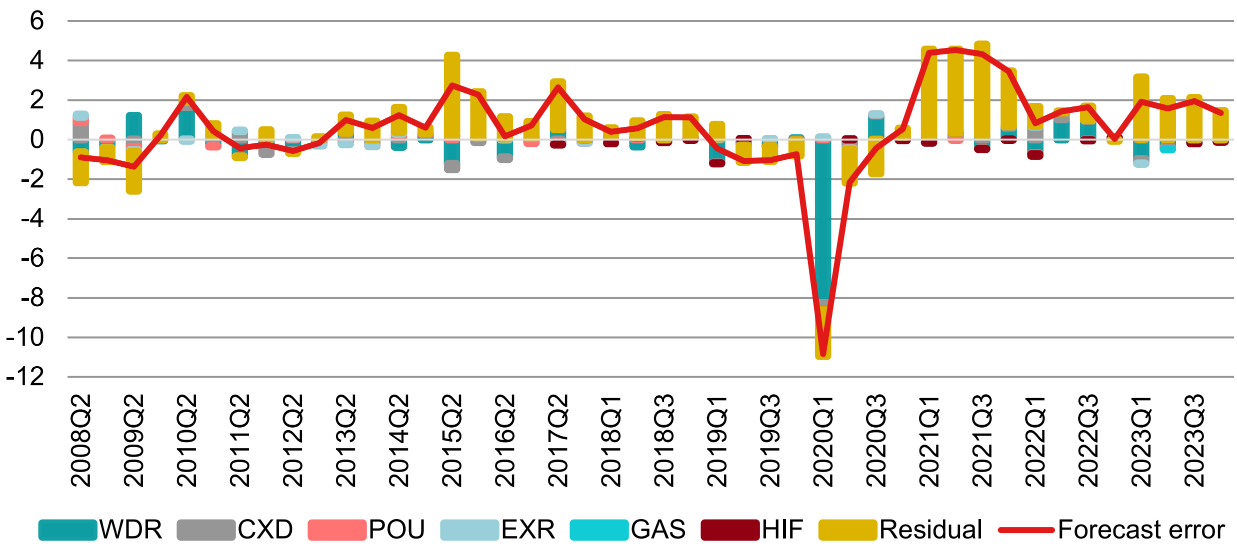
Similarly, Figure 10 outlines the effect of assumption revisions on headline inflation projection errors. In this case, revisions in assumptions on oil prices are the main contributors to forecast errors of HICP throughout the sample, followed by revisions in foreign demand. The effect of revisions in assumptions of oil prices on HICP forecast errors was especially pronounced in the wake of the financial crisis in 2008. Indeed, this was highlighted by Kenny & Morgan (2011) who noted that the overestimation of euro area inflation following the financial crisis was closely tied to fundamental assumption errors regarding oil and other commodity prices, specifically the assumption that these prices would follow the trajectories projected in commodity futures contracts. The latter was also the case for Malta at the start of the COVID-19 pandemic in 2020, whereby revisions in assumptions of oil prices and foreign demand in particular explained most of the overprediction in headline inflation at the time.

At the same time, as outlined by Lane (2024), in the recent period of persistently high inflation, energy prices were consistently anticipated by markets to decline, only for a series of upward

[13] More precisely, assumption revisions from the forecast year and the preceding three years are multiplied by their corresponding BME values, reflecting the potential lagged effects of past revisions. Each result is then adjusted by the proportional change represented by the BME and the adjusted impacts are summed across the four years to estimate the total effect of each assumption revision on the forecast error. This process is repeated for all relevant assumptions, and the individual impacts are aggregated to determine the overall contribution of assumption revisions to the forecast error for each same-year horizon forecast year.

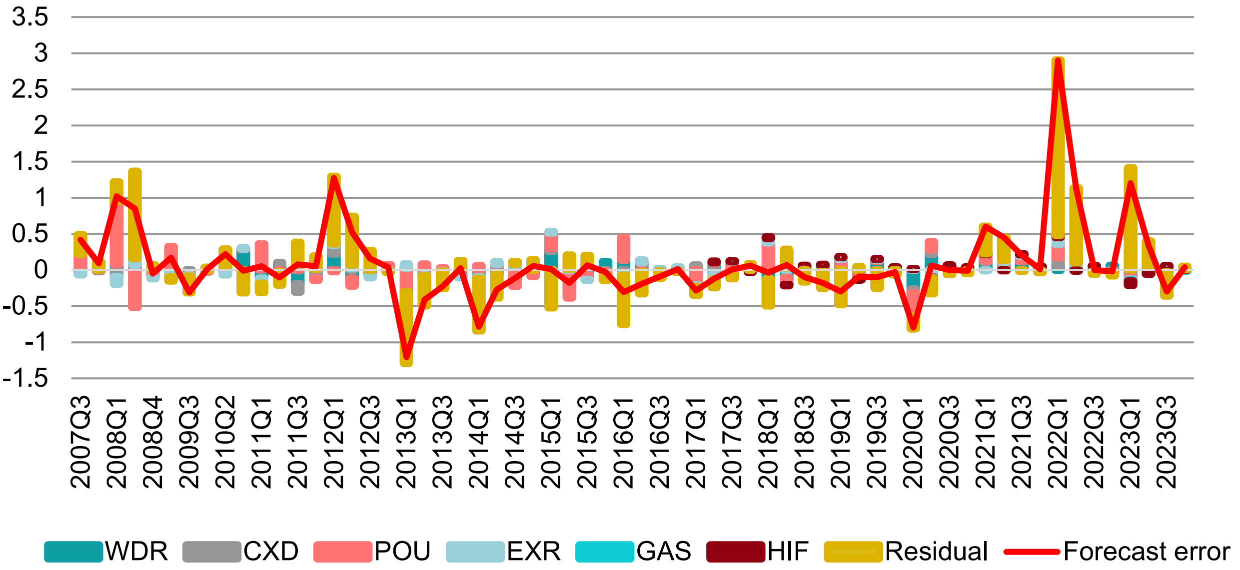
shocks, notably following the Russian invasion of Ukraine, to materialize. In return, these shocks significantly contributed to rising inflation and exacerbating forecast errors over this period, for the euro area. However, in the case of Malta, forecast errors for HICP during the initial surge in inflation were less affected by assumption revisions, as government subsidies shielded the country from direct energy price inflation. Nonetheless, an important caveat needs to be kept in mind, namely that certain indirect effects, such as imported energy inflation, are not fully captured by the BMEs, suggesting that the impact of assumption revisions during this period may be misrepresented.

Figure 9 Contribution of assumption errors to real GDP forecast errors
(percentage point)



Notes: Residual refers to any other factors contributing to the forecast errors other than assumption revisions, including the impact of data revisions, model residuals and error in judgement. The x-axis represents the projection rounds.
Source: CBM, ECB and author's calculations

Figure 10 Contribution of assumption errors to HICP forecast errors
(percentage point)



Notes: Residual refers to any other factors contributing to the forecast errors other than assumption revisions, including the impact of data revisions, model residuals and error in judgement. The x-axis represents the projection rounds.
Source: CBM, ECB and author's calculations

9. Evaluating the impact of data revisions on forecast errors

As outlined throughout this paper, Maltese national accounts data undergoes significant revisions from one release to another. This naturally has an impact on the macroeconomic projections produced by the CBM, given the increased uncertainty surrounding the data on real GDP and its main components. Throughout this forecast evaluation, emphasis was put on limiting this impact on the forecast errors produced by calculating forecast errors on a real-time basis as outlined in previous sections. Nonetheless, the effect of data revisions may still be pronounced and therefore this section aims to provide a better understanding on the extent to which forecast errors are affected by data reliability issues.

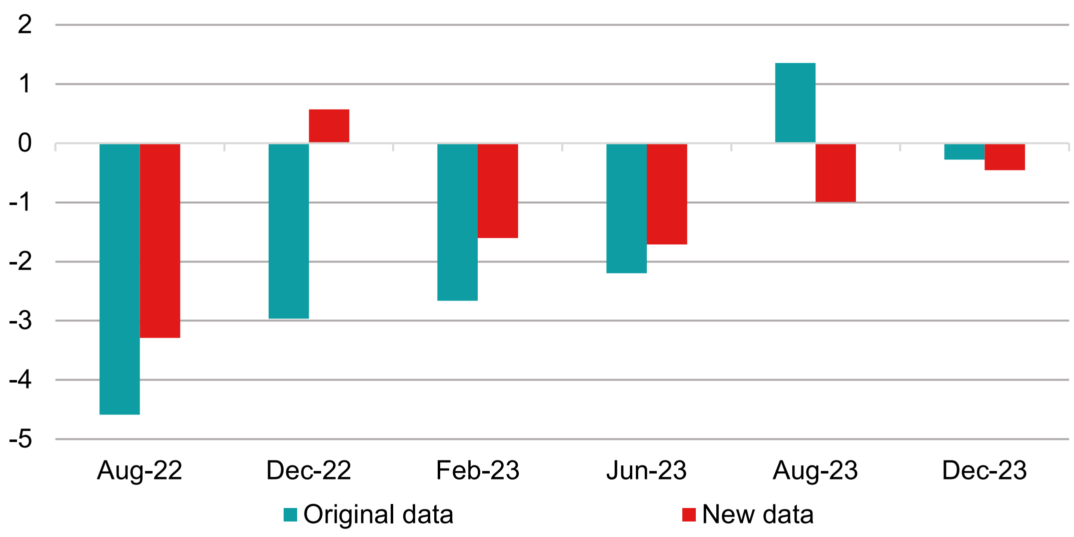
The impact of data revisions on forecast errors is assessed for real GDP and its main components at the same-year forecast horizon. Additionally, as opposed to the previous sections, the analysis in this section focuses on the projection rounds produced in 2022 and 2023 rather than the full sample.

In order to isolate the impact of data revisions, the in-house macro-econometric model STREAM is used to generate two sets of projections for each forecast round. The latter are based on two databases, one of which includes the original macroeconomic data and technical assumptions that were available at the time that the forecast round was produced. Meanwhile, the second database includes the original technical assumptions available at the time of the forecast round and a revised set of (historical) macroeconomic variables which became available in the first quarter of the following year. This allows for the computation of two sets of forecast errors at each round, one reflecting an error based on the original data and another forecast error which reflects the error made when taking into account revised data. [14]

Figure 11 displays the forecast errors for real GDP, consistent with using the two databases as explained above. Based on the forecast rounds analysed, the forecast error would have been significantly lower if the revised set of historical data was available during the period in which the forecast was actually produced. The differences range from -2.4 to 0.2 percentage points and on average, the forecast error for real GDP during this period would have been 0.9 percentage points lower had revised data been available. Similar results are confirmed across all components of real GDP, whereby forecasts incorporating data revisions, on average limit forecast errors between -12.1 and -0.3. The components which display the largest impact of data revisions on forecast errors are private investment, imports and exports. Indeed, the latter components reflect the components with the largest size, bias and volatility of revisions as outlined in Grech (2018) and reiterated in Debono & Mock (2024).

These results tentatively suggest that data revisions have a significant impact on forecast errors. For example, in the December 2022 projection round, the forecast error of around -3 p.p. would have been minimised to close to 0 had the revised data been available during the actual forecast round.

Figure 11 GDP forecast error under different data vintages
(percentage point)



Notes: Original data refers to the original macroeconomic data that was available at the time that the forecast round was produced, while new data reflects the revised set of (historical) macroeconomic variables which became available in the first quarter of the following year. Bars represent forecast errors based on orignal and new data for the same-year forecast horizon.
Source: CBM and author's calculations

[14] Each set of projections are computed with the model equations estimated over the two different databases. In this light, while the analysis tries to isolate the impact of data revisions on forecast errors, part of the error may also reflect model-based uncertainty.

10. Conclusion

Based on the comprehensive analysis provided, several significant conclusions can be drawn from this forecast evaluation. It has been established that forecast accuracy declines as the forecast horizon increases. This is evidenced by the rising standard deviation and RMSE across variables, consistent with theoretical expectations. However, when excluding the years affected by the COVID-19 pandemic (2020 and 2021), a marked improvement in GDP forecast accuracy is observed, with both the standard deviation and RMSE declining significantly. In contrast, the accuracy of HICP and unemployment rate forecasts remains broadly unchanged, indicating that these variables were less impacted by the extraordinary conditions of the pandemic period.

Systematic bias is most evident in the unemployment rate forecasts, as shown by a consistently negative mean error, suggesting a tendency to overestimate this variable across all forecast horizons. This finding is further corroborated by the results of the unbiasedness test, where systematic bias is found for the unemployment rate in both the full and adjusted samples. Meanwhile, no strong evidence of systematic bias is detected for HICP in both samples, whereas some evidence of an overall bias for real GDP growth cannot be ruled out.

In terms of informational efficiency, headline inflation (HICP) forecasts exhibit strong efficiency, as neither past forecast errors nor past observed data significantly predict current forecast errors. Similarly, the unemployment rate forecasts are found to be informationally efficient, but only when considering the full sample period, indicating some sensitivity to the sample timeframe. Conversely, real GDP forecasts show signs of inefficiency when using past observed data, but this inefficiency disappears when excluding the COVID-19 pandemic years, highlighting the deteriorating impact of the pandemic on forecast accuracy and efficiency. Further analysis using robust regressions and Wald tests confirms that subcomponent forecasts for GDP and headline inflation generally do not significantly explain forecast errors, supporting the overall efficiency of these projections.

The scaled RMSE and Diebold-Mariano (DM) tests highlight that the forecasting model outperforms the random walk (RW) benchmark model across all variables. For GDP and HICP, the model shows significantly better performance at shorter horizons, although the difference narrows at longer horizons. For the unemployment rate, the scaled RMSE approaches 1 at outer horizons, indicating similar performance between the two models, a finding supported by the DM test results. Meanwhile, comparisons of CBM's results with those of other institutions, namely the EC and the MFE suggest broadly similar results for forecast accuracy.

Forecast errors are also influenced by technical assumption revisions. For GDP, revisions in foreign demand and competitors' export prices have the largest impact, especially during periods of heightened uncertainty, such as the financial crisis, COVID-19 pandemic and geopolitical tensions. For HICP, oil price assumption revisions dominate, further underscoring the importance of external factors in shaping forecast accuracy. At the same time, data revisions have also been proven to have a marked impact on forecast errors. Results obtained indicate that indeed the timely availability of revised data would have minimised real GDP forecast errors by 0.9 percentage points in the period analysed in section 9. In particular, as expected, data revisions for private investment, imports and exports exhibit the largest impact on forecast errors.

This study has produced an in-depth evaluation of the forecasting performance of the macroeconomic projections produced by the CBM. Building on earlier studies, this work delivers the first comprehensive evaluation done on the CBM's projections. It presents an important step that will facilitate future research in this area. In particular, further work may be done to address a few of the limitations which were exposed in this paper. In this regard, as more data becomes available, future work could focus in much more detail on the impact of data revisions and put greater emphasis on the model's forecasting performance. Additionally, enhancing the forecasting process through regular model improvements and validation will strengthen further the predictive accuracy of the CBM projections.

References

Ager, P., Kappler, M., & Osterloh, S. (2007). The Accuracy and Efficiency of the Consensus Forecasts: A Further Application and Extension of the Pooled Approach. ZEW - Leibniz Centre for European Economic Research.

Bank of England, I. E. (2015). Evaluating forecast performance.

Blanchard, O., & Leigh, D. (2013). Growth Forecast Errors and Fiscal Multipliers. IMF Working Paper.

Borg, I., Cumbo, L., & Rapa, N. (2024). Re-estimation of STREAM - The macro-econometric model of the Maltese economy. *Central Bank of Malta Research Bulletin*.

Borg, I., Farrugia, J., & Ellul, R. (2016). The Forecasting Process at the Central Bank of Malta. *Central Bank of Malta*.

Camilleri, G., & Vella, K. (2015). Interpolating Forecast Errors for Assessing Uncertainty in Macroeconomic Forecasts: An Analysis for Malta. *Economic Policy Department, Working Paper Series*(WP 01/2015).

Central Bank of Malta. (2022). Outlook for the Maltese Economy. (2022:1).

Central Bank of Malta. (2023). Outlook for the Maltese Economy. (2023:1).

Clements, M. P., Joutz, F., & Stekler, H. O. (2007). An evaluation of the forecasts of the Federal Reserve: a pooled approach. *Journal of Applied Econometrics, 22(1)*, 121-136.

Davison, K., Camilleri , G., & Spiteri, K. (2024). Evaluating the Macroeconomic Forecasting Performance of the Ministry for Finance and Employment. *Malta Fiscal Advisory Council,*(Working Paper 01/2024), Working Paper 01/2024.

Debono, N., & Mock, D. (2024). An updated analysis of revisions to Maltese national accounts data. *Central Bank of Malta Quarterly Review*(2024:3), pp. 33-41.

Diebold, F. X. (2007). *Elements of forecasting* (4th edition ed.). South-Western, Mason. .

Eicher, T. S., Kuenzel, D. J., Papageorgiou, C., & Christofides, C. (2018). *Forecasts in Times of Crises* (Vols. Working Paper, WP/18/48). International Monetary Fund.

Elliott, G., & Timmermann, A. (2016). *Economic Forecasting.* Princeton University Press.

Ellul, R., & Ruisi, G. (2022). Nowcasting the Maltese economy with a dynamic factor model. *Central Bank of Malta Working Paper*.

El-Shagi, M., Giesen, S., & Jung, A. C. (2016). Revisiting the relative forecast performances of Fed staff and private forecasters: A dynamic approach. *International Journal of Forecasting*, 32(2): 313-323.

Gavin, W. T., & Mandal, R. J. (2003). Evaluating FOMC forecasts. International Journal of Forecasting, 655-667.

Glas, A., & Drechsel, K. (2021). Conditional macroeconomic forecasts: Disagreement, revisions and forecast errors. (I. D. Papers, Ed.) *Halle Institute for Economic Research (IWH), Halle (Saale), IWH Discussion Papers*, No. 7/2021.

Grech, O. (2018). An Analysis of Revisions to Maltese GDP Data. *Central Bank of Malta Research Bulletin*, 1(1):27–34.

Holden, K., & Peel, D. A. (1990). On Testing for Unbiasedness and Efficiency of Forecasts. *The Manchester School of Economics and Social Studies, 58*(2), 120-27.

Kanngiesser, D., & Willems, T. (2025). Forecast accuracy and efficiency at the Bank of England - and how errors can be leveraged to do better. Staff Working Paper No. 1,078.

Kenny, G., & Morgan, J. (2011). Some lessons from the financial crisis for the economic analysis. *European Central Bank Occasional Paper Series*(NO 130/October 2011).

Kontogeorgos, G., & Lambrias, K. (2019). An analysis of the Eurosystem/ECB projections. *European Central Bank Working Paper Series*(NO 2291/June 2019.).

Lane, P. R. (2024). The 2021-2022 inflation surges and the monetary policy response through the lens of macroeconomic models. *Speech by Philip R. Lane, Member of the Executive Board of the ECB, at the SUERF Marjolin Lecture hosted by the Banca d'Italia.*

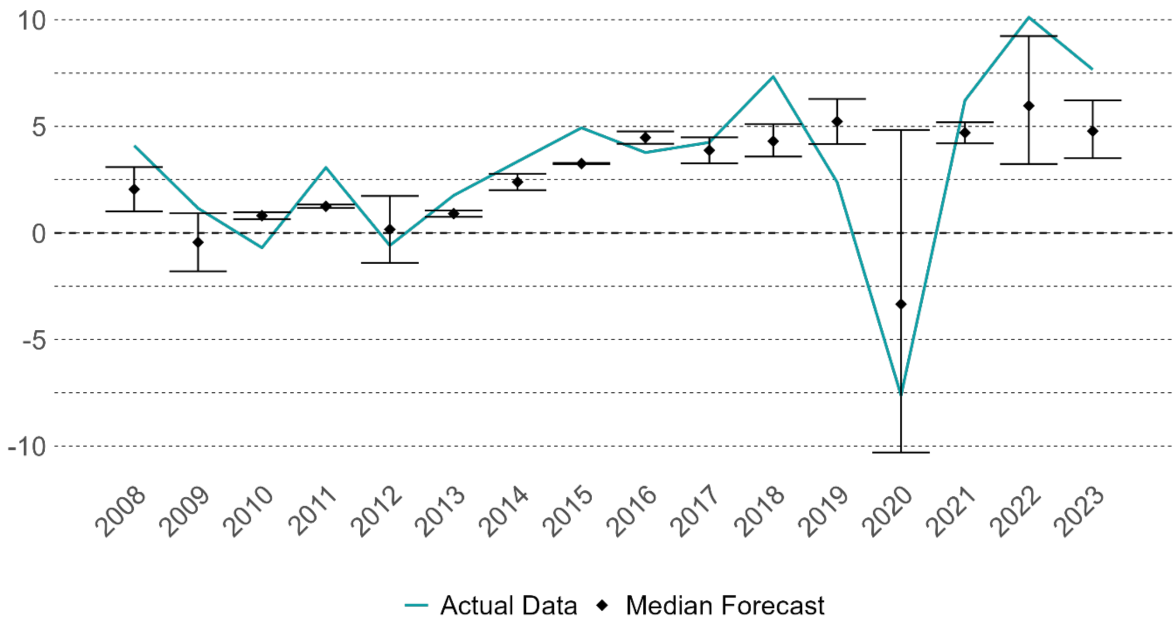
Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts. *NBER, economic forecasts and expectations: Analysis of forecasting behaviour and performance*, 3-46.

Stekler, H. O. (2010). Perspectives on Evaluating Macroeconomic Forecasts. *George Washington University*(RPF Working Paper No. 2010-002.).

11. Appendix

Real Economic Activity Figures

Distribution of Forecasts for Consumption
(annual percentage change)

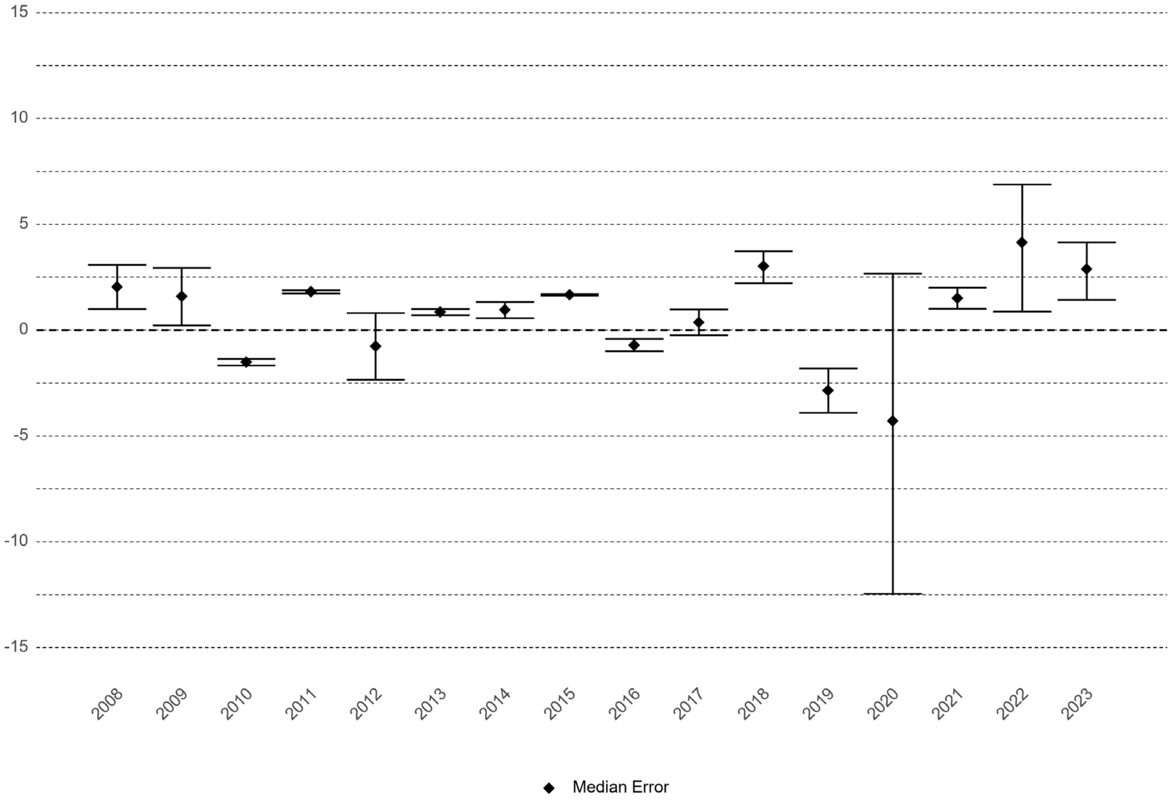


Source: CBM projections, CBM real-time GDP database

Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2017 the range of forecasts includes 2 data points, from 2018-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t + 1), where t is the time-period being forecasted.

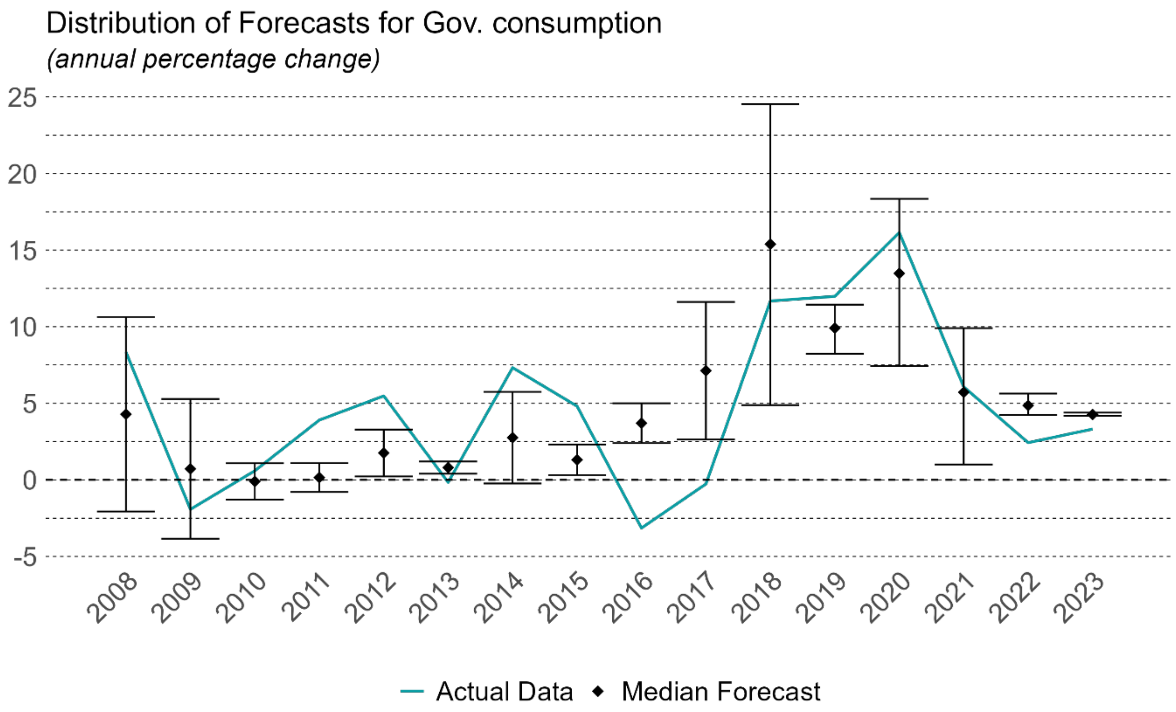
Distribution of Forecast Errors for Consumption

(percentage point)



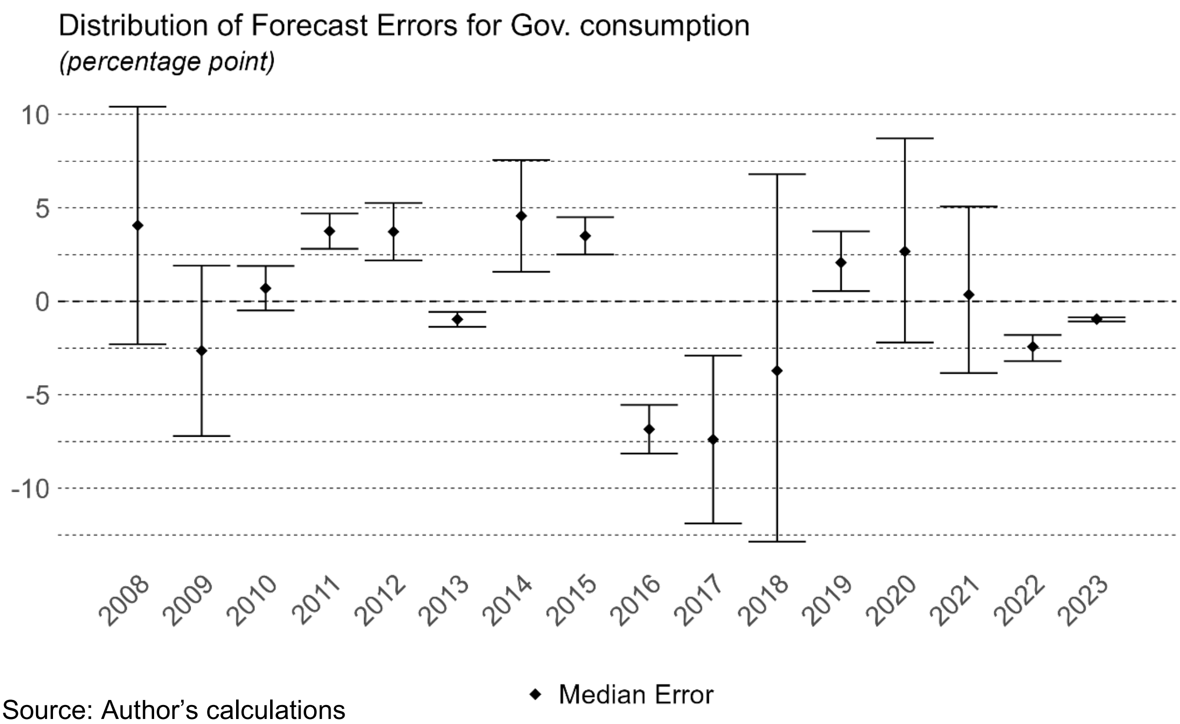
Source: Author’s calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.



Source: CBM projections, CBM real-time GDP database

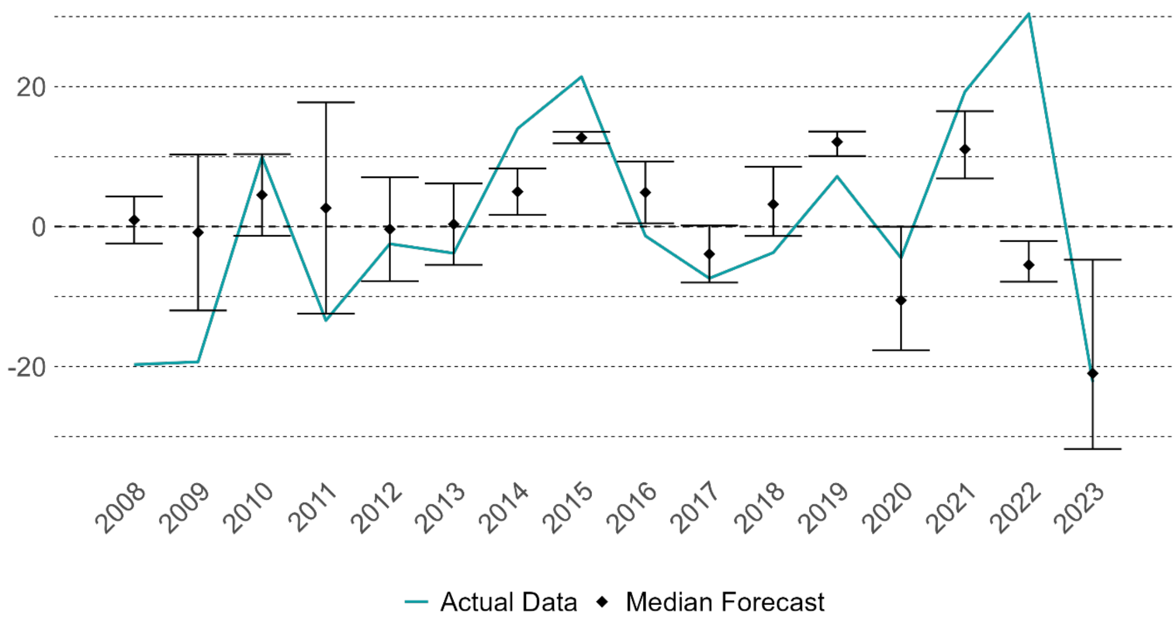
Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2017 the range of forecasts includes 2 data points, from 2018-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t + 1), where t is the time-period being forecasted.



Source: Author's calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.

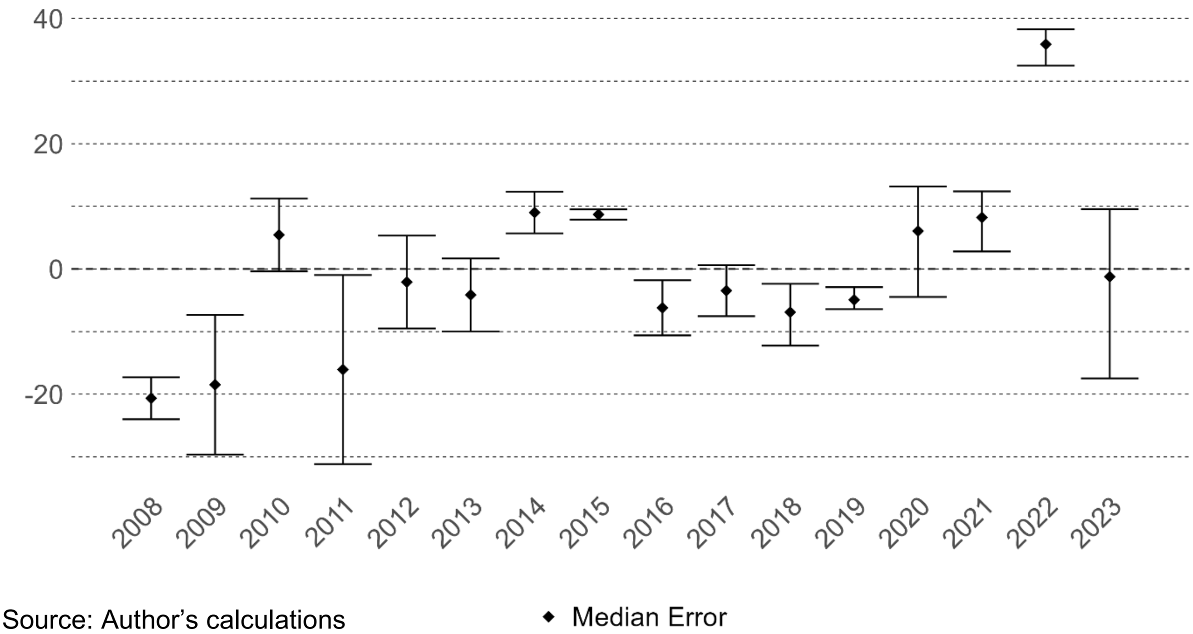
Distribution of Forecasts for Investment
(annual percentage change)



Source: CBM projections, CBM real-time GDP database

Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2017 the range of forecasts includes 2 data points, from 2018-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t + 1), where t is the time-period being forecasted.

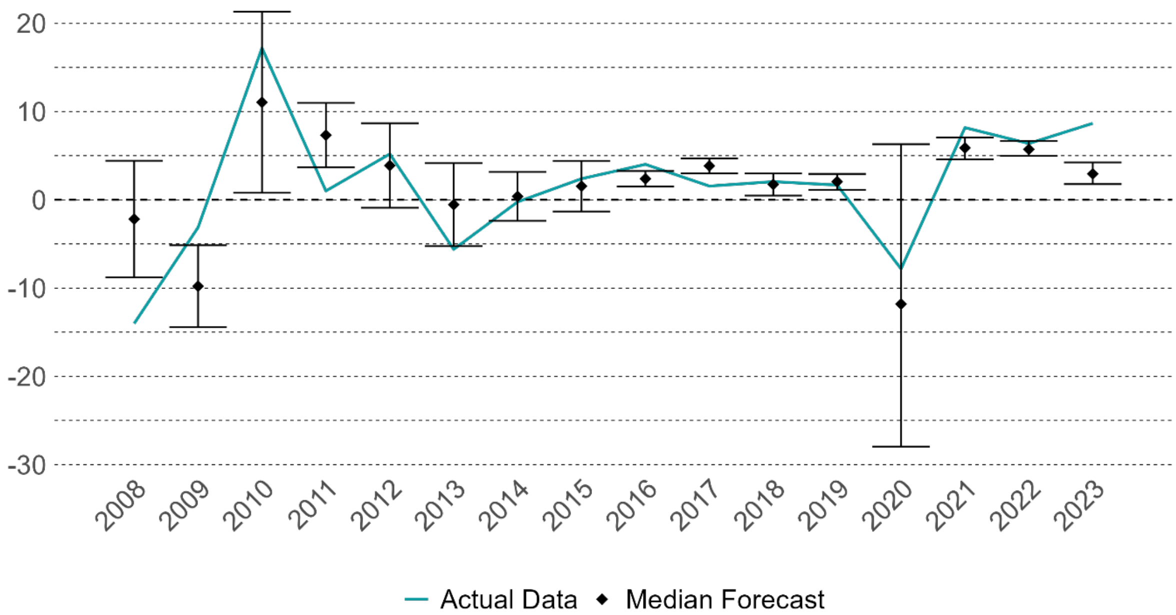
Distribution of Forecast Errors for Investment
(percentage point)



Source: Author’s calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.

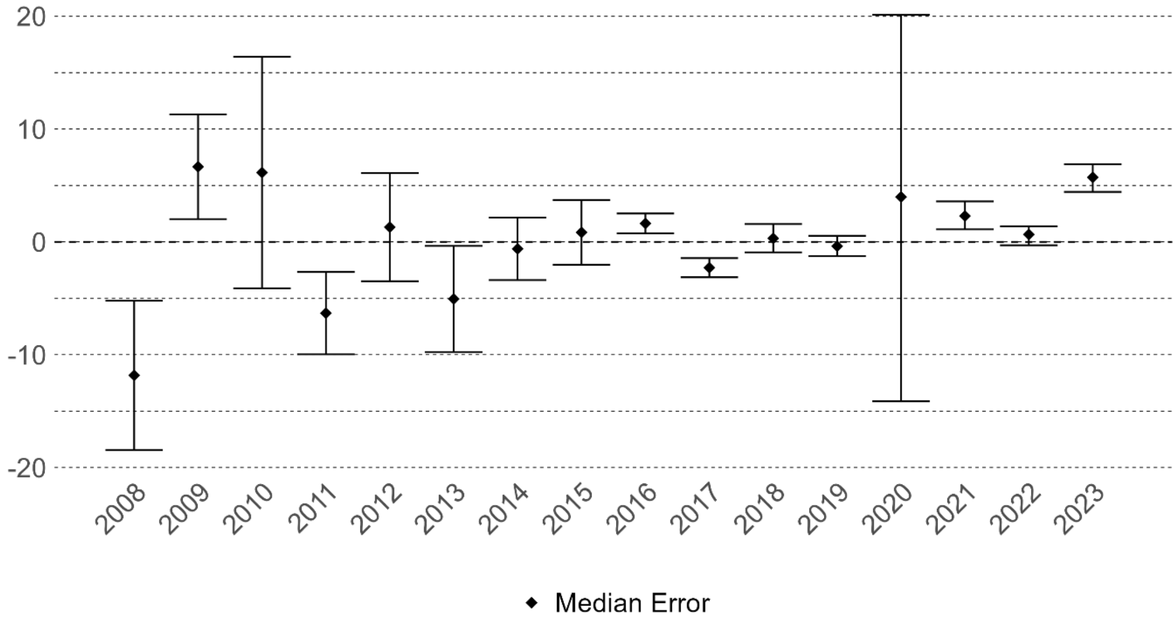
Distribution of Forecasts for Exports
(annual percentage change)



Source: CBM projections, CBM real-time GDP database

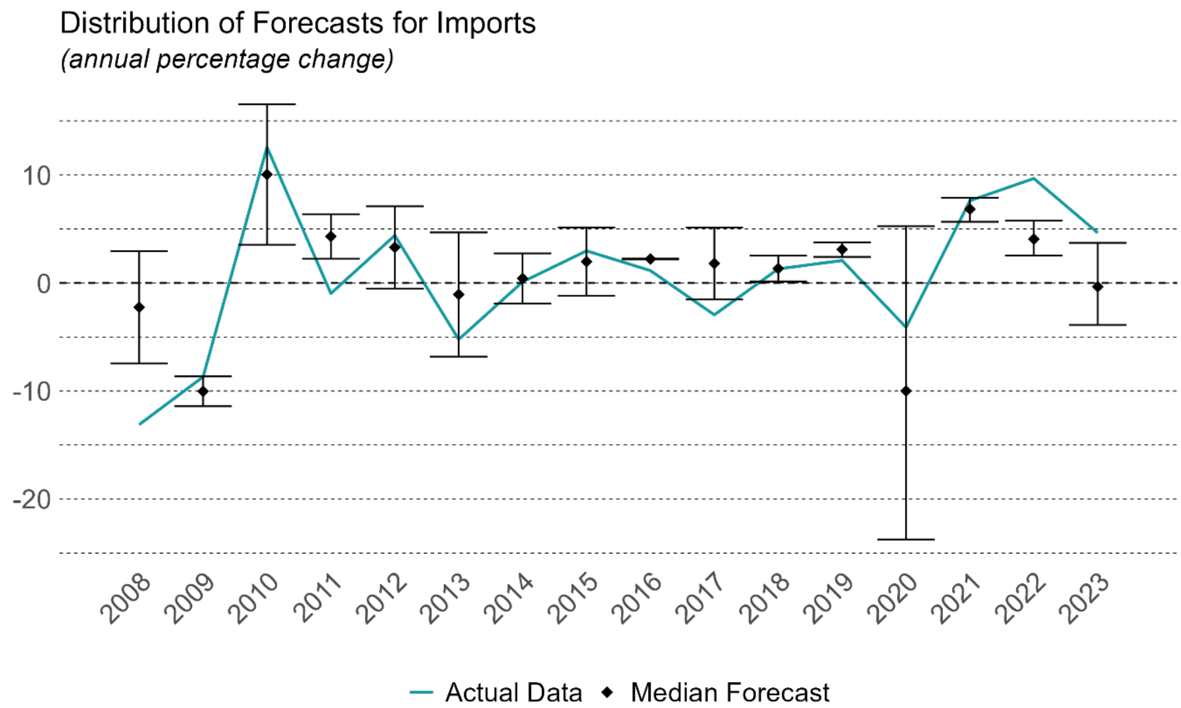
Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2017 the range of forecasts includes 2 data points, from 2018-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t + 1), where t is the time-period being forecasted.

Distribution of Forecast Errors for Exports
(percentage point)



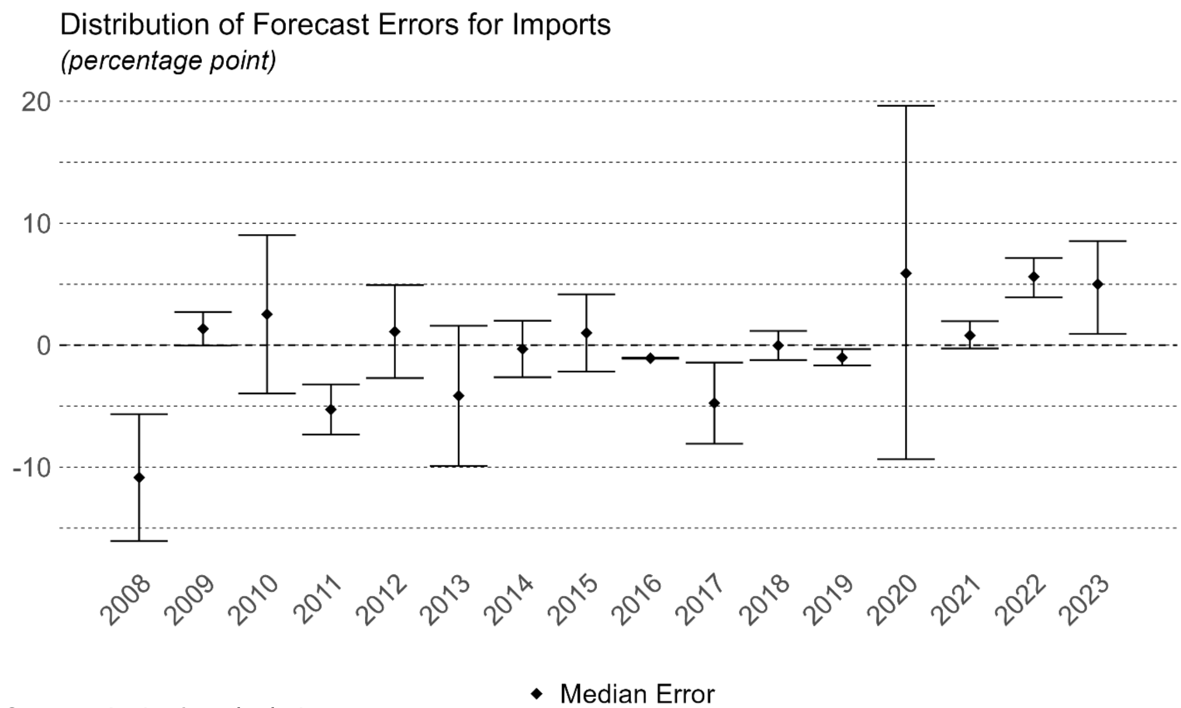
Source: Author’s calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.



Source: CBM projections, CBM real-time GDP database

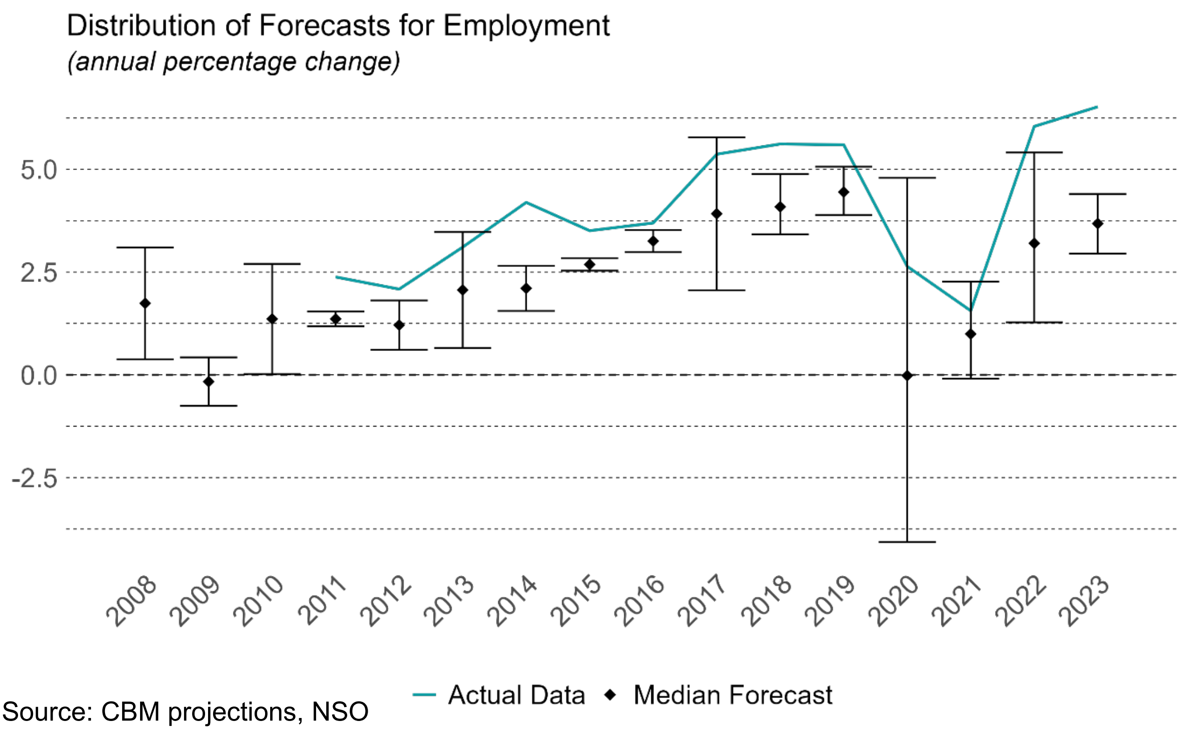
Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2017 the range of forecasts includes 2 data points, from 2018-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t + 1), where t is the time-period being forecasted.



Source: Author's calculations

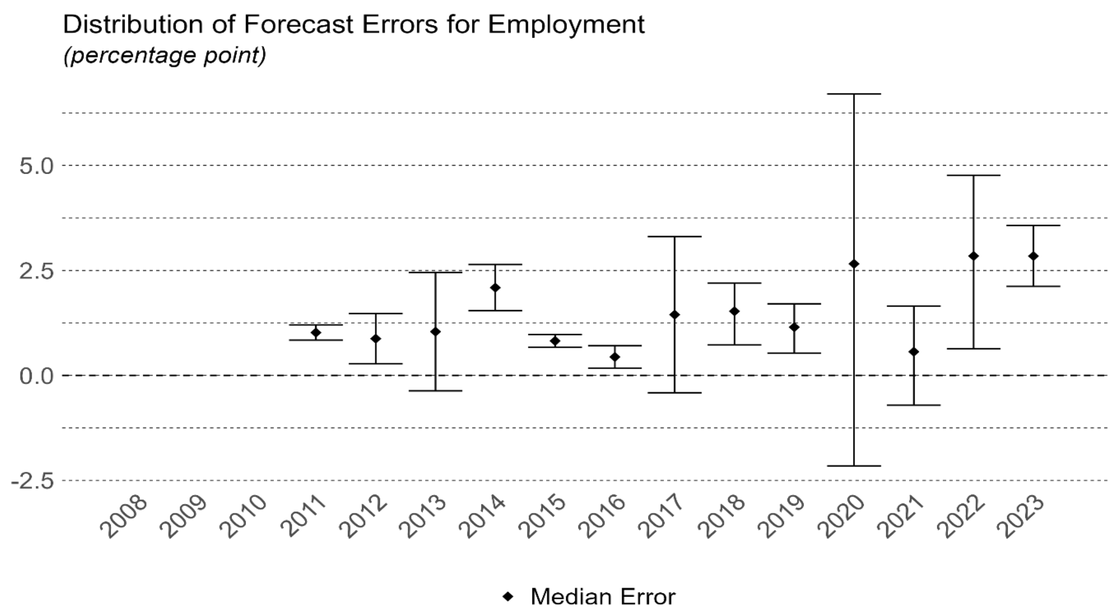
Note: The error bars represent the range in the forecast errors computed for that particular year. From 2008-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.

Labour Market Figures



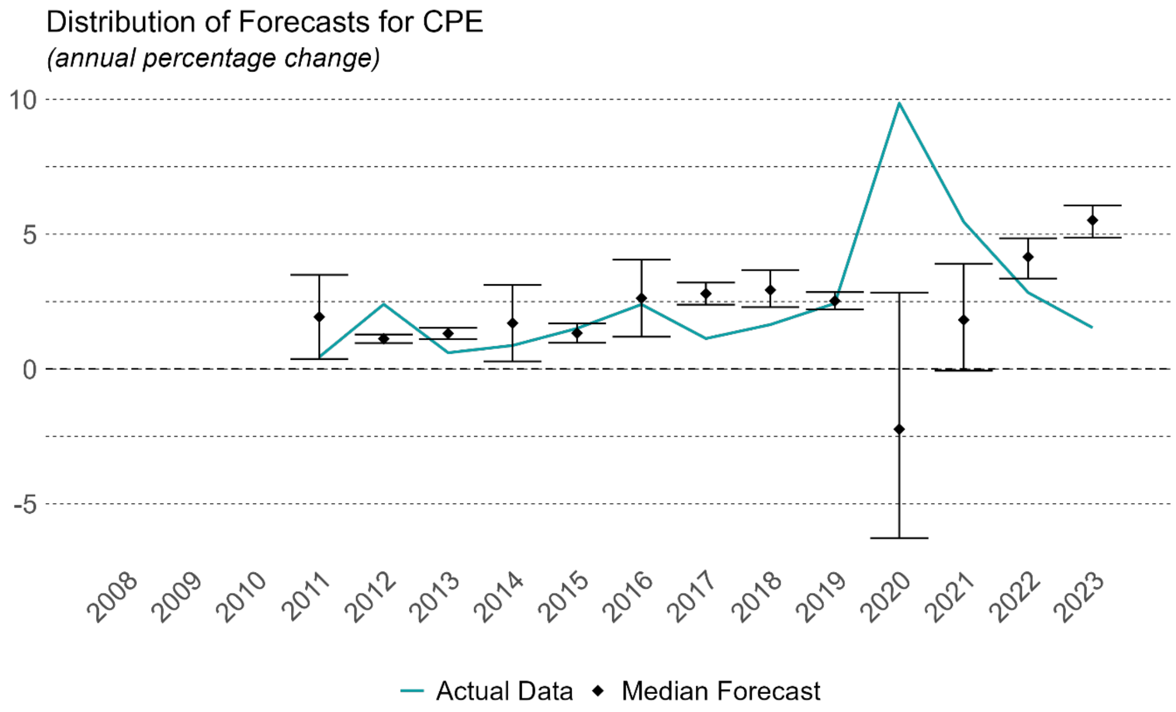
Source: CBM projections, NSO

Note: The bars represent the range in the forecasts conducted in that particular year. From 2008-2017 the range of forecasts includes 2 data points, from 2018-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t + 1), where t is the time-period being forecasted.



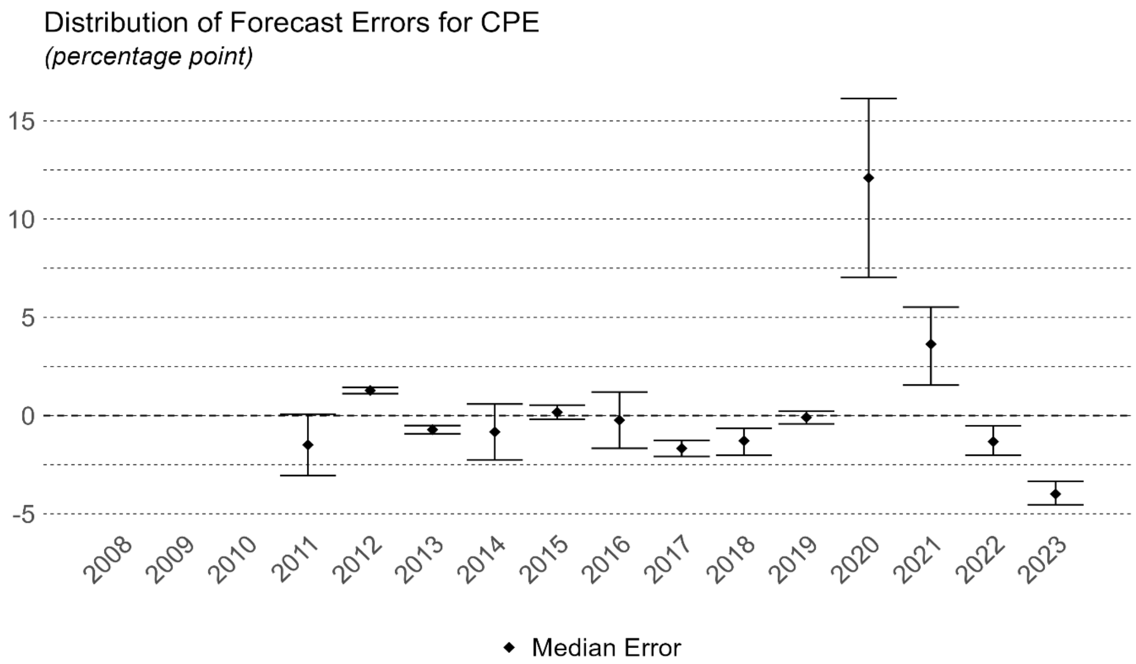
Source: Author’s calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2011-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.



Source: CBM projections, NSO

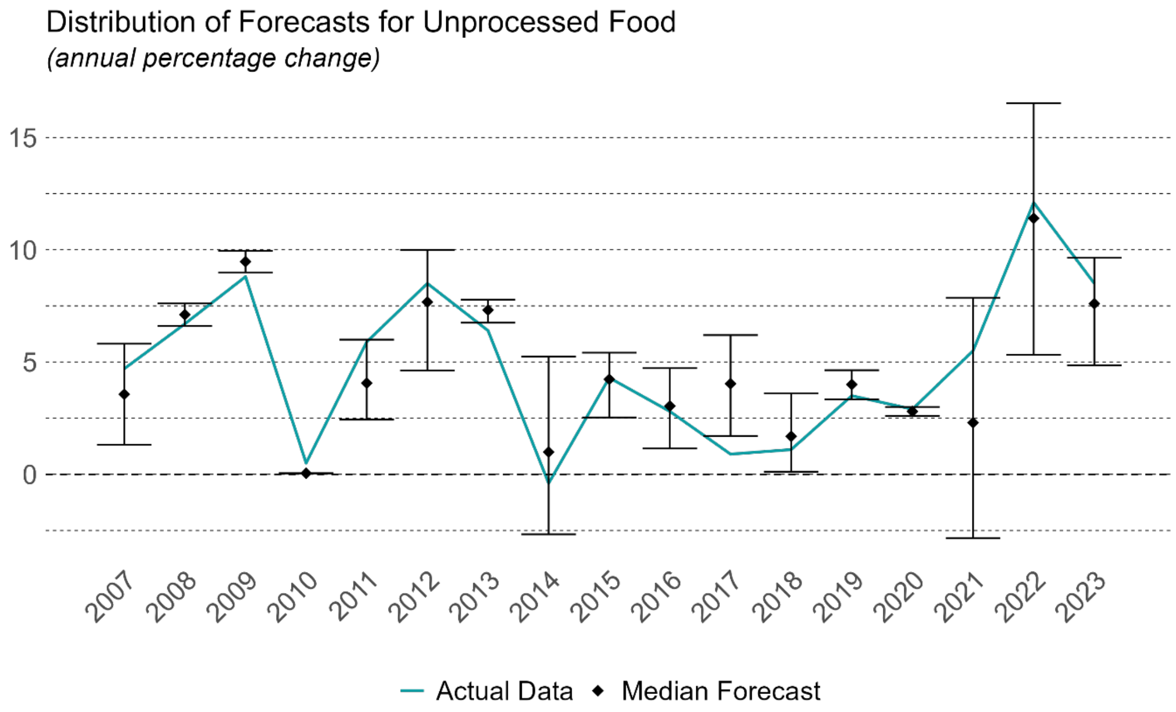
Note: The bars represent the range in the forecasts conducted in that particular year. From 2011-2017 the range of forecasts includes 2 data points, from 2018-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data which became available in the first quarter of (t + 1), where t is the time-period being forecasted.



Source: Author's calculations

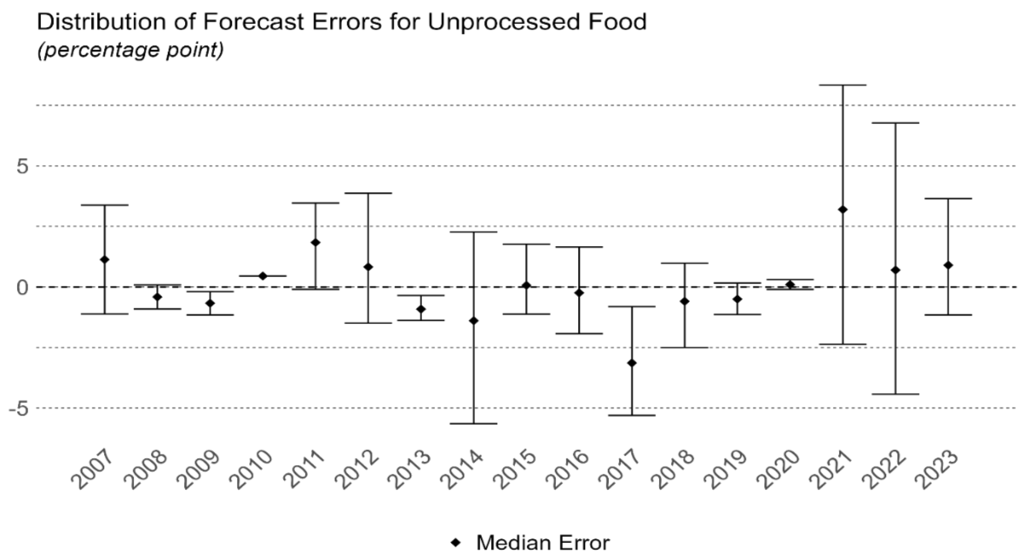
Note: The error bars represent the range in the forecast errors computed for that particular year. From 2011-2017 the range of forecast errors includes 2 data points, from 2018-2023 the range of forecast errors includes 4 data points.

Prices Figures



Source: CBM projections, NSO

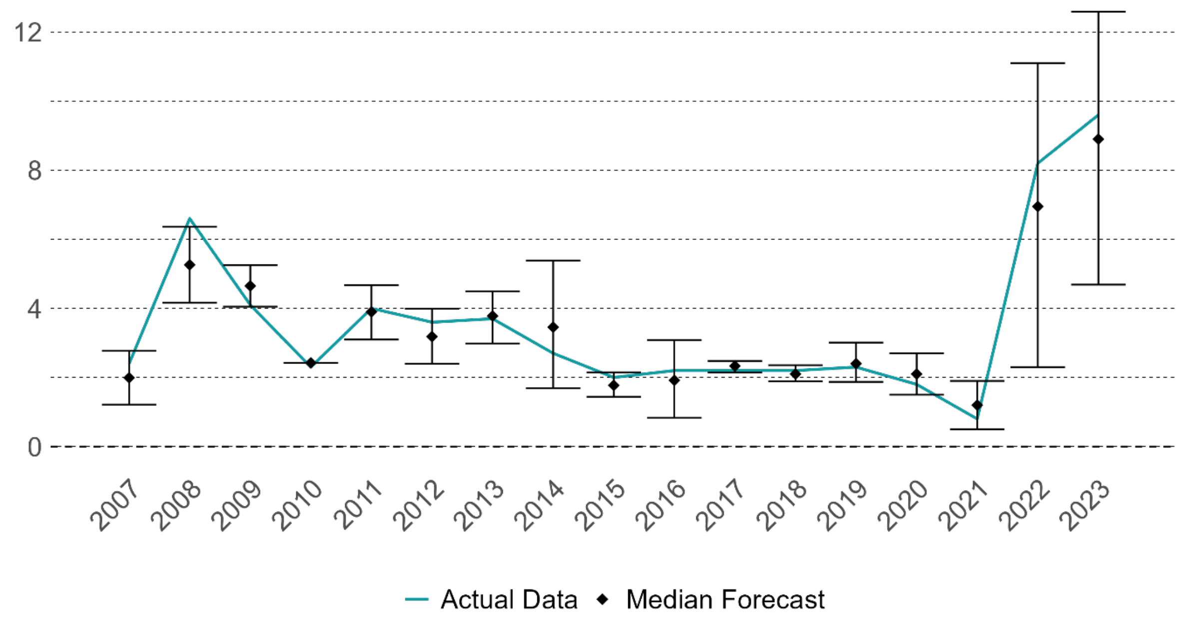
Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2011 the range of forecasts includes 2 data points, from 2012-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data in the latest data release.



Source: Author's calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2011 the range of forecast errors includes 2 data points, from 2012-2023 the range of forecast errors includes 4 data points.

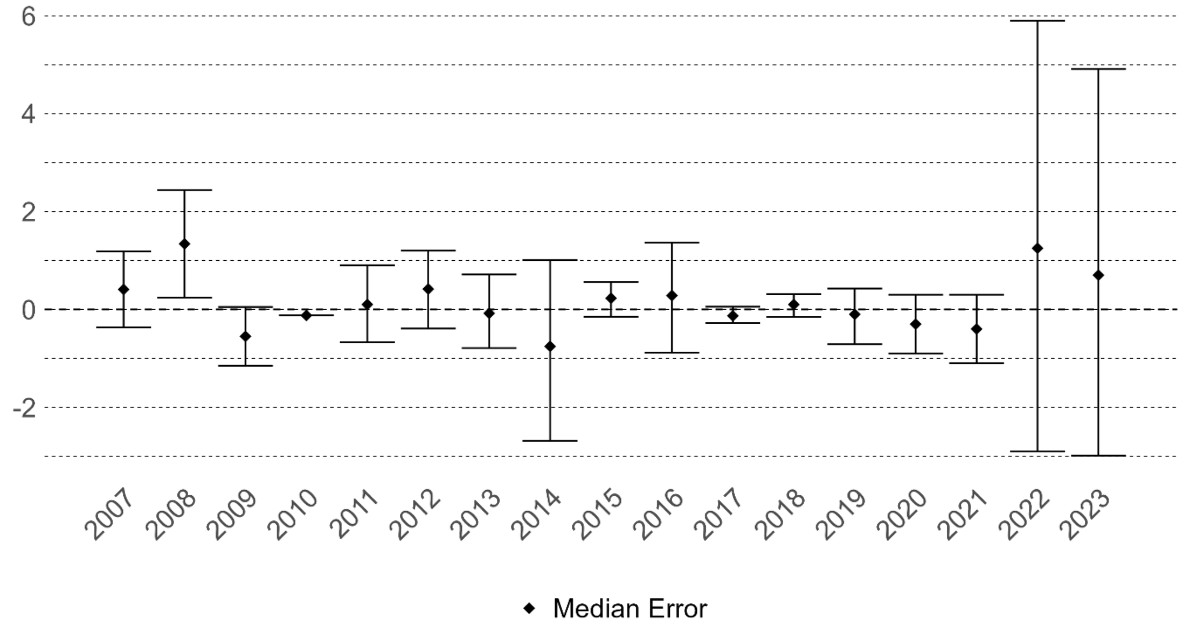
Distribution of Forecasts for Processed Food
(annual percentage change)



Source: CBM projections, NSO

Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2011 the range of forecasts includes 2 data points, from 2012-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data in the latest data release.

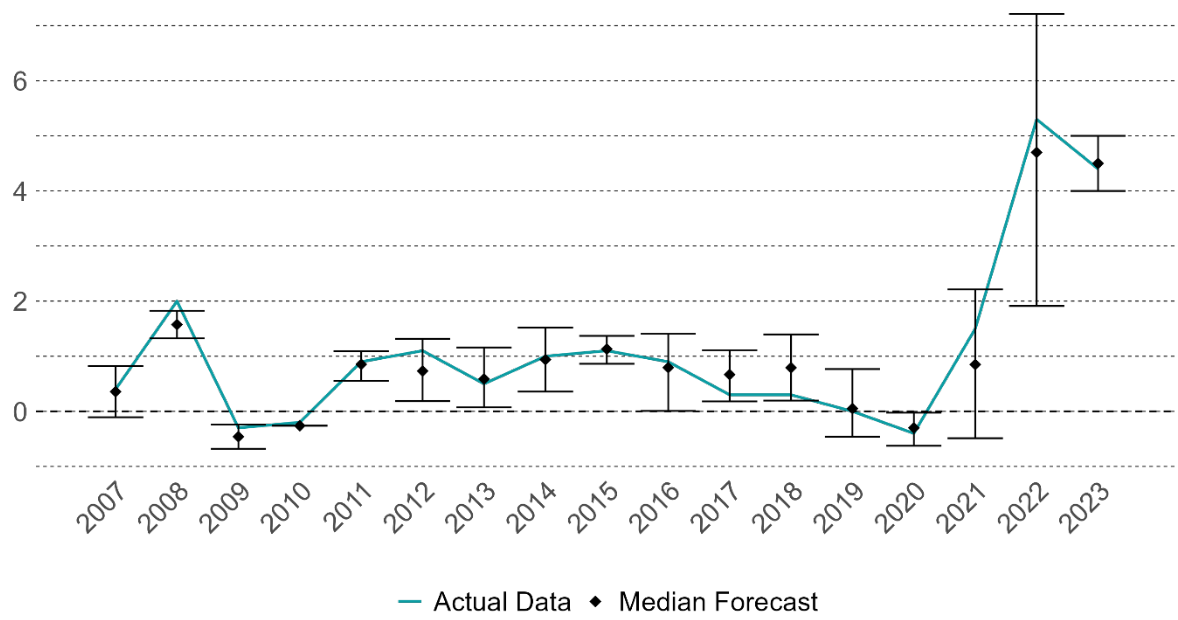
Distribution of Forecast Errors for Processed Food
(percentage point)



Source: Author's calculations

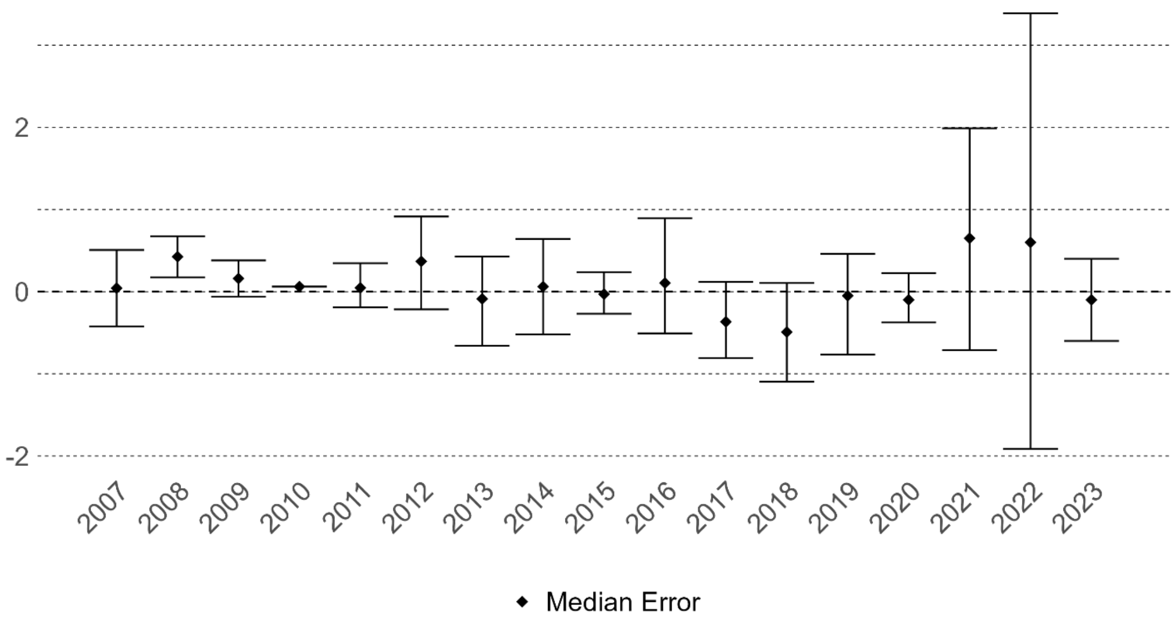
Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2011 the range of forecast errors includes 2 data points, from 2012-2023 the range of forecast errors includes 4 data points.

Distribution of Forecasts for NEIG
(annual percentage change)



Source: CBM projections, NSO Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2011 the range of forecasts includes 2 data points, from 2012-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data in the latest data release.

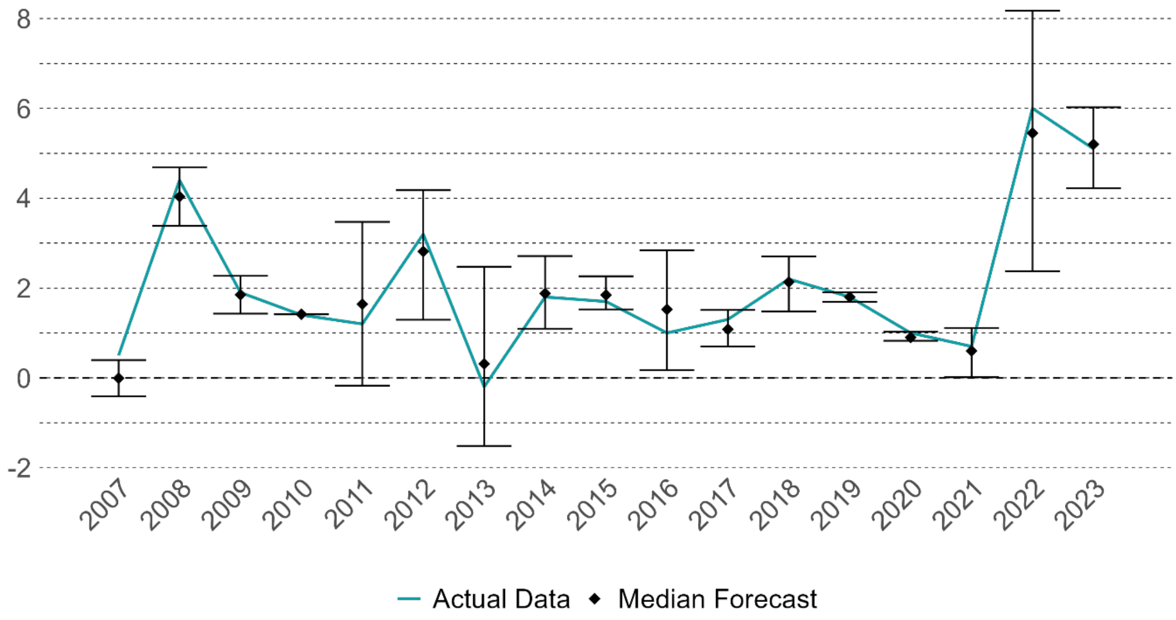
Distribution of Forecast Errors for NEIG
(percentage point)



Source: Author's calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2011 the range of forecast errors includes 2 data points, from 2012-2023 the range of forecast errors includes 4 data points.

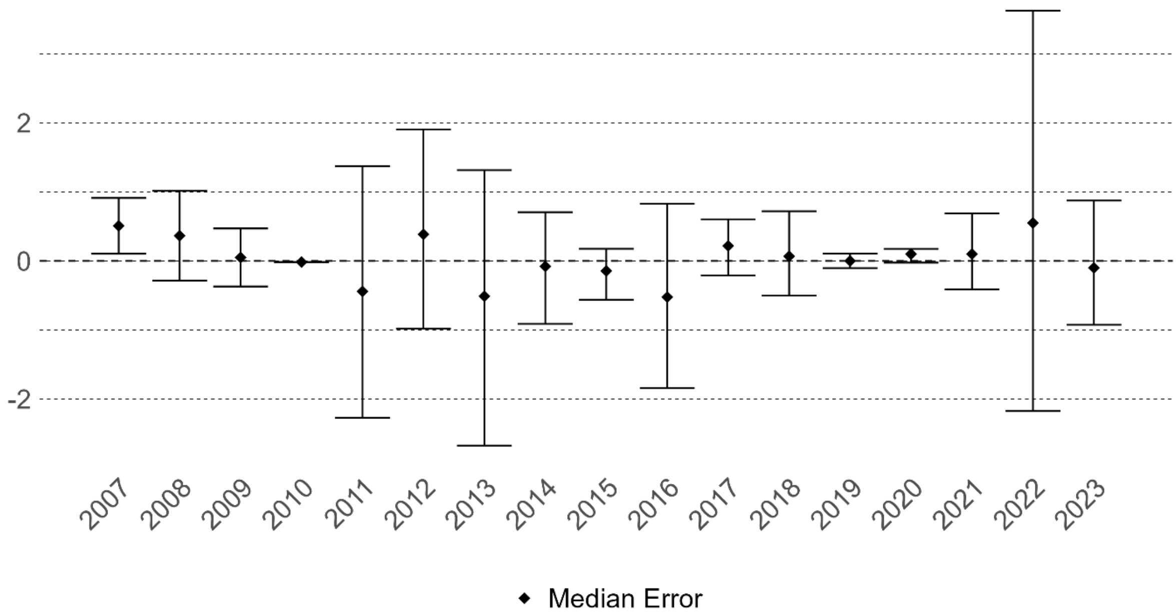
Distribution of Forecasts for Services
(annual percentage change)



Source: CBM projections, NSO

Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2011 the range of forecasts includes 2 data points, from 2012-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data in the latest data release.

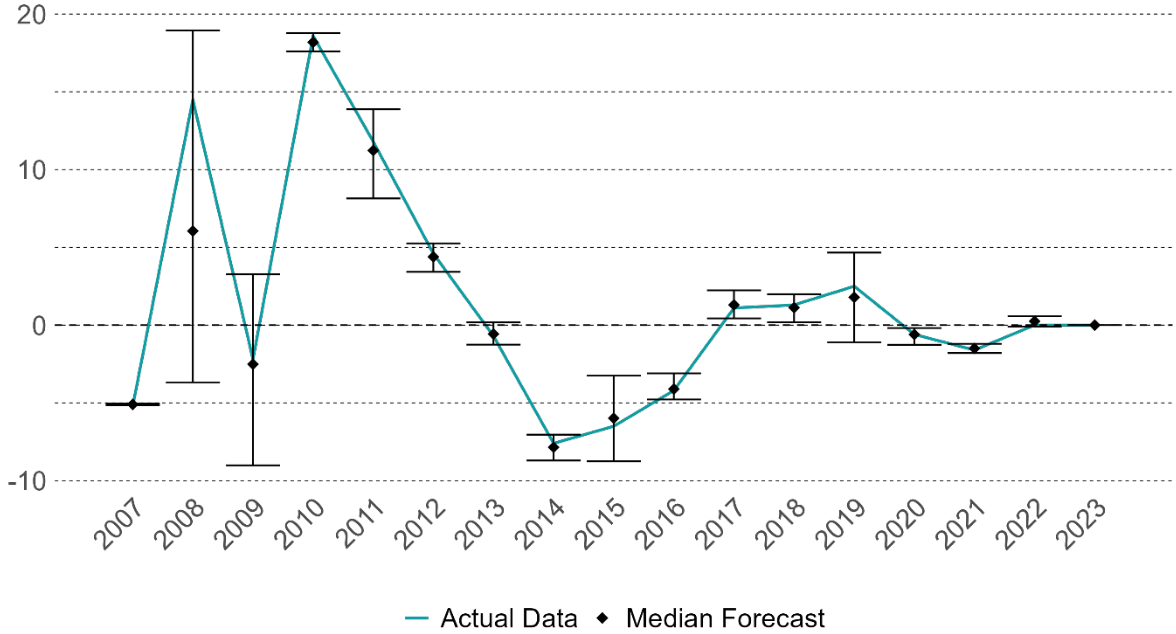
Distribution of Forecast Errors for Services
(percentage point)



Source: Author's calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2011 the range of forecast errors includes 2 data points, from 2012-2023 the range of forecast errors includes 4 data points.

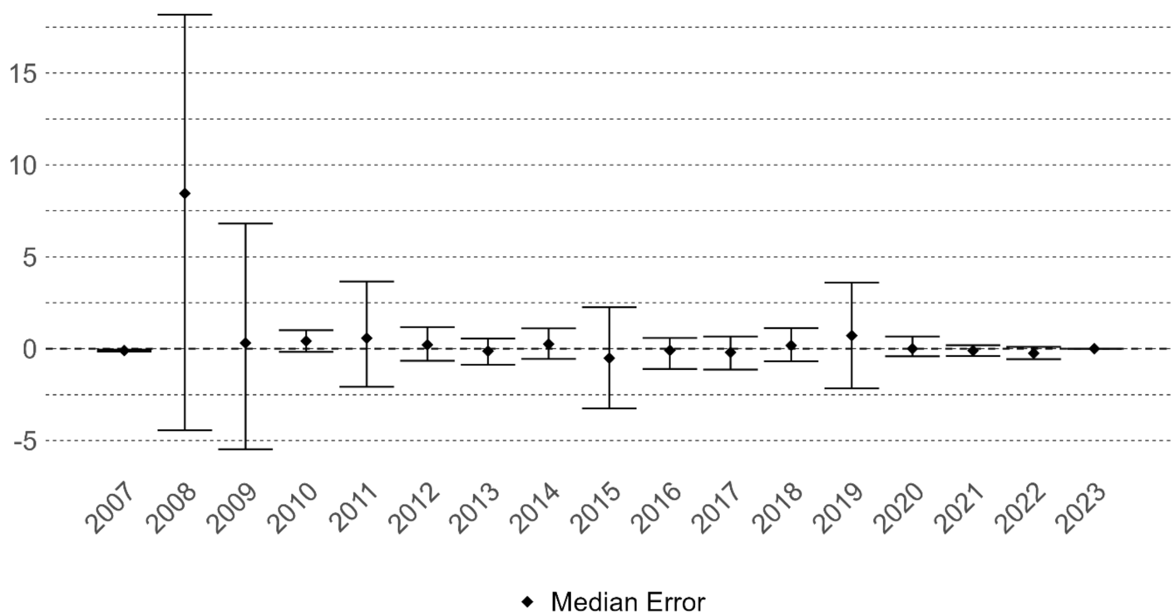
Distribution of Forecasts for Energy
(annual percentage change)



Source: CBM projections, NSO

Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2011 the range of forecasts includes 2 data points, from 2012-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data in the latest data release.

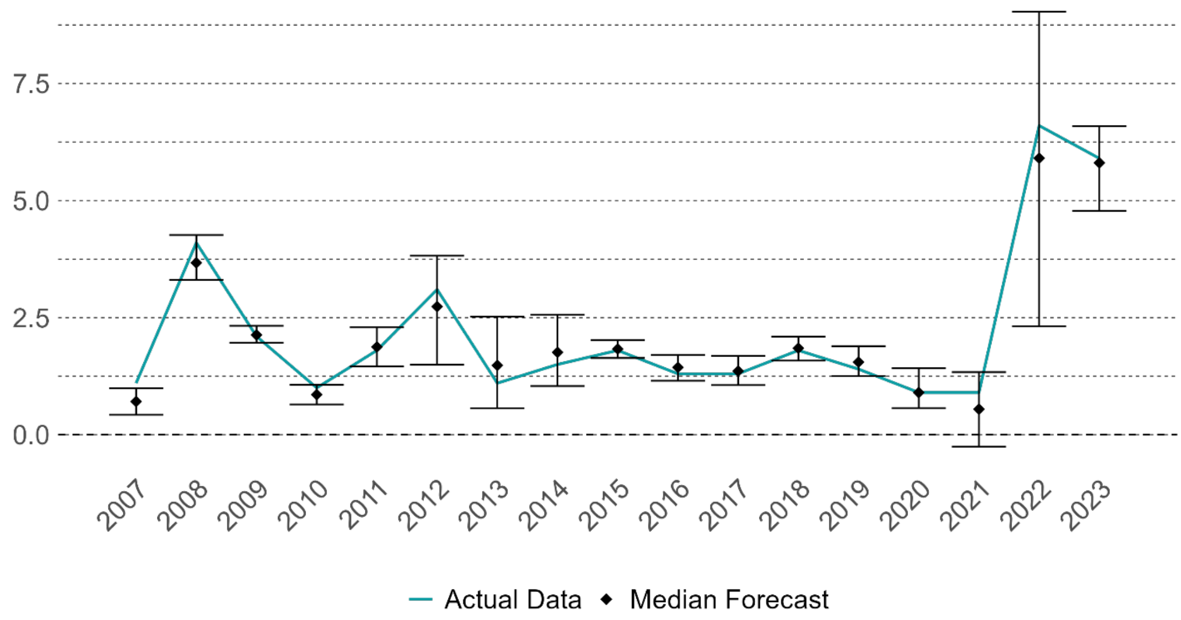
Distribution of Forecast Errors for Energy
(percentage point)



Source: Author's calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2011 the range of forecast errors includes 2 data points, from 2012-2023 the range of forecast errors includes 4 data points.

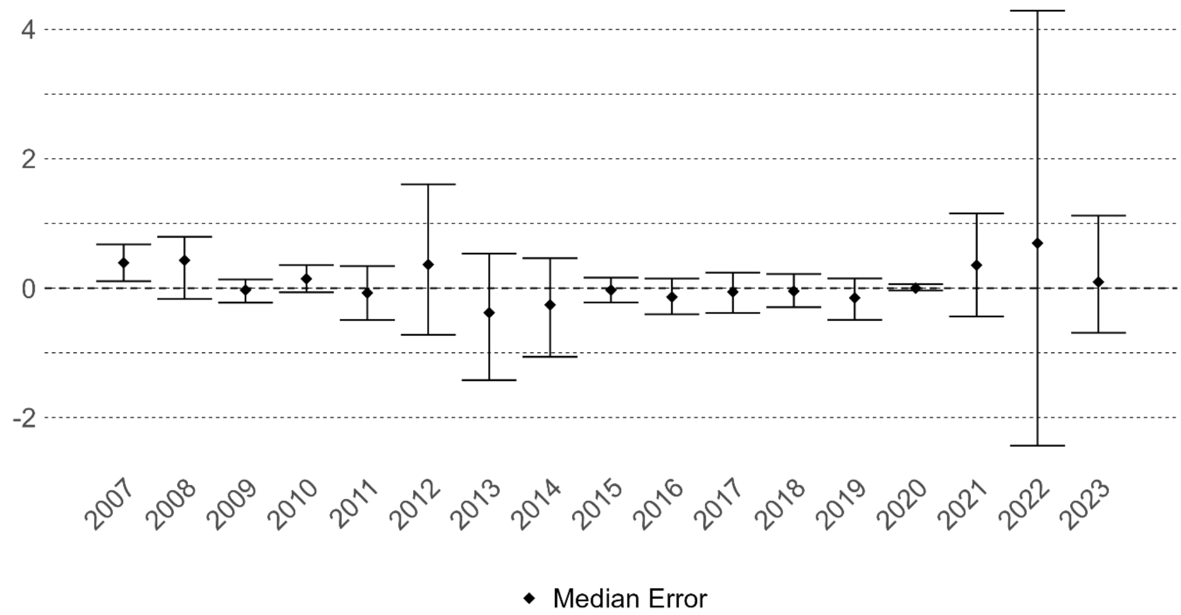
Distribution of Forecasts for HICP exc. Energy
(annual percentage change)



Source: CBM projections, NSO

Note: The bars represent the range in the forecasts conducted in that particular year. From 2007-2011 the range of forecasts includes 2 data points, from 2012-2023 the range of forecasts includes 4 data points. Actual data represents the outturn of the data in the latest data release.

Distribution of Forecast Errors for HICP exc. Energy
(percentage point)



Source: Author's calculations

Note: The error bars represent the range in the forecast errors computed for that particular year. From 2007-2011 the range of forecast errors includes 2 data points, from 2012-2023 the range of forecast errors includes 4 data points.

Tables

Table 14: Mapping of forecast round publication and GDP release

Year of forecast	Forecast round publication	GDP release included in forecast
2007	2007Q2	2006Q4
2007	2007Q4	2007Q2
2008	2008Q2	2007Q4
2008	2008Q4	2008Q2
2009	2009Q2	2008Q4
2009	2009Q4	2009Q2
2010	2010Q2	2009Q4
2010	2010Q4	2010Q2
2011	2011Q2	2010Q4
2011	2011Q4	2011Q2
2012	2012Q2	2011Q4
2012	2012Q4	2012Q2
2013	2013Q2	2012Q4
2013	2013Q4	2013Q2
2014	2014Q2	2013Q4
2014	2014Q4	2014Q2
2015	2015Q2	2014Q4
2015	2015Q4	2015Q2
2016	2016Q2	2015Q4
2016	2016Q4	2016Q2
2017	2017Q2	2016Q4
2017	2017Q4	2017Q2
2018	2018Q1	2017Q3
2018	2018Q2	2017Q4
2018	2018Q3	2018Q1
2018	2018Q4	2018Q2
2019	2019Q1	2018Q3
2019	2019Q2	2018Q4
2019	2019Q3	2019Q1
2019	2019Q4	2019Q2
2020	2020Q1	2019Q3
2020	2020Q2	2019Q4
2020	2020Q3	2020Q1
2020	2020Q4	2020Q2
2021	2021Q1	2020Q3
2021	2021Q2	2020Q4
2021	2021Q3	2021Q1
2021	2021Q4	2021Q2
2022	2022Q1	2021Q3
2022	2022Q2	2021Q4
2022	2022Q3	2022Q1
2022	2022Q4	2022Q2
2023	2023Q1	2022Q3
2023	2023Q2	2022Q4
2023	2023Q3	2023Q1
2023	2023Q4	2023Q2

Source: CBM

Table 15: Forecast performance metrics by horizon and dataset for the main macroeconomic variables

Variable	Horizon	Error Metrics (full sample/adjusted sample)			
		St. deviation	RMSE	Mean error	Mean absolute error
GDP	h = 0	2.57 / 1.02	2.66 / 1.35	0.81 / 0.91	1.79 / 1.16
	h = 1	4.75 / 1.32	4.67 / 1.85	0.02 / 1.33	3.19 / 1.55
	h = 2	5.50 / 1.40	5.45 / 2.71	0.81 / 2.34	4.26 / 2.35
Unemployment Rate	h = 0	0.40 / 0.28	0.45 / 0.32	-0.21 / -0.18	0.34 / 0.25
	h = 1	0.61 / 0.41	0.78 / 0.77	-0.50 / -0.66	0.68 / 0.66
	h = 2	0.68 / 0.44	1.08 / 1.28	-0.85 / -1.21	0.94 / 1.21
HICP	h = 0	0.59 / 0.64	0.59 / 0.63	0.04 / 0.01	0.29 / 0.31
	h = 1	1.70 / 1.80	1.70 / 1.85	0.31 / 0.51	1.07 / 1.17
	h = 2	2.05 / 2.16	2.04 / 2.24	0.28 / 0.71	1.45 / 1.54

Notes: Sample starts from 2013
Source: Author's calculations

Table 16: Forecast performance metrics by horizon and dataset for the sub-components of real GDP

Variable	Horizon	Error Metrics (full sample/adjusted sample)			
		St. deviation	RMSE	Mean error	Mean absolute error
Consumption	h = 0	3.24 / 2.29	3.23 / 2.56	0.48 / 1.23	2.52 / 2.22
	h = 1	4.98 / 2.02	4.90 / 2.65	-0.19 / 1.77	3.39 / 2.34
	h = 2	5.58 / 2.27	5.53 / 3.82	0.80 / 3.12	4.43 / 3.35
Government consumption	h = 0	4.19 / 3.96	4.14 / 4.04	-0.30 / -1.12	3.29 / 3.35
	h = 1	5.11 / 4.75	5.77 / 4.81	2.84 / 1.28	4.96 / 4.18
	h = 2	4.80 / 4.29	5.18 / 4.23	2.17 / 0.71	3.96 / 3.29
Investment	h = 0	13.90 / 15.62	14.07 / 15.57	3.25 / 2.77	9.81 / 10.68
	h = 1	15.53 / 17.05	15.42 / 16.73	-2.14 / -1.51	12.65 / 13.89
	h = 2	17.84 / 19.41	17.56 / 18.90	1.48 / 1.21	14.86 / 15.41
Exports	h = 0	3.64 / 2.93	3.70 / 2.92	0.91 / 0.51	2.64 / 2.06
	h = 1	4.51 / 2.64	4.50 / 2.62	-0.74 / 0.41	3.12 / 1.99
	h = 2	5.48 / 2.96	5.37 / 3.15	0.14 / 1.29	4.34 / 2.67
Imports	h = 0	4.00 / 3.58	4.10 / 3.57	1.12 / 0.65	3.03 / 2.72
	h = 1	3.88 / 3.29	3.95 / 3.23	-1.05 / -0.30	3.17 / 2.66
	h = 2	4.94 / 4.09	4.84 / 4.06	0.20 / 0.81	4.04 / 3.11

Notes: Sample starts from 2013
Source: Author's calculations

Table 17: Forecast performance metrics by horizon and dataset for the sub-components of HICP

Variable	Horizon	Error Metrics (full sample/adjusted sample)			
		St. deviation	RMSE	Mean error	Mean absolute error
Unprocessed food	h = 0	2.13 / 2.02	2.11 / 2.03	-0.05 / -0.37	1.41 / 1.36
	h = 1	4.14 / 4.50	4.12 / 4.43	0.48 / 0.25	3.08 / 3.41
	h = 2	3.85 / 4.31	3.89 / 4.30	0.85 / 0.80	2.77 / 3.12
Processed food	h = 0	1.06 / 1.13	1.06 / 1.14	0.15 / 0.25	0.55 / 0.59
	h = 1	2.59 / 2.72	2.74 / 3.01	1.00 / 1.37	1.58 / 1.77
	h = 2	3.13 / 3.33	3.26 / 3.66	1.05 / 1.65	1.88 / 2.16
NEIG	h = 0	0.56 / 0.55	0.56 / 0.55	0.04 / -0.01	0.34 / 0.33
	h = 1	1.65 / 1.65	1.67 / 1.71	0.39 / 0.52	1.11 / 1.03
	h = 2	1.97 / 2.03	1.97 / 2.13	0.37 / 0.75	1.25 / 1.32
Services	h = 0	0.66 / 0.72	0.66 / 0.71	-0.02 / -0.05	0.37 / 0.41
	h = 1	1.63 / 1.72	1.61 / 1.73	0.16 / 0.38	1.17 / 1.24
	h = 2	1.89 / 1.98	1.86 / 1.99	0.02 / 0.42	1.49 / 1.54
Energy	h = 0	0.57 / 0.61	0.56 / 0.61	-0.03 / -0.03	0.33 / 0.37
	h = 1	2.36 / 2.55	2.51 / 2.72	-0.92 / -1.05	1.73 / 1.91
	h = 2	2.03 / 2.30	2.17 / 2.41	-0.83 / -0.85	1.49 / 1.70

Notes: Sample starts from 2013
Source: Author's calculations

Table 18: Horizon specific bias by dataset for the main macroeconomic variables

Horizon	GDP	Unemployment Rate	HICP
	(full sample/adjusted sample)		
	Common bias		
1	0.63 / 0.40	-0.02 / -0.07	0.004 / 0.003
2	1.26 / 0.04	-0.28 / -0.34	-0.02 / -0.02
3	0.72 / 0.51	-0.42*** / -0.37**	0.15 / 0.04
4	-0.62 / -0.03	-0.16 / -0.22	0.32 / -0.06
5	0.29 / -0.16	-0.24 / -0.26	0.22 / -0.11
6	-1.18 / -5.95*	-0.30 / 0.009	0.14 / -0.48**
7	0.29 / -0.18	-0.58*** / -0.53**	0.32 / -0.37
8	-0.38 / -5.88	-0.36 / -0.02	0.65 / -0.30
9	1.01 / 0.79	-0.56** / -0.58**	0.54 / -0.16
10	0.22 / -2.70	-0.36 / -0.16	0.38 / -0.72**
11	1.08 / 0.86	-0.84*** / -0.83***	0.60 / -0.48
12	0.56 / -2.52	-0.56 / -0.23	0.56 / -0.30

*Notes: *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

Table 19: Horizon specific bias by dataset for the sub-components of real GDP

Horizon	Consumption	Government consumption	Investment	Exports	Imports
	(full sample/adjusted sample)				
	Common bias				
1	0.55 / 0.56	-0.25 / 0.16	0.68 / -3.18	0.27 / -0.85	0.33 / -1.14
2	0.83 / 0.15	-0.61 / -1.59	7.03 / -4.79	2.92 / 0.001	3.74** / -0.44
3	0.74 / 0.47	-0.21 / -0.17	-1.99 / -6.82**	0.13 / -0.61	-0.8 / -2.34*
4	-0.50 / 0.27	3.35* / 3.83	-2.39 / -7.38	-0.20 / -0.77	0.13 / -1.12
5	-0.007 / -0.60	1.69 / 1.91	-5.66 / -7.08**	0.20 / -0.50	-1.32 / -2.22
6	-0.92 / -6.68**	4.37** / 7.72**	-4.06 / -9.54	-0.73 / -5.37	-0.51 / -4.02
7	0.13 / -0.55	2.06 / 2.07	-4.62 / -5.83*	-0.48 / -1.21	-1.76 / -2.65*
8	-0.13 / -6.46**	3.28 / 7.23*	-1.22 / -7.97	0.45 / -5.89	0.86 / -4.43
9	0.67 / -0.18	2.05 / 2.42	-0.93 / -0.62	0.34 / -0.41	-0.41 / -1.22
10	1.45 / -4.45	2.08 / 5.28	-0.13 / 2.69	1.64 / -2.34	2.05 / -0.90
11	0.84 / -0.16	2.38* / 2.95*	0.14 / 0.43	0.13 / -0.84	-0.41 / -1.43
12	1.57 / -4.32	1.26 / 5.29	0.15 / 4.36	1.88 / -2.79	1.97 / -1.24

*Notes: *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

Table 20: Horizon specific bias by dataset for the sub-components of HICP (full sample/adjusted sample)

Horizon	Unprocessed food	Processed food	NEIG	Services	Energy
	Common bias				
1	-0.03 / -0.11	-0.04 / -0.06	-0.03 / -0.03	0.06 / 0.09	-0.003 / 0.00
2	0.004 / -0.18	0.03 / 0.06	-0.02 / -0.03	-0.04 / -0.004	-0.14 / -0.16
3	0.07 / -0.54	0.20 / 0.07	0.12 / -0.03	0.05 / -0.05	0.99 / 1.32*
4	0.78 / -0.44	0.59 / 0.05	0.30 / 0.06	0.03 / -0.38	0.96 / 1.28
5	0.88 / -0.23	0.66 / 0.08	0.25 / -0.09	-0.07 / -0.45*	0.96 / 1.21
6	0.87 / -1.09	1.04 / 0.18	0.23 / -0.38*	-0.12 / -0.75**	-1.26 / -0.83
7	0.71 / -1.47*	1.25 / -0.01	0.36 / -0.42*	-0.10 / -0.84***	0.42 / 0.56
8	1.00 / -1.36	1.85** / -0.02	0.52 / -0.26	0.36 / -0.54	-0.67 / -0.94
9	0.99 / -0.60	1.41 / -0.12	0.39 / -0.26	0.41 / -0.36	1.48 / 1.75
10	0.99 / -1.25	1.44 / -0.41	0.44 / -0.42	0.23 / -0.76*	-1.87 / -1.76
11	1.74 / -0.55	1.86* / -0.46	0.59 / -0.41	0.20 / -0.90**	-0.46 / -0.31
12	1.07 / -0.76	1.80 / -0.26	0.31 / -0.32	0.20 / -0.56	0.16 / 0.02

*Notes: *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

Table 21: Biasedness tests by horizon and dataset for the main macroeconomic variables

Variable	Horizon	Biasedness Tests (full sample/adjusted sample)	
		Test statistic	P-value
Mean Error t-test			
GDP	h = 0	1.65 / 3.24	0.11 / 0.00***
	h = 1	-0.16 / 2.28	0.88 / 0.03**
	h = 2	0.65 / 3.66	0.52 / 0.00***
Unemployment Rate	h = 0	-3.88 / -4.24	0.00*** / 0.00***
	h = 1	-4.88 / -6.79	0.00*** / 0.00***
	h = 2	-6.14 / -7.24	0.00*** / 0.00***
HICP	h = 0	1.47 / 1.18	0.15 / 0.24
	h = 1	1.08 / 1.52	0.29 / 0.14
	h = 2	0.76 / 1.60	0.45 / 0.12
CUSUM test			
GDP	h = 0	0.40 / 0.88	0.85 / 0.08*
	h = 1	0.53 / 1.28	0.57 / 0.00***
	h = 2	0.57 / 1.63	0.47 / 0.00***
Unemployment Rate	h = 0	0.31 / 0.41	0.96 / 0.83
	h = 1	0.51 / 0.68	0.61 / 0.27
	h = 2	1.19 / 1.52	0.01** / 0.00***
HICP	h = 0	0.70 / 0.66	0.25 / 0.30
	h = 1	0.55 / 0.60	0.52 / 0.42
	h = 2	0.56 / 0.73	0.48 / 0.21

*Notes: *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations

Table 22: Ljung-Box test by horizon and dataset for the main macroeconomic variables

Variable	Horizon	LB test (full sample/adjusted sample)	
		Test statistic	P-value
GDP	h = 0	0.29 / 0.17	0.59 / 0.68
Unemployment rate	h = 0	0.02 / 0.30	0.90 / 0.59
HICP	h = 0	1.08 / 0.11	0.30 / 0.74

*Notes: *, **, *** indicate the null hypothesis is rejected at 10%, 5% and 1% significance level respectively.*
Source: Author's calculations