BANK OF GREECE
EUROSYSTEM

Working Paper

Outlier-robust evaluation of fixed-event macroeconomic survey expectations

Panagiotis Delis
Georgios Kontogeorgos

356

JANUARY 2026

BANK OF GREECE
Economic Analysis and Research Department – Special Studies Division
21, E. Venizelos Avenue
GR-102 50 Athens
Tel: +30210-320 3610
Fax: +30210-320 2432

www.bankofgreece.gr

Published by the Bank of Greece, Athens, Greece

ISSN: 2654-1912 (online)
DOI: https://doi.org/10.52903/wp2026356

OUTLIER-ROBUST EVALUATION OF FIXED-EVENT MACROECONOMIC SURVEY EXPECTATIONS

Panagiotis Delis
Bank of Greece and University of Piraeus

Georgios Kontogeorgos
University of Strathclyde

ABSTRACT

Evaluating macroeconomic forecasts for their unbiasedness and efficiency is essential for policymakers, economists, and investors. The degree to which these stakeholders incorporate expectations into their decision-making processes depends heavily on how these forecasts have been formed. Existing methodologies do not explicitly address critical dimensions, such as the variability of bias across target events and forecast horizons, the forecast errors' heteroscedasticity, and the potential state-dependence in bias. More importantly, they encounter difficulties during high-uncertainty periods, which can lead to inaccurate inference due to the presence of outliers. Apart from generalising the unbiasedness tests, this study contributes to the literature on both strong and weak efficiency by incorporating these aspects. Finally, the proposed methods are applied to the expectations of a crucial survey of the US economy, namely, the Survey of Primary Dealers (SPD). The findings from this application indicated that interested parties should investigate unbiasedness and efficiency in an outlier-robust way, while also allowing for greater flexibility in the methods regarding the variables and periods examined.

Keywords: Fixed-event Forecasts; Forecast Outliers; Forecast Evaluation; NY Fed Survey of Primary Dealers

JEL Classifications: C01, C53, E37

Correspondence:
Panagiotis Delis
Economic Analysis and Research Department
Bank of Greece
El. Venizelos 21, 10250 Athens, Greece
Tel.: +30-2103202371
email: pandelis@bankofgreece.gr

1. Introduction

Over recent decades, there has been an increasing interest in evaluating the information content of macroeconomic surveys and how economic agents, i.e., households, firms, economists, investors, and policymakers, form their views about the future path of the economic outlook and financial markets (Clements, 2019; Coibion and Gorodnichenko, 2026). In particular, households adjust their consumption and savings profile, and non-financial firms develop their financial strategies in terms of capital expenditure and optimal allocation of available resources based on their perception of the future macroeconomic environment. In addition, investors and policymakers, especially those in central banks and international organisations, are interested in the perceptions of the surveys' respondents, since they provide valuable information about market expectations regarding the future path of the economy, financial markets, and policy actions.

Financial firms and investors incorporate these expectations to optimise their portfolios from a mean-variance perspective and maintain their desired level of risk exposure. Furthermore, central banks that operate under the primary mandate of maintaining price stability and achieving maximum employment conduct their monetary policy in a forward-looking manner. Consequently, it is essential to incorporate agents' and financial markets' expectations in the policy stance. For example, the importance of the NY Fed Survey of Primary Dealers (SPD) is also apparent in the expected path of monetary policy. More specifically, according to Sinha et al. (2023), while forecasters generally disagree about the expected path of monetary policy, the level of disagreement, as measured in the SPD, has increased substantially since 2022, which provides valuable information to policymakers.

Finally, regarding the importance of the surveys' macroeconomic expectations, they can be seen either as standalone individual forecasts or can be incorporated into macroeconometric models with a certain degree of conviction to enhance their predictability. Expectations capture recent underlying trends, undergoing structural changes, and fast-evolving economic narratives, which may take time to be reflected, even in correctly specified models. For instance, the New York Fed DSGE model forecasts use the federal funds rate SPD expectations (Cho et al., 2024). Due to their importance, they must be evaluated accurately by simultaneously capturing time-critical aspects that result in a more robust statistical inference.

According to Carroll (2003), there is a strong consensus among economists that macroeconomic outcomes and the relevant decision-making process depend on agents' expectations, which urges researchers to investigate whether the provided expectations are rational. A novel study on forecast evaluation is that of Clements et al. (2007). In their paper, they examined whether the FED forecasts exhibit systematic bias and whether information is used efficiently, i.e., whether revisions to these forecasts are predictable. In a later study, Coibion and Gorodnichenko (2015) proposed a novel approach to test the null hypothesis of full-information rational expectations theory. Their empirical specification can be used to quantify the impact of policy changes on the expectations formation process by providing a more theoretical concept than ambiguous concepts (e.g., anchored expectations). Additional studies that contribute to the literature on expectations formation are those of Coibion et al. (2018, 2020), Roth and Wohlfart (2020), and Aastveit et al. (2023).

The aforementioned studies often categorise macroeconomic expectations surveys based on their forecast horizon structure, i.e. fixed-event or fixed-horizon. The first category includes surveys based on fixed-event forecasts, which are used by a wide range of published papers (e.g. Bakhshi et al., 2005). The second category is that of fixed-horizon forecasts (e.g. Clements et al. 2007). Their differences rely on the fact that fixed-event forecasts are made at different points in time with the same target, while fixed-horizon forecasts target a variable at a specific horizon ahead. Moreover, it is highlighted that some of the surveys are not conducted regularly, resulting in irregularly spaced time series. Regarding fixed-event surveys, the literature remains silent on some important features, which are explicitly discussed later. This motivated us to highlight the need for a generic econometric approach.

One of the first studies to evaluate economic forecasts, and more specifically the Consensus Forecasts, is that of Loungani (2001), who compares the forecasts' unbiasedness and efficiency for industrialised and developing countries using a sample from 1989 to 1998. Another study that assesses the aforementioned forecasts for the period of 1990-2005 is that of Batchelor (2007), who finds that GDP forecasts are overoptimistic in the case of Japan, France, Germany, and Italy, while there is no evidence of bias for inflation. However, the latter studies are very preliminary and do not test for unbiasedness at a more granular level.

More recently, Ager et al. (2009) tested for unbiasedness using a pooled approach, while they also enhanced the methodology for testing weak efficiency by capturing horizon-specific behaviours of the economic forecasts. They find that several countries' forecasts are biased and more particularly for forecast horizons greater than 12 months, while the forecasts' weak efficiency property is rejected for almost all cases. Nevertheless, in their paper, they do not test for horizon- or event-specific bias at different frequencies or under the existence of heteroscedasticity (for the case of event-specific bias). In addition, they do not account for the impact of outliers, which may lead to inaccurate inference for bias and efficiency, especially when periods of high uncertainty are examined. Regarding efficiency, according to Nordhaus (1987), if forecasts are efficient in terms of using all available and relevant information about the target variable's future path, then the forecast revision process should behave like a random walk. A large variety of studies use this method, but instead of examining weak efficiency over each target event separately, pool fixed-event forecasts over different target years to conduct the test (e.g., Clements, 1997; Harvey et al., 2001). At a later point in time, Isiklar et al. (2006) evaluate GDP growth forecasts, provided by Consensus Economics, and find that efficiency is rejected for the entire sample of countries.

Another important feature that has been observed when evaluating inflation forecasts refers to state-dependence in bias. More specifically, Granziera et al. (2025) show that the ECB tends to underpredict when the observed inflation rate at the time of forecasting is higher than the estimated threshold. This implies that policymakers' forecasts are subject to systematic forecast errors, depending on the level of inflation observed when producing their forecasts. From a technical point of view, they applied the newly developed state-dependent bias test proposed by Odendahl et al. (2023), which estimates a regression model that allows forecast errors to depend not only on a linear term but also on a non-linear one.

Notably, even if these studies developed enhanced methods to address specific types of bias, a critical gap remains in the literature, which is the evaluation of fixed-event surveys' expectations during highly uncertain periods, such as the COVID-19 pandemic[1]. Such periods are considered crucial to be covered, since the surveys' expectations in extremely uncertain periods[2] play an

[1] According to Goodhart and Pradhan (2023) and Levy (2023), central banks made large forecasting errors and revisions during the COVID-19 pandemic and the subsequent reopening, which shows the magnitude of uncertainty in the economic conditions of those periods.

[2] Rossi and Sekhposyan (2016) proposes forecast rationality tests that can be used in unstable environments for the case of fixed-horizon forecasts. In theory, we could derive an optimal approximation for fixed-horizon forecasts based

important role in policy decisions. A question that could better reflect this issue is the following: "*Do the developed methodologies correctly infer the unbiasedness and efficiency under the existence of outliers?*" In our study, we deal with this challenge by developing an outlier-robust approach that is integrated in the tests for both unbiasedness and efficiency. The extended methodologies apply to fixed-event surveys, independent of the variable's nature.

Another issue that could arise from the existing methodologies is the examination of separate bias for every forecast horizon. Specifically, the use of such a large number of dummies, representing forecast horizons, in the relevant regressions could not only lead to less robust estimates but also to complicated interpretation of the results due to the fact that forecasts revisions might not occur so frequently but at a lower frequency than the one that the survey is conducted (i.e., the forecasts of a monthly survey may be significantly revised on a quarterly basis). To account for such irregularities that may result in misleading findings, we enhance the methodologies by varying the frequency at which unbiasedness is examined. In this way, the assessment is not only outlier-robust but also tests for unbiasedness and weak efficiency in alternative frequencies.

The proposed generalised econometric framework is applied to an important, but insofar not systematically evaluated[3], survey for the US economy, the SPD. As already highlighted, SPD is a critical input into the FOMC decision-making process and is often cited in the FOMC meeting minutes since it is considered a valuable source of information for understanding market participants' expectations. The SPD is a fixed-event, structured survey of macroeconomic expectations that is not conducted on a regular basis, but only prior to each FOMC meeting (eight times per year). This characteristic raises issues in assessing the weak efficiency property, which is designed under the assumption of a regularly spaced time series of the forecasts. Moreover, the SPD provides expectations for a wide range of macroeconomic variables, especially for those on which monetary policy decisions are based on (i.e., GDP growth, inflation, and policy expectations).

on the examined fixed-event forecasts by using the approach of Knüppel and Vladu (2025) and implement the former study afterwards. However, first, this exercise is based on several assumptions or approximations and, second, it is out of the scope of this study.

[3] Selected references of SPD can be found in Correia-Golay et al. (2013), in Diercks et al. (2022), concluding that the asymmetry in policy rate expectations can result in significant implications for the measurement of term premia and in Diercks and Jendoubi (2023).

In summary, although several methodologies have been developed for testing the unbiasedness and efficiency properties of forecasts, this is the first study to propose a generalised framework, which also accounts for a variety of critical aspects required in the assessment of fixed-event macroeconomic survey expectations. The application of the proposed econometric framework is implemented in GDP, inflation, and policy expectations of the SPD, as well as during the COVID-19 period.

The application of the methods to the SPD dataset indicated that a more sophisticated and robust evaluation methodological framework is essential for accurately assessing unbiasedness and efficiency, particularly when high uncertainty periods are included in the sample under examination. In particular, when evaluating the full sample, including the COVID-19 pandemic, the proposed methodology significantly enhances the robustness of the drawn inference. Our findings show evidence of bias and asymmetric behaviour in bias for specific horizons and events, but the results differ across variables. The findings for both strong and weak efficiency indicate room for improvement.

The remainder of the paper is organised as follows. Section 2 provides detailed information on the motivation and the proposed econometric framework for testing unbiasedness and efficiency. Section 3 provides a detailed description of the SPD, applies the proposed methodologies in the SPD dataset, and analyses the results. Finally, Section 4 concludes and summarises the key aspects that matter and should be covered when assessing macroeconomic expectations.

2. From motivation to a generalised econometric framework

In this section, we develop an improved integrated evaluation framework for assessing fixed-event forecasts in an outlier-robust way. The proposed techniques focus on enhancing statistical inference for evaluating optimal forecast properties, and are built primarily on previous methodologies developed by Davies and Lahiri (1995), Clements et al. (2007), and Ager et al. (2009).

We propose a generalised methodology to evaluate the rationality of fixed-event surveys' expectations. Forecasts are considered optimal when they minimise a given loss function. Our framework is based on the standard concept of a quadratic loss function (Capistran and

Timmermann, 2009; Elliott and Timmermann, 2016). For a more generic treatment of the optimal forecast properties under unknown loss, more details are provided in the studies of Elliott et al. (2005), Patton and Timmermann (2007a, 2007b).

The first proposed methodological aspect concerns the existence of outliers, which may lead to false decisions regarding unbiasedness, in a sample of both high and low uncertainty subperiods. Applying the existing methodologies to this sample's forecasts, it is not adequate to extract robust results regarding unbiasedness and efficiency. This is due to the large variation of the standard errors of the corresponding estimated coefficients, which may result in misleading outcomes. In other words, the problem, when testing for biasedness, is attributed to the large errors that are created by large shocks in the actual series and not necessarily from potential forecasts' bias. For example, if an economist or an institution generates, during mid-2019, a forecast for GDP growth for the next year's last quarter (2020 Q4), the forecast error is going to be large, not because the respondent is necessarily biased, but due to the large shocks that the GDP series is subject to during the COVID-19 period. Moreover, the existence of a large number of outliers may cause issues concerning the assumption of normally distributed error terms, as described in detail in Section 2.1. This is a limitation in the current literature that we cover with our study by proposing a stricter statistical inference, which isolates extreme events and narrows the coefficients' confidence intervals. The same concept is also developed for testing weak efficiency, with the only difference being that outliers refer to the forecast revision series, which present large changes during volatile periods.

To test horizon-specific bias at alternative frequencies by isolating the impact of outliers, we provide a generalised framework that gives researchers the ability to implement it on any dataset of fixed-event forecasts with different horizons and variables. This enhancement has been motivated by the fact that a test for specific horizon bias at the frequency where the examined survey is conducted may incorrectly fail to reject the null hypothesis of unbiasedness. Moreover, as mentioned in the first section, in cases where we evaluate forecasts that are sampled at high frequency, there is a high number of dummies that are included as explanatory variables, which could result in less robust estimates[4]. Finally, additional aspects that make the framework more integrated are the state-dependence in bias, which was initially examined for inflation, and the

[4] This issue is visible when evaluating daily forecasts (e.g., interest rate futures).

efficiency tests, which ensure that the available information is efficiently incorporated. In both tests, the proposed extensions of the unbiasedness testing have been appropriately incorporated. Further technical details are provided in the following subsections.

2.1. Outlier-robust unbiasedness test

One of the most widely used tests for unbiasedness is that of Mincer and Zarnowitz (1969), a regression-based test, where the outcome is regressed on a constant, and the forecast and the unbiasedness property are tested under the joint null hypothesis of the constant and the coefficient of the forecast being zero and one, respectively. However, according to Holden and Peel (1990), the Mincer and Zarnowitz test is a sufficient but not necessary condition for unbiasedness. To follow the simplest method, we estimate separate regressions for each target event, which, however, would not be sensible with very few observations available for each horizon. The most recent and comprehensive study about the assessment of fixed-event macroeconomic forecasts is the study of Ager et al. (2009), who conduct analyses to evaluate the performance of the Consensus Forecasts regarding bias and weak efficiency. Therefore, this study's tests are considered a benchmark for our methodological framework, which is enhanced by adapting important aspects that are missing from the existing literature.

Before starting with the theoretical background, it is pointed out that in our study, we follow the notation of Ager et al. (2009) closely in order to facilitate comparison and make the extensions easily recognised. We then start by defining the forecast error for a forecast horizon h of a target variable in year t as:

$$e_{th} = A_t - F_{th} \tag{1}$$

where A_t is the actual value for variable in period t, with $t = 1, \cdots, T$ denoting the reference date and F_{th} is the forecast for the target event at period t that is provided at time $t - h$, with $h_m = 1, \ldots, H$ in months. Under the pooled estimation, the following model is estimated:

$$\boldsymbol{e} = \boldsymbol{X\alpha} + \boldsymbol{v} \tag{2}$$

where $\boldsymbol{e} = \boldsymbol{A} - \boldsymbol{F}$ is the stacked vector of the forecast errors with dimension $(TH \times 1)$. The vector of the forecasts is further defined as $\boldsymbol{F} = [F_{1H}, \cdots, F_{11}, F_{TH}, \cdots, F_{T1}]'$. Correspondingly, the vector of the outturns is defined as $\boldsymbol{A} = \widetilde{\boldsymbol{A}} \otimes \mathbf{i}_H$, where $\widetilde{\boldsymbol{A}} = [A_1, \cdots, A_T]'$ and $\mathbf{i}_H$ is a vector of ones with dimension $(H \times 1)$. Finally, $\boldsymbol{v}$ is normally distributed with zero mean.

As previously mentioned, we not only test uniform bias for every forecast horizon, but we also expand the methodological framework by testing separate bias for every forecast horizon. We generalise further this framework and show that under a suitable definition of matrix $\boldsymbol{X}$ we can estimate separate bias with $\boldsymbol{\alpha}$, for forecast horizons in a lower frequency than the examined survey's sampling frequency (for instance, bias for quarterly or annual forecast horizons at a monthly survey), as well as separate bias for each target event.

We define N, as the number of forecast horizons for which we want to estimate uniform bias. For instance, for a monthly survey, with $N = 1$ we estimate separate bias for each monthly horizon, and with $N = 3$ or $N = 12$ we estimate uniform bias for each separate quarterly or annual horizon, respectively. It is also evident that for the case of $N = H$, we estimate a uniform bias. First, we construct matrix $\boldsymbol{W}$ as a matrix of dimension $(H \times H/N)$, with ones for the corresponding monthly, quarterly, annual, or uniform bias at each horizon and target event, as:

$$\boldsymbol{W} = \mathbf{I}_{H/N} \otimes \mathbf{i}_N \tag{3}$$

where $\mathbf{I}_{H/N}$ is the identity matrix with dimension $(H/N \times H/N)$.

Finally, by applying the Kronecker product to matrix $\boldsymbol{W}$ for all target events $t = 1, \cdots, T$ we obtain the matrix $\boldsymbol{X}$ in eq. (2) with all dummy regressors[5]:

$$\boldsymbol{X} = \mathbf{i}_T \otimes \boldsymbol{W} \tag{4}$$

Depending on the term N, the dimension of the estimated regression coefficients vector $\boldsymbol{\alpha}$ in eq. (2) is of dimension $(H/N \times 1)$, where obviously for the case of $H = N$, i.e., uniform bias, $\boldsymbol{\alpha}$ is a scalar.

Turning to the case, where we estimate event-specific bias, matrix $\boldsymbol{X}$ in eq. (2) is defined directly as:

$$\boldsymbol{X} = \mathbf{I}_T \otimes \mathbf{i}_H \tag{5}$$

As regards the consistent variance of the estimator $\boldsymbol{\alpha}$ in eq. (2), it is defined as:

$$\boldsymbol{var}(\widehat{\boldsymbol{\alpha}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{6}$$

[5] Note that for $N = 3$ or $N = 12$, forecast horizon H should be a multiple of 3.

For the case of horizon-specific bias, under the assumption of homoscedasticity driven by a common macro shock with variance σ^2 and autocorrelated errors with the structure defined by matrix $\boldsymbol{\Psi}$ of dimension $(TH \times TH)$[6], the covariance matrix $\boldsymbol{\Sigma}$ is defined as $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Psi}$. Turning to the empirical counterpart $\widehat{\boldsymbol{\Sigma}}$, the estimation of $\hat{\sigma}^2$ is required. Having obtained $\widehat{\boldsymbol{v}}$ from eq. (2), we estimate $\hat{\sigma}^2$ as φ_0 in the following regression[7]:

$$\widehat{\boldsymbol{v}} \odot \widehat{\boldsymbol{v}} = \boldsymbol{\tau}\varphi_0 + \boldsymbol{\delta}_{outl}\varphi + \omega \tag{7}$$

where $\boldsymbol{\tau} = \mathbf{i}_T \otimes \boldsymbol{\tau_H}$ is a $(TH \times 1)$ vector with $\boldsymbol{\tau_H} = [H, H-1, \cdots, 1]'$, $\boldsymbol{\delta}_{outl}$ is of dimension $(TH \times 1)$ and φ is a scalar. In this case, there is only one common macro-shock across horizons and events. This assumption is extended later on, where different macro shocks appear to exist per event.

As analytically described in the previous sections, one of the key aspects covered by this study is to effectively capture the effect of extreme values' impact on inference. In order to account for these outliers (e.g., the GDP growth deterioration during the COVID-19 pandemic period) and how they affect the common macro shock with variance σ^2 for the entire period, we include dummy $\boldsymbol{\delta}_{outl}$ in eq. (7). Moreover, from a statistical point of view, it is important to detect such outliers in order to better approach the normal distribution of $\boldsymbol{v}$. To be specific, all v_{th} are pooled together and $\delta_{outl,th}$, $\forall t = 1 \cdots T$, and $\forall h = 1 \cdots H$ takes the value of one, if a forecast error v_{th}, according to the Tukey interquartile range (IQR) rule, i.e., observations lying below Q_1−1.5 IQR or above Q_3+1.5 IQR, where it is defined as $IQR = Q_3 - Q_1$.

$$\delta_{outl,th} = \begin{cases} 1, if\ v_{th} < Q_1 - 1.5IQR \\ 1, if\ v_{th} > Q_3 + 1.5IQR \\ 0, otherwise \end{cases} \tag{8}$$

An extension of the pooling approach which we develop addresses a form of heteroscedasticity in macroeconomic shocks. More specifically, when estimating event-specific bias, we allow for heteroscedasticity induced by each individual event. Therefore, eq. (7) becomes:

$$\widehat{\boldsymbol{v}} \odot \widehat{\boldsymbol{v}} = \boldsymbol{\tau}\boldsymbol{\varphi_0} + \boldsymbol{\delta}_{outl}\boldsymbol{\varphi} + \boldsymbol{\omega} \tag{9}$$

[6] For further details, see Section 3.1 of Ager et al (2009).
[7] Operator $\odot$ refers to the element-by-element multiplication (Hadamard product).

where $\boldsymbol{\tau} = \mathbf{I}_T \otimes \boldsymbol{\tau_H}$ is $(TH \times T)$ matrix with $\boldsymbol{\tau_H} = [H, H-1, \cdots, 1]'$, $\boldsymbol{\varphi_0}$ is $(T \times 1)$ with the event-specific innovations variance, $\boldsymbol{\delta}_{outl}$ is $(TH \times T)$ and $\boldsymbol{\varphi}$ is a $(T \times 1)$ vector of coefficients. Regarding the examination of event-specific bias, the identification of outliers is adjusted accordingly.

More specifically, $\boldsymbol{\delta}_{outl}$ is computed as:

$$\boldsymbol{\delta}_{outl} = (\mathbf{I}_T \otimes \mathbf{i}_H) \odot (\mathbf{i}_T \otimes \widetilde{\boldsymbol{\delta}}_{outl}) \tag{10}$$

where $\widetilde{\boldsymbol{\delta}}_{outl}$ is a $(H \times T)$ matrix that takes the value of one, for each event, $t = 1, \cdots, T$ separately, if a forecast error v_{th}, at the corresponding event *t,* exceeds 1.5 times the IQR, where $IQR = Q_3 - Q_1$.

$$\tilde{\delta}_{outl,th} = \begin{cases} 1, if\ v_{th} < Q_1 - 1.5IQR \\ 1, if\ v_{th} > Q_3 + 1.5IQR \\ 0, otherwise \end{cases} \tag{11}$$

where each event t is examined separately.

For robustness, we also estimate variance $\boldsymbol{\varphi_0}$ in eq. (7) and eq. (9) – without the inclusion of the dummy outliers – using an M-estimator[8] which is designed to mitigate the influence of outliers while retaining high efficiency under Gaussian errors. In particular, using the Huber loss (Huber, 1964), the estimator replaces the quadratic loss of ordinary least squares with a piecewise loss function that is quadratic for small residuals and linear for large residuals, such that extreme observations are down-weighted rather than being entirely discarded. More formally, the Huber estimator (Huber, 1973) is defined as:

$$\widehat{\boldsymbol{\varphi}}_{\mathbf{0}} = \arg \min_{\varphi} \sum_{t=1}^{TH} \rho_c\,(y_t - \boldsymbol{x}_t\, \boldsymbol{\varphi}) \tag{12}$$

where $y_t = \hat{v}_{th}^2$, $\boldsymbol{x}_t = \boldsymbol{\tau}_t$, and the Huber loss function is defined as:

[8] An M-estimator is a generalisation of maximum likelihood estimation, which was initially proposed to improve robustness to outliers or model misspecification. See He et al. (2021), and Xing and Zhang (2022) for applications of Huber loss in time series forecasting.

$$\rho_c(u) = \begin{cases} \frac{1}{2}u^2, & |u| \leq c \\ c|u| - \frac{1}{2}c^2, & |u| > c \end{cases} \tag{13}$$

where $c > 0$ is the tuning constant. Following the robustness-efficiency trade-off for M-estimators under the Gaussian model, $F_0 = \Phi$, discussed in Huber (1981, Chapter 4), the Huber tuning constant is selected such that the asymptotic relative efficiency equals 95%, ARE = 0.95, implying $c \approx 1.345$.

The empirical counterpart $\widehat{\boldsymbol{\Sigma}}$ is now defined as:

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{S}\,\boldsymbol{\Psi}\,\boldsymbol{S}' \tag{14}$$

where $\boldsymbol{S} = diag\left(\boldsymbol{\varphi_0}^{1/2}\right) \otimes \boldsymbol{I}_H$. Concerning the matrix $\boldsymbol{\Psi}$, it incorporates the special structures of the overlapping forecasts[9]. For example, having set the maximum horizon $H = 36$ (it corresponds to a 3-year horizon), this implies that the maximum autocorrelation is MA(35), since the January and February survey forecasts for a target event three calendar years[10] ahead share 35 shocks as a common component. In greater detail, matrix $\boldsymbol{\Psi}$ is of dimension $(TH \times TH)$ and the component matrices $\mathbf{A}_{\boldsymbol{i}}$, where *i* refers to each year, and $\mathbf{0}$ are each of dimension $(H \times H)$. More specifically, $\boldsymbol{\Psi}$ is defined as:

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{A_1} & \mathbf{A_2} & \mathbf{A_3} & \cdots & \mathbf{A}_{\boldsymbol{M}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A_2}' & \mathbf{A_1} & \mathbf{A_2} & \cdots & \mathbf{A}_{\boldsymbol{M-1}} & \mathbf{A}_{\boldsymbol{M}} & \cdots & \mathbf{0} \\ \mathbf{A_3}' & \mathbf{A_2}' & \mathbf{A_1} & \cdots & \mathbf{A}_{\boldsymbol{M-2}} & \mathbf{A}_{\boldsymbol{M-1}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{\boldsymbol{M}}' & \mathbf{A}_{\boldsymbol{M-1}}' & \mathbf{A}_{\boldsymbol{M-2}}' & \cdots & \mathbf{A_1} & \mathbf{A_2} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{\boldsymbol{M}}' & \mathbf{A}_{\boldsymbol{M-1}}' & \cdots & \mathbf{A_2}' & \mathbf{A_1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \tag{15}$$

In order to proceed with the specification of the $\mathbf{A}_{\boldsymbol{i}}$ matrices, we first define $K_i = H - (i - 1)m$, where m[11] refers to horizons per year. Moreover, it is noted that each of the below-described matrices is of dimension $(H \times H)$. The first component matrix, namely $\mathbf{A_1}$, is defined as:

[9] This structure, which affects the estimation of error variance, has been extended by Vereda et al. (2024) by proposing a framework in which two types of informational shocks arrive each month.

[10] More specifically, $i = 1, \dots, M$, where *M* denotes the maximum number of forecasted years ahead.

[11] For example, for quarterly data, *m=4* while for monthly, *m=12*.

$$\mathbf{A_1} = \begin{bmatrix} K_1 & K_1-1 & K_1-2 & K_1-3 & \cdots & 3 & 2 & 1 \\ K_1-1 & K_1-1 & K_1-2 & K_1-3 & \cdots & 3 & 2 & 1 \\ K_1-2 & K_1-2 & K_1-2 & K_1-3 & \cdots & 3 & 2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 3 & 3 & 3 & 3 & \cdots & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & \cdots & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix} \tag{16}$$

To continue with the remaining component matrices of $\mathbf{\Psi}$, we use a generalised form that replaces the specific determination of each of the $\mathbf{A_1}, \cdots, \mathbf{A_M}$ matrices and is written as:

$$\mathbf{A_i} = \begin{bmatrix} K_i & K_i-1 & K_i-2 & \cdots & 2 & 1 & 0 & \cdots & 0 \\ K_i & K_i-1 & K_i-2 & \cdots & 2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ K_i & K_i-1 & K_i-2 & \cdots & 2 & 1 & 0 & \cdots & 0 \\ K_i-1 & K_i-1 & K_i-2 & \cdots & 2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 & 0 & \cdots & 0 \end{bmatrix} \tag{17}$$

2.2. Testing for state-dependence in bias

As previously highlighted, this study also captures potential asymmetries in the assessment of forecasts. In this regard, a methodological extension is to examine whether the state of the economy induces bias in the examined forecasts. To do so, we augment eq. (2) using a dummy $\boldsymbol{D}$:

$$\boldsymbol{e} = \boldsymbol{X\alpha} + \gamma \boldsymbol{D} + \boldsymbol{v} \tag{18}$$

where $\boldsymbol{D}$ is a stacked vector of dummies of dimension $(TH \times 1)$ and is written as $\boldsymbol{D} = [D_{1H}, \cdots, D_{11}, \cdots, D_{TH}, \cdots, D_{T1}]'$. We use index th for D_{th} to enable ease of reference to the corresponding forecast error e_{th}. D_{th} takes the values of one if the last available observation for the forecasted variable at period $t-h-1$ is above the longer-run forecast estimate of the variable produced at period $t-h$, or alternatively is above the historical average of the variable defined over the last *K* periods. For quarterly variables, *K* refers to the last *K* quarters, and for monthly variables like the policy expectations, *K* refers to the last K months.

$$D_{th} = \begin{cases} 1, A_{t-h-1} \geq LR_{t-h} \\ 0, otherwise \end{cases} \tag{19}$$

$$D_{th} = \begin{cases} 1, A_{t-h-1} \geq {}^{1}/_{K} \sum_{i=t-h-1-K}^{t-h-1} A_i \\ 0, otherwise \end{cases} \tag{20}$$

The inference is performed under the methodological framework described in the previous section for unbiasedness by redefining $\widetilde{\boldsymbol{X}} = [\boldsymbol{X}\ \boldsymbol{D}]$. By adjusting the above-mentioned method to the previous section, we provide the research community and policymakers with a generalised framework for assessing the macroeconomic expectations of fixed-event surveys.

To summarise, this approach enables the entire framework to be flexible regarding: (1) the inclusion of uncertain periods in the sample, (2) the examination of potential horizon-specific and event-specific bias at different frequencies, and (3) the examination of state-dependence in bias.

2.3. Testing for efficiency

In this section, we test for efficiency by extending the existing methodologies in multiple ways. First, we use outlier-robust methods following the aforementioned proposed econometric framework. Second, we generalise the tests to assess fixed-event surveys' expectations more flexibly, in terms of sampling frequency, examined horizons, and the existence of outliers. Finally, we not only test for the predictability of forecast revisions, but also for the efficient incorporation of the exogenous information set, which is available at the time of forecasting, in the expectations' formulation.

2.3.1. Strong efficiency test

Let us first denote that while weak efficiency refers to an information set that includes only past forecasts, strong efficiency tests whether the information set, which was available at the time that the forecast was made, is incorporated efficiently. The core structure of the proposed methodological framework is based on Timmermann (2007, 2016). In this regard, the strong efficiency property is examined by testing the orthogonality of errors e_{th} to all variables Z_{t-h} that belong to the information set I_{t-h} when the forecasts were produced at *t-h*. Consequently, eq. (2) is augmented as:

$$\boldsymbol{e} = \boldsymbol{Z\gamma} + \boldsymbol{X\alpha} + \boldsymbol{v} \tag{21}$$

where Z_{t-h} includes the information set of the explanatory variables, which should be taken into consideration by forecasts. Regarding our application, the set of explanatory variables is limited to those under examination, namely the real GDP, inflation, and policy expectations. We provide further details in Section 3.2. We test strong efficiency using an outlier-robust method, based on the unbiasedness test framework, that examines strong efficiency at different frequencies.

2.3.2. Weak efficiency test

The concept of weak efficiency is widely analysed in the literature (Loungani, 2001; Ager et al., 2009; Capistran and López-Moctezuma, 2014). In detail, this test examines whether the current period's forecast revision is uncorrelated with the last period's revision. A forecast revision between $h+1$ and h for target event t is defined as: $r_{t;h,h+1} = F_{t;h} - F_{t;h+1}$. By pooling together all forecast revisions $r_{t;h,h+1}$ for all target events $t = 1, \cdots, T$ and forecast horizons, we form the following pooled regression:

$$\boldsymbol{r} = \boldsymbol{r}_{-\mathbf{1}}\gamma + \boldsymbol{X\alpha} + \boldsymbol{\omega} \tag{22}$$

where $\boldsymbol{r}$ is a vector of dimension $((H-2)T \times 1)$ due to the two lost observations. In detail, the first lost observation is due to taking the first differences between subsequent forecasts, and the second is due to using one lag of the forecast revisions as explanatory variables in the regressions. In order to develop a generalised framework, which tests horizon-specific bias at a lower frequency than the sampling one, we increase the horizon H, which is a multiple of 3, by the two lost observations, such that $\widetilde{H} = H + 2$.

In more detail, $\boldsymbol{r} = \left[\boldsymbol{r}_{\widetilde{H}-2,\widetilde{H}-1}, \cdots, \boldsymbol{r}_{1,2}\right]'$, where each subvector $\boldsymbol{r}_{\widetilde{H}-2,\widetilde{H}-1}$ to $\boldsymbol{r}_{1,2}$ of dimension $\left((\widetilde{H}-2) \times 1\right)$ stacks all revisions for individual forecast horizons for all target events, i.e. $\boldsymbol{r}_{\widetilde{H}-2,\widetilde{H}-1} = \left[\boldsymbol{r}_{t;\widetilde{H}-2,\widetilde{H}-1}, \cdots, \boldsymbol{r}_{T;\widetilde{H}-2,\widetilde{H}-1}\right], \cdots, \boldsymbol{r}_{1,2} = \left[\boldsymbol{r}_{t;1,2}, \cdots, \boldsymbol{r}_{T;1,2}\right]$.

To account for horizon-specific bias, conceptually, we follow the proposed approach used for testing unbiasedness. In greater detail, technically, matrix $\boldsymbol{X}$, which contains dummies to control for horizon-specific bias at different frequencies, is defined by using first matrix $\boldsymbol{W}$:

$$\boldsymbol{W} = \boldsymbol{I}_{\widetilde{H}-2/N} \otimes \mathbf{i}_N \tag{23}$$

Therefore, matrix $\boldsymbol{X}$ is defined using the Kronecker product as follows:

$$\boldsymbol{X} = \boldsymbol{W} \otimes \mathbf{i}_T \tag{24}$$

To test event-specific bias, equivalently, matrix $\boldsymbol{X}$ contains dummies for specific target events and is defined as:

$$\boldsymbol{X} = \mathbf{i}_{\widetilde{H}-2} \otimes \boldsymbol{I}_T \tag{25}$$

Turning to the error term $\boldsymbol{\omega}$ in eq. (7), we assume $E[\boldsymbol{\omega}] = 0$ and $E[\boldsymbol{\omega}\boldsymbol{\omega}'] = \boldsymbol{\Omega}$ for the second moment. Regarding $\boldsymbol{\Omega}$, it is of dimension $\left((\widetilde{H}-2)T \times (\widetilde{H}-2)T\right)$.

Also in this case, both approaches for error variance are examined, namely the homoscedastic and the heteroscedastic case, where for each target event, there are different variances. It is highlighted that the generalisation of the methodology used for testing weak efficiency follows, conceptually, the previous section of the paper. All technical details in this regard are included in the Appendix.

3. Application: NY Fed Survey of Primary Dealers

3.1. Description of the SPD dataset

Primary dealers perform three key functions within the US financial system. First, they serve as counterparties to the Federal Reserve in open-market operations, thereby facilitating the implementation and transmission of monetary policy. Second, they underwrite and distribute US Treasury securities, playing a key role in primary market auctions and ensuring liquidity and efficient price discovery in secondary markets. Third, through their continuous engagement with market participants and their close relationship with the Federal Reserve, primary dealers provide valuable market intelligence that enhances policy formulation and contributes to overall financial stability. The current list of SPD comprises 23 financial firms, while over time, there have been a limited number of additions and removals.

One structural characteristic of the SPD is that it is conducted 8 times per year, approximately 2 weeks before each FOMC meeting, and the survey period lasts 5 days. SPD questionnaires are not entirely fixed but change according to the prevalent features regarding the economy, including questions about the policy stance implementation, Fed balance sheet, recession probabilities, and balance of risks. For example, when risk is tilted towards inflation, there are more detailed inflation-related questions, whereas when risk is focused more on growth, quarter-on-quarter

(QoQ) point estimates are asked in addition to year-end. Nonetheless, the questionnaires' key components always contain a set of fixed questions asking the respondents for point estimates for key macroeconomic variables (real GDP, core PCE, PCE inflation, and unemployment) and policy expectations (FFR). SPD questions for these variables are designed similarly to the Summary of Economic Projections structure, i.e., the forecasts always refer to year-end forecasts, and consequently SPD is a fixed-event expectations survey. SPD also has probabilistic questions about FFR, asking respondents to describe their views on the likelihood of FFR falling in specific ranges, but these are not examined in this paper.

The key macroeconomic variables, which are examined in our paper, namely real GDP, core PCE and PCE inflation year-end annual (Q4/Q4) forecast values, are targeted by consecutive surveys spanning from 36 to 48 months ahead (three to four calendar years-end ahead). Regarding surveys that are conducted for the five regular FOMC meetings of January, March, May, June, and July, the forecast horizon is three years ahead, whereas for the September, October, and December meetings' surveys, the forecast horizon is 4 years ahead. Additionally, for these variables, as well as for FFR, SPD questionnaires ask for point estimates of the longer-run forecasts, which are used in the asymmetric behaviour in bias analysis. Regarding FFR, questionnaires ask for point estimates at different frequencies with respect to the forecast horizon. For the short-term forecast horizon after each survey, questionnaires initially ask FFR target rate/mid-point target range[12] point estimates after each one of the seven consecutive FOMC meetings. Subsequently, the surveyed frequency becomes quarterly, asking for five to eight quarters' end FFR values, and then the frequency becomes half-annual or annual, asking for approximately two more years of FFR values.

The aggregated results (median, 25th, and 75th percentiles) are published three weeks after each FOMC meeting, while individual responses are not made publicly available[13]. SPD aggregated results are available since January 2011; however, after 2024, SPD has been merged with a similar, in terms of the respondents' characteristics, survey, which is the Survey of Market Participants (SMP). Thereafter, the combined surveys constitute the Survey of Market

[12] For the surveys conducted between September 2014 and January 2016, FFR was estimated as the weighted average of all responses: explicit mid-point target rates were taken as reported, while for range responses, i.e. top of range and bottom of range, the midpoint was assigned, with weights proportional to the number of respondents in each category.

[13] Several studies (e.g., Engelberg et al., 2011) document sizable disagreements between individual forecasts and aggregate responses. Consequently, other research (e.g., Keane and Runkle, 1990) argues that individual-level data allow more granular inference on how different types of respondents form expectations, which is a useful avenue for future research.

Expectations (SME), which is considered a more market-focused survey. However, we choose to apply the proposed methodologies only to the SPD dataset due to the larger sample starting from 2011 compared to the SMP, which was initially conducted in 2014.

3.2. Implementation settings

Before implementing the proposed methodologies, we first fill in the missing monthly SPD values by following Vereda et al. (2021). For the eight months where the SPD median fixed-event forecasts exist, their values are assigned to the end of the month, and for the remaining four months, missing values are created by linear interpolation. In greater detail, for months in which fixed-event forecasts are constructed by linear interpolation between adjacent surveys, we align the information set with that of the earlier survey. For example, the interpolated February value - constructed from the January and March surveys - is evaluated using the January real-time vintage (i.e., the information available as of the January survey date). Following this approach, 1/3 of the sample size under analysis is created artificially.

We perform our analysis on both the full sample for the target events of 2012-2024 and the subsample 2012-2019, to examine in detail how the COVID-19 and the subsequent inflation surge have affected the results. The maximum examined forecast horizon is 36 months, which remains consistent across all the SPD surveys conducted throughout the year. Target-event 2011 is not included in the analysis because it has a maximum available forecast horizon of 12 months only, and additionally, PCE forecast questions were missing for the January to July 2011 questionnaires.

Regarding the actual data used for estimating forecast errors, we get the values as they were available 2 quarters after the reference period[14], using real-time vintage data, similar to Tulip (2006) and Faust and Wright (2009). Regarding FFR, as actual, we use the federal funds lower and upper range mid-point, which is available since December 2008, whereas before this date, the federal funds effective rate is used. Finally, when testing the properties of state dependence relative to the historical average and the strong efficiency property, we respect the respondents' information set at each survey date. Specifically, for each survey, we use the most recent observations from the real-time data vintage that would have been available to participants on that date. When state-

[14] For robustness, we also perform the analysis using the first and the latest releases of the reference periods as actual data, and the findings are qualitatively similar. The results are available upon request.

dependence is examined, the PCE longer-run forecasts are used for the case of core PCE because questionnaires do not ask for core PCE inflation longer-run forecasts.

As regards the strong efficiency test, the selection of the exogenous variables is limited to the variables examined in this study (i.e., GDP, inflation, and policy expectations)[15] based on the importance of these variables for the monetary policy decision-making process. More specifically, a common VAR structure, which is used to analyse the monetary policy transmission mechanism in the US, includes the macroeconomics (non-policy) variables (output and prices) and variables controlled by the monetary policymakers (interest rates). There is a long literature that extensively explains how these variables are included in the model design in order to investigate the transmission of monetary policy shocks (e.g., Bagliano and Favero, 1998; Koop et al., 2009; Gertler and Karadi, 2015).

3.3. Empirical findings: Unbiasedness

Figure 1 shows the evolution of macroeconomic variables' expectations and the corresponding actual series, with the boxes reflecting target events' values. It is obvious from all variables' graphs that the period since the beginning of the COVID-19 pandemic is characterised by extremely high uncertainty, which results in cases of extremely large forecast errors. It is visually clear that SPD's aggregate responses overpredict inflation before 2021, whereas evidence of underprediction is provided for the remaining period. A similar behaviour is observed for the case of FFR, which is interconnected to inflation due to the monetary policy objective to stabilise inflation at the target level.

[FIGURE 1 HERE]

As regards the case of real GDP, the mean values, depicted in Figure 2, show a clear overprediction for the 2020 and 2022 year-end, while the SPD's responses appear to underpredict GDP for the remaining post-COVID-19 target events. From Figure 2, it is also illustrated that the most recent target events are subject to increased forecast errors as well as higher variation compared to the pre-COVID period. The motivation of one of the contributions of the paper, i.e., to assume heteroscedasticity when testing for event-specific bias, is clearly enhanced by this

[15] Expanding the information set to more variables, it requires to apply the one covariate at the time statistical method as in Iregui et al. (2025).

visualisation. Therefore, there is a need for a stricter methodological framework that will isolate the outliers, which is at the same time aligned with the normality assumed for the error term in eq. (2). The detection of outliers is also in effect to the examination of horizon-specific bias, with the difference that these are detected from the joint distribution of forecast errors after pooling all horizons. As shown in the right-hand side subplot, there is a large number of detected outliers across horizons during the full sample, compared to the pre-COVID period. This also holds true for the remaining variables, as visualised in Figures 3, 4 and 5. According to the left-hand side subplots, it is observed that core PCE and PCE expectations overpredict the target events' actual data points during the pre-COVID period, while a significant underprediction appears at the remaining period. The same behaviour is also noticed for the case of FFR.

[FIGURES 2-5 HERE]

Regarding the event-specific bias, we observe, from the left-hand side graph of Figure 6, that SPD expectations are biased for the majority of years, with the nature of biasedness being characterised by overprediction for some years, such as 2015, 2020, 2022, and by underprediction for others, such as 2021, 2023, and 2024. Moving forward to the statistical inference, the above-mentioned results are confirmed from the estimated coefficients in Table 3, which are statistically different from zero, with signs consistent with the findings of the preliminary visualisation analysis. To proceed with the horizon-specific unbiasedness test for real GDP, from the right-hand side subplot of Figure 6, it is shown that bias appears not only at longer horizons, which is in line with the findings of previous studies (Davies, 2006; Ager et al., 2009), but also at the very short ones. This outcome relies only on the estimated coefficient itself and not on the statistical inference. To the best of our knowledge, this is the first study presenting such an outcome, which contradicts the results of the previous studies that find biasedness only at longer horizons. The argument behind their results is that it is easier to generate forecasts a few horizons ahead, when a large part of the information set is already available, than to produce forecasts some years ahead. Regarding the short-term horizon bias, it can be explained by the fact that the respondents tend to underestimate GDP growth in order to cancel out any potential overestimation of previous releases. This is particularly relevant in the context of target events that exhibit significant changes, due to the uncertainty surrounding their final outlook. Finally, regarding the two samples analysed, the direction of bias appears consistent, indicating that the SPD respondents tend to overpredict GDP at longer forecast horizons, while they underpredict it in the months preceding the realisation of the target event.

[FIGURE 6 HERE]

From Table 1, it is obvious that by employing the proposed outlier-robust methodology in eq. (7) and (9) the standard deviation of the common macro shock in eq. (2) are reduced significantly for the full sample analysis (2012-2024), which includes the COVID-19 period, while in the pre-COVID-19 period (2012-2019), the impact of the outliers remained muted.

[TABLES 1-3 HERE]

From a statistical point of view, the estimation results, which are presented in Table 2, show that there is no common bias. However, when testing for horizon-specific bias, the results are statistically significant at short horizons. This specific bias remains significant for both samples examined. Moreover, under the examination of horizon-specific bias at different frequencies, we get a clearer picture of the forecast horizons (in quarterly or annual windows) before the realisation of the target event, for which the level of bias is significant. In greater detail, the exact quarters and years that are characterised by bias are revealed, which also serves as a robustness check. As regards the strong evidence of bias at short horizons, and as mentioned in the previous paragraph, it might be driven by the availability of more reliable data to respondents, which forces them to correct any potential forecast errors by strongly revising their forecasts. Finally, it is notable to mention that our proposed outlier-robust methodology is stricter, regarding the statistical inference, than the existing methodologies. Hence, by using our proposed methodology, the confidence intervals narrow towards the pre-COVID period's levels. Therefore, although the statistical outcome remains the same, the developed outlier-robust methodology makes the inference stricter in terms of capturing potential bias. For example, the confidence intervals are narrowed during the shorter horizons, which results in a clearer result for bias.

To continue with the inflation variable, represented by core PCE and PCE, Figures 7 and 8 indicate a clear overprediction during the pre-COVID period, whereas the SPD responses seem to underpredict inflation after 2020. This is consistent with the economic conditions that occurred during this period. More specifically, the inflation surge was unprecedented, and it was mainly driven by both increased supply and demand shocks. According to Bernanke and Blanchard (2025), the primary drivers of pandemic-era inflation are high energy prices and disruptions of global supply chains. On the other hand, among others, Giannone and Primiceri (2024) conclude that the pandemic-era inflation surge was driven mainly by demand shocks. These findings are confirmed by the estimation results that are presented in columns 2 and 3 of Table 3. Regarding the results of horizon-specific bias, we observe no specific differences across horizons. The SPD overpredicts inflation consistently during the pre-COVID period, while the opposite happens when the full sample is examined. This is due to the large impact of the post-COVID years. However, due to the large errors of this uncertain period, the statistical inference is not straightforward. The results of Table 2 show that even if there is indeed statistically significant common bias before the COVID-19 period, as well as significant bias across all forecast horizons, this does not remain when the full sample is examined. Upon examining the right-hand side subplots, it becomes evident that the outlier-robust methodology we have developed is much stricter in terms of statistical inference. It is important to note that the purpose of our study is not to necessarily reject unbiasedness, but rather to assess forecasts more robustly.

Concerning FFR, SPD expectations present a very similar behaviour to that of both inflation measures. The only difference is that during the pre-COVID years, the level of bias was not definite. This is confirmed by the respective results of Table 2, where the null hypothesis of common unbiasedness is not rejected. Nevertheless, there is evidence of bias at longer horizons, which holds only for the pre-COVID period.

[FIGURES 7-9 HERE]

The last aspect that is tested is the state-dependence in bias, which is a very recently documented pattern. In our study, we investigate whether this asymmetric behaviour of bias holds for the entire set of variables. Regarding the horizon-specific bias during the pre-COVID sample, evidence of state-dependence is provided (from Table 4) for the case of core PCE and only when long-run forecasts are used as a threshold. This implies that the SPD aggregate responses are presented as biased, with that level depending on the long-run forecasts for this variable. The results are consistent with the previous studies claiming that there is a systematic, state-dependent bias, both statistically and economically significant (Granziera et al. 2025). However, these results are not confirmed when implementing the methodology on the full sample, which includes the post-COVID period. For the event-specific test, using the full sample, the state-dependence in bias is present for all variables, and the results are qualitatively similar under both historical average and long-run forecasts used as thresholds. The positive coefficients indicate that SPD responses exhibit greater underprediction when the variable's value at the forecast date exceeds the threshold, whether defined by the historical average or the long-run forecast, consistent with the findings of Granziera et al. (2025) for inflation.

Moreover, when implementing the horizon-specific test, although there is evidence of bias, the coefficient that reflects asymmetric behaviour is not statistically significant. The only cases of both significant biasedness and asymmetric behaviour in bias are for core PCE and FFR for the pre-COVID sample. The latter results, however, cannot be considered robust due to inconsistencies across the defined thresholds and the frequencies at which the horizon-specific test is examined.

[TABLE 4 HERE]

3.4. Empirical findings: Efficiency

Proceeding to the strong efficiency results, Table 5 reports that when generating FFR forecasts, using the pre-COVID sample, the last available real GDP data has not been efficiently considered, which is confirmed from each test, i.e. testing for common bias, horizon-specific bias at different frequencies, and event-specific bias. Results of a similar fashion are presented for the remaining variables, but only when testing for event-specific bias. For example, it is shown that when forecasting real GDP, the last available information of GDP itself is not efficiently considered. This could also be considered as evidence of bias, which, together with the strong efficiency results, leads to the irrationality of SPD expectations. This specific finding appears for PCE, as well. When implementing the methodology using the full period, the results are more robust to reject strong efficiency in more cases, i.e., when testing event-specific, horizon-specific, and the common.

[TABLE 5 HERE]

To continue with weak efficiency, which is based on an alternative methodological framework, we test for common, horizon- and event-specific weak efficiency by using an outlier-robust (based on the identification of large revisions) method that provides a higher level of granularity regarding the statistical inference. Regarding the results of the common test in pooled regressions, Table 6 shows that weak efficiency is rejected for GDP, core PCE, PCE and FFR during both periods examined[16]. In greater detail, from Table 6, it is observed that none of the table's entries, which reflect variables, horizon-specific weak-efficiency at different frequencies as well as different types of error variance, are significant. Therefore, the general finding is that we reject the null hypothesis that the forecast revisions of SPD are not predictable, which is in line with the outcome of previous studies. A plausible explanation for the rejection of weak efficiency is that SPD is conducted several months within the year, but it is unlikely that the respondents revise their forecasts so frequently. Therefore, it is rational to say that forecasts take time to incorporate the new information. This could be justified by the publication lag that most macroeconomic series are subject to and because the international organisations provide forecasts for most macroeconomic variables at a lower frequency (e.g., the IMF publishes its projections for the macroeconomic outlook twice per year). The overall finding is that there is still room for improvement regarding the SPD forecast quality by incorporating the available information in a more efficient way.

[TABLE 6 HERE]

4. Conclusions

In our proposed methodology, we extend the existing approaches in an integrated framework so that several drawbacks that could lead to less robust results are efficiently captured. From a policymaking point of view, this paper contributes in several ways. First and more importantly, international institutions, like central banks and governmental authorities, can identify more accurately unbiased and efficient forecasts provided by several surveys, which are used to support their policy decisions or even to enhance the predictability of their models.

Regarding the results of unbiasedness, it is crucial that the assessment of the macroeconomic expectations should cover a wide range of aspects. First, it is found that more granularity when revealing the forecast horizons that appear to be characterised by biasedness is value-added. Moreover, testing for event-specific bias, under the assumption of heteroscedasticity across target events, is found to help policymakers identify specific properties of fixed-event surveys' expectations more robustly. From the implementation of the proposed methodology to SPD, it is

[16] Ager et al. (2019) also use two lags of past revisions as explanatory variables in the model specification, since they are interested in whether the forecast revision's predictability remains, but their findings are similar.

found that GDP forecasts are subject to specific characteristics, such as short-term bias and consistency over different periods, including COVID-19. In this regard, the outlier-robust approach results in stricter inference, which does not significantly affect the decision on unbiasedness. Finally, regarding state-dependence bias, it is reported to exist for specific events.

The second property examined by this study is efficiency. We not only test for both weak and strong efficiency, which is not the case according to the literature, but we also enhance the existing methodologies by integrating the adjustments developed for tracing horizon- and event-specific efficiency in an outlier-robust framework. From the application conducted in SPD, it is shown that the forecast revisions are predictable (weak efficiency) and, more particularly, the exogenous information that is available at the time of forecasting is not efficiently incorporated in the provided expectations (strong efficiency).

The contribution of the proposed framework for evaluating fixed-event forecasts is of significant importance for policymakers and investors, who incorporate these surveys' responses in their decision-making process. In conclusion, it is important that both the irrationality in forecasts and its sources should be examined, to the greatest extent possible, before proceeding with their use. Future research could explore the time variation of coefficients in the regression-based tests, as well as the extension of this econometric framework to fixed-horizon surveys, which are also extensively employed by policymakers.

References

Aastveit, K. A., Bjørnland, H. C., & Cross, J. L. (2023). Inflation expectations and the pass-through of oil prices. The Review of Economics and Statistics, 105 (3), 733–743.

Ager, P., Kappler, M., & Osterloh, S. (2009). The accuracy and efficiency of the consensus forecasts: A further application and extension of the pooled approach. International Journal of Forecasting, 25(1), 167–181.

Bagliano, F. C., & Favero, C. A. (1998). Measuring monetary policy with VAR models: An evaluation. European Economic Review, 42, 1069–1112.

Bakhshi, H., Kapetanios, G., & Yates, T. (2005). Rational expectations and fixed-event forecasts: An application to UK inflation. *Empirical Economics,* 30, 539–553.

Batchelor, R. A. (2007). Bias in macroeconomic forecasts. International Journal of Forecasting, 23 (2), 189–203.

Bernanke, B. & Blanchard, O. (2025). What caused the US pandemic-era inflation? American Economic Journal: Macroeconomics, 17 (3), 1–35.

Capistran, C. & Timmermann, A. (2009). Disagreement and biases in inflation expectations. Journal of Money, Credit and Banking, 41, 365-396.

Capistran, C. & López-Moctezuma, G. (2014). Forecast revisions of Mexican inflation and GDP growth. International Journal of Forecasting, 30 (2), 177-191.

Carroll, C.D. (2003). Macroeconomic expectations of households and professional forecasters. The Quarterly Journal of Economics, 118 (1), 269-298.

Cho, S., Del Negro, M., Diagne, I., Gundam, P., Lee, D., & Pacula, B. (2024). The New York Fed DSGE model forecast - September 2024. Federal Reserve Bank of New York Liberty Street Economics, September 20, 2024.

Clements, M. P. (1997). Evaluating the rationality of fixed-event forecasts. Journal of Forecasting, 16, 225–239.

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.

Clements, M. P. (2019). Macroeconomic survey expectations. Palgrave Macmillan.

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. American Economic Review, 105 (8), 2644–78.

Coibion, O., Gorodnichenko, Y., & Kumar, S. (2018). How do firms form their expectations? New survey evidence. American Economic Review, 108 (9), 2671–2713.

Coibion, O., Gorodnichenko, Y., Kumar, S., & Pademonte, M. (2020). Inflation expectations as a policy tool? Journal of International Economics, 124, 103297.

Coibion, O., & Gorodnichenko, Y. (2026). Expectations matter: The new causal macroeconomics of surveys and experiments. Princeton University Press, forthcoming.

Correia-Golay, E., Friedman, S., & McMorrow, M. (2013). Understanding the New York Fed's Survey of Primary Dealers. Current Issues in Economics and Finance, 19(6).

Davies, A., & Lahiri, K. (1995). A new framework for analysing survey forecasts using three-dimensional panel data. Journal of Econometrics, 68(1), 205-227.

Davies, A. (2006). A framework for decomposing shocks and measuring volatilities derived from multi-dimensional panel data of survey forecasts. International Journal of Forecasting, 22 (2), 373–393.

Diercks, A.M., Tanaka, H., & Cordova, P., (2022). Asymmetric monetary policy expectations. Available at SSRN 3930267.

Diercks, A. M., & Jendoubi, H. (2023). Expectations of financial market participants. In Handbook of Economic Expectations (pp. 385-410). Academic Press.

Elliott, G., Komunjer, I., & Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. Review of Economic Studies, 72(4), 1107–1125.

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

Engelberg, J., Manski, C. F., & Williams, J. (2011). Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition. Journal of Applied Econometrics, 26 (7), 1059–1078.

Faust, J., & Wright, J. H. (2009). Comparing Greenbook and reduced form forecasts using a large real-time dataset. Journal of Business & Economic Statistics, 27(4), 468-479.

Gertler, M., & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. American Economic Journal: Macroeconomics, 7 (1), 44–76.

Giannone, D., & Primiceri, G. (2024). The Drivers of Post-Pandemic Inflation. NBER Working Paper No. 32859.

Goodhart, C., & Pradhan, M. (2025). A snapshot of central bank (two-year) forecasting: A mixed picture. Journal of Forecasting, 44 (3), 1097-1131.

Granziera, E., Jalasjoki, P., & Paloviita, M. (2025). The bias of the ECB inflation projections: A state-dependent analysis. Journal of Forecasting, 44(3), 922-940.

Harvey, D. I., Leybourne, S. J., & Newbold, P. (2001). Analysis of a panel of UK macroeconomic forecasts. Econometrics Journal, 4, 37–55.

He, M., Hao, X., Zhang, Y., & Meng, F. (2021). Forecasting stock return volatility using a robust regression model. Journal of Forecasting, 40 (8), 1463-1478.

Holden, K., & Peel, D.A. (1990). On testing for unbiasedness and efficiency of forecasts. The Manchester School, 58, 120-127.

Huber, P. J. (1964). Robust estimation of a location parameter. Annals of Mathematical Statistics, 35 (1), 73–101.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. Annals of Statistics, 1 (5), 799–821.

Huber, P. J. (1981). Robust Statistics, Wiley.

Iregui, A.M., Nunez, H.M., & Otero, J. (2025). Testing the efficiency of oil price forecast revisions in times of COVID-19 and the Russia–Ukraine conflict. Journal of Commodity Markets, 40, 100513.

Isiklar, G., Lahiri, K., & Loungani, P. (2006). How quickly do forecasters incorporate news? Evidence from cross-country surveys. Journal of Applied Econometrics, 21 (6), 703–725.

Keane, M. P., & Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. American Economic Review, 80 (4), 714–735.

Knüppel, M., & Vladu, A.L. (2025). Approximating fixed-horizon forecasts using fixed-event forecasts. Journal of Applied Econometrics, 40, 359-379.

Koop, G., Leon-Gonzalez, R., & Strachan, R.W. (2009). On the evolution of monetary policy. Journal of Economic Dynamics and Control, 33, 997–1017.

Levy, M. D. (2023). The Fed: Bad forecasts and misguided monetary policy. Economics Working Paper 23103, Hoover Institution.

Loungani, P. (2001). How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth. International Journal of Forecasting, 17 (3), 419–432.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. NBER economic forecasts and expectations: Analysis of forecasting behaviour and performance. 1–46.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55 (3), 703–708.

Nordhaus, W. D. (1987). Forecasting efficiency: Concepts and applications. The Review of Economics and Statistics, 69 (4), 667–674.

Odendahl, F., B. Rossi, & Sekhposyan, T. (2023). Evaluating forecast performance with state dependence. Journal of Econometrics, 237 (2), 105220.

Patton, A. J., & Timmermann, A. (2007a). Properties of optimal forecasts under asymmetric loss and nonlinearity. Journal of Econometrics, 140 (2), 884–918.

Patton, A. J., & Timmermann, A. (2007b). Testing forecast optimality under unknown loss. Journal of the American Statistical Association, 102 (480), 1172–1184.

Romer C. D., & Romer, D. H. (2000). Federal Reserve information and the behaviour of interest rates. American Economic Review, 90 (3), 429–457.

Rossi, B., & Sekhposyan, T. (2016). Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts. Journal of Applied Econometrics, 31, 507–532.

Roth, C., & Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behaviour? The Review of Economics and Statistics, 102 (4), 731–748.

Sinha, A., Topa, G., & Torralba, F. (2023). Why do forecasters disagree about their monetary policy expectations? Federal Reserve Bank of New York Liberty Street Economics, August 2, 2023.

Timmermann, A. (2006). An evaluation of the World Economic Outlook forecasts. IMF Working Paper, 059 (2006).

Tulip, P. (2009). Has the economy become more predictable? Changes in Greenbook forecast accuracy. Journal of Money, Credit and Banking, 41(6), 1217-1231.

Vereda, L., Savignon, J., & da Silva, T. G. (2021). A new method to assess the degree of information rigidity using fixed-event forecasts. International Journal of Forecasting, 37 (4), 1576-1589.

Vereda, L., de Mendonça, H. F., & Morcerf, G. (2025). Exploring the role of heterogeneous informational shocks in bias testing of consensus forecasts. Applied Economics, 57 (56), 9557–9572.

Xing, L. M., & Zhang, Y. J. (2022). Forecasting crude oil prices with shrinkage methods: Can nonconvex penalty and Huber loss help?. Energy Economics, 110, 106014.

TABLES

Table 1: Estimated standard deviations of common macro shocks for the homoscedastic specification.

Common macro shocks				
	RGDP	**CPCE**	**PCE**	**FFR**
Pre-COVID-19 (2012-2019)	0.13	0.04	0.09	0.13
Pre-COVID-19 - Outlier Dummy	0.13	0.04	0.07	0.13
Full Sample (2012-2024)	0.35	0.24	0.32	0.39
Full Sample - Outlier Dummy	0.20	0.13	0.18	0.19

Table 2: Unbiasedness results under the horizon-specific test.

Horizon-specific bias				
***Pre-COVID sample* (2012-2019)**				
	RGDP	**CPCE**	**PCE**	**FFR**
Common	-0.11	-0.30***	-0.52***	-0.34
Months	3, 6	1-36	1-36	20-21, 31-36
Quarters	1	1-12	1-12	7, 11-12
Years	-	1-3	1-3	-
***Full sample* (2012-2024)**				
Common	-0.05	0.20	0.16	0.08
Months	1-7	-	-	-
Quarters	1-2	-	-	-
Years	-	-	-	-

Note: Common shows uniform bias, whereas months, quarters and years rows show for which individual forecast horizons, bias is significant in monthly, quarterly and annual survey frequency, respectively, at 5% significance level. The panel regression is estimated with the outliers dummy in the variance estimation equation. *, **, *** denote significance at the 10%, 5% and 1% level, respectively.

Table 3: Results of unbiasedness under the event-specific test.

Event-specific bias				
Event	**RGDP**	**CPCE**	**PCE**	**FFR**
2012	-0.73*	-0.17*	-0.21**	-0.22
2013	0.13	-0.50***	-0.71***	0.00
2014	-0.23	-0.34***	-0.71***	-0.09
2015	-0.80***	-0.36***	-1.09***	-0.31***
2016	-0.53**	-0.13	-0.38***	-0.85**
2017	0.29**	-0.36***	-0.27***	-0.43
2018	0.57**	-0.07	-0.14**	0.21
2019	0.19*	-0.43***	-0.53***	-0.89***
2020	-2.68**	-0.44***	-0.51**	-1.69**
2021	1.63	2.34***	3.09***	-0.74
2022	-1.42***	2.15***	2.73***	3.14***
2023	1.63***	0.39	0.16	2.25
2024	1.02***	0.40***	0.17**	0.64

Note: Target event rows show bias estimates. The panel regression is estimated with the outlier dummy in the variance estimation. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 4: Results of state-dependence in bias using both the historical average and longer-run forecasts as a threshold.

State-dependence in bias								
***Pre-COVID sample* (2012-2019)**								
	Historical average				**Long-run forecasts**			
	RGDP	**CPCE**	**PCE**	**FFR**	**RGDP**	**CPCE**	**PCE**	**FFR**
Common	0.11	0.05	0.08	0.28	-0.1	-0.14***	0.02	-0.17
Common -Bias	-0.19	-0.35***	-0.59***	-0.6**	0	-0.14***	-0.53***	-0.17
Months	0.16	0.1	0.05	0.58**	-0.1	-0.29***	0.08	-0.38
Months-Bias	-	2-36	6-36	1-36	6	1-9, 15-23, 25-30	4-36	4, 31-34
Quarters	0.15	0.1	0.05	0.58**	0.15	0.1***	0.05	0.58
Quarters-Bias	-	1-12	3-12	1-12	-	1-10	2-12	11-12
Years	0.15	0.1	0.05	0.53**	-0.08	-0.22***	0.05	-0.28
Years-Bias	-	1-3	1-3	1-3	-	2-3	1-3	-
Events	0.22***	0.08***	0.1*	-0.15	-0.02	-0.26***	0.1*	-0.26**
Events-Bias	12, 14-18	12-19	12-19	18-19	13, 15-18	18-19	13-19	12-14, 18-19
***Full sample* (2012-2024)**								
	Historical average				**Long-run forecasts**			
Common	0.33	-0.38	-0.11	0.63	0.12	-0.43	-0.05	-0.14
Common -Bias	-0.25	0.47**	0.23	-0.41	-0.04	0.59**	0.25	0.24
Months	0.36	-0.42	-0.13	0.65	0.12	-0.5	-0.1	-0.19
Months-Bias	-	11-19, 24	-	-	5-7	12-13, 19-21, 24, 31	-	-
Quarters	0.35	-0.42	-0.13	0.65	0.12	-0.5	-0.1	-0.19
Quarters-Bias	-	4-6	-	-	2	5-7, 11	-	-
Years	0.35	-0.41	-0.12	0.63	0.13	-0.48	-0.09	-0.18
Years-Bias	-	2	-	-	-	2	-	-
Events	0.32**	0.61***	0.68***	1.6***	0.09	1.52***	1***	2.72***
Events-Bias	12-16, 20, 22-24	12-22, 24	12-22, 24	12-22	13, 15-18, 20, 22-24	12-22, 24	12-22, 24	12-21

Note: Common, months, quarters, years and event rows show coefficient $\hat{\gamma}$ for the dummy regressor. Dummy regressor takes values of one when the last observed value is larger than the specified threshold when the survey is conducted, and zero otherwise. The second rows, where *Bias* is included in the name, show for which individual forecast horizons bias is significant in monthly, quarterly and annual survey frequency, respectively, at 5% significance level. The panel regression is estimated with the outlier dummy in the variance estimation. *, **, *** denote significance at the 10%, 5% and 1% level, respectively.

Table 5: Results of strong efficiency test.

		Common	Months	Quarters	Years	Events
Strong efficiency						
***Pre-COVID sample* (2012-2019)**						
RGDP	**RGDP**	-0.1	-0.14	-0.13	-0.13	-0.21***
	CPCE	0.53	0.6	0.58	0.58	-0.05
	PCE	-0.12	-0.09	-0.09	-0.1	-0.31***
	FFR	0.33	0.29	0.29	0.3	-0.21*
CPCE	**CPCE**	0	-0.02	-0.02	-0.01	-0.08
	RGDP	-0.01	-0.03	-0.03	-0.03	0.01
	PCE	-0.09	-0.08	-0.08	-0.08	-0.06*
	FFR	0.04	-0.01	-0.01	-0.01	0.12**
PCE	**PCE**	-0.16	-0.15	-0.15	-0.15	-0.19***
	RGDP	0.09	0.03	0.03	0.03	0.08*
	CPCE	0.2	0.21	0.2	0.21	-0.21**
	FFR	0.19	0.12	0.12	0.13	0.17***
FFR	**FFR**	-0.08	-0.2	-0.2	-0.19	0.21
	RGDP	-0.44**	-0.54***	-0.54***	-0.53***	-0.27***
	CPCE	0.04	0.08	0.07	0.06	0.03
	PCE	-0.01	0.02	0.02	0.01	-0.19*
***Full sample* (2012-2024)**						
RGDP	**RGDP**	-0.02	-0.02	-0.02	-0.02	-0.06
	CPCE	0.36*	0.36*	0.35*	0.36*	0.06
	PCE	0.26*	0.26*	0.26*	0.26*	0.11
	FFR	0.28***	0.27**	0.27**	0.27**	0.15
CPCE	**CPCE**	0.15	0.16	0.16	0.16	-0.43***
	RGDP	-0.09***	-0.09***	-0.09***	-0.09***	-0.06***
	PCE	0.06	0.06	0.06	0.06	-0.25***
	FFR	-0.02	0	0	-0.01	-0.18***
PCE	**PCE**	0.03	0.04	0.04	0.03	-0.36***
	RGDP	-0.12***	-0.12***	-0.12***	-0.12***	-0.08***
	CPCE	0.13	0.14	0.14	0.14	-0.58***
	FFR	-0.02	0	0	0	-0.15***
FFR	**FFR**	-0.19	-0.19	-0.19	-0.19	-0.67***
	RGDP	-0.05	-0.05	-0.05	-0.05	-0.07**
	CPCE	0.51***	0.51***	0.51***	0.51***	-0.46**
	PCE	0.27**	0.27**	0.27**	0.27**	-0.38***

Note: Common, months, quarters, years and event columns show coefficient $\hat{\gamma}$ for the regressor of a selected variable from the second column regressed on the forecast error of a variable in the first column, when bias is accounted for as uniform, in monthly, quarterly, annual frequency, or as event-specific. The panel regression is estimated with the outlier dummy in the variance estimation. *, **, *** denote significance at the 10%, 5% and 1% level, respectively.

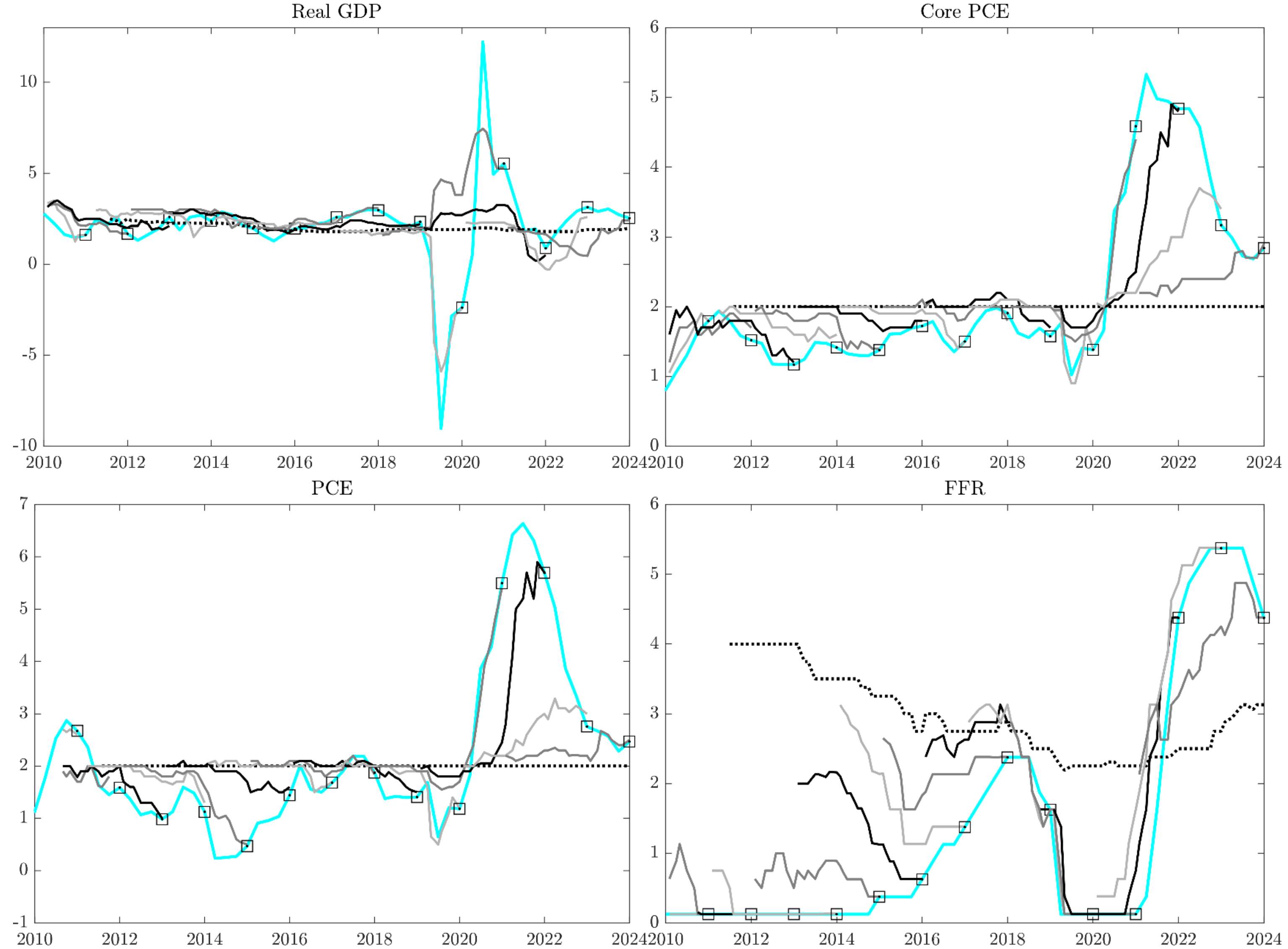
Table 6: Results of weak efficiency test for both homoscedastic and heteroscedastic cases.

Weak efficiency								
***Pre-COVID sample* (2012-2019)**								
	Homoscedastic				**Heteroscedastic**			
	RGDP	**CPCE**	**PCE**	**FFR**	**RGDP**	**CPCE**	**PCE**	**FFR**
Common	0.31*** (8.82)	0.23*** (10.24)	0.32*** (7.57)	0.24*** (13.5)	0.31*** (4.12)	0.23*** (3.52)	0.32*** (4.05)	0.24*** (3.6)
Months	0.31*** (6.96)	0.24*** (6.3)	0.32*** (7.12)	0.25*** (8.08)	0.31*** (4.49)	0.24*** (3.74)	0.32*** (3.75)	0.25*** (3.56)
Quarters	0.28*** (7.02)	0.2*** (5.3)	0.27*** (6.54)	0.22*** (10.05)	0.28*** (4)	0.2*** (3.18)	0.27*** (3.22)	0.22*** (3.23)
Years	0.31*** (8.86)	0.22*** (6.2)	0.29*** (6.01)	0.24*** (16.32)	0.31*** (4.12)	0.22*** (3.27)	0.29*** (3.53)	0.24*** (3.56)
Events	0.27*** (6.64)	0.21*** (7.55)	0.3*** (6.68)	0.23*** (14.12)	0.27*** (3.68)	0.21*** (3.28)	0.3*** (3.82)	0.23*** (3.3)
***Full sample* (2012-2024)**								
	Homoscedastic				**Heteroscedastic**			
	RGDP	**CPCE**	**PCE**	**FFR**	**RGDP**	**CPCE**	**PCE**	**FFR**
Common	0.28*** (30.7)	0.42*** (41.96)	0.46*** (22.81)	0.41*** (623.31)	0.28** (2.52)	0.42*** (6.05)	0.46*** (5.73)	0.41*** (6.28)
Months	0.27*** (23.54)	0.43*** (21.74)	0.47*** (23.56)	0.42*** (29.18)	0.27*** (2.68)	0.43*** (6.45)	0.47*** (6.04)	0.42*** (6.32)
Quarters	0.25*** (20.21)	0.41*** (23)	0.46*** (23.36)	0.39*** (25.31)	0.25** (2.32)	0.41*** (6.02)	0.46*** (5.71)	0.39*** (6.09)
Years	0.28*** (29.5)	0.41*** (24.71)	0.46*** (22.58)	0.41*** (130.62)	0.28** (2.52)	0.41*** (6.06)	0.46*** (5.73)	0.41*** (6.22)
Events	0.26*** (21.64)	0.35*** (16.49)	0.38*** (14.25)	0.31*** (12.32)	0.26** (2.37)	0.35*** (5.35)	0.38*** (4.9)	0.31*** (4.99)

Note: Values show the coefficient $\hat{\gamma}$ for the lagged forecast revisions. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. T-statistics are given in parentheses.

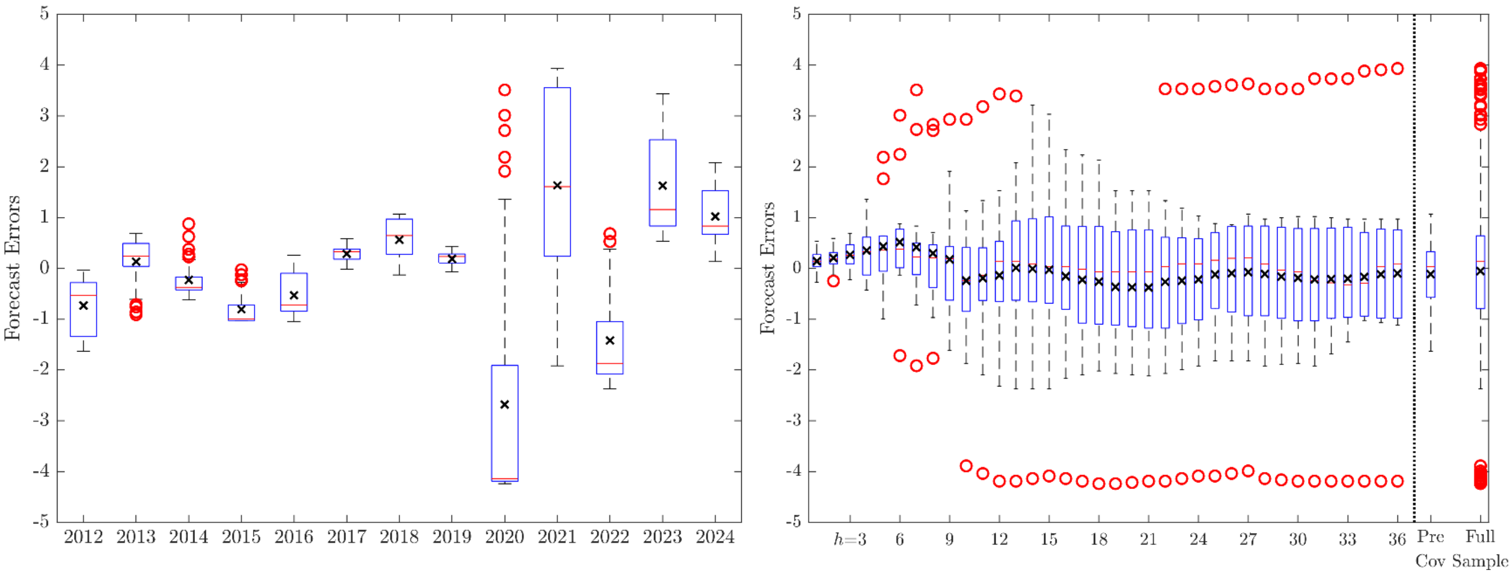
FIGURES

Figure 1: Macroeconomic forecasts (real GDP, core PCE, PCE inflation and FFR) and the corresponding targeted actual values.



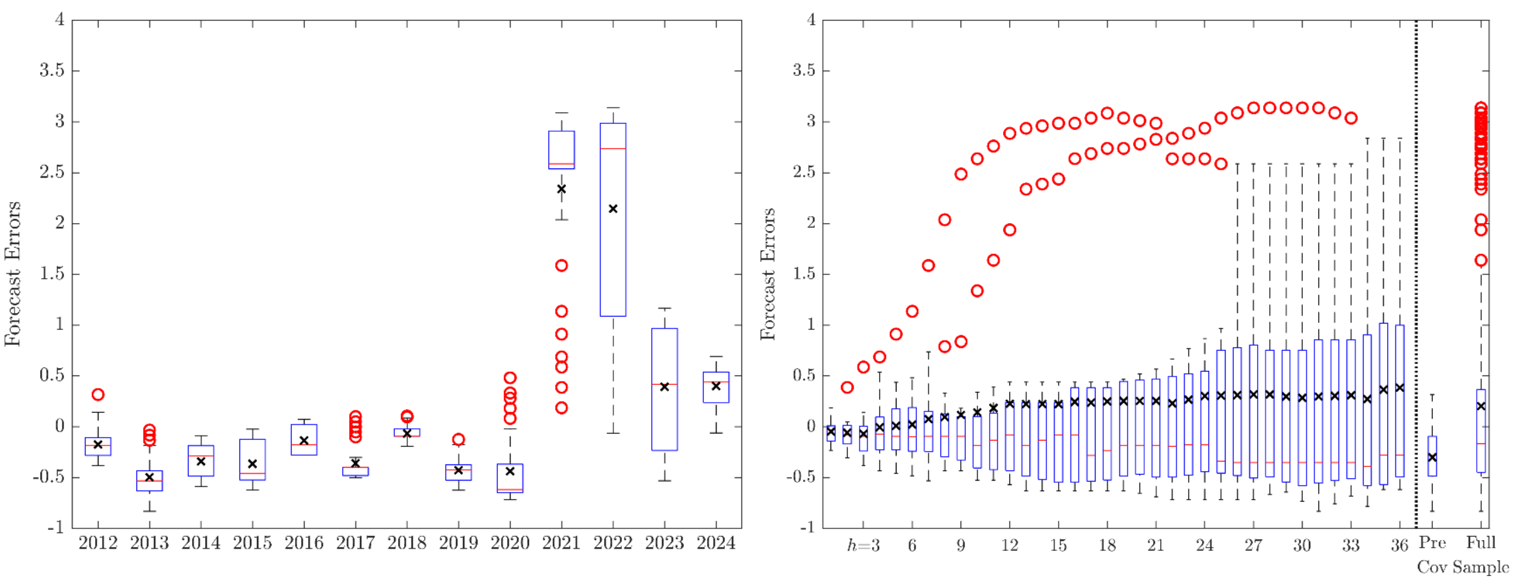
Note: Light blue lines show the actual as available 2 quarters after the reference period. Real GDP, core PCE and PCE are in annual growth rates. Boxes denote the actual Q4/Q4, which are the target events from the consecutive surveys. Forecast paths starting 36 months prior to each target event are represented in black, dark grey and light grey. Dotted lines show the longer-run forecast at the time each survey was conducted.

Figure 2: Forecast errors for real GDP.



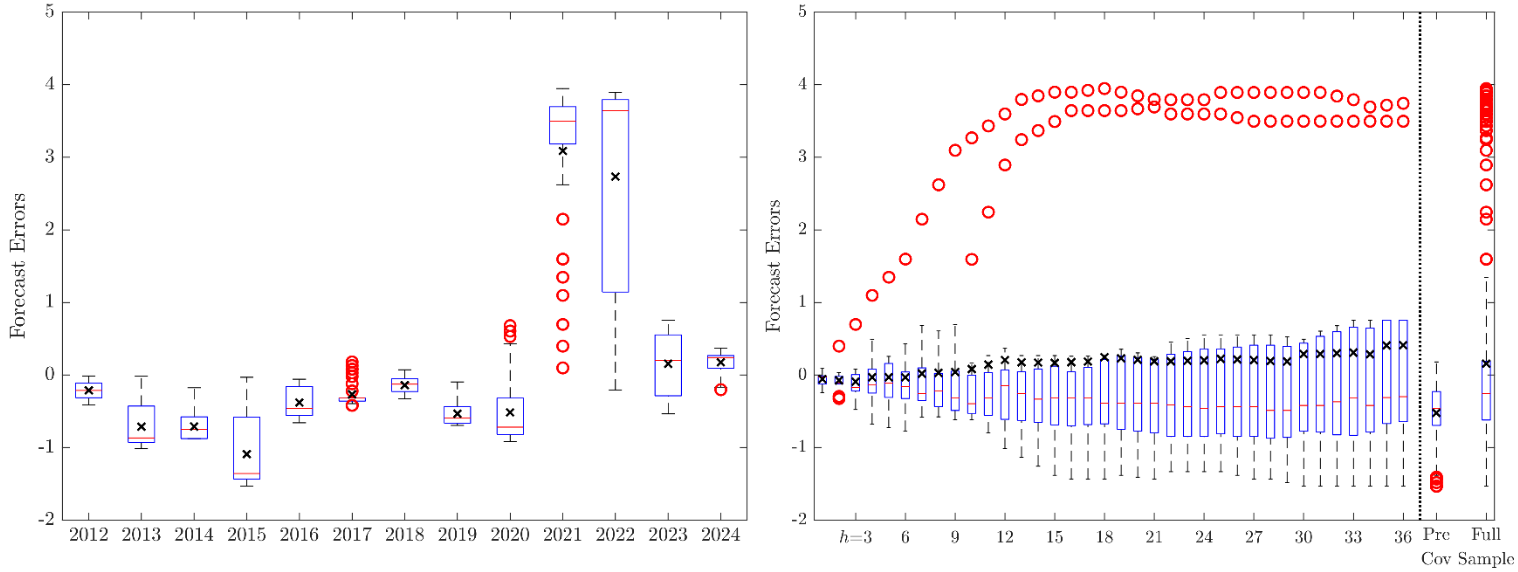
Note: LHS: subplot: The boxplots visualise the distribution of the forecast errors for each target event. RHS subplot: The boxplots visualise the distribution of the forecast errors for each forecast horizon at monthly frequency, the Pre-Covid (2012-2019) and the Full Sample (2012-2024) forecast errors pooled samples. The outliers (shown in red) are identified using the Tukey *IQR* rule: observations lying below $Q_1 - 1.5\ IQR$ or above $Q_3 + 1.5\ IQR$. Crosses show the mean forecast error for each target event.

Figure 3: Forecast errors for core PCE.



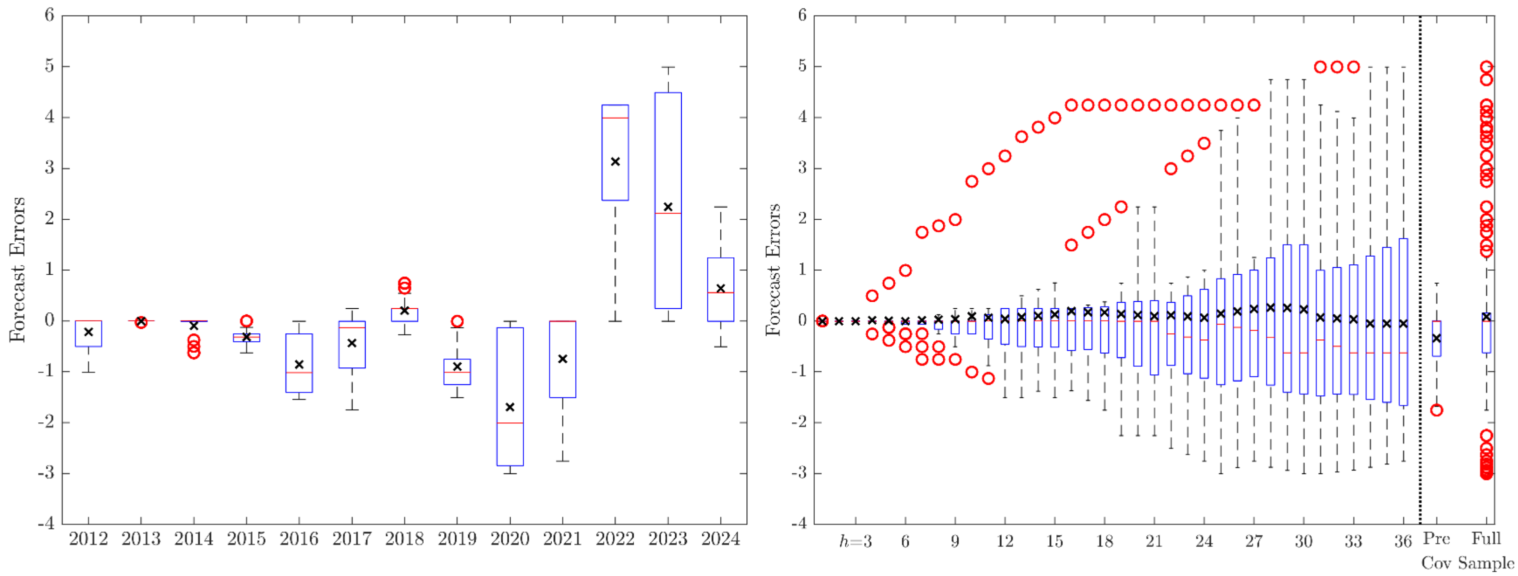
Note: See Note in Figure 2.

Figure 4: Forecast errors for PCE.



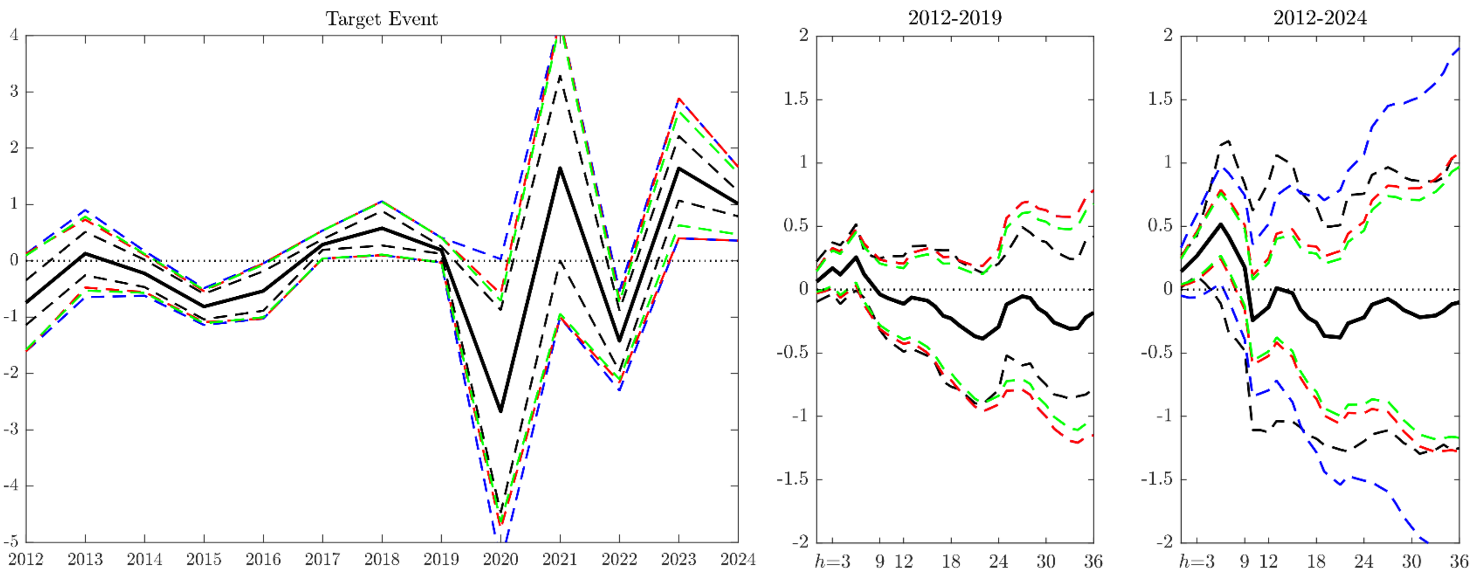
Note: See Note in Figure 2.

Figure 5: Forecast errors for FFR.



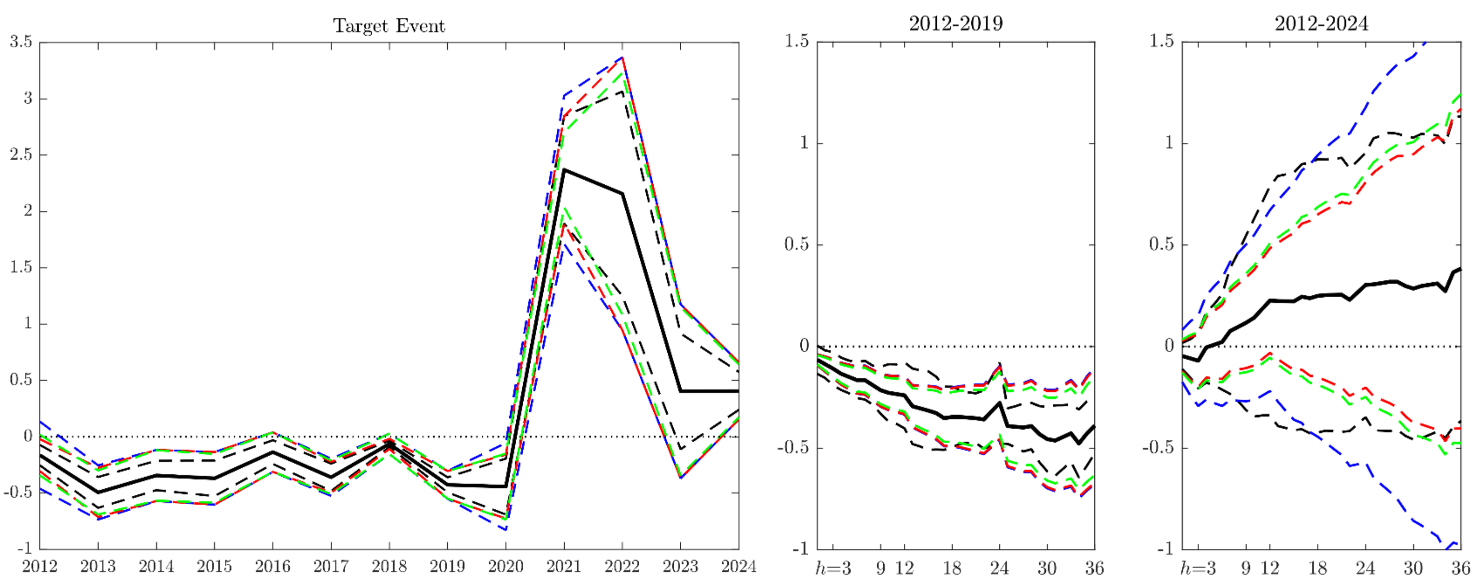
Note: See Note in Figure 2.

Figure 6: Unbiasedness test results for real GDP per target event and per horizon at monthly frequency.



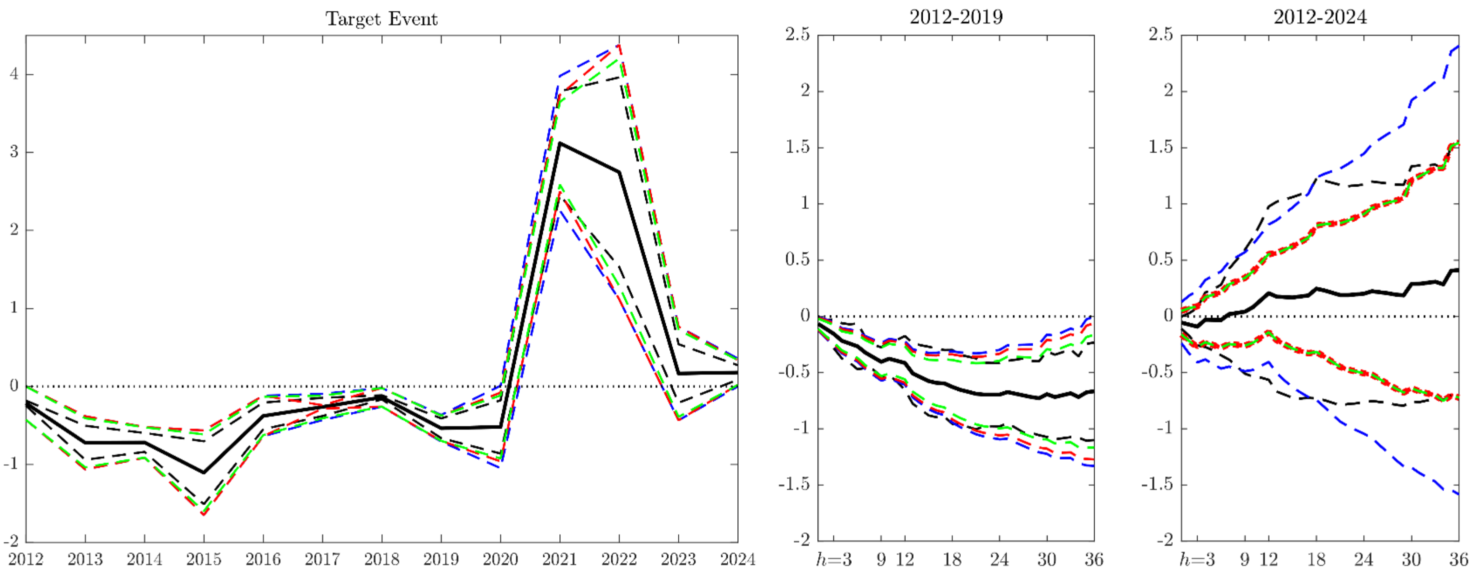
Note: LHS subplot: Black continuous lines show event-specific bias. RHS subplots: Black continuous lines show the 2012-2019 and 2012-2024 sample bias for each forecast horizon at monthly frequency. Black dashed lines show HAC robust confidence intervals (based on Bartlett window, with bandwidth set as *H*-1; Newey and West, 1987). Blue and red dashed lines show confidence intervals estimated without and with the outliers dummy in the variance estimation, respectively. Green dashed lines show confidence intervals estimated using the Huber estimator. All confidence intervals are at the 5% significance level.

Figure 7: Unbiasedness test results for core PCE inflation per target event and per horizon at monthly frequency.



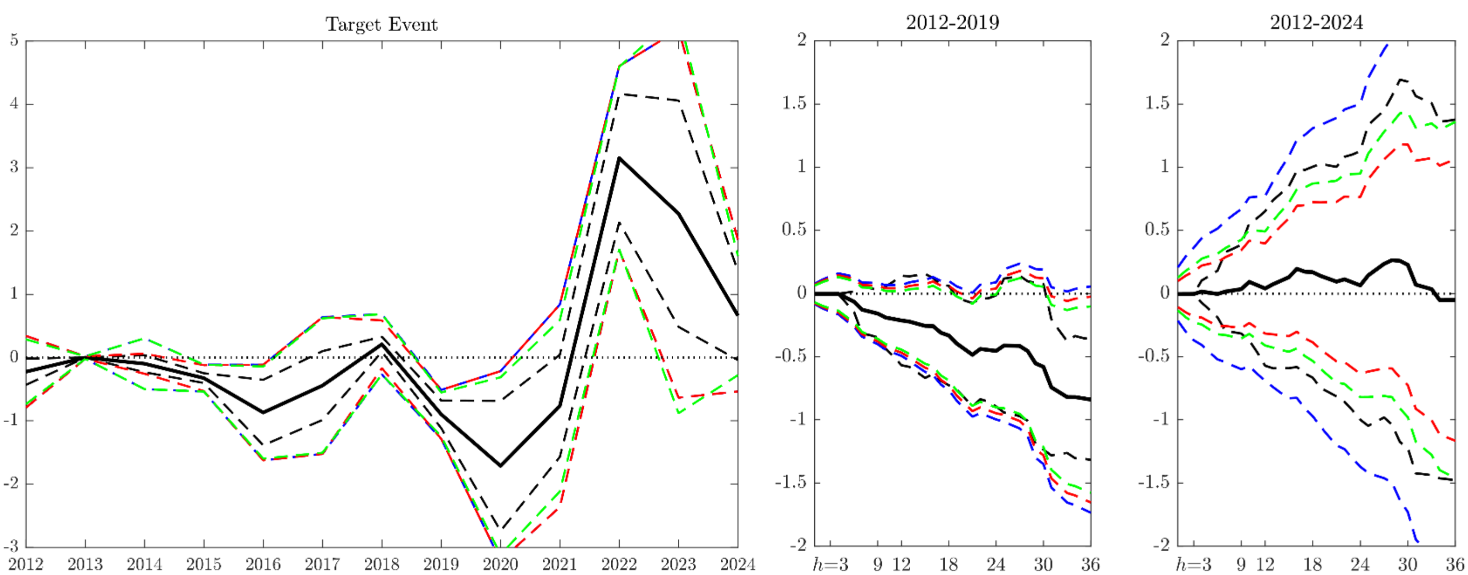
Note: See Note in Figure 6.

Figure 8: Unbiasedness test results for PCE inflation per target event and per horizon at monthly frequency.



Note: See Note in Figure 6.

Figure 9: Unbiasedness test results for FFR per target event and per horizon at monthly frequency.



Note: See Note in Figure 6.

APPENDIX

Weak efficiency

Homoscedastic case

The covariance matrix $\mathbf{\Omega}$ of dimension $\left((\widetilde{H}-2)T \times (\widetilde{H}-2)T\right)$ is estimated as $\widehat{\mathbf{\Omega}} = \hat{\sigma}^2\mathbf{\Lambda}$.

The estimated common variance $\hat{\sigma}^2$ is estimated using the bias-adjusted revisions $\tilde{\boldsymbol{r}}_{adj}$ from the following regression:

$$\boldsymbol{r} = \boldsymbol{X\alpha} + \boldsymbol{\eta} \tag{A.1}$$

where $\tilde{\boldsymbol{r}}_{adj} \equiv \widehat{\boldsymbol{\eta}} = \boldsymbol{r} - \boldsymbol{X}\widehat{\boldsymbol{\alpha}}$. Having obtained $\tilde{\boldsymbol{r}}_{adj}$ from (A.1), we estimate $\hat{\sigma}^2$ as φ_0 in the following regression:

$$\tilde{\boldsymbol{r}}_{adj} \odot \tilde{\boldsymbol{r}}_{adj} = \boldsymbol{\tau}\varphi_0 + \boldsymbol{\delta}_{outl}\varphi + \boldsymbol{\zeta} \tag{A.2}$$

where $\boldsymbol{\tau} = \mathbf{i}_{(\widetilde{H}-2)T}$, and $\boldsymbol{\delta}_{outl}$ are $\left((\widetilde{H}-2)T \times 1\right)$ vectors and φ is a scalar.

The dummy $\delta_{outl,th}$, $\forall t = 1 \cdots T, \forall h = 1 \cdots \widetilde{H}-2$ takes the value of one, if the bias-adjusted forecast revision $\tilde{r}_{adj,th}$ exceeds 1.5 times the interquartile range: $\delta_{outl,th} = \mathbf{1}\left(\tilde{r}_{adj,th} < Q_1 - 1.5IQR \vee \tilde{r}_{adj,th} > Q_3 + 1.5IQR\right)$.

Turning to the matrix $\mathbf{\Lambda}$ of dimension $\left((\widetilde{H}-2)T \times (\widetilde{H}-2)T\right)$, it is defined as follows:

$$\mathbf{\Lambda} = \begin{bmatrix}
\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}' & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{B}' & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{A} & & & & \mathbf{B}' & & & & & \mathbf{0} & \mathbf{B}' & \ddots & \vdots \\
\vdots & \mathbf{0} & \ddots & & & & \ddots & & & & & & \ddots & \mathbf{0} \\
\mathbf{0} & \vdots & & & & & & \ddots & & & & & & \mathbf{B} \\
\mathbf{B} & \mathbf{0} & & & & & & & \mathbf{B}' & & & & & \mathbf{0} \\
\mathbf{0} & \mathbf{B} & & & & \ddots & & & & \ddots & & & & \vdots \\
\vdots & \mathbf{0} & \ddots & & & & \mathbf{A} & & & & & & & \mathbf{0} \\
\mathbf{0} & \vdots & & & & & & \ddots & & & & \ddots & & \vdots \\
\vdots & & & \ddots & \mathbf{B} & & & & & & & & \mathbf{B}' & \mathbf{0} \\
\mathbf{0} & & & & & \ddots & & & & & & & & \mathbf{B} \\
\mathbf{B} & \mathbf{0} & & & & & & & & & & & & \mathbf{0} \\
\mathbf{0} & \mathbf{B} & & & & & \ddots & & & & & & & \vdots \\
\vdots & \ddots & \ddots & & & & & \mathbf{B} & & & & & \mathbf{A} & \mathbf{0} \\
\mathbf{0} & \cdots & \mathbf{0} & \mathbf{B} & \mathbf{0} & \cdots & & \mathbf{0} & \mathbf{B} & \mathbf{0} & & \cdots & \mathbf{0} & \mathbf{A}
\end{bmatrix}$$

where $\mathbf{A} = \boldsymbol{I}_T$ is the identity matrix of dimension T, and $\mathbf{B}$ is a $(T \times T)$ matrix, which has ones on the first diagonal below the main diagonal, and it is written as:

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & & \cdots & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{1} & \ddots & \ddots & \\ \vdots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1} & \mathbf{0} \end{bmatrix}$$

Matrix **B** is placed in $\mathbf{\Lambda}$ matrix in the $1/3\left(\widetilde{H}-2\right)$ and $2/3\left(\widetilde{H}-2\right)$ diagonals above and below the main diagonal.

Heteroscedastic case

Similar to the homoscedastic case, firstly, we obtain the bias-adjusted revisions $\tilde{\boldsymbol{r}}_{adj}$ using (A.2). In addition, we regress $\tilde{\boldsymbol{r}}_{adj} \odot \tilde{\boldsymbol{r}}_{adj}$ on a matrix of event-specific dummies $\boldsymbol{\tau}$ and event-specific outlier dummies $\boldsymbol{\delta}_{outl}$:

$$\tilde{\boldsymbol{r}}_{adj} \odot \tilde{\boldsymbol{r}}_{adj} = \boldsymbol{\tau}\boldsymbol{\varphi}_{\mathbf{0}} + \boldsymbol{\delta}_{outl}\boldsymbol{\varphi} + \boldsymbol{\zeta} \tag{A.4}$$

where $\boldsymbol{\tau} = \mathbf{i}_{\widetilde{H}-2} \otimes \mathbf{I}_T$ is a $\left(\left(\widetilde{H}-2\right)T \times T\right)$ matrix, $\boldsymbol{\varphi}_{\mathbf{0}}$ is $T \times 1$, $\boldsymbol{\delta}_{outl}$ is of dimension $\left(\left(\widetilde{H}-2\right)T \times T\right)$, and $\boldsymbol{\varphi}$ is $(T \times 1)$ correspondingly.

The estimated coefficients $\widehat{\boldsymbol{\varphi}}_{\mathbf{0}}$ from (A.4) represent the $\hat{\sigma}_1^2, \cdots, \hat{\sigma}_T^2$ for each target event $t = 1, \cdots, T$. Matrix $\boldsymbol{\delta}_{outl}$ of dimension $(T(\widetilde{H}-2) \times T)$ contains the outliers dummies and is defined as:

$$\boldsymbol{\delta}_{outl} = \mathrm{diag}\left(\mathrm{vec}(\widetilde{\boldsymbol{\delta}}_{outl})\right) \times (\mathbf{i}_{\widetilde{H}-2} \otimes \mathbf{I}_T) \tag{A.5}$$

where the dummy $\widetilde{\boldsymbol{\delta}}_{outl}$, which is a $\left(T \times (\widetilde{H}-2)\right)$ matrix, $\forall t = 1 \cdots T$ separately, takes the value of one if the bias-adjusted forecast revision $\tilde{r}_{adj,th}$ exceeds 1.5 times the interquartile range: $\tilde{\delta}_{outl,th} = \mathbf{1}\left(\tilde{r}_{adj,th} < Q_1 - 1.5IQR \vee \tilde{r}_{adj,th} > Q_3 + 1.5IQR\right)$.

Now the covariance matrix $\widehat{\mathbf{\Omega}}$ for the weak efficiency regression in eq. (22) is estimated as:

$$\widehat{\mathbf{\Omega}} = \mathbf{I}_{\widetilde{H}-2} \otimes \mathrm{diag}(\hat{\sigma}_1^2, \cdots, \hat{\sigma}_T^2) + \sum_{i=1}^{(\widetilde{H}-2)/12-1} \mathbf{C}_{\boldsymbol{i}} \otimes \mathrm{diag}(\hat{\sigma}_1^2, \cdots, \hat{\sigma}_{T-\mathrm{i}}^2; -\mathrm{i}) \tag{A.6}$$

where $\mathrm{diag}(\hat{\sigma}_1^2, \cdots, \hat{\sigma}_T^2)$ is a $(T \times T)$ matrix that contains the variances $\mathrm{diag}(\hat{\sigma}_1^2, \cdots, \hat{\sigma}_{T-1}^2)$ on the i^{th} diagonal below the main diagonal.

$\mathbf{C}_{\boldsymbol{i}}$ is a $(\widetilde{H}-2) \times (\widetilde{H}-2)$ matrix defined as $\mathbf{C}_{\boldsymbol{i}} = \mathrm{diag}(1, +i * m) + \mathrm{diag}(1, -i * m)$.

BANK OF GREECE WORKING PAPERS

337. Papadopoulos, G., Ojea-Ferreiro, J., and R. Panzica, "Climate stress test of the global supply chain network: the case of river floods", February 2025
338. Papaoikonomou, D., "Stochastic debt sustainability analysis: a methodological note", March 2025
339. Dellas, H. and G. Tavlas, "The great dollar shortage debate: a modern perspective", March 2025
340. Hall, S. and G. Tavlas, "Quantifying Federal Reserve credibility", April 2025
341. Bragoudakis, Z. and E.T. Gazilas, "Does primary and secondary education contribute to environmental degradation? Evidence from the EKC framework", April 2025
342. Delis, P., Degiannakis, S., and G. Filis, "Navigating crude oil volatility forecasts: assessing the contribution of geopolitical risk", May 2025
343. Angelis, A. and A. Tagkalakis, "Formation, heterogeneity and theory consistency of inflation expectations in the euro area", June 2025
344. Konstantinou, P., Rizos, A., and A. Stratopoulou, "The effectiveness of macroprudential policies in curbing operational risk exposures", July 2025
345. Vilerts, K., Anyfantaki, S., Beņkovskis, K., Bredl, S., Giovannini, M., Horky, F. M., Kunzmann, V., Lalinsky, T., Lampousis, A., Lukmanova, E., Petroulakis, F., and K. Zutis, "Details matter: loan pricing and transmission of monetary policy in the euro area", July 2025
346. Hondroyiannis, G., Papapetrou, E., and P. Tsalaporta, "Exploring the role of technological innovation and fertility on energy intensity: Is a fresh narrative unfolding?", August 2025
347. Anyfantaki, S., Blix Grimaldi, M., Madeira, C., Malovana, S., and G. Papadopoulos, "Decoding climate-related risks in sovereign bond pricing: a global perspective", September 2025
348. Brissimis, S. and E. Georgiou, "Assessing the impact of unconventional monetary policy on long-term interest rates in the euro area with the use of a macro-finance model", September 2025
349. Economides, G., Malley, J., Philippopoulos, A., and A. Rizos, "Policy interventions to mitigate the long run costs of Brexit", October 2025
350. Angelopoulos, G., Bragoudakis, Z., Dimitriou, D., and A. Tsioutsios, "A new proposal for forecasting inflation in the eurozone. A global model", October 2025
351. Delikouras, S., Kontinopoulos, A., Malliaropulos, D., and P. Migiakis, "Bond portfolio rebalancing during dash-for-cash events: Evidence from the COVID-19 outbreak", October 2025.
352. Bragoudakis, Z. and A. Tsioutsios, "Bank concentration and asymmetric interest rate spreads pass through: evidence from selected euro area countries", November 2025.
353. Andreou, P., Anyfantaki, S., Dellis, K., and C. Makridis, "Doing matters more than knowing: evidence from environmental preferences", December 2025.
354. Chrysanthakopoulos, C. and A. Tagkalakis, "Natural disasters and the effects of reconstruction expenditure on output", December 2025.
355. Gazilas, E.T. and Z. Bragoudakis, "Can higher education levels improve tax compliance?", January 2026.