Essays on Forecast Evaluation of the European Central Bank Survey of Professional Forecasters

Fabio Profumo

PhD

University of York

Economics

September 2020

Abstract

This thesis consists of three essays on the evaluation of the European Central Bank Survey of Professional Forecasters (ECB SPF).

Chapter 1 serves as an introductory chapter and it describes the European Central Bank Survey of Professional Forecasters and revisions in macroeconomic data.

Chapter 2 examines the rationality of ECB SPF aggregate point forecasts using a test that allows less restrictive assumptions on the loss function and fixed-smoothing asymptotics to avoid small sample size distortion. ECB SPF forecasts are shown to be rational.

Chapter 3 explores the evaluation of ECB SPF aggregate point forecasts using the Diebold and Mariano test for the null hypothesis of equal forecast accuracy with asymmetric loss functions. I address the small sample size distortion of the Diebold and Mariano test with fixed-smoothing asymptotics. Results show that ECB SPF are slightly more accurate than benchmark models in some cases, especially for short horizons, while their advantage decreases as the horizon extends.

Chapter 4 evaluates the accuracy of ECB SPF aggregate density forecasts with the Diebold and Mariano test with loss functions which accommodate the fact that forecasts are reported as histograms. As in the previous Chapters, small sample size distortion of the test is taken into account using fixed-smoothing asymptotics. ECB SPF density forecasts show a good predictive ability, especially for unemployment and real GDP growth forecasts. For inflation, there is no evidence that the ECB SPF forecasts are better than simple benchmark forecasts because inflation was anchored to the ECB target.

Contents

List of Figures

List of Tables

Introduction

This thesis is the outcome of my Ph.D. studies at the University of York and presents the development of my research interests in forecast evaluation. What follows are three Chapters that contribute to the empirical literature on the European Central Bank Survey of Professional Forecasters (ECB SPF) forecast evaluation, plus an introductory Chapter describing the ECB SPF and the revision in macroeconomic data.

The ECB SPF is a quarterly survey collecting point and density forecasts of inflation, unemployment and real GDP growth for the euro area. This survey is usually neglected in the literature due to the reduced amount of rounds available and the small sample size distortion of available evaluation tests. However, assessing the quality of SPF forecasts is vital because they influence the monetary policy of central banks and the expectations of other economic agents, see Carroll (2003).

In general, forecast evaluation has always been a field of interest in the economic literature. Informal graphic evaluation methods were considered by Theil (1958), which suggests using scatter plots of the forecast against the outcome to understand the magnitude of forecast errors. More formal approaches were proposed by Wilson (1934), which uses correlation between forecasts and realisations, Mincer and Zarnowitz (1969) that proposes a test for forecast rationality and Fair and Shiller (1989, 1990) that examine the information content of ex-ante forecasts. When two competing forecasts are available for the same variable of interest, Chong and Hendry (1986) proposes a test for forecast encompassing while Diebold and Mariano (1995) suggests a test for equal forecast accuracy. In small samples, these tests suffer from size distortion (Clark, 1999). This issue can be alleviated using fixed-smoothing asymptotics (Sun, 2013, 2014; Hualde and Iacone, 2015a), which proved capable of delivering correctly sized tests in small samples (Harvey, Leybourne and Whitehouse, 2017; Coroneo and Iacone, 2020).

The second Chapter assesses the rationality of point ECB SPF aggregate forecasts. Forecasters are considered rational when they use all the information available to them to make a forecast, see Muth (1961). Early studies test forecast rationality under a Mean Squared Error (MSE) loss function but evidence suggests that economic agents can adopt different loss functions. To take this into account, I test rationality using the test developed by Patton and Timmermann (2007), that does not rely on a specific loss function, with fixed-smoothing critical values to eliminate small sample size distortion. Before the empirical exercise, I verify that tests are correctly sized using a Monte Carlo inspired by the one in Patton and Timmermann (2007). This Chapter also provides a comparison of the ECB SPF to the Philadelphia FED SPF. ECB SPF forecasts appear rational both under a general loss function and a MSE loss function. The fact that ECB SPF inflation forecasts are rational under a symmetric loss function (MSE) despite the ECB having an asymmetric inflation target, supports the growing literature claiming that the ECB is losing credibility.

In the third Chapter, I evaluate the predictive ability of ECB SPF aggregate point forecasts using the Diebold and Mariano test (Diebold and Mariano, 1995) with different loss functions. The classic approach is to consider symmetric loss functions when evaluating forecasts but in practice, the true loss function used by the forecaster is not known and the ECB has an asymmetric inflation target. Including asymmetric loss functions that weight forecast errors differently according to their sign in my empirical exercise allows me to conduct a comprehensive evaluation of survey forecasts. Before turning to the empirical exercise, I check that fixed-smoothing asymptotics provide correctly sized Diebold and Mariano test with asymmetric loss functions using the Monte Carlo setting of Coroneo and Iacone (2020) adapted for asymmetric loss functions. Empirical results show that ECB SPF forecasts are generally as good as simple benchmark forecasts but their predictive ability decreases as the forecast horizon increases. My findings are robust to the type of loss function used because forecasters are interested in producing quality forecasts independently of the asymmetric target of the European Central Bank.

The fourth Chapter focuses on ECB SPF aggregate density forecasts. These forecasts are reported as histograms, meaning that forecasters are given a series of intervals, or bins, and they are asked to predict the probability that the target variable will fall in each specific bin. Hence, they give a wider understanding of the uncertainty associated with predictions. In this Chapter, I assess forecast accuracy using the Diebold and Mariano test with specific loss functions that accommodate the bin structure of forecasts, such as the Quadratic Probability Score (QPS) and the Ranked Probability Score (RPS). Differently from the existing literature, I take into account the small sample size distortion of the Diebold and Mariano test using fixed-smoothing asymptotics. I use an original Monte Carlo simulation to check that the test with QPS and RPS is correctly sized under fixed-smoothing asymptotics. Results indicate that ECB SPF density forecasts of unemployment rate and real GDP growth beat simple benchmarks at one-year horizon. Inflation forecasts, instead, do not outperform simple benchmarks as inflation expectations and realisations are close to the ECB target. After the 2008 financial crisis, the predictive ability of SPF increases for all variables indicating that professional forecasters adopted more sophisticated models to predict the variables of interest.

Chapter five summarises all results and conclusions from previous Chapters, and suggests possible future research in forecast evaluation.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Where individual Chapters were co-authored with other researchers, this is indicated with the necessary specifications in this declaration. The work was financially supported by the Economic and Social Research Council, grant number ES/J500215/1.

I am the sole author of Chapter 1. Parts of it were used in my working paper '*The Accuracy of the Survey of Professional Forecasters for the Euro Area: an Heteroscedasticity Autocorrelation Robust Assessment*' and in the working paper '*A Real-time Density Forecast Evaluation of the ECB Survey of Professional Forecasters*' with Dr Laura Coroneo (University of York) and Dr Fabrizio Iacone (Università degli Studi di Milano and University of York).

I am the sole author of Chapter 2 and 3 and appropriate references are made throughout the Chapters to other works. I presented an earlier version of Chapter 3 titled '*The Accuracy of the Survey of Professional Forecasters for the Euro Area: an Heteroscedasticity Autocorrelation Robust Assessment*' at the 59th Riunione Scientifica Annuale of the Società Italiana degli Economisti, at the 6th Workshop for PhD Students in Econometrics and Empirical Economics by Società Italiana di Econometria and as a poster at the 2018 Asset Pricing Workshop of the University of York.

Chapter 4 is adapted from the working paper 19/14 in the University of York Department of Economics discussion papers series, titled '*A Real-time Density Forecast Evaluation of the ECB Survey of Professional Forecasters*' with Dr Laura Coroneo (University of York) and Dr Fabrizio Iacone (Università degli Studi di Milano and University of York). I contributed coding and executing the Monte Carlo simulations, designing and conducting all the empirical analysis and writing the paper. My co-authors contributed at various points with revisions and comments on design and methodology.

Fabio Profumo,

York, September 2020

Chapter 1

The ECB Survey of Professional Forecasters

This introductory Chapter describes the European Central Bank Survey of Professional Forecasters (ECB SPF) that I use in the empirical sections of the following Chapters and the revisions that systematically occur in macroeconomic variables that need to be kept into account in forecast evaluation.

1.1 The ECB Survey of Professional Forecasters

The European Central Bank Survey of Professional Forecasters (ECB SPF) is a quarterly survey that was started in 1999 with the aim to gather information about private sector expectations and assess the credibility of the new European Central Bank founded the year before. It contains forecasts about four main economic indicators:

1. Inflation defined as the year-on-year percentage change of the Harmonised Index of Consumer Prices (HICP) published by Eurostat;

2. Core inflation, defined as the year-on-year percentage change in the euro area HICP special aggregate 'all items excluding energy, food, alcohol and tobacco' published

by Eurostat (only available from the 2016.Q4 survey round);

3. Real GDP growth defined as the year-on-year percentage change of real GDP according to the standardised European System of National and Regional Accounts (ESA) 2010 definition;

4. Unemployment rate, which refers to the International Labour Organisation's (ILO) definition and it is calculated as the percentage of the labour force.

Respondents are also asked to provide their point forecasts about other variables on which their main forecasts are based, such as labour cost, Brent crude oil price in dollars, the dollar/euro exchange rate and some qualitative comments on factors affecting the outlook of each variable.

The ECB SPF questionnaire is regularly submitted to a panel of expert forecasters, about 80 institutions with an average of 60 responses each round. All of the participants are experts affiliated with financial or non-financial institutions based within the European Union and have been chosen with the assistance of National Central Banks to form an heterogeneous group in order to guarantee the representativeness and independence of the expectations collected. Panellists need to be experts in macroeconomics ad have previous forecasting experience for the euro area. The survey is about the euro area but respondents can be also based in the whole European Union, including countries which are not using the euro as their currency.

Professional forecasters are asked to provide their point and density forecasts for several horizons: current calendar year, the following calendar year, the calendar year after that, a long term horizon (5 years ahead), rolling horizons one year ahead and two years ahead of the latest data available. To report their density forecasts, participants are given a set of specific ranges and they are asked to predict the probability that the target variable will fall in each specific range or bin, with the first and the last being open intervals. The number of ranges given in every survey round can change but their width is fixed. The ECB SPF reports both the anonymised individual density forecasts

Table 1.1: ECB SPF timing

Inflation and Core inflation					
Survey	Deadline	Info available	Forecast 1 year	Forecast 2 years	Forecast 5 years
Q1.Y	M1.Y	M12.Y-1	M12.Y	M12.Y+1	M12.Y+4
Q2.Y	M4.Y	M3.Y	M3.Y+1	M3.Y+2	M3.Y+5
Q3.Y	M7.Y	M6.Y	M6.Y+1	M6.Y+2	M6.Y+5
Q4.Y	M10.Y	M9.Y	M9.Y+1	M9.Y+2	M9.Y+5

Unemployment Rate					
Survey	Deadline	Info available	Forecast 1 year	Forecast 2 years	Forecast 5 years
Q1.Y	M1.Y	M11.Y-1	M11.Y	M11.Y+1	M11.Y+4
Q2.Y	M4.Y	M2.Y	M2.Y+1	M2.Y+2	M2.Y+5
Q3.Y	M7.Y	M5.Y	M5.Y+1	M5.Y+2	M5.Y+5
Q4.Y	M10.Y	M8.Y	M8.Y+1	M8.Y+2	M8.Y+5

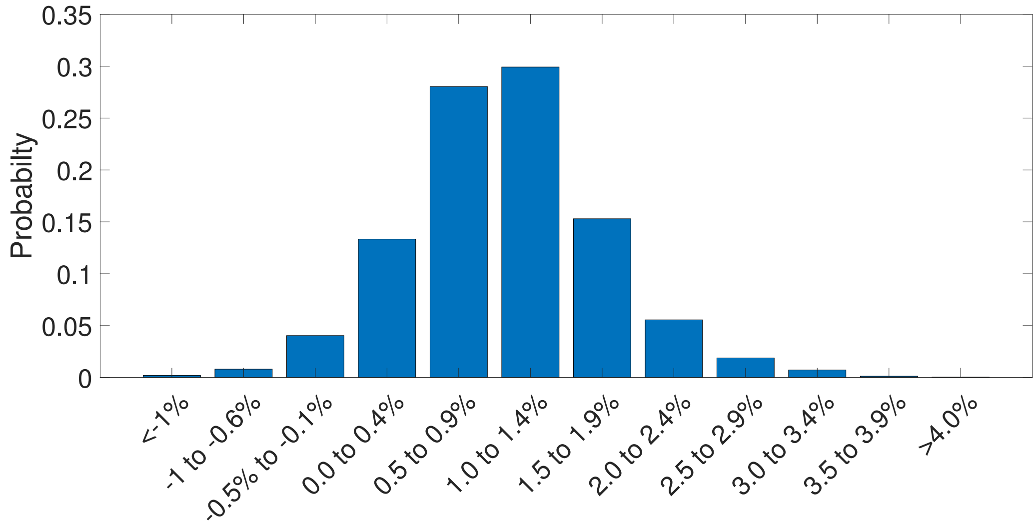
Real GDP Growth					
Survey	Deadline	Info available	Forecast 1 year	Forecast 2 years	Forecast 5 years
Q1.Y	M1.Y	Q3.Y-1	Q3.Y	Q3.Y+1	Q3.Y+4
Q2.Y	M4.Y	Q4.Y-1	Q4.Y	Q4.Y+1	Q4.Y+4
Q3.Y	M7.Y	Q1.Y	Q1.Y+1	Q1.Y+2	Q1.Y+5
Q4.Y	M10.Y	Q2.Y	Q2.Y+1	Q2.Y+2	Q2.Y+5

Note: for each variable and survey round, the Table reports the survey deadline, the latest information available to respondents at the deadline and the forecasts requested. M, Q and Y refer to the month, quarter and year considered respectively. Each survey is produced quarterly but forecasts about inflation and unemployment are about a specific month: end of quarter month and middle of quarter month respectively. For real GDP growth, forecasts are about quarters. Forecasts rolling horizons are 1 year, 2 years and 5 years ahead from the latest information available and not from the date of the survey.

and the aggregated (consensus) forecast which is obtained from the simple average of individual forecasts. For example, Figure 1.1 shows the aggregate density one-year ahead rolling horizon forecast for the Harmonised Index of Consumer Prices for September 2017 produced in the 2016.Q4 survey round.

The ECB SPF survey is conducted four times per year, in the second half of the middle month of each quarter and, from the last quarter of 2001, in the second half of the first month of the quarter. A list of deadlines for reply to the survey is available on the ECB website.

Figure 1.1: ECB SPF density one-year rolling horizon forecast for December 2016 HICP inflation



Note: the histogram reports one-year ahead rolling horizon aggregate density forecast for HICP inflation from the 2016.Q1 survey round. Participants are asked to report a probability for the realisation in December 2016 to fall in each bin.

Table 1.1 shows timings, information available to forecasters and forecasts requested for each quarterly survey: for HICP inflation and unemployment rate, forecasters are asked to forecast a specific month one year, two years and five years ahead from the latest available realisation of the target variable and not from the survey date. For real GDP growth, forecasts are collected for one and two years ahead of the latest information available but these forecasts are about quarters and not about a specific month.

In addition, special questionnaires are sent periodically asking participants about their forecasting practices. Responses from 2008, 2013 and 2018 special questionnaires indicate that forecasts are updated regularly according to the frequency of the target variable and are based on one or more models to cross check results but, especially for long-term forecasts, judgement plays an important role with one third of respondents reporting that their forecasts are essentially judgement based. Moreover, the majority of participants reported the importance of judgement has increased following the 2008 financial crisis. For more information on surveys see Garcia (2003) and Bowles, Friz, Genre, Kenny, Meyler and Rautanen (2007). A thorough analysis of responses is provided by Garcia and Manzanares (2007) in which a bias towards favourable predictions is discovered for all forecast horizons.

Table 1.2: ECB SPF sample statistics. 2002.Q1 - 2016.Q4

HICP Inflation							
Horizon	Mean	Median	Variance	Skewness	Kurtosis	Min	Max
1	1.62	1.70	0.12	-0.28	2.38	0.80	2.40
2	1.75	1.80	0.04	-0.70	2.99	1.20	2.10

Unemployment Rate							
Horizon	Mean	Median	Variance	Skewness	Kurtosis	Min	Max
1	9.51	9.60	2.38	0.08	2.00	6.70	12.40
2	9.21	9.30	2.18	0.09	1.95	6.60	12.00

Real GDP Growth							
Horizon	Mean	Median	Variance	Skewness	Kurtosis	Min	Max
1	1.29	1.45	0.79	-1.71	6.57	-2.00	2.40
2	1.78	1.80	0.19	-0.02	2.19	0.90	2.60

Note: the table reports the 2002.Q1 - 2016.Q4 ECB SPF surveys sample statistics, $T = 60$. The rolling horizons are expressed in years.

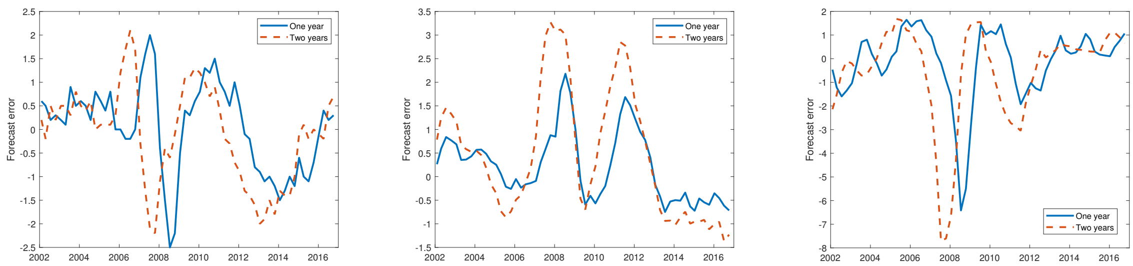
Table 1.2 reports basic statistics for the sample of 60 surveys from 2002.Q1 to 2016.Q4 for one-year and two-years rolling horizon forecasts of the three main variables in the survey. For HICP inflation, the sample mean is below the 2% inflation target of the ECB and so is the median. However, the one-year ahead forecast reached a maximum of 2.4 % in the 2008.Q3 survey; in the same survey, the two-years ahead forecast reached the maximum for the sample of 2.1 %. The minimum for both horizons is reached in the 2015.Q1 survey round. The values of skewness and kurtosis also confirm that forecasters tend to make forecasts that are slightly below but close to the ECB inflation target. The variance is quite small, in particular for two-years ahead forecasts, confirming that forecasters tend to predict values of inflation close to the ECB target in all survey rounds in my sample. This phenomenon is also a possible indication of overconfidence of forecasters as documented in Giordani and Söderlind (2006) for the US SPF. Mitchell and Wallis (2011) speculate about how the judgemental adjustment of forecasts could be a possible cause of overconfidence.

For the unemployment rate, the maximum rate was forecasted in 2013.Q3 for both horizons, while the minimum was forecasted in the third and fourth quarters of 2007 for two-years ahead forecasts and in the third quarter of the same year for one-year ahead forecasts. This variable has the biggest variance compared to the inflation and real GDP growth indicating that bigger variations occurred in the sample with respect to other variables. Skewness and kurtosis suggest that forecasts for extreme values are not frequent and equally below or above the mean.

Real GDP growth forecasts reached their maximum values in 2002 and, for one-year ahead forecasts, in 2007.Q3 too. Minimum values are coming from the 2009.Q1 and 2009.Q2 surveys. The variance is low and it gets smaller the longer the forecast horizon becomes, indicating that for two-years ahead forecasts, surveys report more similar forecasts when the horizon is longer. However, skewness and kurtosis indicate that surveys tend to report values bigger than the mean for one-year ahead forecasts.

Figure 1.2: Forecast errors in ECB SPF

(a) HICP Inflation | (b) Unemployment Rate | (c) Real GDP Growth



Note: the three plots show forecast errors for the ECB SPF forecasts from 2002.Q1 to 2016.Q4 of HICP inflation (a), unemployment rate (b) and real GDP growth (c). The series of the realised data is taken at the latest release available on 31/01/2019. The solid blue line refers to rolling horizon one-year ahead forecasts and the dashed orange line to rolling horizon two-years ahead forecasts. The horizontal axis reports the SPF survey dates.

Figure 1.2 reports ECB SPF forecast errors calculated subtracting the SPF forecasts from the latest release available on 31/01/2019 of the target variables. The blue solid line represents forecast errors from one-year ahead forecasts, while the orange dashed line plots the two-years ahead forecast errors. The horizontal axis reports survey dates. Forecast errors appear similar for both horizons, in fact, the sign of the error is always the same for both horizons while magnitude is slightly bigger for two-years forecasts. It is also clear the effect of the 2008 financial crisis, which took forecasters by surprise and made forecast errors increase in magnitude. However, in surveys immediately succeeding the financial crisis, forecast errors decrease as forecasters quickly adjust their models. A similar pattern is visible during the European sovereign debt crisis.

Table 1.3: ECB SPF forecast errors statistics. 2002.Q1 - 2016.Q4

HICP Inflation							
Horizon	Mean	Median	Variance	Skewness	Kurtosis	Min	Max
1	0.05	0.20	0.90	-0.46	2.92	-2.50	2.00
2	-0.11	0.05	1.06	-0.24	2.39	-2.20	2.10

Unemployment Rate							
Horizon	Mean	Median	Variance	Skewness	Kurtosis	Min	Max
1	0.22	0.15	0.53	0.72	2.78	-0.75	2.18
2	0.48	0.24	1.84	0.57	2.17	-1.36	3.27

Real GDP Growth							
Horizon	Mean	Median	Variance	Skewness	Kurtosis	Min	Max
1	-0.18	0.18	2.39	-1.91	7.81	-6.42	1.64
2	-0.56	0.23	4.40	-1.85	6.37	-7.72	1.68

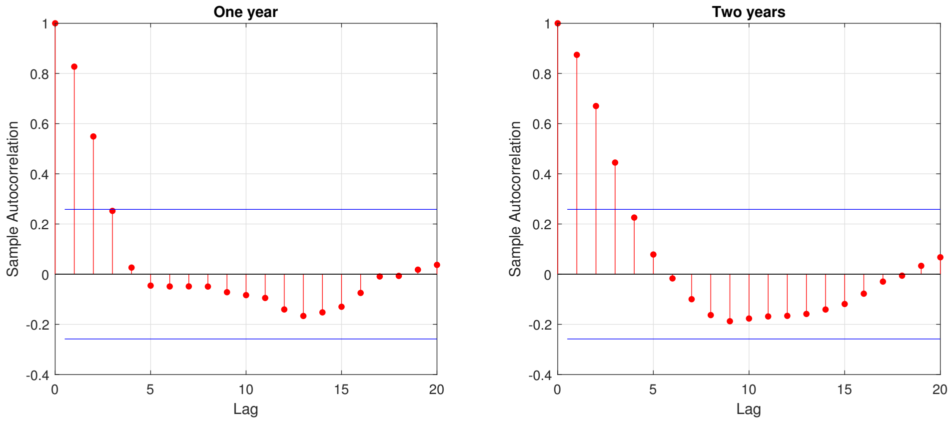
Note: the table reports the 2002.Q1 - 2016.Q4 ECB SPF surveys forecast errors statistics, $T = 60$. The rolling horizons are expressed in years. The series of the realised data is taken at the latest release available on 31/01/2019.

Table 1.3 reports basic statistics of forecast errors in the sample from 2002.Q1 to 2016.Q4. For all variables, the sample mean and median of forecast errors are close to zero and the sample variance grows as the forecast horizon increases. For HICP inflation and unemployment rate, skewness and kurtosis indicate that forecast errors are approximately equally distributed around the mean. While, for real GDP growth, forecast errors tend to be positive, indicating the tendency of forecasters to underpredict GDP growth in the sample.

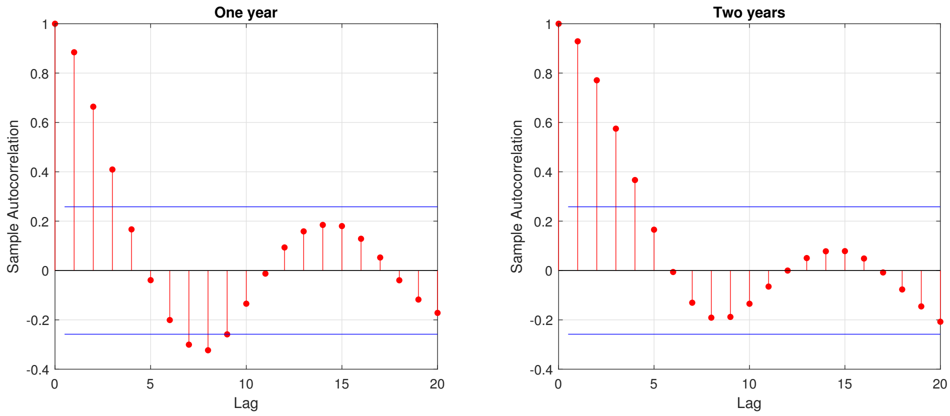
Figure 1.3 depicts the autocorrelation function of forecast errors in the sample from 2002.Q1 to 2016.Q4. For every variable, plots on the left report the autocorrelation for rolling horizon one-year ahead forecasts while plots on the right report autocorrelation for rolling horizon two-years ahead forecasts. Forecast errors for all the three major variables of ECB SPF exhibit quite a high level of serial correlation that persists up to 10 lags for unemployment rate and real GDP growth. For HICP inflation forecast errors,

Figure 1.3: Forecast errors sample autocorrelation in ECB SPF

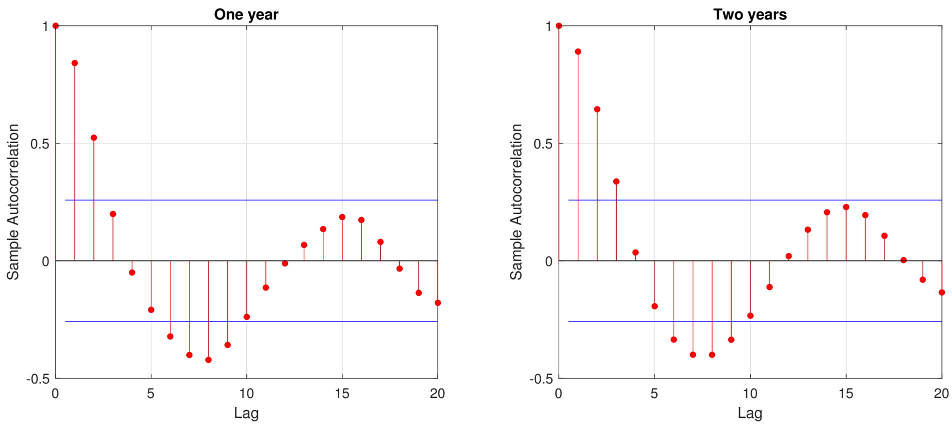
(a) HICP Inflation



(b) Unemployment Rate



(c) Real GDP Growth



Note: the three plots show forecast errors sample autocorrelation for the ECB SPF forecasts from 2002.Q1 to 2016.Q4 of HICP inflation (a), unemployment rate (b) and real GDP growth (c). Each plot consists of two sub-plots; one for one-year ahead forecast errors and one for two-years ahead forecast errors. The series of the realised data is taken at the latest release available on 31/01/2019.

autocorrelation is high until lag 4 for both forecast horizons.

1.2 Revision of Historical Data

Historical data is subject to revision, scheduled or not, caused by new data available, changes in definitions and classifications or correction of clerical mistakes. Mankiw and Shapiro (1986) argue that revision is most likely caused by unforecastable new information not known at time forecasts were made. To record the effect of revision, real-time databases collect historical data realisations and their revisions for a series of variables of interest. For the United States, the real-time database was created by Croushore and Stark (2001) with data from November 1965 while, for the euro area, the real-time database was built by Giannone, Henry, Lalik and Modugno (2012) starting from January 2001. Their European database contains data for 230 indicators and takes into account the political changes of the European Union providing fixed composition series, that use the same group of countries throughout all periods, and changing composition series using the euro area composition at the time to which the statistics relate; i.e. data includes information from new Countries joining the Union time after time.

To illustrate the structure of the real-time database, Table 1.4 reports an extract of the European real-time database for the unemployment rate available on 31/01/2019. The top panel shows that for every data point, the database reports different revisions and different series can be extracted as reported in the bottom panel. Cells shaded in yellow represent the first release for each data point. Circled cells form the series of the four releases after the first and purple cells are those of the current release, the most recent release at the time the database was accessed. In some cases, the same release can be part of different series. For example, the realised value 8.23545 of May 2018 at the vintage of 30/01/2019 is used for both the series of the current release and for the series of four releases after the first. In this extract, unemployment gets revised quiet often after the first release and then remains stable.

Table 1.4: European real-time database for the Unemployment rate.

	Revisions					
	13/06/2018	**25/07/2018**	**12/09/2018**	**24/10/2018**	**12/12/2018**	**30/01/2019**
Mar-18	8.55121	8.49874	8.46137	8.44998	8.45589	8.45589
Apr-18	8.50715	8.43756	8.36587	8.35995	8.37293	8.37293
May-18		8.35278	8.22947	8.22495	8.23545	8.23545
Jun-18			8.23391	8.21701	8.21438	8.21438
Jul-18			8.19096	8.16539	8.13847	8.13847
Aug-18				8.09932	8.07029	8.07029
Sep-18					8.06130	8.0613
Oct-18					8.05632	8.05632

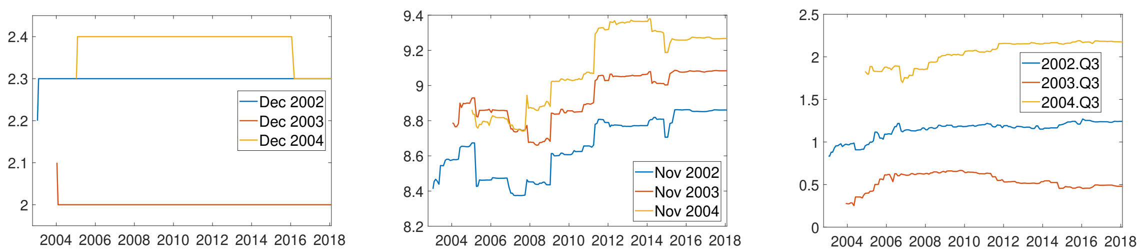
	Series		
	First release	Four releases after the first	Current release
Mar-18	8.55121	8.45589	8.45589
Apr-18	8.50715	8.37293	8.37293
May-18	8.35278	8.23545	8.23545
Jun-18	8.23391	N/A	8.21438
Jul-18	8.19096	N/A	8.13847
Aug-18	8.09932	N/A	8.07029
Sep-18	8.0613	N/A	8.0613
Oct-18	8.05632	N/A	8.05632

Note: the top panel reports an extract of the European real-time database for the Unemployment rate. The names of the columns represent the release dates of the vintages, the rows represent the reference period of each data point. Cells shaded in yellow represent the first release of each data point, circled cells represent the fourth revision of the data after the first release and cells shaded in purple represent the current release available at the date the database was accessed (31/01/2019). The same realisation can be used in different series, i.e. the data for May 2018 at the current release is used also for the series of four releases after the first and its cell is both circled and shaded in blue. The bottom panel reports the examples of series obtained from the extract of the real-time database. "N/A" indicates that the specific observation is not available for a vintage series.

Figure 1.4 shows the effect of revision in realised European data for the variables in the ECB SPF. For instance, the HICP annual growth rate for December 2002 was initially released on the 15/01/2003 at 2.2 and after revisions, it has been amended and kept to 2.3 until the 15/01/2018. The November 2002 unemployment first release was 8.41 on the 15/01/2003 and the last release available for the same month in my dataset is 8.86. For real GDP growth, the first release of the third quarter of 2002 was 0.83, this figure has been changed and my latest release is 1.24. The effect of revision is quite important in unemployment and real GDP growth but minor in inflation confirming the findings of Giannone, Henry, Lalik and Modugno (2012). The revision patter is similar for US data: there is small or no revision for inflation, smaller revision than in Europe for unemployment and bigger revision than in Europe for real GDP growth.

Figure 1.4: Revision in historical European data

(a) HICP Inflation (b) Unemployment Rate (c) Real GDP Growth



Note: the three plots show the effect of revision in the realised series, from January 2003 to February 2018, of HICP inflation (a), unemployment rate (b) and real GDP growth (c) from the first release to the latest available vintage on the 30/01/2019. The horizontal axis denotes the vintage of the data.

The vintage of the latest observation on which SPF participants base their forecasts is usually a first release. However, there are some exceptions, i.e. in the case of inflation, at the 2007.Q1 survey deadline, the latest realisation available to forecasters was January 2007 instead of December 2006. In my exercise, I always use the exact information forecasters had available at the survey deadline as my aim is to perform a fully real-time exercise.

In the case of unemployment, there is more revision and forecasters may not use newly available information because they deem it not reliable. In this case, I try both keeping and ignoring the additional information and results, which are available upon request, do not change. Surveys affected by this phenomenon are 2007.Q1 for which forecasters had one more realisation and 2004.Q2, 2008.Q4 and 2009.Q3 for which forecasters had one realisation less.

For real GDP growth, the latest information available is the one expected but it has generally already been revised once except in the case of survey 2002.Q2 and I always use the latest revision available at the survey deadline.

Chapter 2

The Rationality of the ECB SPF

ABSTRACT: I assess the rationality of the forecasts of the rate of inflation, real GDP growth and unemployment rate provided by the European Central Bank Survey of Professional Forecasters (ECB SPF) using tests based on Mean Squared Error loss function and on unknown loss functions homogeneous in the forecast errors. Fixed-m asymptotics critical values deliver correctly sized tests in small samples allowing me to test rationality of the ECB SPF. I find that ECB SPF forecasts are rational, in particular, inflation forecasts appear rational under Mean Squared Error symmetric loss despite the ECB inflation target being asymmetric possibly because of the dubious ECB's reputation across professional forecasters. In comparison, US SPF forecasts are rational under a loss function different from a Mean Squared Error one confirming existing literature's results.

Keywords: forecast evaluation, ECB SPF, rationality test, fixed-smoothing asymptotics

JEL Classification: C12, C22, E17

2.1 Introduction

Under the rational expectation hypothesis, agents are assumed to use all the available information when making their forecasts, see Muth (1961). As macroeconomic forecasts from trusted sources, such as professional forecasters, are influential in the economic environment (Carroll, 2003), it is of interest to test this assumption on surveys of professional forecasters. Early tests of this kind were predicated on forecasters minimising a Mean Squared Error (MSE) loss function. This assumption is convenient as properties of forecast errors under MSE are well known and easily testable: forecast errors should have zero mean and be uncorrelated with all the variables in the information set at the time the forecasts were made. Empirical studies, such as Fama (1975), Zarnowitz (1985) and Bonham and Cohen (1995), among others, all test for rationality under a MSE loss; other frameworks based on MSE loss include Rossi and Sekhposyan (2016) test, which is designed to detect rationality in sub-samples of the data. Pesaran and Weale (2006) provide a detailed survey of this strand of the literature and concluded that rationality is frequently rejected.

Over time, several studies, such as Granger and Newbold (1986), on the fact that forecasters and policy makers could have other loss functions apart from a MSE loss set in and the idea that forecasts may be rational under other types of loss functions, such as asymmetric quadratic loss, takes hold. Economic agents usually face different costs for overpredicting or underpredicting a target. For example, Elliott, Komunjer and Timmermann (2008) suggest that the cost of overpredicting output growth in the US is higher than the cost of underpredicting it, while for inflation they find the opposite happens. This indicates that agents are averse to negative outcomes, such as lower than expected output or higher than expected inflation and this is reflected in their forecasts. This is coherent to the fact that in reality, firms face different costs for reducing rather than increasing production, wages and prices. Financial firms, instead, bear different costs according to unexpectedly low or unexpectedly high inflation while they adapt their interest rates. Investigating the behaviour of policymakers, Aguiar and Martins (2008) notice that the ECB's definition of price stability is essentially an indication of asymmetric preferences because the central bank is aiming for a year-on-year increase of the Harmonised Index of Consumer Prices for the euro area below two percent. Nobay and Peel (2003) note that both the European Central Bank and the Bank of England may have asymmetric loss functions which suggests forecast errors with the same magnitude have different consequences according to their sign. Also, Capistrán (2008) after recovering the Federal Reserve Bank's loss function from its forecasts, speculates about it having asymmetric preferences. Ruge-Murcia (2000), instead, provides empirical evidence that the Canadian central bank may weight differently positive and negative deviations from the inflation target and, in a subsequent paper (Ruge-Murcia, 2003), they find evidence that asymmetric preferences of central banks are also present in Sweden and the UK. If policymakers are thought to potentially have asymmetric loss functions, it is highly likely that also private sector professional forecasters may base their predictions on similar loss functions. Evidence supporting asymmetric loss functions also come from the psychology literature as Weber (1994) maintains forecasters minimise an asymmetric loss when they are concerned about their reputation and can adjust future forecasts according to past forecast errors. Along the same lines, Ehrbeck and Waldmann (1996) and Laster, Bennett and Geoum (1999) claim subjects that want to signal their forecast ability tend to have an asymmetric loss function. Elliott, Komunjer and Timmermann (2008) provide empirical evidence that participants in the Federal Reserve of Philadelphia Survey of Professional Forecasters (US SPF) minimise an asymmetric loss function to produce forecasts of inflation and output growth. Surveys like this one do not specify the objective of the forecasting exercise and hence it is not clear that forecasters minimise a quadratic loss function and report the conditional mean as forecast. Moreover, forecasts should closely reflect the participants' underlying loss function because anonymous responses allow them to report the same predictions used internally or for their customers. However, participants are still interested in showing good forecast abilities and are able to adjust their forecasts in line with their previous forecast errors. Furthermore, consensus survey forecasts are the result of the aggregation of forecasts coming from different agents with possibly heterogeneous loss functions and this can exacerbate the problem of evaluating forecasts under a specific loss function. For these reasons, the ranking and evaluation of predictions is sensitive to the choice of the loss function and tests of rationality should allow for different and general types of loss function (Patton, 2019).

Elliott, Komunjer and Timmermann (2005, 2008) propose a GMM framework to test rationality under a loss function in which the asymmetry is driven by an estimated parameter. Their empirical exercise rarely rejects rationality of the International Monetary Fund and of the Organisation for Economic Co-operation and Development forecasts of budget deficit for G7 countries. Following the same approach, Wang and Lee (2014) cannot reject the null hypothesis of rationality for the US Survey of Professional Forecasters and the Greenbook forecasts. Ahn and Tsuchiya (2019) test rationality of forecasts in emerging markets and find that these are rational under asymmetric loss while they are not under symmetric loss. Other empirical studies include Auffhammer (2007) on energy consumption forecasts, Pierdzioch, Reid and Gupta (2016) about the South African inflation rate and Naghi (2015) testing for the rationality of US SPF forecasts.

Patton and Timmermann (2007) develop a more general approach in which the loss function has to be a function of only the forecast error or it has to be homogeneous in the forecast error while relaxing conditions on the DGP. Forecast rationality is tested assuming that the series of the indicator function of the sign of the forecast error is independent of all the variables in the forecasters' information set at the time the forecast was made. In this way, the loss function can be unknown and their assumptions are satisfied by many common loss function families such as Mean Squared Error (MSE), Mean Absolute Error, Lin-Lin and asymmetric quadratic losses but not Linex which is not homogeneous in the forecast error. They test Federal Reserve Greenbook forecasts of output growth over the period 1968.Q4 - 1999.Q4 and they reject rationality.

Rationality tests suffer from severe size distortion in small samples, for this reason, the existing empirical literature concentrates on testing the Survey of Professional Forecasters of the Federal Bank of Philadelphia (US SPF) for which the sample is substantial. I demonstrate with a Monte Carlo exercise that fixed-smoothing asymptotics and in particular fixed-m asymptotics eliminate size distortion in small samples, in line with results on Diebold and Mariano type tests by Harvey, Leybourne and Whitehouse (2017); Coroneo, Iacone and Profumo (2019); Coroneo and Iacone (2020). Moreover, Hualde and Iacone (2017) show that when the long run variance is estimated using a Weighted Periodogram Estimate, under fixed-m asymptotics, the distribution of a t-test is standard under the null hypothesis and critical values do not need to be simulated. This enables me to test the rationality of the European Central Bank Survey of Professional Forecasters (ECB SPF), which has been overlooked due to its small sample. As a comparison, I also test US SPF over the same period and split both samples in the middle of 2009 to check whether the 2008 financial crisis impacted forecasters' views and forecasting practices.

Results indicate that ECB SPF forecasts are overall rational. I can only reject rationality under MSE loss for unemployment in the first sub-sample and, despite the fact that the ECB inflation target is asymmetric, rationality for inflation forecasts is not rejected under MSE loss. This could be due to the fact that professional forecasters do not believe the ECB is able to keep inflation levels within the target and, as a consequence, they adopt a simple and symmetric loss function to form their forecasts.

US SPF forecasts also appear rational, however, rejections of the null hypothesis of rationality is more frequent under MSE loss. This is especially noticeable for civilian unemployment forecasts in the second sub-sample while it is not visible in the first sub-sample indicating that forecasters may have changed their loss function over time and review their forecasting practices after the 2008 financial crisis. The converse happens for real GDP growth forecasts while inflation forecasts are rational across sub-samples. Rationality is only rejected for nowcast forecasts using fixed-smoothing asymptotics confirming findings of the existing literature.

The rest of the Chapter is structured as follows. Section 2.2 gives an overview of all the tests I use, in Section 2.3 I present the Monte Carlo design and results. Section 2.4 is devoted to the empirical exercise and Section 2.5 concludes.

2.2 Forecast Rationality Tests and Fixed-smoothing Asymptotics

Forecasts are considered rational if, given a loss function, they fully capture all available information in the sense that a better forecast cannot be produced given the information available at the time.

An h steps ahead forecast f_{t+h} for the target variable y_{t+h} made at time $t = 1, \dots, T$ is a rational forecast if it satisfies

$$f^*_{t+h} = argmin_f E[L(f_{t+h}, y_{t+h})|Z_t], \tag{2.1}$$

where Z_t is the information set available to forecasters in t when the forecast was produced and $L(.)$ is the loss function which represents the cost associated with an incorrect forecast.

If the loss function is once differentiable with respect to the forecast, the first order condition for forecast rationality is

$$E[L'(f_{t+h}, y_{t+h})|Z_t] = 0. \tag{2.2}$$

Several tests are available depending on assumptions on the loss function made.

Most of the time, the loss function assumed is a Mean Squared Error loss (MSE), $L(f_{t+h}, y_{t+h}) = (y_{t+h} - f_{t+h})^2$, because of its mathematically appealing properties: the rational forecast is the conditional expectation of y_{t+h} and the forecast is always unbiased, the h-step-ahead forecast error has zero serial correlation beyond lag $h-1$ and the unconditional variance of the forecast is a non-decreasing function of the forecast horizon (Granger and Newbold, 1986; Diebold and Lopez, 1996).

In this case, defining the forecast error $e_{t+h} = y_{t+h} - f_{t+h}$, the first order condition in

(2.2) becomes $E[e_{t+h}|Z_t] = 0$ and the rationality property can be checked on forecast errors having zero mean and being not serially correlated after the $h-1$ lag. That is testing $H_0 : \beta = 0$ in

$$e_{t+h} = \beta' v_t + u_{t+h}, \tag{2.3}$$

where $v_t = g(Z_t)$ is a set of variables, including a constant, known to the forecaster at time t and so contained in the information set Z_t.

Setting $v_t = (1, f_{t+h})$ and adding f_{t+h} to both sides of the equation leads to the Mincer and Zarnowitz (1969) regression

$$y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}. \tag{2.4}$$

Here, unbiasedness and efficiency can be tested setting $H_0 : \beta_0 = 0$ and $H_0 : \beta_1 = 1$ respectively, while the null of forecast rationality is tested with the joint hypothesis $H_0 : \beta_0 = 0 \cap \beta_1 = 1$. The intuition behind this test comes from the covariance between the forecast error and the forecast

$$Cov(y_{t+h} - f_{t+h}, f_{t+h}) = Cov((\beta_1 - 1)f_{t+h} + u_{t+h}, f_{t+h}). \tag{2.5}$$

Except for $\beta_1 = 1$, the forecast error and the forecast are related and this correlation can be used to make better forecasts; when $\beta_1 = 1$, the forecast error is biased unless $\beta_0 = 0$.

Agents usually adopt symmetric loss functions, such as MSE loss, when their cost of over predicting and under predicting the target variable is the same. However, this cost and the agent that made the predictions are usually unknown during ex-post evaluation and so are their loss functions. This is especially true of forecasts that are the result of an aggregation of several forecasts coming from different agents such as consensus survey forecasts.

Patton and Timmermann (2007) suggest a test for rationality that is valid when the loss function is essentially unknown. Their test is based on the assumptions that the DGP has dynamics in the conditional mean and variance and the loss function is a homogeneous function in the forecast error. This latter assumption of homogeneity covers a wide range of functions but it could be eventually relaxed at the cost of introducing conditional homoscedasticity in the DGP allowing for great flexibility of the test. This test is constructed on the regression of the indicator function $I_{t+h} \equiv \mathbb{1}_{t+h}(e_{t+h} \leq 0)$ on elements v_t of the information set Z_t. Patton and Timmermann (2007) show that under forecast rationality, the series of I_{t+h} is independent of the variables in the forecaster's information set. The test statistics is based on the regression

$$I_{t+h} = \alpha + \beta' v_t + u_{t+h}, \tag{2.6}$$

testing $H_0 : \beta = 0$. Commonly used regressors include the forecast tested, the lagged indicator function and the realisation of the target variable available to forecasters.

Tests of multiple linear restrictions are implemented using a Wald-type test statistics

$$W = \frac{(R\hat{\beta} - c)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - c)}{\sigma^2} \rightarrow_d \chi^2_k, \tag{2.7}$$

where R is a $k \times l$ matrix of known constants with full rank (no redundant restrictions), k is the number of restrictions, $\hat{\beta}$ is the l dimensional vector of parameter estimates, X is the $l \times l$ matrix of regressors, c is a $k \times 1$ vector of known constants and σ^2 is the unknown long run variance. Instead, tests of a single linear restriction are implemented using a t-test

$$t = \sqrt{T}\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2}} \rightarrow_d N(0, 1). \tag{2.8}$$

Both (2.7) and (2.8) are infeasible because σ^2 is unknown and it should be replaced by an estimate. Priestley (1981) suggests several semiparametric techniques for estimation based on a weighted sum of covariances or periodograms. If the estimate is consistent,

$\hat{\sigma} - \sigma = o_p(1)$, the feasible test statistic retains the same asymptotic distribution.

One of the possible approaches to estimate the long run variance σ^2 and calculate the test statistics is to use a Weighted Covariance Estimate (WCE)

$$\widehat{\sigma}^2_{WCE} = \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{M(T)}\right)\widehat{\gamma}_j = \widehat{\gamma}_0 + 2\sum_{j=1}^{T-1} k\left(\frac{j}{M(T)}\right)\widehat{\gamma}_j, \tag{2.9}$$

where $\hat{\gamma}_j = \frac{1}{T}\sum_{t=1}^{T-j}(\hat{u}_t - \bar{u})(\hat{u}_{t+j} - \bar{u})$ is the sample autocovariance, $\hat{u}_t$ is the residual from one of the regressions before, $\bar{u} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t$ and $M(T) \in [1, T]$ is the bandwidth or lag truncation, which increases with T but at a slower rate. As Andrews (1991) suggests, for consistency of the variance estimator, regularity conditions include $M(T) \to \infty$ and $M(T)/T \to 0$ as $T \to \infty$. $k(.)$ is the weighting scheme or kernel function; commonly used kernel functions include the rectangular kernel

$$k_{RECT}\left(\frac{j}{M(T)}\right) = \begin{cases} 1 & \text{if } j \leq M(T); \\ 0 & \text{otherwise,} \end{cases} \tag{2.10}$$

which gives the following WCE estimator

$$\widehat{\sigma}^2_{WCE-R} = \widehat{\gamma}_0 + 2\sum_{j=1}^{M(T)} \widehat{\gamma}_j. \tag{2.11}$$

This estimator may generate zero or negative estimates in finite samples which is not desirable for an estimate of a variance.

Newey and West (1987) use the Bartlett kernel (Bartlett, 1950),

$$k_{BART}\left(\frac{j}{M(T)}\right) = \begin{cases} 1 - \left|\frac{j}{M(T)}\right| & \text{if } \left|\frac{j}{M(T)}\right| \leq 1; \\ 0 & \text{otherwise,} \end{cases} \tag{2.12}$$

in their WCE estimators to get robust standard errors and the estimator becomes

$$\widehat{\sigma}^2_{WCE-B} = \widehat{\gamma}_0 + 2\sum_{j=1}^{M(T)} \left(\frac{M(T) - j}{M(T)} \right) \widehat{\gamma}_j. \tag{2.13}$$

Another method to obtain long run variance estimators is available in the frequency domain using a Weighted Periodogram Estimate (WPE). The long run variance is a function of the spectral density at frequency 0, that is $\sigma^2 = 2\pi f(0)$ (Müller, 2014), and an estimate of it can be obtained as a weighted average of periodograms:

$$\widehat{\sigma}^2_{WPE} = 2\pi \sum_{j=1}^{T/2} K_M(\lambda_j) I(\lambda_j), \tag{2.14}$$

where $K_M(\lambda_j)$ is a symmetric kernel (weighting) function, $I(\lambda_j) = |\frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \hat{u}_t e^{i\lambda_j t}|^2$ is the periodogram of the residuals evaluated at the Fourier frequencies $\lambda_j = \frac{2\pi j}{T}$ for $j = 1, \ldots, T/2$ and i is the imaginary unit. Using the Daniell kernel (Daniell, 1946)

$$K_M^D(j) = \begin{cases} m^{-1} & \text{if } j \leq m; \\ 0 & \text{otherwise,} \end{cases}$$

where m is a user chosen parameter that is a function of the bandwidth $M(T)$, the estimator becomes a weighted sum of periodograms evaluated at the first m Fourier frequencies

$$\widehat{\sigma}^2_{WPE-D} = \frac{2\pi}{m} \sum_{j=1}^{m} I(\lambda_j). \tag{2.15}$$

Similarly to the WCE case, when $m \to \infty$ and $m/T \to 0$ as $T \to \infty$, the estimate is consistent (Hannan, 1970; Koopmans, 1995).

Assumptions on bandwidth parameters $M(T)$ and m affect the asymptotic distribution of test statistics. Standard asymptotics are based on the idea that the bandwidth parameter grows with the sample size but at a slower rate to get $M(T)/T \to 0$ and $m/T \to 0$. Neave (1970) was the first to argue that, in finite samples, T is fixed and the ratio of the bandwidth parameter and the sample size may not be negligible, they argue that considering M/T fixed provides a better approximation of reality. This approach is referred to as fixed-smoothing asymptotics in contrast to the classic approach of standard asymptotics.

Kiefer and Vogelsang (2002a,b, 2005) implement this approach for hypothesis testing assuming $M(T)/T \to b \in (0,1]$ when $T \to \infty$ and find that the non standard limiting distribution depends on both b and the choice of the kernel function. This approach is referred to as 'fixed-b' asymptotics. Jansson (2004); Sun, Phillips and Jin (2008); Zhang and Shao (2013) and Sun (2014) showed that fixed-b asymptotics provide a higher order approximation over the traditional small-b standard asymptotics, in particular Sun (2014) develops an asymptotics expansion of their limit distribution under small-b and fixed-b asymptotics and finds that the latter captures some terms in the high order expansion of the small-b asymptotics. Therefore, they show that critical values from fixed-b asymptotics are second order correct under the small-b asymptotics.

In the same way, fixed-smoothing asymptotics in frequency domain are obtained under the assumption that m is kept fixed. Results from Hualde and Iacone (2017) and others show that in this case the limiting distribution is standard and critical values do not need to be simulated providing a more convenient way to perform inference with respect to the fixed-b case. In this case, the long run variance estimator in (2.15) is not consistent but it is asymptotically unbiased.

For fixed-m, under regularity conditions, such as strict stationarity of $\hat{u}_t$, existence of the second moment and $0 < \sum_{j=-\infty}^{\infty} cov(\hat{u}_t, \hat{u}_{t+j}) < \infty$, the test statistics in (2.8), with a WPE-D estimator for the long run variance, has a t_{2m} asymptotic distribution (Brillinger, 1975). This result allows not to simulate critical values for the Patton and Timmermann (2007) rationality test, when only one regressor is used. Sun (2013), Müller (2014), Hualde and Iacone (2017) and others obtain the same asymptotic distribution under different assumptions ensuring the suitability of this approach for a wide range of settings in forecast evaluation. In particular, Coroneo and Iacone (2020) use a Functional Central Limit Theorem (FCLT) under the same framework of Giacomini and White (2006) which satisfy the conditions of Corollary 3.1 of Wooldridge and White (1988) FCLT and Central Limit Theorem for mixing processes which is appropriate for regression residuals as in forecast rationality tests.

Critical values for Wald-type tests (the Mincer and Zarnowitz (1969) test, the test based on forecast errors in (2.3) and the Patton and Timmermann (2007) when more than one regressor is used) under fixed-b asymptotics are not available. Fixed-m critical values can be obtained using results previously mentioned for t-tests and the continous mapping theorem from a $k \times F_{k,2m}$ distribution, where k is the number of linear restrictions, see Appendix 2.B for a detailed derivation.

2.3 Monte Carlo Study for Size and Power

It is well established in the literature that fixed-smoothing asymptotics provide a better approximation of the empirical size of tests. Coroneo, Iacone and Profumo (2019) and Coroneo and Iacone (2020) provide results on the Diebold and Mariano (1995) test in small samples while Harvey, Leybourne and Whitehouse (2017) also use fixed-smoothing asymptotics in forecast encompassing tests. A thorough study on this topic is available in Lazarus, Lewis, Stock and Watson (2018) however, to the best of my knowledge, there is no evidence in the literature that size improvements are confirmed in rationality tests. Thus, in this Section, I perform a Monte Carlo study of the size and power of rationality tests in small samples to check that size improvements are still present for several kind of rationality tests that I use in my empirical application. The setting of this experiment is similar to the one in Patton and Timmermann (2007) but it uses smaller sample sizes, more Monte Carlo replications and a AR(2)-GARCH(1,1) model for the Data Generating Process (DGP) instead of a AR(1)-GARCH(1,1). The DGP presented in this Section allows to study the behaviour of augmented versions of rationality tests that include the past forecast and the latest observation of the target variable, which are both available to SPF respondents in reality, and avoid multicollinearity under the null hypotesis. See Appendix 2.C for a detailed description of the multicollinearity issue and Monte Carlo results with fixed-smoothing asymptotics of the original Patton and Timmermann (2007) setting.

The Data Generating Process follows a AR(2)-GARCH(1,1)

$$\begin{aligned} y_t &= 0.5y_{t-1} - 0.2y_{t-2} + \sigma_t\epsilon_t, \\ \sigma_t^2 &= 0.1 + 0.8\sigma_{t-1}^2 + 0.1\sigma_{t-1}^2\epsilon_{t-1}^2, \\ \epsilon_t &\sim iidN(0,1), \\ \sigma_1^2 &= 1; \\ \sigma_2^2 &= 0.1 + 0.8 + 0.1\epsilon_1^2; \\ y_1 &= \epsilon_1; \\ y_2 &= \sigma_2\epsilon_2. \end{aligned} \tag{2.16}$$

Recalling that a rational forecast fully captures all available information at the time it was made and following Proposition 2 of Patton and Timmermann (2007), the rational one step ahead forecast for a general loss function is

$$f_{t+1}^* = \mu_{t+1} + \gamma\sigma_{t+1}, \tag{2.17}$$

where $\mu_{t+1} = E_t[y_{t+1}]$, $\sigma_{t+1}^2 = V_t[y_{t+1}]$ and γ is a constant that depends only on the loss function and the conditional distribution of the target variable. It measures the deviation from the conditional expectation μ, which is the rational forecast under MSE loss. As long as γ is set different from zero, the rational forecast is coming from an agent with a loss function different from a MSE loss and its value is related to the asymmetry parameter of the loss function selected. Patton and Timmermann (2007) show the relation between the loss function and the value of γ in the case of a general loss which is function of the forecast error in the proof of their Proposition 2. Also, Christoffersen and Diebold (1997) provide an analytic solution for γ under a Linex and Lin-Lin loss functions obtaining

$\gamma = \dfrac{a}{2}$ and $\gamma = F_t^{-1}\left(\dfrac{a}{a+b}\right)$ respectively, where a and b are asymmetry parameters of the loss functions and $F_t^{-1}(.)$ is a generic conditional inverse cumulative distribution function. In this Monte Carlo exercise, $\gamma = 0$ under MSE loss and $\gamma = 0.25$ under the asymmetric quadratic loss function with $a = 1.84$ as in Patton and Timmermann (2007), see Appendix 2.A for a detailed derivation of the relation between a and γ for this specific asymmetric loss function. A description of several asymmetric loss functions is available in the following Chapter of this thesis.

Non-rational forecasts are obtained by adding independent noise to the rational forecasts

$$\begin{aligned} f_{t+1} =& f^*_{t+1} + \xi\varepsilon_{t+1}, \\ \varepsilon_{t+1} \sim & \, iidN(0,1), \end{aligned} \tag{2.18}$$

where ξ is the standard deviation of the noise. When ξ is equal to zero, forecasts are rational and the null hypothesis is true. This is used to inspect the finite-sample size of tests.

I use $10,000$ Monte Carlo repetitions and the sample size is set at $T = \{30, 60\}$ to replicate the sample sizes available for the empirical exercise. Long run variance estimators are obtained using Weighted Periodogram Estimates with Daniell kernel and bandwidths $m = \lfloor T^{1/3} \rfloor$ and $m = \lfloor T^{1/4} \rfloor$ for the fixed-smoothing asymptotics case. This bandwidth choice was driven by the existing literature, in particular results from Coroneo and Iacone (2020) suggest using $m = \lfloor T^{1/3} \rfloor$ and Hualde and Iacone (2015b) show that the best size performances are obtained using small bandwidths such as $m = \lfloor T^{1/4} \rfloor$. Lazarus, Lewis, Stock and Watson (2018) suggest using $m = \lfloor 0.2 \times T^{2/3} \rfloor$ however, for samples as small as the ones I use in this work, bandwidths obtained with their rule are very close to those I use, i.e. when $T = 60$, $m = 3$ with both methods. For the standard asymptotics case, Patton and Timmermann (2007) do not mention any bandwidth or rule and so I assume that estimates are obtained using a Weighted Covariace Estimate with Bartlett kernel and the 'textbook' bandwidth $M = \lfloor 0.75 \times T^{1/3} \rfloor$ as suggested in Andrews (1991). This is also in line with Newey and West (1994) which advise to use bandwidths proportional to $T^{1/3}$.

The study is performed for five different tests and results are reported in Table 2.1 under standard asymptotics and in Tables 2.2 and 2.3 using fixed-m asymptotics. The 'MZ' test is based on the Mincer and Zarnowitz (1969) regression as in (2.4) testing $H_0 : \beta_0 = 0 \cap \beta_1 = 1$ which is equivalent to testing $H_0 : \beta_0 = \beta_1 = 0$ in $e_{t+1} = \beta_o + \beta_1 f_{t+1} + u_{t+1}$. The 'EA' test is the augmented version of the MZ test and it is based on the regression

$$e_{t+1} = \beta_o + \beta_1 f_{t+1} + \beta_2 e_t + u_{t+1}, \tag{2.19}$$

testing $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$ as suggested in Diebold and Lopez (1996).

The test 'EA2' is another augmented version of the previous one testing $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$ on the regression

$$e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}. \tag{2.20}$$

The last three tests are different versions of the original Patton and Timmermann (2007): 'PT' is the test based on the regression

$$I_{t+1} = \alpha + \beta_1 f_{t+1} + u_{t+1}, \tag{2.21}$$

testing $H_0 : \beta_1 = 0$, 'PT2' is the test based on the regression

$$I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+1}, \tag{2.22}$$

testing $H_0 : \beta_1 = \beta_2 = 0$ and 'PT2A' is based on the regression

$$I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}, \tag{2.23}$$

testing $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

The first three tests presented (MZ, E and EA2) are designed to detect rationality under MSE loss while the last three tests (PT, PT2 and PT2A) allow for the loss function to be unknown as long as it is a function of the forecast error.

Table 2.1: Empirical size and power of rationality tests under standard asymptotics

	MSE loss						**Asymmetric loss**					
ξ	MZ	EA	EA2	PT	PT2	PT2A	MZ	EA	EA2	PT	PT2	PT2A
						T = 30						
0	0.083	0.098	0.125	0.075	0.093	0.119	0.272	0.255	0.276	0.074	0.089	0.118
0.25	0.127	0.222	0.229	0.123	0.136	0.167	0.296	0.377	0.379	0.122	0.138	0.166
0.5	0.416	0.570	0.589	0.370	0.383	0.427	0.540	0.666	0.674	0.365	0.382	0.423
0.75	0.763	0.844	0.867	0.674	0.666	0.702	0.814	0.881	0.897	0.671	0.662	0.696
1	0.927	0.956	0.964	0.866	0.849	0.873	0.943	0.966	0.972	0.864	0.844	0.865
1.25	0.981	0.988	0.992	0.948	0.938	0.946	0.984	0.990	0.993	0.947	0.937	0.945
1.5	0.994	0.997	0.998	0.980	0.974	0.979	0.995	0.997	0.998	0.981	0.975	0.979
1.75	0.999	0.999	1.000	0.993	0.990	0.992	0.999	0.999	1.000	0.993	0.991	0.992
2	1.000	1.000	1.000	0.997	0.997	0.997	1.000	1.000	1.000	0.997	0.996	0.997
						T = 60						
0	0.074	0.080	0.096	0.063	0.070	0.079	0.433	0.392	0.378	0.064	0.065	0.074
0.25	0.141	0.300	0.294	0.133	0.151	0.184	0.467	0.587	0.560	0.133	0.147	0.178
0.5	0.607	0.816	0.826	0.545	0.578	0.643	0.776	0.902	0.901	0.535	0.566	0.630
0.75	0.934	0.980	0.984	0.889	0.899	0.926	0.962	0.989	0.991	0.886	0.890	0.918
1	0.994	0.998	0.999	0.984	0.983	0.989	0.997	0.999	1.000	0.982	0.983	0.989
1.25	0.999	1.000	1.000	0.997	0.997	0.999	1.000	1.000	1.000	0.998	0.998	0.999
1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: this Table reports finite-sample size ($\xi = 0$) and power ($\xi > 0$) for rationality tests in small samples ($T = 30, 60$) under MSE loss and asymmetric loss with standard asymptotics. The theoretical size is 5%. 'MZ' is the test based on the regression $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; 'EA' is the test based on the regression $e_{t+1} = \beta_o + \beta_1 f_{t+1} + \beta_2 e_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; 'EA2' is the test based on the regression $e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; 'PT' is the test based on $I_{t+1} = \alpha + \beta_1 f_{t+1} + u_{t+1}$, $H_0 : \beta_1 = 0$; 'PT2' is the test based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = 0$ and 'PT2A' is based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Long run variance estimators are obtained using WCE with Bartlett kernel, bandwidth $M = \lfloor 0.75 \times T^{1/3} \rfloor$.

Table 2.1 reports finite-sample size ($\xi = 0$) and power ($\xi > 0$) when standard asymptotics are used. All tests, except EA2, are clearly oversized especially for a sample as small as $T = 30$ but also for the slightly bigger sample of 60 observations, results are still not satisfactory. EA2 under MSE loss is, instead, undersized.

MZ, EA and EA2 tests are constructed to detect rationality under MSE loss where the

Table 2.2: Empirical size and power of rationality tests under Mean Square Error loss and fixed-smoothing asymptotics

	$\mathbf{m} = \lfloor \mathbf{T^{1/3}} \rfloor$						$\mathbf{m} = \lfloor \mathbf{T^{1/4}} \rfloor$					
ξ	MZ	EA	EA2	PT	PT2	PT2A	MZ	EA	EA2	PT	PT2	PT2A
						T = 30						
0	0.037	0.036	0.019	0.048	0.047	0.037	0.039	0.037	0.021	0.048	0.046	0.034
0.25	0.057	0.091	0.071	0.072	0.065	0.069	0.057	0.085	0.067	0.070	0.063	0.067
0.5	0.196	0.273	0.283	0.227	0.198	0.219	0.160	0.217	0.221	0.203	0.168	0.184
0.75	0.466	0.534	0.623	0.474	0.406	0.444	0.376	0.417	0.495	0.407	0.329	0.351
1	0.720	0.750	0.852	0.695	0.608	0.651	0.617	0.623	0.747	0.611	0.496	0.517
1.25	0.875	0.884	0.951	0.835	0.760	0.785	0.783	0.778	0.897	0.759	0.634	0.651
1.5	0.949	0.951	0.983	0.911	0.852	0.871	0.895	0.879	0.963	0.853	0.734	0.744
1.75	0.980	0.980	0.994	0.954	0.914	0.921	0.949	0.936	0.986	0.911	0.810	0.808
2	0.992	0.991	0.999	0.977	0.947	0.948	0.977	0.970	0.994	0.943	0.859	0.851
2.25	0.997	0.997	1.000	0.989	0.968	0.967	0.989	0.984	0.998	0.965	0.896	0.885
2.5	0.998	0.998	1.000	0.994	0.979	0.976	0.995	0.992	0.999	0.977	0.920	0.905
2.75	0.999	0.999	1.000	0.997	0.987	0.983	0.997	0.996	1.000	0.985	0.939	0.922
3	1.000	1.000	1.000	0.998	0.990	0.987	0.999	0.997	1.000	0.990	0.953	0.932
3.25	1.000	1.000	1.000	0.999	0.994	0.991	0.999	0.999	1.000	0.994	0.963	0.939
3.5	1.000	1.000	1.000	1.000	0.995	0.993	1.000	0.999	1.000	0.995	0.970	0.945
3.75	1.000	1.000	1.000	1.000	0.997	0.994	1.000	1.000	1.000	0.996	0.974	0.949
4	1.000	1.000	1.000	1.000	0.998	0.995	1.000	1.000	1.000	0.997	0.978	0.954
						T = 60						
0	0.048	0.047	0.026	0.047	0.048	0.039	0.051	0.049	0.035	0.051	0.049	0.043
0.25	0.083	0.167	0.157	0.092	0.096	0.120	0.076	0.139	0.133	0.092	0.092	0.112
0.5	0.347	0.533	0.616	0.377	0.366	0.450	0.282	0.416	0.477	0.324	0.304	0.356
0.75	0.731	0.837	0.928	0.730	0.707	0.778	0.619	0.718	0.833	0.650	0.590	0.645
1	0.927	0.960	0.990	0.920	0.893	0.929	0.858	0.889	0.965	0.860	0.793	0.824
1.25	0.985	0.991	0.999	0.979	0.965	0.978	0.954	0.959	0.992	0.950	0.900	0.913
1.5	0.997	0.998	1.000	0.994	0.988	0.993	0.987	0.987	0.999	0.982	0.957	0.960
1.75	1.000	1.000	1.000	0.999	0.997	0.997	0.997	0.996	1.000	0.995	0.979	0.977
2	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	1.000	0.998	0.989	0.988
2.25	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	0.999	0.995	0.992
2.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.995
2.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998
3.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
3.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
3.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999

Note: this Table reports finite-sample size ($\xi = 0$) and power ($\xi > 0$) for rationality tests in small samples ($T = 30, 60$) under MSE loss with fixed-smoothing asymptotics. The theoretical size is 5%. 'MZ' is the test based on the regression $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; 'EA' is the test based on the regression $e_{t+1} = \beta_o + \beta_1 f_{t+1} + \beta_2 e_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; 'EA2' is the test based on the regression $e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; 'PT' is the test based on $I_{t+1} = \alpha + \beta_1 f_{t+1} + u_{t+1}$, $H_0 : \beta_1 = 0$; 'PT2' is the test based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = 0$ and 'PT2A' is based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Long run variance estimators are obtained using WPE with Daniell kernel, bandwidths $m = \lfloor T^{1/3} \rfloor$ and $m = \lfloor T^{1/4} \rfloor$.

Table 2.3: Empirical size and power of rationality tests under asymmetric loss and fixed-smoothing asymptotics

	$\mathbf{m = \lfloor T^{1/3} \rfloor}$						$\mathbf{m = \lfloor T^{1/4} \rfloor}$					
ξ	MZ	EA	EA2	PT	PT2	PT2A	MZ	EA	EA2	PT	PT2	PT2A
						$\mathbf{T = 30}$						
0	0.128	0.102	0.052	0.047	0.042	0.035	0.108	0.085	0.044	0.047	0.043	0.033
0.25	0.132	0.159	0.127	0.073	0.064	0.067	0.112	0.138	0.111	0.067	0.062	0.063
0.5	0.267	0.343	0.355	0.222	0.196	0.222	0.213	0.266	0.273	0.195	0.164	0.186
0.75	0.521	0.581	0.671	0.474	0.403	0.438	0.420	0.460	0.540	0.403	0.320	0.342
1	0.748	0.777	0.871	0.687	0.599	0.627	0.644	0.647	0.771	0.604	0.488	0.508
1.25	0.887	0.896	0.959	0.831	0.753	0.777	0.799	0.794	0.908	0.758	0.628	0.641
1.5	0.955	0.954	0.985	0.911	0.850	0.865	0.902	0.887	0.964	0.849	0.729	0.734
1.75	0.982	0.982	0.996	0.955	0.907	0.915	0.954	0.941	0.988	0.908	0.804	0.802
2	0.993	0.992	0.999	0.977	0.943	0.946	0.979	0.971	0.995	0.940	0.856	0.847
2.25	0.997	0.997	1.000	0.987	0.964	0.966	0.990	0.985	0.998	0.962	0.896	0.880
2.5	0.999	0.998	1.000	0.994	0.978	0.975	0.995	0.992	1.000	0.976	0.922	0.901
2.75	1.000	0.999	1.000	0.996	0.985	0.983	0.997	0.996	1.000	0.984	0.937	0.915
3	1.000	1.000	1.000	0.998	0.990	0.988	0.999	0.998	1.000	0.988	0.951	0.930
3.25	1.000	1.000	1.000	0.999	0.994	0.991	0.999	0.999	1.000	0.992	0.960	0.939
3.5	1.000	1.000	1.000	0.999	0.996	0.992	1.000	0.999	1.000	0.994	0.968	0.943
3.75	1.000	1.000	1.000	1.000	0.997	0.993	1.000	1.000	1.000	0.996	0.973	0.949
4	1.000	1.000	1.000	1.000	0.998	0.995	1.000	1.000	1.000	0.997	0.979	0.954
						$\mathbf{T = 60}$						
0	0.252	0.194	0.121	0.049	0.045	0.036	0.205	0.161	0.105	0.052	0.045	0.040
0.25	0.255	0.342	0.313	0.095	0.094	0.117	0.210	0.270	0.246	0.089	0.088	0.104
0.5	0.483	0.647	0.722	0.369	0.362	0.438	0.393	0.520	0.576	0.323	0.294	0.337
0.75	0.795	0.880	0.951	0.724	0.692	0.761	0.681	0.765	0.869	0.646	0.582	0.618
1	0.942	0.971	0.992	0.914	0.890	0.924	0.881	0.907	0.972	0.851	0.786	0.815
1.25	0.988	0.993	0.999	0.978	0.964	0.977	0.962	0.966	0.993	0.947	0.899	0.913
1.5	0.998	0.998	1.000	0.995	0.988	0.992	0.989	0.989	0.999	0.981	0.952	0.958
1.75	1.000	1.000	1.000	0.998	0.996	0.997	0.997	0.997	1.000	0.993	0.976	0.977
2	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999	1.000	0.998	0.988	0.988
2.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.994	0.993
2.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.995
2.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.996
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998
3.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998
3.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
3.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999

Note: this Table reports finite-sample size ($\xi = 0$) and power ($\xi > 0$) for rationality tests in small samples ($T = 30, 60$) under asymmetric loss with fixed-smoothing asymptotics. The theoretical size is 5%. 'MZ' is the test based on the regression $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; 'EA' is the test based on the regression $e_{t+1} = \beta_o + \beta_1 f_{t+1} + \beta_2 e_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; 'EA2' is the test based on the regression $e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; 'PT' is the test based on $I_{t+1} = \alpha + \beta_1 f_{t+1} + u_{t+1}$, $H_0 : \beta_1 = 0$; 'PT2' is the test based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = 0$ and 'PT2A' is based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Long run variance estimators are obtained using WPE with Daniell kernel, bandwidths $m = \lfloor T^{1/3} \rfloor$ and $m = \lfloor T^{1/4} \rfloor$.

rational forecast is $f^*_{t+1} = \mu_{t+1}$. When the Monte Carlo is set under the asymmetric loss function, rational forecasts are produced using $f^*_{t+1} = \mu_{t+1} + \gamma\sigma_{t+1}$ and this causes the null hypothesis not to be satisfied even when $\xi = 0$ for the mentioned tests. As a consequences, MZ, EA and EA2 appear poorly sized.

Table 2.2 reports empirical size and power under Mean Squared Error loss function comparing the two different WPE badwidth $m = \lfloor T^{1/3} \rfloor$ and $m = \lfloor T^{1/4} \rfloor$. All tests, for both sample sizes, are undersized but PT tests still show an empirical size closer to the theoretical size than MZ and EA tests. However, the two latter tests have slightly more power than PT tests. In terms of bandwidths, $m = \lfloor T^{1/3} \rfloor$ seems to be preferred as it leads to almost identical size performances but to higher power.

Table 2.3 reports results of the Monte Carlo under an asymmetric loss function with the usual WPE bandwidths. Tests MZ, EA and EA2 appear oversized but this result is not related to fixed-smoothig asymptotics and their ability to provide correctly sized tests. It shows that these tests are likely to reject rationality when forecast are made under a different loss function than a MSE loss because the null hypothesis is never satisfied even when $\xi = 0$ as discussed in Elliott, Komunjer and Timmermann (2005, 2008) and Patton and Timmermann (2007) among others. PT tests still look slightly undersized but all show an acceptable empirical size. In the sample of size $T = 30$, PT tests struggle to reach 1 indicating that in very small samples they might lack power.

Considering results on the whole, empirical size and power under fixed-m asymptotics in small samples are acceptable and enable reliable inference in comparison to standard asymptotics which always provide oversized tests. Tests based on MSE loss with fixed-m critical values show very poor size when the Monte Carlo is conducted under MSE loss, however, tests based on unknown loss are correctly sized under the same circumstances. This behaviour highlights that the test introduced by Patton and Timmermann (2007) for unknown loss should be preferred to those based on MSE loss not only for the fact that the loss function of forecasters may be unknown but also for their size and power performances in small samples when paired with fixed-smoothing critical values.

Lazarus, Lewis and Stock (2019) and Coroneo and Iacone (2020) highlight the presence of a size-power trade off when fixed-smoothing asymptotics are used. This feature is also visible in my results but the choice of the WPE bandwidth $m = \lfloor T^{1/3} \rfloor$ gives satisfactory results, providing correctly sized tests with adequate power.

2.4 Empirical Results

Surveys of professional forecasters are commonly used to test rationality because only professional forecasters have the appropriate incentives to reveal their true beliefs and are usually better informed than other economic agents such as households (Keane and Runkle, 1990). In this light, I test the rationality of forecasts taken from the European Central Bank Survey of Professional Forecasters (ECB SPF) from 2002.Q1 to 2016.Q4 for a total of 60 observations and compare the results to those obtained from the Survey of Professional Forecasters of the Federal Bank of Philadelphia (US SPF). Because the ECB SPF third special questionnaire conducted in 2018 suggests participants changed their forecasting practices after the 2008 financial crisis, I also perform the analysis splitting the data in two sub-samples between quarters 2009.Q2 and 2009.Q3, each sub-sample containing 30 observations. The choice of the start of the sample was made to avoid including early surveys when participants were not used to respond to the questionnaire and the ECB had just started collecting and aggregating professional forecasts while the end date was chosen to include the largest number of survey rounds possible allowing the realisations of the target variable to be sufficiently revised and amended. Data is taken from the Federal Bank of Philadelphia and the ECB Statistical Data Warehouse.

For this empirical exercise, I consider HICP inflation, unemployment rate and real GDP growth rolling horizon one-year and two-years ahead aggregate point forecasts from ECB SPF, and CPI inflation rate, civilian unemployment rate and real GDP growth aggregate point forecasts for all horizons available in the US SPF. Aggregate forecasts reported in the survey are obtained from the unweighted average of individual point forecasts. I concentrate on the evaluation of aggregate forecasts only because a very large literature maintains that the simple average of single point forecasts is an effective way of obtaining predictions more accurate than the single ones, see Clemen (1989); Aiolfi, Capistrán and Timmermann (2011); Manski (2011); Genre, Kenny, Meyler and Timmermann (2013) and Meyler (2020) among others.

The series of realisations of target variables are taken at the current release (available on 31/01/2019) except when the realisations are regressors. In this case, the appropriate vintage and realisation are selected according to the point in time forecasters had to produce each forecast such that all regressors are part of forecasters' information set at each survey round deadline. For example, in the ECB SPF 2002.Q1 survey, the latest information available to forecasters was the first release of the December 2001 inflation. However, following Clark and McCracken (2009), which show that revisions have usually no effect on asymptotic distribution of tests, I neglect its influence on critical values.

Rationality should be checked using information available in the moment the forecast was made and so the last forecast error/indicator function observable by forecasters is the one relative to the one-year ahead prediction that matches December 2001 inflation they made for the 2001.Q1 survey and not for the survey immediately preceding the one being tested (2001.Q4).

In line with Diebold and Rudebusch (1991) and Croushore and Stark (2001), I also repeat this exercise using the series of first-ever released actuals of the target variable notwithstanding using real-time data when realisations are taken as regressors as described earlier. Results of this exercise are similar to those obtained using the current release and are available in Appendix 2.D.

Forecast rationality is tested under MSE loss with Equations (2.4), (2.19) and (2.20) and under unknown loss function with (2.21), (2.22) and (2.23) described in Section 2.2. The long run variance of residuals is estimated using a Weighted Periodogram Estimator with Daniell kernel and, given the results obtained from the Monte Carlo exercise, the bandwidth adopted is $m = \lfloor T^{1/3} \rfloor$.

Under fixed-m asymptotics, limit distributions under the null hypothesis for every test statistic are derived according to Hualde and Iacone (2017) and are reported in Table 2.4. Wald type tests have a $k \times F_{k,2m}$ distribution and t test has distribution t_{2m} where k is the number of linear restrictions tested and m is the bandwidth parameter involved in the estimation of the long run variance.

Table 2.4: Asymptotic distributions under the null for rationality tests

Test	MZ	EA	EA2	PT	PT2	PT2A
Distribution	$2F_{2,2m}$	$3F_{3,2m}$	$4F_{4,2m}$	t_{2m}	$2F_{2,2m}$	$3F_{3,2m}$

Note: the Table shows asymptotics distribution for the rationality tests used in the empirical exercise derived following results from Hualde and Iacone (2017). m is the bandwidth used in the long run variance estimator calculated with Weighted Periodogram Estimator using the Daniell kernel. MZ test: $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; EA test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; EA2 test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; PT test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_1 = 0$; PT2 test: $I_{t+h} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = 0$; PT2A test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + \beta_2 I_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. h is the forecast horizon.

Test statistics values are reported in Tables 2.5 - 2.7. Rejections using fixed-m asymptotics critical values are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level. Rejections from standard asymptotics critical values are indicated shading the appropriate cell; and indicate, respectively, two-sided significance at the 5% and 10% level.

Table 2.5 shows test statistics values for the full sample of 60 observations from 2002.Q1 to 2016.Q4, h is the forecast horizon expressed in years for the ECB SPF and in quarters for the US SPF. ECB SPF forecasts appear to be rational. Especially in ECB SPF forecasts for HICP inflation , the null hypothesis of forecast rationality is never rejected also by Patton and Timmermann (2007) tests (PT, PT2 and PT2A) which allow for an unknown loss function. In addition, as tests based on a MSE loss (MZ, EA and EA2) do not reject the null, despite being oversized for asymmetric loss functions, it is highly likely that forecasters have a MSE loss function for this variable although the ECB has an asymmetric inflation target. Also in US SPF forecasts for CPI inflation, rationality it is not rejected under unknown loss except for the current quarter where tests based on MSE loss strongly reject rationality and also Patton and Timmermann (2007) test PT2A

Table 2.5: Rationality tests, full sample 2002.Q1 - 2016.Q4, T = 60

Horizon	MZ Test	EA Test	EA2 Test	PT Test	PT2 Test	PT2A Test
Years			**ECB SPF**			
			HICP			
1	0.24	0.36	0.95	-0.16	0.33	0.76
2	0.11	0.53	0.62	0.03	0.08	0.34
			Unemployment rate			
1	4.70	9.17	8.98	1.70	7.25*	7.36
2	4.43	4.70	6.09	1.46	2.77	3.34
			Real GDP growth			
1	1.64	2.24	7.11	-0.86	0.96	2.41
2	2.27	4.23	4.08	1.46	6.69	8.63
Quarters			**US SPF**			
			CPI			
0	85.14**	73.28**	89.43**	-4.40**	18.78**	22.03**
1	0.37	1.78	2.21	-0.48	0.99	1.26
2	0.88	0.99	1.37	-0.11	0.02	0.94
3	0.08	0.42	0.58	-0.56	1.37	1.46
4	0.39	1.16	2.13	-0.27	0.16	0.89
			Civilian unemployment rate			
0	2.05	4.33	87.89**	0.46	0.26	26.30**
1	0.55	5.78	49.13**	0.32	2.78	15.22**
2	0.20	3.84	34.06**	0.38	0.67	9.34
3	0.19	2.16	19.13*	0.48	2.98	14.33*
4	0.23	1.30	13.33	0.84	2.17	9.27
			Real GDP growth			
0	4.08	4.01	15.55*	0.56	0.56	2.90
1	12.80**	16.19*	27.91**	0.70	0.65	1.78
2	7.87 *	8.15	9.07	-0.01	0.10	0.73
3	4.99	5.00	5.41	0.24	0.06	0.48
4	4.67	5.10	5.19	-0.54	1.62	1.65

Note: the Table reports test statistics values for ECB SPF one-year and two-years ahead rolling horizon forecasts. For US SPF, it reports test statistics for current quarter, one quarter ahead, two quarters ahead, three quarters ahead and four quarters ahead forecasts. Results are obtained on full sample from 2002.Q1 to 2016.Q4 ($T = 60$) using the current release of the realised data. Rationality tests used are: MZ test: $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; EA test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; EA2 test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; PT test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_1 = 0$; PT2 test: $I_{t+h} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = 0$; PT2A test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + \beta_2 I_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. h is the forecast horizon expressed in years (quarters) for ECB SPF (US SPF). Long run variances are estimated using WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$, 5% fixed-$m$ critical values (asymptotics distributions) are 10.28 ($2F_{2,2m}$), 15.42 ($3F_{3,2m}$), 20.56 ($4F_{4,2m}$), 2.45 (t_{2m}), 10.28 ($2F_{2,2m}$), 14.28 ($3F_{3,2m}$) respectively for each column, 10 % fixed-m critical values are 6.92, 10.38, 13.84, 1.94, 6.92, 9.87; rejections are reported using ** and * to indicate, respectively, significance at the 5% and 10% level. 5% standard asymptotics critical values (asymptotics distributions) are 5.99 (χ^2_2), 7.81 (χ^2_3), 9.49 (χ^2_4), 1.96 ($N(0,1)$), 5.99 (χ^2_2), 7.81 (χ^2_3) respectively for each column, 10 % standard asymptotics critical values are 4.61, 6.25, 7.78,1.65, 4.61, 6.25; rejections are reported using and to indicate, respectively, significance at the 5% and 10% level.

Table 2.6: Rationality tests, first sample 2002.Q1 - 2009.Q2, T = 30

Horizon	MZ Test	EA Test	EA2 Test	PT Test	PT2 Test	PT2A Test
Years			**ECB SPF**			
			HICP			
1	1.17	8.99	13.24	1.43	2.30	3.58
2	1.88	1.90	2.04	3.11**	9.11*	9.16
			Unemployment rate			
1	6.00	94.38**	93.83**	0.16	3.31	3.46
2	9.06*	9.34	41.15**	0.92	1.22	3.12
			Real GDP growth			
1	2.95	3.15	5.12	-0.93	1.11	1.61
2	1.59	7.99	8.55	1.58	7.71*	25.98**
Quarters			**US SPF**			
			CPI			
0	112.40**	224.41**	158.87**	-3.45**	20.64**	23.00**
1	0.85	2.05	3.32	0.14	0.56	4.67
2	1.55	1.57	1.56	1.40	1.91	4.09
3	0.60	1.20	2.04	0.03	6.61	7.53
4	0.39	0.64	2.26	-0.36	0.48	1.38
			Civilian unemployment rate			
0	3.08	6.27	14.63*	-1.77*	2.89	8.03
1	2.39	5.54	10.01	-1.04	2.33	8.70
2	2.21	2.27	9.66	-1.32	2.16	5.17
3	1.74	1.74	6.87	-1.24	1.73	4.40
4	1.45	1.50	6.52	-0.36	0.32	4.47
			Real GDP growth			
0	2.47	3.38	20.88**	0.98	1.69	5.94
1	9.19*	20.33**	33.86**	0.52	12.40**	14.55**
2	7.45*	17.51**	17.00*	-0.11	1.18	3.83
3	3.22	3.55	4.22	0.41	0.20	0.20
4	4.09	7.31	10.05	-1.90	5.90	6.53

Note: the Table reports test statistics values for ECB SPF one-year and two-years ahead rolling horizon forecasts. For US SPF, it reports test statistics for current quarter, one quarter ahead, two quarters ahead, three quarters ahead and four quarters ahead forecasts. Results are obtained on first sub-sample from 2002.Q1 to 2009.Q2 ($T = 30$) using the current release of the realised data. Rationality tests used are: MZ test: $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; EA test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; EA2 test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; PT test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_1 = 0$; PT2 test: $I_{t+h} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = 0$; PT2A test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + \beta_2 I_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. h is the forecast horizon expressed in years (quarters) for ECB SPF (US SPF). Long run variances are estimated using WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$, 5% fixed-$m$ critical values (asymptotics distributions) are 10.28 ($2F_{2,2m}$), 15.42 ($3F_{3,2m}$), 20.56 ($4F_{4,2m}$), 2.45 (t_{2m}), 10.28 ($2F_{2,2m}$), 14.28 ($3F_{3,2m}$) respectively for each column, 10 % fixed-m critical values are 6.92, 10.38, 13.84, 1.94, 6.92, 9.87; rejections are reported using ** and * to indicate, respectively, significance at the 5% and 10% level. 5% standard asymptotics critical values (asymptotics distributions) are 5.99 (χ^2_2), 7.81 (χ^2_3), 9.49 (χ^2_4), 1.96 ($N(0,1)$), 5.99 (χ^2_2), 7.81 (χ^2_3) respectively for each column, 10 % standard asymptotics critical values are 4.61, 6.25, 7.78,1.65, 4.61, 6.25; rejections are reported using and to indicate, respectively, significance at the 5% and 10% level.

Table 2.7: Rationality tests, second sample 2009.Q3 - 2016.Q4, T = 30

Horizon	MZ Test	EA Test	EA2 Test	PT Test	PT2 Test	PT2A Test
Years			**ECB SPF**			
			HICP			
1	0.73	1.23	1.89	-0.46	1.03	1.74
2	1.22	2.70	2.71	0.26	1.45	1.83
			Unemployment rate			
1	1.00	1.74	2.47	0.88	1.93	2.84
2	1.58	2.75	2.75	1.04	3.43	3.62
			Real GDP growth			
1	0.63	5.27	13.42	-1.61	4.88	13.47*
2	0.78	1.60	1.60	0.42	2.33	2.43
Quarters			**US SPF**			
			CPI			
0	7.93*	7.46	8.14	-5.43**	22.85**	19.50**
1	1.94	1.85	2.16	0.88	1.67	3.33
2	3.28	3.37	3.97	0.58	0.55	0.68
3	2.17	2.94	4.77	1.11	1.27	1.64
4	1.25	1.33	2.16	2.35*	5.72	6.97
			Civilian unemployment rate			
0	35.03**	23.84**	33.80**	1.30	1.64	3.63
1	21.41**	18.14**	21.04**	0.98	2.22	3.73
2	19.82**	21.84**	21.82**	0.88	1.45	1.65
3	17.82**	19.87**	20.50*	1.95*	3.78	4.62
4	14.38**	17.34**	18.88*	1.96*	3.89	4.12
			Real GDP growth			
0	0.93	1.44	1.46	-0.90	0.80	0.99
1	2.78	4.39	5.80	1.34	5.71	5.71
2	3.72	5.31	6.19	-0.15	2.09	2.22
3	2.98	3.41	7.46	-0.33	0.11	2.13
4	4.87	5.94	8.57	1.12	1.53	1.54

Note: the Table reports test statistics values for ECB SPF one-year and two-years ahead rolling horizon forecasts. For US SPF, it reports test statistics for current quarter, one quarter ahead, two quarters ahead, three quarters ahead and four quarters ahead forecasts. Results are obtained on second sub-sample from 2009.Q3 to 2016.Q4 ($T = 30$) using the current release of the realised data. Rationality tests used are: MZ test: $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; EA test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; EA2 test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; PT test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_1 = 0$; PT2 test: $I_{t+h} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = 0$; PT2A test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + \beta_2 I_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. h is the forecast horizon expressed in years (quarters) for ECB SPF (US SPF). Long run variances are estimated using WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$, 5% fixed-$m$ critical values (asymptotics distributions) are 10.28 ($2F_{2,2m}$), 15.42 ($3F_{3,2m}$), 20.56 ($4F_{4,2m}$), 2.45 (t_{2m}), 10.28 ($2F_{2,2m}$), 14.28 ($3F_{3,2m}$) respectively for each column, 10 % fixed-m critical values are 6.92, 10.38, 13.84, 1.94, 6.92, 9.87; rejections are reported using ** and * to indicate, respectively, significance at the 5% and 10% level. 5% standard asymptotics critical values (asymptotics distributions) are 5.99 (χ^2_2), 7.81 (χ^2_3), 9.49 (χ^2_4), 1.96 ($N(0,1)$), 5.99 (χ^2_2), 7.81 (χ^2_3) respectively for each column, 10 % standard asymptotics critical values are 4.61, 6.25, 7.78,1.65, 4.61, 6.25; rejections are reported using and to indicate, respectively, significance at the 5% and 10% level.

rejects the null hypothesis of rationality.

Rationality tests on ECB SPF unemployment rate reject the null hypothesis for the one-year rolling horizon case only with standard asymptotics. Using fixed-m critical values, leads to rejection only for the PT2 test at 10% significance. This is due to the fact that these tests in small samples are oversized with standard asymptotics as shown in the Monte Carlo simulation of the previous Section. In US SPF, testing unemployment forecasts leads to a strong rejection both under MSE loss and unknown loss only in tests with the lagged target variable as additional regressor (EA2 and PT2A tests) for all forecast horizons even though these two tests seems undersized. This suggests that forecasters did not consider the previous unemployment rate when making their forecasts even though it was available to them maybe because they did not update their previously made forecasts once the information was made available.

Real GDP growth ECB SPF forecasts appear to be rational, rejection of rationality emerges only occasionally using standard asymptotics critical values. For US SPF, the null hypothesis of rationality can only be rejected under MSE loss function supporting the idea that assessing rationality of forecasts has to be carried out considering the possibility that the loss function of forecasters can be different from a MSE loss. Under unknown loss tests, the only rejection is obtained with the PT2A test two quarters ahead forecasts.

Focusing on the first sub-sample of 30 observations from 2002.Q1 to 2009.Q2, ECB SPF forecasts appear rational especially when the more general test of Patton and Timmermann (2007) is used corroborating the assumption that forecasters may not use a MSE loss function. The only exception is for real GDP growth rolling horizon two-years ahead forecasts as rejection of the null hypothesis with tests based on MSE loss is confirmed by test of rationality for unknown loss function (PT2 and PT2A tests) under standard asymptotics and some rejections become visible with fixed-m asymptotics too. In this occasion, and also for HICP rolling horizon two-years ahead forecasts, the particularly small sample, makes both MZ and EA tests largely undersized and with low power under fixed-smoothing asymptotics. This prevents MSE loss based tests to detect non rationality. Forecasts for the unemployment rate appear non-rational under a MSE loss function with strong rejections of the null hypothesis of rationality with the EA2 test but they are rational under a general loss function. This highlight the fact that forecasters may not always minimise a quadratic loss function when making forecasts.

In US SPF, the null hypothesis of forecast rationality cannot be rejected except for nowcast forecasts of CPI. It looks like forecasters did not take into account past information when making forecasts for horizons up to one quarter ahead. In the Civilian unemployment case, it is clear the benefit from fixed-m asymptotics because standard asymptotics lead to strong rejections for EA2 and PT2A tests while with fixed smoothing asymptotics rejections are milder or not present. For real GDP growth, it is clear that forecasters do not minimise a quadratic loss function. Forecasts are only rational when assessed with Patton and Timmermann (2007) tests.

In the second sub-sample, ECB SPF forecasts appear all rational as both tests based on MSE loss and unknown loss do not reject the null hypothesis of rationality. Only real GDP growth rolling horizon one-year ahead forecasts tested including the observed realisation of the target variable (colums EA2 and PT2A) leads to strong rejections of the null with standard asymptotics but only to mild rejections with fixed-smoothing asymptotics. In this case, forecasters may have not taken into account the latest realization available to them when making the forecast or they did not update a previous forecast as soon as new information became available and reported that. This results can also be due to the fact that these augmented tests are very oversized under standard asymptotics and slightly undersized under fixed-smoothing asymptotics.

In the US, all forecasts appear rational under unknown loss function. In particular, civilian unemployment shows strong rejections of the null hypothesis of rationality with tests for MSE loss function but there are no rejections with asymmetric loss. This case well represent the problem of testing for rationality under a specific loss function. Using such tests may lead to rejecting the null hypothesis of rationality and considering forecasts as not optimal while forecasters were only using a different loss function than the one the researcher assumed. CPI forecasts of the current quarter appear strongly non rational under a general loss function but evidence against rationality is milder under MSE loss. This result could be due to the poor power performances of tests based on a MSE loss function.

The contribution of fixed-smoothing asymptotics arises in several cases in which standard asymptotics critical values lead to rejection while fixed-smoothing asymptotics critical values do not, for instance in the one-year ahead ECB SPF full sample forecasts of unemployment rate and first sample of US SPF real GDP growth four quarters ahead forecasts. In other instances, using fixed-smoothing asymptotics critical values led to less strong rejections, such as for US SPF real GDP growth full sample forecasts.

These findings confirm and are coherent with Romer and Romer (2000), Sims (2002), Capistrán (2008) and Capistrán and Timmermann (2009) which ultimately provides evidence of asymmetric loss in professional and institutional forecasts of real output and inflation. Rossi and Sekhposyan (2016) and El-Shagi (2019) perform the EA test on sub-samples of US SPF and strongly reject the null of rationality for inflation. Patton and Timmermann (2012) use the Optimal Revision Regression test on Federal Reserve Greenbook forecasts for GDP, the GDP deflator and CPI and find even more evidence against rationality than the MZ tests. Also, Elliott, Komunjer and Timmermann (2008) and Wang and Lee (2014) get to the same conclusions exhibiting robustness to the vintage used for the realisation of the target variable.

Given the limited sample, there is not much literature about rationality of forecasts in Europe; to the best of my knowledge, there are no previous studies of rationality of the ECB SPF point forecasts. The only empirical analysis available on European data is the one by Ulu (2015) which examines Money Market Survey (MMS) of several countries and rejects rationality under asymmetric loss of the European MMS for the sample of inflation between 2000 and 2011. This result is confirmed by my study of the first subsample in which tests for rationality under unknown loss reject rationality for two-years ahead forecasts of inflation. In the full sample and in the second subsample, rationality of the

ECB SPF inflation forecasts is not rejected using tests based on a MSE loss function despite the asymmetric inflation target of the ECB therefore, professional forecasters might base their predictions on a MSE loss. This could be due to the low credibility of the ECB in maintaining the target inflation level. With this in mind, several authors highlight how the ECB has gradually lost credibility among institutions and the general public: Howarth and Loedel (2003) first claimed that credibility of the ECB has suffered since its inception; Fourçans and Vranceanu (2007) notice that credibility is hard to achieve if the central bank focuses on unemployment or real activity in the short term instead of inflation; Geraats (2008) observes how trust in the ECB to maintain inflation lower than 2% has fallen from 1999 to 2008 analysing SPF density forecasts. SPF respondents reported lower and lower probabilities that inflation could be below 2% in the medium term clearly signalling that professional forecasters do not believe in the ECB's ability to meet its target. Other authors, such as Weber and Forschner (2014), claim that the ECB is at risk of losing independence and this is not promoting credibility. Also Gros and Roth (2010) reports that citizens' trust in the ECB is decreasing from the 2008 financial crisis and more recently Bergbauer, Hernborg, Jamet, Persson and Schölermann (2020) notice that the ECB is facing diminishing public trust. In these conditions, professional forecasters may not be persuaded by the asymmetric ECB target and they just want to produce the best forecast they can irrespective that it is below or above 2%.

2.5 Conclusions

I test for rationality of the ECB SPF and US SPF aggregate point forecasts over the period 2002.Q1 - 2016.Q4 using tests based on a MSE loss, like the Mincer and Zarnowitz (1969) test, and tests that do not assume a particular loss function, such as the test of Patton and Timmermann (2007). To perform reliable inference in small samples like those available in the euro area, I use critical values from fixed-smoothing asymptotics.

Results indicate that forecasts are rational and, particularly for ECB SPF forecasts of unemployment rate and real GDP growth, the benefit coming from fixed-m asymptotics stands out as, while standard asymptotics lead to rejection of the null of rationality, fixed-smoothing asymptotics do not. Interestingly, rationality of European inflation forecasts is not rejected under MSE loss although the ECB has an asymmetric inflation target and it was plausible to assume that forecasters had an asymmetric loss as well. This could be due to the low level of credibility the ECB has among professional forecasters about maintaining inflation below the 2% target.

US professional forecasters seem to change their loss function over time as only in the second sub-sample rationality is rejected under MSE loss but not under unknown loss. All results are robust to the vintage of the realised series.

2.A Appendix: Relation Between the Rational Forecast and the Loss Function

This Appendix explores the relation between the asymmetry parameter a of the asymmetric quadratic loss function and γ in the rational forecast equation from the Monte Carlo set-up borrowed from Patton and Timmermann (2007).

Assuming that the DGP is of the form $y_{t+h} = \mu_{t+h} + \sigma_{t+h}\eta_{t+h}$, where h is the forecast horizon, $\mu_{t+h} = E_t[y_{t+h}]$, $\sigma^2_{t+h} = V_t[y_{t+h}]$ and $\eta_{t+h}|Z_t \sim N(0,1)$ in this particular setting. In fact, $\eta_{t+h}|Z_t$ could come from any distribution with mean 0 and unit variance that may depend on the forecast horizon but not on the information set.

The rational forecast takes the form $f^*_{t+h} = \mu_{t+h} + \gamma\sigma_{t+h}$, where γ is a constant that depends only on the distribution of η_{t+h} and the loss function.

The loss function of the Monte Carlo setting, in which subscript notation is suppressed, except where essential, for ease of exposition is

$$L(y,f,a) = \begin{cases} a(y-f)^2 & \text{if } y-f>0; \\ (y-f)^2 & \text{if } y-f \leq 0, \end{cases}$$

which can be expressed in terms of η_{t+h} and γ as

$$L(\eta,\gamma,a) = \begin{cases} a(\eta-\gamma)^2 & \text{if } \gamma<\eta<\infty; \\ (\eta-\gamma)^2 & \text{if } -\infty<\eta \leq \gamma. \end{cases}$$

Starting from the definition of rational forecast $f^*_{t+h} = argmin_\gamma E[L(\eta,\gamma,a)|Z_t]$, it is possible to recover the relation between γ and a:

$$f^*_{t+h} = argmin_\gamma \left[\int_{-\infty}^{\gamma} (\eta-\gamma)^2 \phi(\eta) d\eta + \int_{\gamma}^{\infty} a(\eta-\gamma)^2 \phi(\eta) d\eta \right], \qquad (2.24)$$

where $\phi(\eta)$ is the probability density function of a Standard Normal distribution, assuming conditional normality of y.

For the first integral, using Leibniz's rule, I get

$$\begin{aligned}\frac{\partial}{\partial\gamma}\int_{-\infty}^{\gamma}(\eta-\gamma)^2\phi(\eta)d\eta = 2\int_{-\infty}^{\gamma}(-1)(\eta-\gamma)^2\phi(\eta)d\eta = \\ 2\left[-\int_{-\infty}^{\gamma}\eta\phi(\eta)d\eta+\gamma\int_{-\infty}^{\gamma}\phi(\eta)d\eta\right] = 2\left[-\int_{-\infty}^{\gamma}\eta\frac{1}{\sqrt{2\pi}}e^{-1/2\eta^2}d\eta+\gamma\Phi(\gamma)\right] = \\ 2\left[\phi(\gamma)+\gamma\Phi(\gamma)\right],\end{aligned} \tag{2.25}$$

where $\Phi(\gamma)$ is the Cumulative Distribution Function of the Standard Normal distribution of η.

For the second integral, using Leibniz's rule, I get

$$\begin{aligned}\frac{\partial}{\partial\gamma}a\int_{\gamma}^{\infty}(\eta-\gamma)^2\phi(\eta)d\eta = 2a\int_{\gamma}^{\infty}(-1)(\eta-\gamma)\phi(\eta)d\eta = \\ 2a\left[-\int_{\gamma}^{\infty}\eta\phi(\eta)d\eta+\gamma\int_{\gamma}^{\infty}\phi(\eta)d\eta\right] = 2a\left[-\phi(\gamma)+\gamma(1-\Phi(\gamma))\right]\end{aligned} \tag{2.26}$$

The FOC becomes

$$0 = 2\left[\phi(\gamma)+\gamma\Phi(\gamma)\right]+2a\left[-\phi(\gamma)+\gamma(1-\Phi(\gamma))\right], \tag{2.27}$$

$$a(\gamma-\gamma\Phi(\gamma)-\phi(\gamma)) = -\gamma\Phi(\gamma)-\phi(\gamma), \tag{2.28}$$

to finally get

$$a = \frac{-\gamma\Phi(\gamma)-\phi(\gamma)}{\gamma-\gamma\Phi(\gamma)-\phi(\gamma)}. \tag{2.29}$$

Christoffersen and Diebold (1997) show the relation between γ and the loss function for a Lin-Lin loss function and a Linex loss function. Clements (2019) derives the rational forecast under Linex loss function and highlights that it depends linearly on the conditional variance while, under homogeneity, the rational forecast depends on the conditional standard deviation. This rules out the Linex loss from the pool of homogeneous loss functions.

2.B Appendix: Asymptotic Distributions of Test Statistics

This Appendix shows the derivation of test statistics asymptotic distributions under fixed-m.

Considering a stationary time series $d_1, \ldots, d_T$, inference on $\mu = E[d_t]$ can be conducted using the test statistic

$$\sqrt{T}\frac{\bar{d}-\mu}{\sqrt{\sigma^2}}. \tag{2.30}$$

where σ^2 is the unknown long run variance and $\bar{d} = \dfrac{1}{T}\sum_{t=1}^{T} d_t$ is the sample mean of d_t.

The unknown long run variance can be replaced with a Weighted Periodogram Estimator such as $\widehat{\sigma}^2_{WPE-D} = \dfrac{2\pi}{m}\sum_{j=1}^{m} I(\lambda_j)$, where $I(\lambda_j) = |\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T} d_t e^{i\lambda_j t}|^2$ is the periodogram of d_t evaluated at the Fourier frequencies $\lambda_j = \frac{2\pi j}{T}$ for $j = 1, \ldots, T/2$ and i is the imaginary unit.

The feasible test statistics under $H_0 : \mu = 0$ is

$$t = \sqrt{T}\frac{\bar{d}}{\sqrt{\hat{\sigma}^2_{WPE-D}}} = \frac{\sqrt{T}\dfrac{1}{T}\sum_{t=1}^{T} d_t}{\sqrt{\hat{\sigma}^2_{WPE-D}}} = \frac{\dfrac{1}{\sqrt{T}}\dfrac{1}{\sigma}\sum_{t=1}^{T} d_t}{\sqrt{\dfrac{1}{\sigma^2}\dfrac{2\pi}{m}\sum_{j=1}^{m} I(\lambda_j)}}. \tag{2.31}$$

Under regularity conditions such as d_t being a linear process with iid or martingale difference innovations, existence of the second moment and $0 < \sum_{j=-\infty}^{\infty} cov(d_t, d_{t+j}) < \infty$, the numerator of (2.31) has limiting Standard Normal distribution.

For the denominator of (2.31), the joint distribution of $2\pi I(\lambda_j)$ converges to m independent $\frac{\sigma^2}{2}\chi^2_2$ and so $2\pi\sum_{j=1}^m I(\lambda_j)$ converges to $\frac{\chi^2_{2m}}{2m}$, see Hannan (1970) or Theorem 5.4.3 in Brillinger (1975).

It is also possible to show that numerator and denominator of (2.31) are independent (Hualde and Iacone, 2017) so, using the continous mapping theorem, the ratio t converges to a t_{2m} distribution. This asymptotic limit can be obtained under different regularity conditions and assumptions for different processes making it widely applicable in forecast evaluation. Müller (2014) includes the linearity condition while Sun (2013) uses milder assumptions.

Theorem 2 of Hualde and Iacone (2017) shows that the Functional Central Limit Theorem (FCLT)

$$\frac{1}{\sqrt{T}}\frac{1}{\sigma}\sum_{t=1}^{\lfloor rT\rfloor} d_t \Rightarrow W(r), \tag{2.32}$$

where $\lfloor . \rfloor$ is the integer part of a number, $r \in [0,1]$ and $W(r)$ is a standard Brownian motion, is a sufficient condition for the results of the numerator, the denominator and the independence of the two.

Coroneo and Iacone (2020) use the same setting as Giacomini and White (2006) showing that the FCLT in (2.32) holds for mixing processes too; this is convenient when dealing with regression residuals. Also, Wu and Shao (2006) give a FCLT for non-linear processes which also nests, for example heteroscedastic processes.

For Wald-type tests, by previous results and the continuous mapping theorem,

$$W = \frac{\hat{\theta}'[R(X'X)^{-1}R']^{-1}\hat{\theta}}{\hat{\sigma}^2_{WPE-D}} = k\frac{\hat{\theta}'[\sigma^2 R(X'X)^{-1}R']^{-1}\hat{\theta}}{\frac{k\hat{\sigma}^2_{WPE-D}}{\sigma^2}} \rightarrow_d k\frac{\chi^2_k/k}{\chi^2_{2m}/2m} = kF_{k,2m}, \tag{2.33}$$

where $\hat{\theta} = R\hat{\beta} - c$ with R, $\hat{\beta}$ and c defined in Section 2.2 and $\hat{\sigma}^2_{WPE-D}$ is the WPE estimator with Daniell kernel of the long run variance σ^2.

2.C Appendix: Original Monte Carlo Setting Results

This Appendix presents Monte Carlo results for the original setting in Patton and Timmermann (2007). Under the original data generating process, regressions EA2 and PT2A exhibit multicollinearity between the variables f and y, however, there results are reported to support benefits coming from fixed-m asymptotics with a variety of DGP.

The Data Generating Process is a AR(1)-GARCH(1,1)

$$\begin{aligned} y_t &= 0.5y_{t-1} + \sigma_t\epsilon_t, \\ \sigma_t^2 &= 0.1 + 0.8\sigma_{t-1}^2 + 0.1\sigma_{t-1}^2\epsilon_{t-1}^2, \\ \epsilon_t &\sim iidN(0,1), \\ Y_1 &= 0, \sigma_1^2 = 1, \end{aligned} \tag{2.34}$$

the rational one step ahead forecast is

$$f_{t+1}^* = \mu_{t+1} + \gamma\sigma_{t+1}, \tag{2.35}$$

where $\mu_{t+1} = E_t[y_{t+1}]$, $\sigma_{t+1}^2 = V_t[y_{t+1}]$ and $\gamma = 0$ under MSE loss or $\gamma = 0.25$ under asymmetric quadratic loss.

In this setting, for the one step ahead forecast, $\mu_{t+1} = E_t[y_{t+1}] = 0.5y_t$.

Non-rational forecasts are obtained by adding independent noise to the rational forecasts

$$\begin{aligned} f_{t+1} =& f_{t+1}^* + \xi\varepsilon_{t+1}, \\ \varepsilon_{t+1} \sim& iidN(0,1), \end{aligned} \tag{2.36}$$

where ξ is the standard deviation of the noise. When ξ is equal to zero, forecasts are rational and the null hypothesis is true. This is used to inspect the finite-sample size of tests, however, for the tests based on MSE loss, when the simulation is conducted under the asymmetric quadratic loss, the null hypotheses is not satisfied despite ξ being zero, thus a complete size study cannot be performed for these tests.

The two tests with the multicollinearity problem in the Monte Carlo setting of Patton and Timmermann (2007) are:

'EA2', which tests $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$ on the regression

$$e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}, \tag{2.37}$$

and 'PT2A', which is based on the regression

$$I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}, \tag{2.38}$$

testing $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

Notice that when $\gamma = 0$ (simulation under MSE loss) and $\xi = 0$ (size study), the rational forecast becomes

$$f_{t+1} = f^*_{t+1} = \mu_{t+1} = E_t[y_{t+1}] = 0.5y_t, \tag{2.39}$$

So in Equation (2.37) and (2.38) the vector f_{t+1} is equal to $0.5y_t$ which is a linear combination of y_t. For this reason, the size cannot be computed and it is impossible to find the right size. These cases are presented in the following Tables with '—' which are still informative in the sense that show how fixed-smoothing asymptotics deliver correctly sized tests also in the original Monte Carlo setting.

Table 2.8: Empirical size and power of rationality tests under standard asymptotics

	MSE loss						Asymmetric loss					
ξ	MZ	EA	EA2	PT	PT2	PT2A	MZ	EA	EA2	PT	PT2	PT2A
						T = 30						
0	0.083	0.112	—	0.080	0.097	—	0.265	0.273	0.309	0.076	0.095	0.132
0.25	0.150	0.239	0.296	0.144	0.152	0.223	0.320	0.390	0.442	0.139	0.148	0.220
0.5	0.432	0.570	0.671	0.382	0.385	0.493	0.548	0.670	0.745	0.385	0.379	0.487
0.75	0.756	0.838	0.910	0.676	0.664	0.757	0.803	0.873	0.930	0.668	0.658	0.753
1	0.923	0.953	0.981	0.862	0.850	0.905	0.939	0.962	0.985	0.856	0.838	0.895
1.25	0.979	0.987	0.997	0.947	0.938	0.968	0.982	0.990	0.998	0.947	0.935	0.964
1.5	0.995	0.997	1.000	0.981	0.975	0.989	0.996	0.998	1.000	0.979	0.974	0.988
1.75	0.999	1.000	1.000	0.994	0.990	0.996	0.999	1.000	1.000	0.993	0.989	0.995
2	1.000	1.000	1.000	0.998	0.997	0.998	1.000	1.000	1.000	0.998	0.996	0.998
						T = 60						
0	0.070	0.087	—	0.063	0.069	—	0.423	0.402	0.396	0.064	0.068	0.081
0.25	0.158	0.266	0.385	0.151	0.160	0.279	0.479	0.550	0.629	0.144	0.155	0.267
0.5	0.597	0.764	0.885	0.533	0.554	0.732	0.758	0.868	0.938	0.523	0.539	0.721
0.75	0.918	0.969	0.991	0.869	0.877	0.953	0.953	0.982	0.996	0.864	0.870	0.948
1	0.989	0.997	1.000	0.977	0.976	0.994	0.994	0.998	1.000	0.976	0.975	0.995
1.25	0.999	1.000	1.000	0.997	0.996	0.999	0.999	1.000	1.000	0.997	0.996	0.999
1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
1.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note: this Table reports finite-sample size ($\xi = 0$) and power ($\xi > 0$) for rationality tests in small samples ($T = 30, 60$) under MSE loss and asymmetric loss with standard asymptotics. The theoretical size is 5%. 'MZ' is the test based on the regression $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; 'EA' is the test based on the regression $e_{t+1} = \beta_o + \beta_1 f_{t+1} + \beta_2 e_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; 'EA2' is the test based on the regression $e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; 'PT' is the test based on $I_{t+1} = \alpha + \beta_1 f_{t+1} + u_{t+1}$, $H_0 : \beta_1 = 0$; 'PT2' is the test based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = 0$ and 'PT2A' is based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Long run variance estimators are obtained using WCE with Bartlett kernel, bandwidth $M = \lfloor 0.75 \times T^{1/3} \rfloor$. '—' indicates that size could not be calculated because of the presence of multicollinearity.

Table 2.9: Empirical size and power of rationality tests under Mean Square Error loss and fixed-smoothing asymptotics

	$\mathbf{m = \lfloor T^{1/3} \rfloor}$						$\mathbf{m = \lfloor T^{1/4} \rfloor}$					
ξ	MZ	EA	EA2	PT	PT2	PT2A	MZ	EA	EA2	PT	PT2	PT2A
						T = 30						
0	0.028	0.023	—	0.046	0.041	—	0.028	0.023	—	0.044	0.042	—
0.25	0.048	0.052	0.058	0.071	0.059	0.081	0.045	0.044	0.047	0.066	0.054	0.074
0.5	0.156	0.179	0.262	0.207	0.155	0.228	0.120	0.125	0.171	0.172	0.121	0.173
0.75	0.368	0.394	0.600	0.418	0.326	0.427	0.271	0.262	0.419	0.337	0.245	0.326
1	0.616	0.629	0.847	0.629	0.520	0.626	0.462	0.438	0.691	0.522	0.390	0.490
1.25	0.801	0.799	0.949	0.785	0.680	0.773	0.649	0.603	0.862	0.675	0.525	0.623
1.5	0.905	0.900	0.985	0.879	0.796	0.865	0.786	0.745	0.947	0.784	0.634	0.724
1.75	0.959	0.952	0.995	0.931	0.869	0.915	0.879	0.843	0.980	0.858	0.725	0.793
2	0.982	0.978	0.999	0.965	0.918	0.949	0.937	0.906	0.992	0.909	0.795	0.847
2.25	0.993	0.991	1.000	0.982	0.949	0.968	0.965	0.946	0.998	0.941	0.843	0.881
2.5	0.997	0.996	1.000	0.989	0.968	0.979	0.982	0.967	0.999	0.961	0.879	0.904
2.75	0.999	0.998	1.000	0.994	0.980	0.985	0.990	0.982	1.000	0.974	0.907	0.920
3	1.000	0.999	1.000	0.996	0.986	0.990	0.996	0.990	1.000	0.982	0.926	0.931
3.25	1.000	1.000	1.000	0.998	0.990	0.992	0.998	0.996	1.000	0.987	0.941	0.938
3.5	1.000	1.000	1.000	0.999	0.993	0.994	0.999	0.998	1.000	0.992	0.951	0.946
3.75	1.000	1.000	1.000	0.999	0.995	0.996	1.000	0.999	1.000	0.994	0.959	0.952
4	1.000	1.000	1.000	0.999	0.997	0.996	1.000	0.999	1.000	0.995	0.965	0.956
						T = 60						
0	0.032	0.025	—	0.041	0.042	—	0.037	0.031	—	0.043	0.044	—
0.25	0.060	0.062	0.116	0.085	0.067	0.127	0.057	0.063	0.099	0.076	0.067	0.115
0.5	0.234	0.262	0.529	0.298	0.236	0.419	0.192	0.199	0.380	0.253	0.191	0.318
0.75	0.555	0.582	0.896	0.614	0.525	0.733	0.431	0.428	0.749	0.525	0.409	0.588
1	0.812	0.821	0.987	0.834	0.763	0.907	0.682	0.656	0.938	0.749	0.617	0.784
1.25	0.936	0.939	0.999	0.941	0.894	0.973	0.847	0.820	0.989	0.878	0.776	0.893
1.5	0.982	0.981	1.000	0.982	0.961	0.992	0.938	0.916	0.998	0.949	0.874	0.947
1.75	0.995	0.994	1.000	0.995	0.984	0.997	0.976	0.964	1.000	0.976	0.929	0.972
2	0.999	0.998	1.000	0.998	0.994	0.999	0.991	0.985	1.000	0.990	0.961	0.985
2.25	1.000	1.000	1.000	0.999	0.997	1.000	0.997	0.994	1.000	0.996	0.979	0.991
2.5	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.997	1.000	0.998	0.988	0.995
2.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.992	0.997
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.998
3.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.998
3.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.999
3.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999

Note: this Table reports finite-sample size ($\xi = 0$) and power ($\xi > 0$) for rationality tests in small samples ($T = 30, 60$) under MSE loss with fixed-smoothing asymptotics. The theoretical size is 5%. 'MZ' is the test based on the regression $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; 'EA' is the test based on the regression $e_{t+1} = \beta_o + \beta_1 f_{t+1} + \beta_2 e_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; 'EA2' is the test based on the regression $e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; 'PT' is the test based on $I_{t+1} = \alpha + \beta_1 f_{t+1} + u_{t+1}$, $H_0 : \beta_1 = 0$; 'PT2' is the test based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = 0$ and 'PT2A' is based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Long run variance estimators are obtained using WPE with Daniell kernel, bandwidths $m = \lfloor T^{1/3} \rfloor$ and $m = \lfloor T^{1/4} \rfloor$. '—' indicates that size could not be calculated because of the presence of multicollinearity.

Table 2.10: Empirical size and power of rationality tests under asymmetric loss and fixed-smoothing asymptotics

	$\mathbf{m} = \lfloor \mathbf{T}^{1/3} \rfloor$						$\mathbf{m} = \lfloor \mathbf{T}^{1/4} \rfloor$					
ξ	MZ	EA	EA2	PT	PT2	PT2A	MZ	EA	EA2	PT	PT2	PT2A
						$\mathbf{T = 30}$						
0	0.093	0.059	0.070	0.042	0.043	0.060	0.072	0.046	0.061	0.043	0.039	0.060
0.25	0.107	0.096	0.106	0.070	0.057	0.082	0.085	0.073	0.077	0.062	0.054	0.071
0.5	0.214	0.230	0.329	0.204	0.153	0.221	0.158	0.158	0.210	0.168	0.122	0.170
0.75	0.418	0.447	0.650	0.419	0.326	0.426	0.304	0.291	0.466	0.337	0.240	0.319
1	0.649	0.661	0.863	0.621	0.518	0.618	0.488	0.465	0.719	0.516	0.380	0.473
1.25	0.815	0.819	0.956	0.777	0.672	0.762	0.668	0.624	0.876	0.667	0.513	0.614
1.5	0.914	0.909	0.986	0.877	0.790	0.858	0.797	0.756	0.952	0.784	0.628	0.715
1.75	0.962	0.954	0.995	0.930	0.868	0.915	0.884	0.850	0.983	0.856	0.718	0.789
2	0.983	0.980	0.999	0.963	0.917	0.946	0.939	0.910	0.993	0.908	0.786	0.841
2.25	0.993	0.992	1.000	0.979	0.947	0.965	0.967	0.947	0.998	0.940	0.835	0.876
2.5	0.997	0.996	1.000	0.988	0.966	0.976	0.981	0.970	0.999	0.960	0.872	0.900
2.75	0.999	0.998	1.000	0.994	0.977	0.984	0.991	0.983	1.000	0.972	0.899	0.916
3	1.000	0.999	1.000	0.996	0.985	0.989	0.996	0.991	1.000	0.981	0.921	0.929
3.25	1.000	1.000	1.000	0.998	0.990	0.992	0.998	0.996	1.000	0.987	0.938	0.939
3.5	1.000	1.000	1.000	0.998	0.992	0.994	0.999	0.998	1.000	0.990	0.949	0.947
3.75	1.000	1.000	1.000	0.999	0.996	0.995	1.000	0.999	1.000	0.993	0.959	0.951
4	1.000	1.000	1.000	0.999	0.997	0.995	1.000	0.999	1.000	0.995	0.965	0.954
						$\mathbf{T = 60}$						
0	0.200	0.126	0.125	0.043	0.038	0.054	0.165	0.107	0.113	0.045	0.043	0.058
0.25	0.191	0.159	0.231	0.082	0.067	0.121	0.158	0.127	0.175	0.079	0.064	0.105
0.5	0.342	0.366	0.646	0.298	0.239	0.410	0.269	0.264	0.467	0.256	0.195	0.319
0.75	0.623	0.648	0.925	0.613	0.521	0.728	0.486	0.481	0.794	0.523	0.407	0.572
1	0.841	0.849	0.990	0.833	0.757	0.903	0.713	0.687	0.948	0.739	0.614	0.775
1.25	0.947	0.949	0.999	0.943	0.893	0.969	0.864	0.837	0.990	0.880	0.772	0.888
1.5	0.984	0.984	1.000	0.980	0.956	0.990	0.943	0.924	0.999	0.947	0.869	0.944
1.75	0.996	0.995	1.000	0.994	0.984	0.996	0.979	0.967	1.000	0.978	0.926	0.975
2	0.999	0.999	1.000	0.998	0.994	0.998	0.992	0.987	1.000	0.991	0.959	0.987
2.25	1.000	1.000	1.000	0.999	0.998	1.000	0.997	0.994	1.000	0.996	0.977	0.991
2.5	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.997	1.000	0.998	0.988	0.995
2.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.999	0.992	0.996
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.998
3.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.999
3.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
3.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999

Note: this Table reports finite-sample size ($\xi = 0$) and power ($\xi > 0$) for rationality tests in small samples ($T = 30, 60$) under asymmetric loss with fixed-smoothing asymptotics. The theoretical size is 5%. 'MZ' is the test based on the regression $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; 'EA' is the test based on the regression $e_{t+1} = \beta_o + \beta_1 f_{t+1} + \beta_2 e_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; 'EA2' is the test based on the regression $e_{t+1} = \beta_0 + \beta_1 f_{t+1} + \beta_2 e_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; 'PT' is the test based on $I_{t+1} = \alpha + \beta_1 f_{t+1} + u_{t+1}$, $H_0 : \beta_1 = 0$; 'PT2' is the test based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = 0$ and 'PT2A' is based on the regression $I_{t+1} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + \beta_3 y_t + u_{t+1}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. Long run variance estimators are obtained using WPE with Daniell kernel, bandwidths $m = \lfloor T^{1/3} \rfloor$ and $m = \lfloor T^{1/4} \rfloor$.

2.D Appendix: Results Using First Release of Realised Series

This Appendix reports empirical results of rationality tests as in Section 2.4 when the vintage of the realised variable is taken at its first unrevised release. In particular, Tables 2.11 - 2.13 report results of rationality tests when the series of realised values are taken at the first release except when realisations are regressors. In this case, the appropriate vintage and realisation are selected according to the point in time forecasts had to produce their forecasts so that all regressors are part of forecasters' information set.

Results are robust to the vintage used and confirm those obtained in Section 2.4. For ECB SPF, there are even less rejections than the case when the current release is used, confirming the rationality of European surveys. For US SPF, there are essentially no variations confirming previous results except for real GDP growth in the second subsample where rationality is strongly rejected under MSE loss.

Table 2.11: Rationality tests, full sample 2002.Q1 - 2016.Q4, T = 60, first release of the realised variable

Horizon	MZ Test	EA Test	EA2 Test	PT Test	PT2 Test	PT2A Test
Years			**ECB SPF**			
			HICP			
1	0.24	0.33	0.96	-0.18	0.22	0.89
2	0.14	0.58	0.65	0.03	0.08	0.34
			Unemployment			
1	1.11	2.41	2.40	1.15	2.46	2.45
2	2.34	2.34	3.30	1.30	1.70	2.30
			Real GDP growth			
1	3.91	4.09	9.98	-0.67	1.31	3.22
2	4.02	5.67	5.38	0.78	1.45	1.92
Quarters			**US SPF**			
			CPI			
0	95.86**	76.43**	106.11**	-2.15	5.21	12.06*
1	0.37	1.75	2.29	-0.59	1.41	1.80
2	0.65	0.86	0.93	-0.13	0.03	0.25
3	0.06	0.57	0.66	-0.54	0.57	0.62
4	0.42	0.96	2.35	-0.25	0.07	1.90
			Civilian unemployment rate			
0	2.34	3.24	74.83**	-0.12	0.02	28.03**
1	0.48	5.28	44.72**	-0.25	1.47	17.37**
2	0.15	3.66	35.58**	-0.01	1.48	14.42**
3	0.15	2.13	19.19*	0.88	2.67	10.15*
4	0.19	1.29	13.56	0.36	1.97	12.48*
			Real GDP growth			
0	0.66	1.29	14.35*	-0.47	1.45	5.48
1	14.99**	13.72*	42.50**	1.22	1.50	4.71
2	12.68**	16.33**	18.14*	1.61	2.99	11.24*
3	8.25*	8.64	8.66	0.98	1.50	1.88
4	8.80*	9.02	9.11	1.37	3.66	3.93

Note: the Table reports test statistics values for ECB SPF one-year and two-years ahead rolling horizon forecasts. For US SPF, it reports test statistics for current quarter, one quarter ahead, two quarters ahead, three quarters ahead and four quarters ahead forecasts. Results are obtained on full sample from 2002.Q1 to 2016.Q4 ($T = 60$) using the first release of the realised data. Rationality tests used are: MZ test: $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; EA test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; EA2 test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; PT test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_1 = 0$; PT2 test: $I_{t+h} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = 0$; PT2A test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + \beta_2 I_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. h is the forecast horizon expressed in years (quarters) for ECB SPF (US SPF). Long run variances are estimated using WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$, 5% fixed-$m$ critical values (asymptotics distributions) are 10.28 ($2F_{2,2m}$), 15.42 ($3F_{3,2m}$), 20.56 ($4F_{4,2m}$), 2.45 (t_{2m}), 10.28 ($2F_{2,2m}$), 14.28 ($3F_{3,2m}$) respectively for each column, 10 % fixed-m critical values are 6.92, 10.38, 13.84, 1.94, 6.92, 9.87; rejections are reported using ** and * to indicate, respectively, significance at the 5% and 10% level. 5% standard asymptotics critical values (asymptotics distributions) are 5.99 (χ^2_2), 7.81 (χ^2_3), 9.49 (χ^2_4), 1.96 ($N(0,1)$), 5.99 (χ^2_2), 7.81 (χ^2_3) respectively for each column, 10 % standard asymptotics critical values are 4.61, 6.25, 7.78,1.65, 4.61, 6.25; rejections are reported using and to indicate, respectively, significance at the 5% and 10% level.

Table 2.12: Rationality tests, first sample 2002.Q1 - 2009.Q2, T = 30, first release of the realised variable

Horizon	MZ Test	EA Test	EA2 Test	PT Test	PT2 Test	PT2A Test
Years			**ECB SPF**			
			HICP			
1	1.07	8.46	12.65	1.22	1.96	3.76
2	1.93	1.93	2.02	3.11**	9.11*	9.16
			Unemployment			
1	1.08	37.62**	37.59**	0.87	1.74	2.29
2	5.56	6.09	38.23**	1.18	1.40	6.56
			Real GDP growth			
1	4.99	5.05	8.27	-0.64	0.42	0.67
2	2.94	9.46	10.47	1.95*	8.27*	33.38**
Quarters			**US SPF**			
			CPI			
0	133.96**	260.88**	176.92**	-1.93	7.37*	12.95*
1	0.78	2.00	3.27	-0.62	4.61	7.95
2	1.24	1.28	1.44	0.81	2.48	4.12
3	0.63	1.42	1.76	0.74	8.24*	11.93*
4	0.30	0.47	2.94	-0.21	0.40	1.64
			Civilian unemployment rate			
0	1.93	2.63	9.99	-2.00*	3.59	10.63*
1	2.32	5.79	10.60	-1.31	3.63	8.72
2	2.43	2.51	10.67	-1.34	2.82	9.14
3	1.70	1.70	6.93	-0.32	0.32	4.20
4	1.44	1.49	6.54	-1.05	1.14	4.32
			Real GDP growth			
0	0.02	0.02	19.29*	-0.85	5.22	13.85*
1	10.84**	16.90**	52.48**	0.89	1.07	3.39
2	4.16	19.50**	18.94*	0.51	1.11	4.72
3	1.87	2.94	3.42	0.99	3.42	4.33
4	2.54	3.82	6.79	0.90	2.62	2.71

Note: the Table reports test statistics values for ECB SPF one-year and two-years ahead rolling horizon forecasts. For US SPF, it reports test statistics for current quarter, one quarter ahead, two quarters ahead, three quarters ahead and four quarters ahead forecasts. Results are obtained on first sub-sample from 2002.Q1 to 2009.Q2 ($T = 30$) using the first release of the realised data. Rationality tests used are: MZ test: $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; EA test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; EA2 test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; PT test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_1 = 0$; PT2 test: $I_{t+h} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = 0$; PT2A test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + \beta_2 I_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. h is the forecast horizon expressed in years (quarters) for ECB SPF (US SPF). Long run variances are estimated using WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$, 5% fixed-$m$ critical values (asymptotics distributions) are 10.28 ($2F_{2,2m}$), 15.42 ($3F_{3,2m}$), 20.56 ($4F_{4,2m}$), 2.45 (t_{2m}), 10.28 ($2F_{2,2m}$), 14.28 ($3F_{3,2m}$) respectively for each column, 10 % fixed-m critical values are 6.92, 10.38, 13.84, 1.94, 6.92, 9.87; rejections are reported using ** and * to indicate, respectively, significance at the 5% and 10% level. 5% standard asymptotics critical values (asymptotics distributions) are 5.99 (χ^2_2), 7.81 (χ^2_3), 9.49 (χ^2_4), 1.96 ($N(0,1)$), 5.99 (χ^2_2), 7.81 (χ^2_3) respectively for each column, 10 % standard asymptotics critical values are 4.61, 6.25, 7.78,1.65, 4.61, 6.25; rejections are reported using and to indicate, respectively, significance at the 5% and 10% level.

Table 2.13: Rationality tests, second sample 2009.Q3 - 2016.Q4, T = 30, first release of the realised variable

Horizon	MZ Test	EA Test	EA2 Test	PT Test	PT2 Test	PT2A Test
Years			**ECB SPF**			
			HICP			
1	0.78	1.41	2.08	-0.46	1.03	1.74
2	1.22	2.69	2.70	0.26	1.45	1.83
			Unemployment			
1	0.74	1.22	2.51	0.35	1.44	2.52
2	1.46	2.18	2.18	1.06	2.52	2.53
			Real GDP growth			
1	0.36	6.14	12.77	-0.08	2.22	9.48
2	1.21	2.09	2.17	-0.68	1.06	1.10
Quarters			**US SPF**			
			CPI			
0	11.97**	10.85*	11.00	-2.51**	6.62	8.14
1	1.33	1.28	1.67	0.81	0.71	2.68
2	2.85	3.09	4.15	0.56	1.64	2.45
3	2.28	2.89	5.57	0.49	1.82	1.84
4	0.94	1.07	1.64	1.45	2.93	3.04
			Civilian unemployment rate			
0	28.87**	21.13**	28.57**	-0.30	0.27	2.46
1	17.96**	14.84*	16.98*	0.18	0.04	1.84
2	18.02**	18.84**	18.73*	-0.04	0.09	0.83
3	15.79**	18.50**	19.06*	1.95*	3.78	4.62
4	13.23**	16.40**	17.11*	0.84	1.52	2.37
			Real GDP growth			
0	1.07	2.15	3.28	0.23	0.05	0.34
1	6.72	10.67*	35.36**	1.47	2.35	4.45
2	12.44**	13.76*	25.55**	1.71	7.78*	36.60**
3	15.53**	17.46**	21.22**	1.35	2.14	2.27
4	20.97**	21.10**	27.10**	1.62	2.65	2.67

Note: the Table reports test statistics values for ECB SPF one-year and two-years ahead rolling horizon forecasts. For US SPF, it reports test statistics for current quarter, one quarter ahead, two quarters ahead, three quarters ahead and four quarters ahead forecasts. Results are obtained on second sub-sample from 2009.Q3 to 2016.Q4 ($T = 30$) using the first release of the realised data. Rationality tests used are: MZ test: $y_{t+h} = \beta_0 + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_0 = 0 \cap \beta_1 = 1$; EA test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$; EA2 test: $e_{t+h} = \beta_0 + \beta_1 f_{t+h} + \beta_2 e_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$; PT test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + u_{t+h}$, $H_0 : \beta_1 = 0$; PT2 test: $I_{t+h} = \alpha + \beta_1 f_{t+1} + \beta_2 I_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = 0$; PT2A test: $I_{t+h} = \alpha + \beta_1 f_{t+h} + \beta_2 I_t + \beta_3 y_t + u_{t+h}$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. h is the forecast horizon expressed in years (quarters) for ECB SPF (US SPF). Long run variances are estimated using WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$, 5% fixed-$m$ critical values (asymptotics distributions) are 10.28 ($2F_{2,2m}$), 15.42 ($3F_{3,2m}$), 20.56 ($4F_{4,2m}$), 2.45 (t_{2m}), 10.28 ($2F_{2,2m}$), 14.28 ($3F_{3,2m}$) respectively for each column, 10 % fixed-m critical values are 6.92, 10.38, 13.84, 1.94, 6.92, 9.87; rejections are reported using ** and * to indicate, respectively, significance at the 5% and 10% level. 5% standard asymptotics critical values (asymptotics distributions) are 5.99 (χ^2_2), 7.81 (χ^2_3), 9.49 (χ^2_4), 1.96 ($N(0,1)$), 5.99 (χ^2_2), 7.81 (χ^2_3) respectively for each column, 10 % standard asymptotics critical values are 4.61, 6.25, 7.78,1.65, 4.61, 6.25; rejections are reported using and to indicate, respectively, significance at the 5% and 10% level.

Chapter 3

The Accuracy of Point Aggregate Forecasts of the ECB SPF

ABSTRACT: I perform the real-time forecast evaluation of inflation, unemployment and real GDP growth aggregate point forecasts from the European Central Bank Survey of Professional Forecasters (ECB SPF) using the Diebold and Mariano test for equal forecast accuracy with asymmetric loss functions and competing forecasts from different simple benchmark models. As macroeconomic historical data is subject to revision, I take it into account when constructing competing forecasts and using several different vintages of the target variables realised series. To account for the small sample available that affect the size of the Diebold and Mariano tests, I use fixed-smoothing asymptotics that proved to alleviate size distortion in small samples. The ECB SPF does not seem to outperform benchmark models in short and medium term forecasts but it shows moderate advantages for long term forecasts of inflation. However, long term benchmarks forecasts of unemployment and real GDP growth outperform SPF and no significant difference in accuracy emerges for short and medium term forecasts. Results are generally robust to revision of historical data and loss functions.

Keywords: Diebold and Mariano test, long run variance estimation, fixed-smoothing asymptotics, Heteroscedasticity Autocorrelation Robust (HAR) inference, SPF, Real-time forecast evaluation, Hypothesis testing

JEL Classification: C12, C32, C51, C53, E17

This Chapter is adapted from the working paper '*The Accuracy of the Survey of Professional Forecasters for the Euro Area: an Heteroscedasticity Autocorrelation Robust Assessment*' presented at the 59th Riunione Scientifica Annuale of the Società Italiana degli Economisti, at the 6th Workshop for PhD Students in Econometrics and Empirical Economics by Società Italiana di Econometria and as a poster at the 2018 Asset Pricing Workshop of the University of York.

3.1 Introduction

Macroeconomic forecasts are essential as they play a key role in the decision process of economic agents and, as a consequence, in the outcome of monetary policies. Economic forecasts that usually get a great deal of attention are those from surveys of professional forecasters because they are deemed to come from better informed agents which have incentives to give their true or carefully considered views (Keane and Runkle, 1990). From this perspective, and because the collection of professional forecasts involves a great amount of time and economic effort, it is of interest to assess the quality of these surveys. One of the longest-running survey is the Federal Reserve of Philadelphia Survey of Professional Forecasters (US SPF) which provide forecasts for major US macroeconomic variables and it is the reference for many private and institutional agents. Its European analogue is the European Central Bank Survey of Professional Forecasters (ECB SPF).

Most empirical studies focus on US SPF because of the large sample available; for example, D'Agostino, Giannone and Surico (2006) use relative mean square errors to evaluate US SPF and find that predictive ability declined after the 80s, Stark (2010) performs a real-time forecast evaluation with the Diebold and Mariano test finding a general good predictive ability which deteriorates as the forecast horizon gets longer. Also, Coroneo and Iacone (2020) perform an evaluation of US SPF with the Diebold and Mariano test finding that SPF generally outperforms random walk forecasts and Demetrescu, Hanck and Kruse (2018) finds that after the Great Moderation the predictive accuracy has sensibly decreased.

Literature on ECB SPF is far less developed because of the scarcity of survey rounds: Bowles, Friz, Genre, Kenny, Meyler and Rautanen (2011) assess real GDP growth and unemployment and see a moderate superiority of surveys over benchmarks however, authors report that findings may be subject to small sample bias; Genre, Kenny, Meyler and Timmermann (2013) conclude that ECB SPF outperform simple benchmarks, Coroneo and Iacone (2020), in addition to US SPF, evaluate ECB SPF correcting the small sample bias and they cannot find strong evidence of ECB SPF superiority. Grothe and Meyler (2015) focus on inflation and they conclude that SPF forecasts have significantly smaller forecast errors than a random walk model or an AR(1). However, one of the fundamental assumptions underlying this existing literature is that the loss function used for evaluation is symmetric. In spite of that, the European Central Bank set its inflation target as an asymmetric objective aiming for a year-on-year increase of the Harmonised Index of Consume Prices (HICP) for the euro area below 2% (Nobay and Peel, 2003; Aguiar and Martins, 2008). Also, there is increasing evidence that many central banks around the world have asymmetric preferences, for example, Capistrán (2008) claims that the loss function of the Federal Reserve is asymmetric, Ruge-Murcia (2000) discuss the asymmetric inflation target of the Bank of Canada and Ruge-Murcia (2003) finds evidence of asymmetric preferences for the Bank of England and the Bank of Sweden. Moreover, evidence coming from the psychology literature suggests that also other economic agents actually may have asymmetric loss functions (Weber, 1994; Ehrbeck and Waldmann, 1996; Laster, Bennett and Geoum, 1999). With this in mind, it is highly likely that professional forecasters use an asymmetric loss function to make their predictions and the same should be done when evaluating them.

This work complements literature on the real-time evaluation of the accuracy of ECB SPF forecasts and it is closely related to Coroneo and Iacone (2020) in using the Diebold and Mariano (1995) test with fixed-smoothing asymptotics critical values to correct small sample bias. In addition, it takes into account competing forecasts from different benchmark models, the possibility that forecasts were made under asymmetric loss functions and that the outcome of the evaluation can change if different vintages of the target variables are considered as true outcome. In fact, euro macroeconomic data is revised regularly and this should be taken into account both when constructing the competing forecasts and when selecting the vintage of the realised data.

Among several evaluation methods available in the literature, my evaluation framework relies on the Diebold and Mariano test for the null hypothesis of equal forecast accuracy of the ECB SPF forecasts and competing forecasts from three simple benchmarks: a random walk without drift, an indirect autoregressive model (IAR) and a direct autoregressive model (DAR). IAR forecasts are obtained using the standard chain rule after estimating the autoregressive parameters, whereas DAR forecasts are obtained directly from an horizon-specific estimated model. These benchmark forecasts should be easily beaten by the apparently high quality forecasts coming from professional forecasters. To carry out a comprehensive evaluation, in addition to the two widely adopted symmetric loss functions, quadratic and absolute loss, which can only take into account the magnitude of forecast errors, I employ different types of asymmetric functions in light of the asymmetric ECB target. The Lin-Lin loss, squared Lin-Lin loss and Linex loss (Varian, 1975) allow to weight positive and negative forecast errors in different ways.

Before evaluating forecasts of inflation, unemployment rate and real GDP growth, I verify that fixed-smoothing asymptotics still provide correctly sized Diebold and Mariano test statistics with sample sizes comparable to those available for the empirical exercise when asymmetric loss functions are employed in a Monte Carlo exercise which setting is borrowed from Coroneo and Iacone (2020).

Empirical results show that long term inflation ECB SPF forecasts moderately outperform benchmark forecasts, however, competing forecasts and, in particular, indirect autoregressive forecasts perform better for unemployment and real GDP growth. For other horizons, there is no marked rejection of the null hypotheses of equal forecast accuracy. In terms of benchmark models, the random walk and the indirect autoregressive model seem to perform better than the direct autoregressive model especially for long horizon forecasts. These results are generally robust to the revision in historical data although, the effect of revision is more pronounced in unemployment and GDP. Also, using asymmetric loss functions does not impact much on the outcome of the test possibly because forecasters are interested in getting their forecasts right no matter the asymmetric target of the European Central Bank.

The remainder of this Chapter is organised as follows. In Section 3.2 I describe the models used to generate competing forecasts, Section 3.3 discusses the loss functions involved in the evaluation process and Section 3.4 gives a short outline of Diebold and Mariano test with fixed-smoothing asymptotics. Section 3.5 presents the results of the Monte Carlo experiment, Section 3.6 sets out the evaluation exercise and Section 3.7 concludes.

3.2 Benchmark Forecasts

In addition to participants' forecasts and their associated forecast errors, the US SPF reports forecast errors from benchmark models as a comparison. The Federal Reserve Bank uses statistical models to generate alternative forecasts for the same horizons included in the survey using the same vintage of data professional forecasters had at their disposal when they were making forecasts and then calculate forecast errors for these benchmark models. Benchmark models include a random walk, an Indirect Autoregressive model (IAR) and a Direct Autoregressive model (DAR). Along the same lines, Stark (2010) bases his US SPF forecast accuracy study on these three models. The rationale behind the choice of these benchmark models resides in the fact that a professional forecaster should be able to produce better forecasts than these simplistic models.

As the ECB SPF does not provide the additional benchmark forecasts information like the US SPF does, I generate competing forecasts from the same three naive benchmark models taking into account the vintage V of historical data available at the time t forecasters had to submit their h-steps ahead forecasts each quarter in order to construct credible competing forecasts and perform a fair and consistent evaluation. For every quarter, I check the deadline for replying to the survey round and I estimate benchmark models using the same information set forecasters had available before that date. The following sub-sections contain a description of the three models used in this study.

3.2.1 Random Walk

The first simple benchmark model is a random walk without drift

$$y_t^V = y_{t-1}^V + u_t, \tag{3.1}$$

and the forecast h steps ahead is given by

$$f_{t+h}^{RW} = y_t^V + \sum_{j=t+1}^{t+h} \epsilon_j, \tag{3.2}$$

where y_t^V is the last historical realization available at the newest vintage V when the forecast is produced for each survey round considered, such that $V \leq t$, and ϵ_j is an error term with mean zero. The forecast of this benchmark is the same no matter the forecast horizon.

Despite its uncomplicated design, Atkeson and Ohanian (2001) and Balcilar, Gupta, Majumdar and Miller (2015) suggest that the random walk model is hard to beat in forecasting inflation and real GDP.

3.2.2 Indirect Autoregressive Model

The second benchmark model is a univariate, indirect autoregressive model (IAR)

$$y_t^V = \theta_0 + \sum_{j=1}^{P(V)} \theta_j y_{t-j}^V + u_t. \tag{3.3}$$

For each survey, parameters are estimated using a rolling window of the last 30 quarterly observations available at vintage V. The lag length $P(V)$ is chosen using the Bayesian Information Criterion and re-estimated each survey round. Because of the data available in the real-time database, the maximum lag is set to 4.

After the autoregressive parameters are estimated, forecasts are obtained recursively according to

$$f_{t+h}^{IAR} = \hat{\theta}_0 + \sum_{j=1}^{P(V)} \hat{\theta}_j f_{t-j+h}^{IAR}, \tag{3.4}$$

given that $f_i^{IAR} = y_i^V$, for $i \leq t$, so that forecasts are constructed recursively with the standard chain rule using observations and vintage V available at the deadline of each survey.

Marcellino, Stock and Watson (2006) suggest that the IAR model works best when the model is correctly specified.

3.2.3 Direct Autoregressive Model

To account for model misspecification, I also use a Direct Autoregressive model (DAR) as a third benchmark

$$y_t^V = \theta_0^{(h)} + \sum_{j=1}^{P(h,V)} \theta_j^{(h)} y_{t-j+1-h}^V + u_t^{(h)}, \tag{3.5}$$

which tends to be more robust to model misspecification, but it produces less efficient parameter estimates than the IAR model, see among others Schorfheide (2005) and Bhansali (2002).

For each survey, parameters are estimated using a rolling window of the last $30-h$ quarterly observations available at vintage V. The lag length $P(h,V)$ is chosen using the Bayesian Information Criterion and re-estimated at each survey round and every forecast horizon; the maximum lag is 4.

Forecasts are obtained directly from

$$f_{t+h}^{DAR} = \hat{\theta}_0^{(h)} + \sum_{j=1}^{P(h,V)} \hat{\theta}_j^{(h)} y_{t-j+1}^V. \tag{3.6}$$

3.3 Loss Functions

Forecast evaluation with the Diebold and Mariano test for the null hypothesis of equal forecast accuracy involves the use of a loss function of forecast errors $e_t^V(h) = y_{t+h}^V - f_{t+h}$ where h is the forecast horizon, y_{t+h}^V is the realisation of the target variable at vintage V at time $t+h$ and f_{t+h} is its forecast for time $t+h$; the most common loss functions are the quadratic loss, defined as

$$L\left[e_t^V(h)\right] = e_t^V(h)^2, \tag{3.7}$$

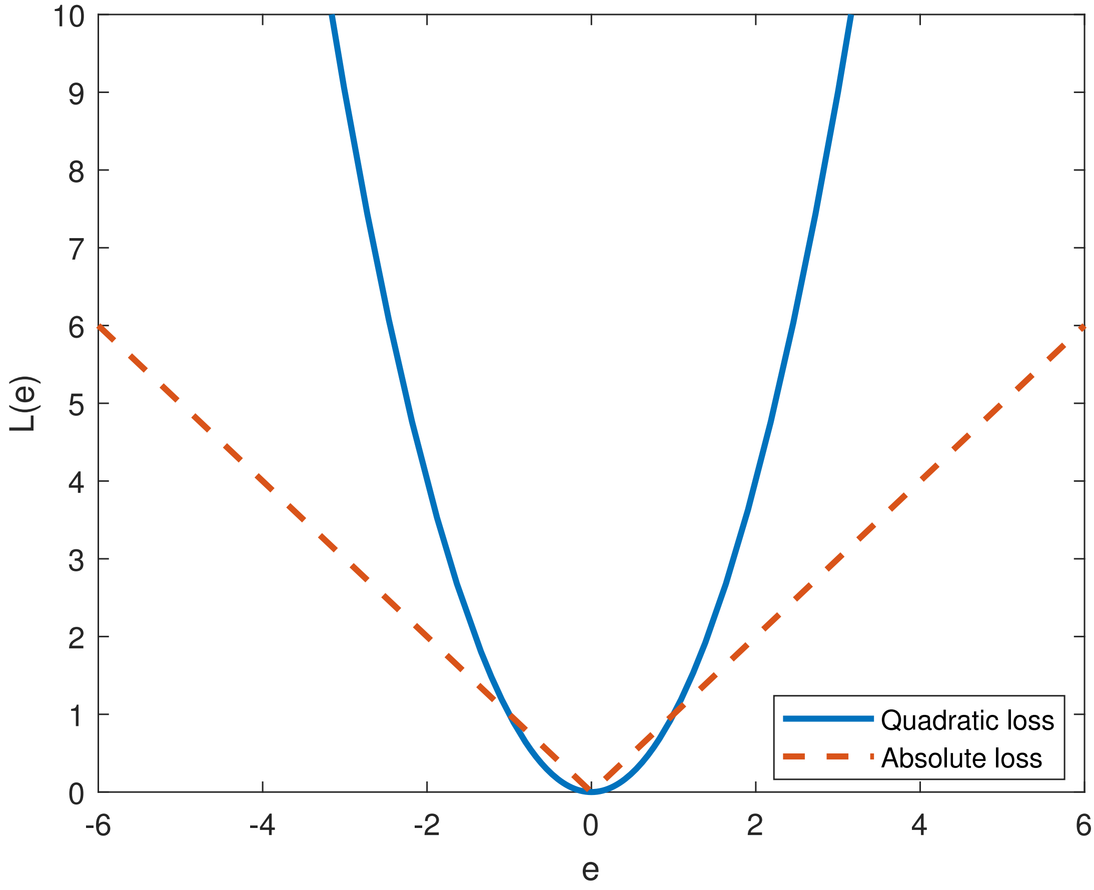
and the absolute loss

$$L\left[e_t^V(h)\right] = \left|e_t^V(h)\right|. \tag{3.8}$$

Both these functions are depicted in Figure 3.1 and commonly used in the literature as they are well known and easy to deal with, they are both symmetric, bowl shaped, differentiable everywhere (except in zero for the Absolute loss) and unbounded from above. They also satisfy all three Granger (1999) properties a good loss function should have: minimal loss of zero, loss always positive or equal to zero, non increasing for negative forecast errors and non decreasing for positive forecast errors. Large forecast errors are highly penalised but while for the quadratic loss the penalty increases quadratically, for the absolute loss, it increases linearly.

The existing literature provides evidence that economic agents can have an asymmetric loss function, in particular Nobay and Peel (2003) and Capistrán (2008) discuss the asymmetry of the Federal Reserve loss function, while Aguiar and Martins (2008) notices that the ECB's definition of price stability is essentially an indication of asymmetric preferences. On this assumption, it is highly likely that also forecasters base their prediction on an asymmetric loss function which weights forecast errors differently according to their

Figure 3.1: Quadratic and absolute loss functions plot



Note: the plot depicts the Quadratic and the Absolute loss functions in a solid blue line and a dashed red line respectively. The horizontal axis denotes the forecast error while the vertical axis is the loss associated.

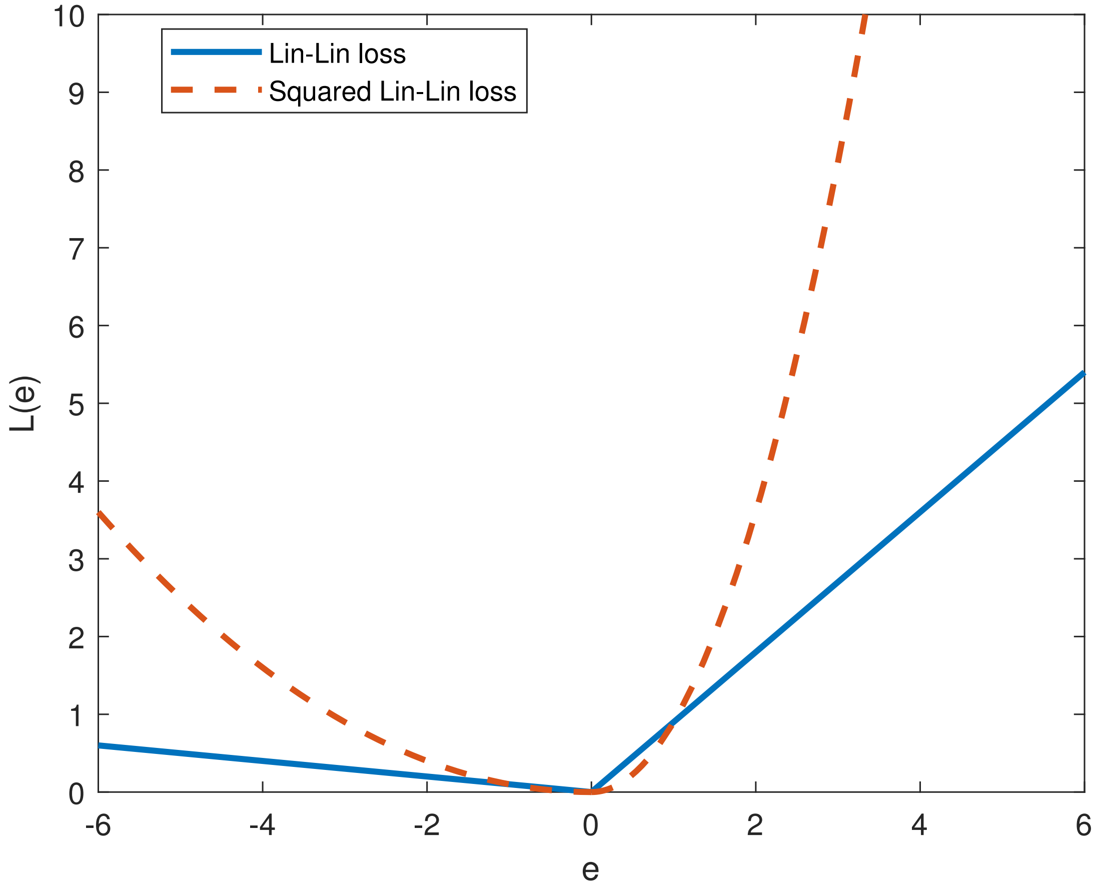
sign;

Common asymmetric loss functions include the Lin-Lin loss (Granger and Newbold, 1986)

$$L\left[e_t^V(h)\right] = \begin{cases} (1-\alpha)\left|e_t^V(h)\right| & \text{if } e_t^V(h) \leq 0; \\ \alpha\left|e_t^V(h)\right| & \text{if } e_t^V(h) > 0, \end{cases} \tag{3.9}$$

where $0 < \alpha < 1$ is the asymmetry parameter, the greater is α, the bigger the loss from positive forecast errors and the smaller the loss from negative errors. When $\alpha = 0.5$ the function becomes an absolute loss function. The Lin-Lin loss function is named after the fact that it is linear on each side of the origin. It is also the base for quantile regression

Figure 3.2: Lin-lin loss functions plot



Note: the plot depicts the Lin-Lin and the squared Lin-Lin loss functions in a solid blue line and a dashed red line respectively when $\alpha = 0.9$. The horizontal axis denotes the forecast error while the vertical axis is the loss associated.

as the optimal forecast under this loss is the α quantile.

Several authors, such as Newey and Powell (1987), Weiss (1996) and Artis and Marcellino (2001), use a squared version of Lin-Lin

$$L\left[e_t^V(h)\right] = \begin{cases} (1-\alpha)\left[e_t^V(h)\right]^2 & \text{if } e_t^V(h) \leq 0; \\ \alpha\left[e_t^V(h)\right]^2 & \text{if } e_t^V(h) > 0, \end{cases} \tag{3.10}$$

for $0 < \alpha < 1$. Figure 3.2 depicts the two versions of Lin-Lin functions when $\alpha = 0.9$. The blue solid line is the classic Lin-Lin function, while the red dashed line is the squared Lin-Lin.

Another popular asymmetric loss function is the Linex function (Varian, 1975) which is asymmetric but it is still differentiable everywhere and it takes the form

$$L\left[e_t^V(h)\right] = \exp\left[\alpha e_t^V(h)\right] - \alpha e_t^V(h) - 1; \tag{3.11}$$

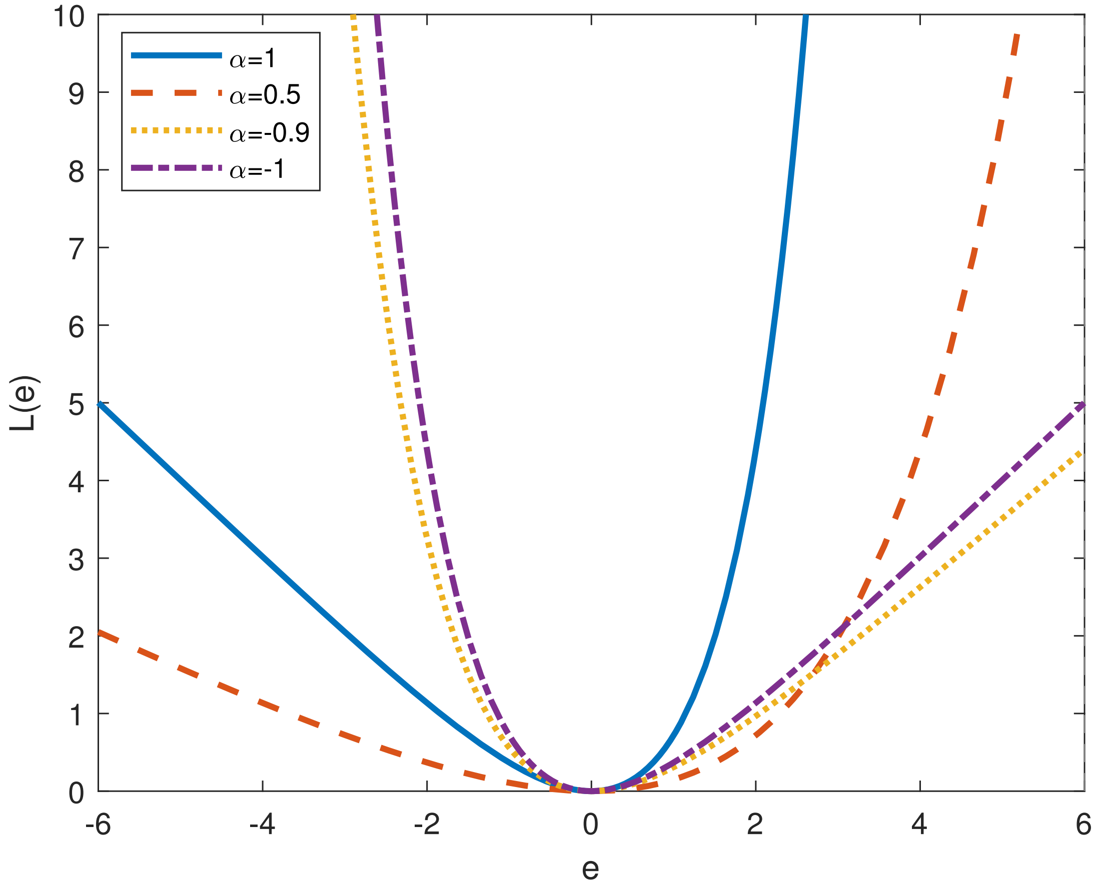
where α is a scalar that controls the aversion towards positive ($\alpha > 0$) or negative forecast errors ($\alpha < 0$). The choice of this parameter has to be done according to costs arising from overpredicting or underpredicting the target variables in the situation of interest.

Figure 3.3 depicts with a dash-dot purple line the case where $\alpha = -1$ and with a yellow dotted line the function with $\alpha = -0.9$. In these two cases, the agent is more adverse to negative forecast errors, while in the opposite cases when $\alpha = 1$ (blue solid line) and $\alpha = 0.5$ (red dashed line), an agent is more adverse to positive forecast errors. Positive forecast errors occurs when the realised value for the target variable is bigger than its corresponding forecast and vice-versa for negative forecast errors.

3.4 The Diebold and Mariano Test

Several techniques are available to evaluate forecasts: Theil (1958) suggests an informal graphics method using scatter plots of the forecast against the true outcome to understand the magnitude of forecast errors. More formal approaches were proposed by Wilson (1934) using the correlation between forecasts and realisations, Mincer and Zarnowitz (1969) proposes a test for forecast rationality and Fair and Shiller (1989, 1990) examine the information content of ex-ante forecasts among others. When two competing forecasts are available for the same variable of interest, Chong and Hendry (1986) propose a test for forecast encompassing while Diebold and Mariano (1995) suggest a test for equal forecast

Figure 3.3: Linex loss function plot



Note: the plot depicts the Linex function by Varian (1975). $L\left[e_t^V(h)\right] = \exp(\alpha e_t^V(h)) - \alpha e_t^V(h) - 1$ where the forecast error is defined as $e_t^V(h) = f_{t+h} - y_{t+h}^V$ with y_{t+h}^V the realisation of the target variable at vintage V and f_{t+h} its forecast. This loss function is asymmetric as it weights positive and negative forecast errors in different ways according to the parameter α. When $\alpha > 0$ the agent is averse to positive forecast errors while with $\alpha < 0$ the agent is averse to negative forecast errors. The horizontal axis denotes the forecast error while the vertical axis is the associated loss.

accuracy (DM test).

To evaluate the forecast performance of ECB SPF, I use the DM test for the null hypothesis of equal forecast accuracy.

Given forecast errors defined as $e_t^V(h) = y_{t+h}^V - f_{t+h}$, f_{t+h} the forecast h steps ahead of the actual y_{t+h}, $t = 1, \ldots, T$, the loss differential is

$$d_t^V(L) = L\left[e_t^{V,B}(h)\right] - L\left[e_t^{V,SPF}(h)\right], \tag{3.12}$$

with $L\left[e_t^{V,B}(h)\right]$ the loss function evaluated at forecast errors from benchmark models and $L\left[e_t^{V,SPF}(h)\right]$ the loss function evaluated at forecast errors from SPF.

The infeasible test statistic is

$$DM = \sqrt{T}\frac{\bar{d}^V - \mu^V}{\sqrt{\sigma^2}} \xrightarrow{d} N(0,1), \tag{3.13}$$

where $\bar{d}^V = \frac{1}{T}\sum_{t=1}^{T} d_t^V(L)$, $\mu^V = E\left[d_t^V(L)\right]$ and σ^2 is the long run variance. The null hypothesis of equal forecast accuracy is $H_0 : \mu^V = 0$.

To obtain the feasible test, the long run variance σ^2 can be estimated using a Weighted Corvariance Estimate (WCE) or a Weighted Periodogram Estimate (WPE) and, fixed-smoothing asymptotics are capable to reduce small sample size distortion. See Section 2.2 of Chapter 1 for a detailed presentation of long run variance estimators and fixed-smoothing asymptotics.

3.5 Monte Carlo Study for Size and Power of the DM Test under Asymmetric Loss Functions

Several studies show that fixed-smoothing asymptotics provide a better approximation of the empirical size in Diebold and Mariano type tests with symmetric loss functions (Harvey, Leybourne and Whitehouse, 2017; Coroneo and Iacone, 2020). To verify that the same improvements apply with asymmetric loss functions, this Section presents a Monte Carlo experiment borrowed from the Online Appendix of Coroneo and Iacone (2020) (which they, in turn, borrowed from Clark (1999)) about size and power of the Diebold and Mariano test with WCE and WPE long run variance estimators and fixed-smoothing asymptotics adapted to accommodate the asymmetric loss functions described in Section 3.3. The sample sizes $T = \{40, 120\}$ reflect the ones available in the empirical exercise. To investigate the empirical size, the vector of forecast error is taken from a bivariate standard normal distribution $(v_{1t}, v_{2t})' \sim N(0_2, I_2)$ and contemporaneous correlation is introduced according to the following transformation

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \tag{3.14}$$

with $\rho = 0.5$.

The two competing forecast errors are serially correlated $MA(q)$ processes constructed as

$$e_{1t} = \frac{\sum_{j=0}^{q} \theta^j u_{1t-j}}{\sqrt{\sum_{j=0}^{q} \theta^{2j}}}, \tag{3.15}$$

$$e_{2t} = \frac{\sum_{j=0}^{q} \theta^j u_{2t-j}}{\sqrt{\sum_{j=0}^{q} \theta^{2j}}}, \tag{3.16}$$

with moving average coefficient $\theta = 0.75$ and lag order $q \in [1, 5]$, indicating the degree of the serial correlation. Clark (1999) provides evidence related to the robustness of the empirical size to changes in θ and ρ, so these are both kept fixed allowing only the serial correlation q to increase.

The theoretical size is set to 5% and I use $10,000$ replications. WCE estimates are performed using the DM estimator in Equation (2.11) with $M(T) = q$ and the one in Equation (2.13) with bandwidths $M = \lfloor T^{1/3} \rfloor$, $M = \lfloor T^{1/2} \rfloor$ and $M = T$ for the Bartlett kernel. WPE estimates follow Equation (2.15) with bandwidths $m = \lfloor T^{1/4} \rfloor$, $m = \lfloor T^{1/3} \rfloor$, $m = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{2/3} \rfloor$ for the Daniell kernel.

Tables 3.1 - 3.8 present results for the empirical size of the Diebold and Mariano Test using standard asymptotics and fixed-smoothing asymptotics. Fixed-smoothing asymptotics provide better empirical size for all loss functions and size improves as the sample size grows but it worsens as the serial correlation q grows. Also, the empirical size is better the larger the bandwidth for WCE estimates and the smaller the bandwidth for WPE estimates confirming the findings of Coroneo and Iacone (2020). Results are robust for

Table 3.1: Empirical size of the DM test with quadratic loss function

Standard Asymptotics

	WCE				WPE			
				$T = 40$				
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.075	0.094	0.116	0.359	0.115	0.093	0.076	0.079
2	0.093	0.103	0.120	0.371	0.114	0.091	0.081	0.102
3	0.121	0.118	0.132	0.387	0.119	0.096	0.094	0.138
4	0.150	0.130	0.137	0.386	0.116	0.097	0.106	0.161
5	0.176	0.136	0.141	0.383	0.119	0.099	0.113	0.179

				$T = 120$				
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.057	0.071	0.080	0.340	0.092	0.081	0.061	0.065
2	0.067	0.084	0.090	0.342	0.096	0.086	0.068	0.081
3	0.067	0.085	0.089	0.347	0.092	0.083	0.067	0.092
4	0.075	0.096	0.095	0.357	0.097	0.086	0.071	0.114
5	0.082	0.101	0.095	0.353	0.094	0.083	0.076	0.126

Fixed-smoothing Asymptotics

	WCE			WPE			
			$T = 40$				
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.056	0.051	0.054	0.045	0.045	0.049	0.063
2	0.061	0.052	0.052	0.039	0.043	0.053	0.083
3	0.073	0.062	0.057	0.046	0.048	0.063	0.114
4	0.079	0.059	0.057	0.041	0.043	0.071	0.137
5	0.083	0.059	0.059	0.040	0.043	0.075	0.152

			$T = 120$				
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.053	0.048	0.047	0.045	0.046	0.048	0.058
2	0.065	0.055	0.052	0.048	0.048	0.053	0.074
3	0.065	0.050	0.050	0.041	0.043	0.049	0.084
4	0.076	0.058	0.054	0.049	0.047	0.057	0.104
5	0.082	0.058	0.052	0.044	0.046	0.056	0.117

Note: the Table reports the empirical size of the Diebold and Mariano test under a quadratic loss function with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.2: Empirical size of the DM test with absolute loss function

Standard Asymptotics

$T = 40$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.080	0.097	0.117	0.351	0.121	0.098	0.080	0.081
2	0.099	0.108	0.123	0.365	0.124	0.097	0.086	0.108
3	0.132	0.126	0.138	0.373	0.130	0.109	0.101	0.139
4	0.157	0.138	0.145	0.379	0.129	0.110	0.115	0.164
5	0.185	0.146	0.153	0.371	0.131	0.116	0.122	0.178

$T = 120$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.058	0.070	0.081	0.332	0.092	0.082	0.063	0.064
2	0.066	0.082	0.093	0.342	0.099	0.088	0.070	0.081
3	0.072	0.085	0.090	0.343	0.094	0.084	0.070	0.091
4	0.080	0.096	0.099	0.351	0.101	0.090	0.076	0.113
5	0.084	0.100	0.097	0.343	0.099	0.088	0.078	0.125

Fixed-smoothing Asymptotics

$T = 40$

	WCE			**WPE**			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.057	0.054	0.055	0.051	0.049	0.055	0.064
2	0.065	0.060	0.057	0.048	0.049	0.058	0.085
3	0.081	0.073	0.066	0.055	0.057	0.070	0.119
4	0.090	0.073	0.065	0.052	0.055	0.082	0.142
5	0.097	0.075	0.069	0.050	0.056	0.088	0.156

$T = 120$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.055	0.048	0.047	0.048	0.049	0.050	0.057
2	0.066	0.058	0.055	0.052	0.053	0.054	0.073
3	0.069	0.054	0.051	0.043	0.047	0.052	0.083
4	0.079	0.062	0.057	0.054	0.053	0.060	0.104
5	0.083	0.062	0.055	0.049	0.051	0.062	0.115

Note: the Table reports the empirical size of the Diebold and Mariano test under an absolute loss function with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.3: Empirical size of the DM test with Lin-Lin loss function $\alpha = 0.9$

Standard Asymptotics

$T = 40$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.074	0.103	0.123	0.364	0.119	0.096	0.081	0.089
2	0.089	0.117	0.128	0.366	0.111	0.092	0.092	0.137
3	0.114	0.138	0.145	0.387	0.121	0.106	0.116	0.189
4	0.128	0.153	0.150	0.398	0.115	0.099	0.130	0.214
5	0.153	0.168	0.152	0.395	0.111	0.103	0.149	0.238

$T = 120$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.054	0.072	0.082	0.335	0.091	0.081	0.060	0.067
2	0.063	0.087	0.091	0.342	0.096	0.084	0.069	0.093
3	0.067	0.095	0.093	0.347	0.095	0.083	0.072	0.116
4	0.070	0.104	0.096	0.347	0.093	0.080	0.076	0.142
5	0.074	0.119	0.102	0.347	0.094	0.084	0.083	0.169

Fixed-smoothing Asymptotics

$T = 40$

	WCE			**WPE**			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.058	0.053	0.055	0.045	0.046	0.053	0.071
2	0.070	0.055	0.054	0.037	0.043	0.061	0.114
3	0.088	0.066	0.060	0.044	0.043	0.078	0.163
4	0.095	0.062	0.060	0.037	0.040	0.090	0.188
5	0.107	0.066	0.062	0.038	0.044	0.107	0.211

$T = 120$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.055	0.047	0.048	0.047	0.046	0.047	0.060
2	0.070	0.057	0.056	0.045	0.048	0.053	0.086
3	0.076	0.058	0.053	0.046	0.045	0.056	0.108
4	0.085	0.059	0.055	0.042	0.046	0.060	0.132
5	0.098	0.062	0.051	0.043	0.046	0.065	0.156

Note: the Table reports the empirical size of the Diebold and Mariano test under a Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.4: Empirical size of the DM test with squared Lin-Lin loss function $\alpha = 0.9$

Standard Asymptotics

T = 40

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.063	0.086	0.105	0.379	0.105	0.084	0.066	0.071
2	0.073	0.091	0.106	0.384	0.092	0.076	0.069	0.104
3	0.097	0.110	0.119	0.403	0.100	0.079	0.081	0.151
4	0.120	0.122	0.123	0.411	0.092	0.079	0.098	0.180
5	0.146	0.131	0.125	0.404	0.092	0.079	0.109	0.199

T = 120

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.053	0.067	0.078	0.334	0.089	0.078	0.057	0.063
2	0.058	0.078	0.084	0.354	0.089	0.079	0.061	0.081
3	0.058	0.085	0.086	0.353	0.087	0.077	0.062	0.102
4	0.060	0.092	0.087	0.355	0.082	0.074	0.063	0.126
5	0.067	0.101	0.089	0.351	0.081	0.071	0.066	0.142

Fixed-smoothing Asymptotics

T = 40

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.047	0.041	0.044	0.035	0.036	0.040	0.054
2	0.047	0.039	0.040	0.028	0.029	0.039	0.083
3	0.055	0.042	0.045	0.031	0.029	0.047	0.124
4	0.066	0.042	0.044	0.025	0.028	0.059	0.149
5	0.069	0.041	0.042	0.025	0.027	0.067	0.164

T = 120

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.051	0.045	0.044	0.041	0.044	0.042	0.055
2	0.060	0.047	0.048	0.041	0.042	0.044	0.073
3	0.065	0.050	0.046	0.040	0.042	0.048	0.095
4	0.071	0.045	0.045	0.034	0.036	0.048	0.114
5	0.078	0.046	0.044	0.034	0.035	0.049	0.131

Note: the Table reports the empirical size of the Diebold and Mariano test under a squared Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.5: Empirical size of the DM test with Linex loss function $\alpha = 1$

Standard Asymptotics

$T = 40$

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.055	0.071	0.092	0.375	0.092	0.070	0.055	0.054
2	0.068	0.076	0.092	0.383	0.087	0.069	0.056	0.078
3	0.095	0.090	0.106	0.397	0.091	0.069	0.065	0.111
4	0.123	0.101	0.111	0.398	0.090	0.075	0.078	0.137
5	0.143	0.106	0.110	0.395	0.091	0.075	0.085	0.151

$T = 120$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.047	0.060	0.069	0.345	0.080	0.069	0.052	0.056
2	0.050	0.066	0.074	0.358	0.079	0.070	0.054	0.066
3	0.052	0.070	0.074	0.360	0.075	0.065	0.054	0.081
4	0.055	0.074	0.075	0.363	0.072	0.065	0.052	0.096
5	0.060	0.077	0.071	0.360	0.073	0.062	0.051	0.107

Fixed-smoothing Asymptotics

$T = 40$

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.036	0.034	0.038	0.032	0.027	0.032	0.039
2	0.038	0.034	0.037	0.026	0.027	0.033	0.059
3	0.043	0.036	0.041	0.028	0.025	0.037	0.085
4	0.050	0.036	0.040	0.025	0.026	0.045	0.107
5	0.055	0.038	0.040	0.025	0.027	0.050	0.120

$T = 120$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.043	0.038	0.038	0.035	0.035	0.038	0.048
2	0.049	0.039	0.042	0.035	0.034	0.037	0.057
3	0.051	0.037	0.039	0.031	0.031	0.038	0.072
4	0.055	0.037	0.037	0.030	0.030	0.038	0.087
5	0.055	0.036	0.036	0.027	0.027	0.038	0.096

Note: the Table reports the empirical size of the Diebold and Mariano test under a Linex function, asymmetry parameter $\alpha = 1$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.6: Empirical size of the DM test with Linex loss function $\alpha = 0.5$

Standard Asymptotics

$T = 40$

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.068	0.086	0.109	0.364	0.111	0.086	0.069	0.072
2	0.083	0.094	0.109	0.367	0.105	0.082	0.071	0.093
3	0.115	0.109	0.128	0.388	0.111	0.091	0.086	0.129
4	0.140	0.123	0.131	0.389	0.109	0.090	0.099	0.158
5	0.165	0.129	0.131	0.385	0.106	0.091	0.104	0.170

$T = 120$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.052	0.066	0.075	0.335	0.087	0.078	0.060	0.060
2	0.059	0.077	0.083	0.348	0.089	0.080	0.064	0.075
3	0.063	0.081	0.086	0.352	0.084	0.076	0.061	0.088
4	0.069	0.088	0.087	0.354	0.086	0.079	0.066	0.107
5	0.073	0.092	0.086	0.352	0.084	0.074	0.066	0.122

Fixed-smoothing Asymptotics

$T = 40$

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.050	0.045	0.048	0.040	0.038	0.043	0.054
2	0.051	0.047	0.047	0.036	0.038	0.044	0.074
3	0.063	0.053	0.050	0.039	0.039	0.054	0.103
4	0.070	0.054	0.053	0.036	0.038	0.062	0.131
5	0.071	0.051	0.053	0.036	0.037	0.064	0.143

$T = 120$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.052	0.046	0.044	0.044	0.042	0.045	0.053
2	0.060	0.049	0.049	0.042	0.044	0.048	0.067
3	0.062	0.047	0.045	0.039	0.038	0.048	0.079
4	0.068	0.049	0.048	0.040	0.042	0.050	0.098
5	0.072	0.049	0.046	0.039	0.036	0.049	0.111

Note: the Table reports the empirical size of the Diebold and Mariano test under a Linex loss function, asymmetry parameter $\alpha = 0.5$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.7: Empirical size of the DM test with Linex loss function $\alpha = -0.9$

Standard Asymptotics

$T = 40$

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.056	0.074	0.098	0.369	0.094	0.075	0.057	0.058
2	0.076	0.086	0.098	0.381	0.092	0.070	0.062	0.085
3	0.098	0.094	0.108	0.391	0.095	0.078	0.071	0.115
4	0.120	0.101	0.107	0.394	0.090	0.073	0.080	0.135
5	0.148	0.106	0.112	0.402	0.090	0.074	0.085	0.149

$T = 120$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.048	0.064	0.073	0.355	0.088	0.074	0.057	0.054
2	0.053	0.071	0.080	0.357	0.086	0.075	0.056	0.071
3	0.056	0.072	0.075	0.355	0.079	0.069	0.055	0.081
4	0.062	0.084	0.082	0.366	0.082	0.071	0.058	0.104
5	0.070	0.089	0.084	0.373	0.080	0.070	0.063	0.119

Fixed-smoothing Asymptotics

$T = 40$

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.037	0.033	0.037	0.030	0.029	0.034	0.041
2	0.042	0.037	0.038	0.030	0.029	0.037	0.064
3	0.049	0.041	0.043	0.031	0.032	0.042	0.090
4	0.053	0.040	0.041	0.026	0.029	0.047	0.111
5	0.058	0.040	0.043	0.029	0.028	0.053	0.123

$T = 120$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.046	0.041	0.043	0.037	0.039	0.043	0.047
2	0.052	0.042	0.044	0.039	0.038	0.040	0.063
3	0.053	0.042	0.039	0.033	0.037	0.039	0.071
4	0.062	0.044	0.043	0.035	0.035	0.043	0.093
5	0.069	0.045	0.043	0.034	0.035	0.046	0.109

Note: the Table reports the empirical size of the Diebold and Mariano test under a Linex loss function, asymmetry parameter $\alpha = -0.9$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.8: Empirical size of the DM test with Linex loss function $\alpha = -1$

Standard Asymptotics

T = 40

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.054	0.071	0.093	0.368	0.091	0.072	0.054	0.055
2	0.071	0.081	0.095	0.381	0.088	0.067	0.059	0.082
3	0.093	0.091	0.104	0.392	0.090	0.073	0.067	0.110
4	0.116	0.097	0.103	0.395	0.085	0.070	0.075	0.132
5	0.145	0.103	0.107	0.404	0.086	0.069	0.082	0.143

T = 120

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.046	0.061	0.071	0.357	0.084	0.072	0.055	0.051
2	0.051	0.069	0.077	0.360	0.084	0.071	0.054	0.069
3	0.054	0.071	0.073	0.359	0.076	0.067	0.054	0.080
4	0.060	0.079	0.079	0.370	0.080	0.069	0.055	0.101
5	0.068	0.088	0.082	0.376	0.078	0.067	0.061	0.118

Fixed-smoothing Asymptotics

T = 40

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.035	0.031	0.034	0.028	0.027	0.031	0.039
2	0.039	0.034	0.036	0.028	0.026	0.034	0.061
3	0.046	0.039	0.040	0.029	0.030	0.038	0.086
4	0.049	0.036	0.039	0.025	0.028	0.044	0.107
5	0.054	0.038	0.041	0.027	0.026	0.049	0.119

T = 120

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.044	0.039	0.041	0.035	0.037	0.041	0.045
2	0.050	0.040	0.043	0.036	0.037	0.039	0.062
3	0.052	0.040	0.038	0.032	0.036	0.038	0.070
4	0.060	0.041	0.042	0.032	0.032	0.040	0.091
5	0.065	0.043	0.043	0.031	0.033	0.044	0.106

Note: the Table reports the empirical size of the Diebold and Mariano test under a Linex loss function, asymmetry parameter $\alpha = -1$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

all loss functions but tests tend to be slightly undersized for asymmetric loss functions and especially for the Linex loss function.

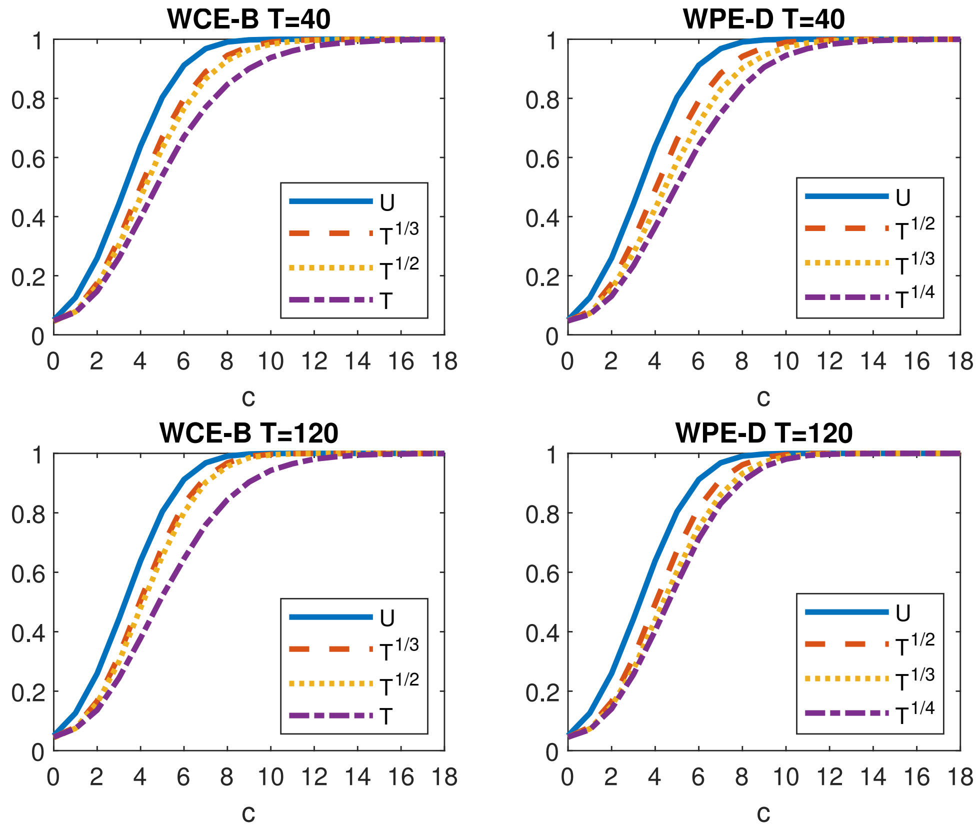
The power analysis is now only conducted under fixed-smoothing asymptotics since they proved capable of delivering correctly sized tests. Differently from the size study, in this simulation, forecast errors are generated directly from a bivariate Normal distribution $(e_{1t}, e_{2t})' \sim N(0_2, I_2)$, not considering contemporaneous and serial correlation. The theoretical size is set to 5% and the exercise involves $10,000$ replications.

The loss differential d_t, from Equation (3.12), is specified according to the loss functions described earlier and $cT^{-1/2}$ is added to generate local alternatives and to obtain $\mu = cT^{-1/2}$ testing the null hypothesis $H_0 : \mu = 0$. c is an integer number that starts from zero (H_0 true) and increases until the power reaches 1 for each loss function.

Figures 3.4 - 3.11 report power performances of the DM test under fixed-smoothing asymptotics. In every Figure, the solid blue line identified by 'U' represents the infeasible case in which the real unknown variance is used and critical values are taken from a standard Normal distribution. As expected, tests have the highest possible power when the true long run variance is used. The true unknown variance σ^2 is simulated using $500,000$ replications and two sets of forecast errors of $10,000$ observations each generated from a standard bivariate Normal distribution. The other lines report power for fixed-smoothing asymptotics when the long run variance estimate is obtained from a Weighted Covariance Estimator using the Bartlett kernel bandwidths $M = \lfloor T^{1/3} \rfloor$, $M = \lfloor T^{1/2} \rfloor$ and $M = T$ or a Weighted Periodogram Estimator with bandwidths $m = \lfloor T^{1/4} \rfloor$, $m = \lfloor T^{1/3} \rfloor$ and $m = \lfloor T^{1/2} \rfloor$ for the Daniell kernel.

In general, all loss functions exhibit good power, the best performances are obtained by asymmetric loss functions and in particular Linex with $\alpha = 0.5$. Also the absolute loss function shows good power. For WCE estimators, the orange dashed line, which represents the bandwidth $M = \lfloor T^{1/3} \rfloor$, is always the closest to the infeasible solid blue line, while the purple dot-dashed line of bandwidth $M = T$ is always the furthest indicating poor power performances. The yellow dotted line for the bandwidth $M = \lfloor T^{1/2} \rfloor$ remains

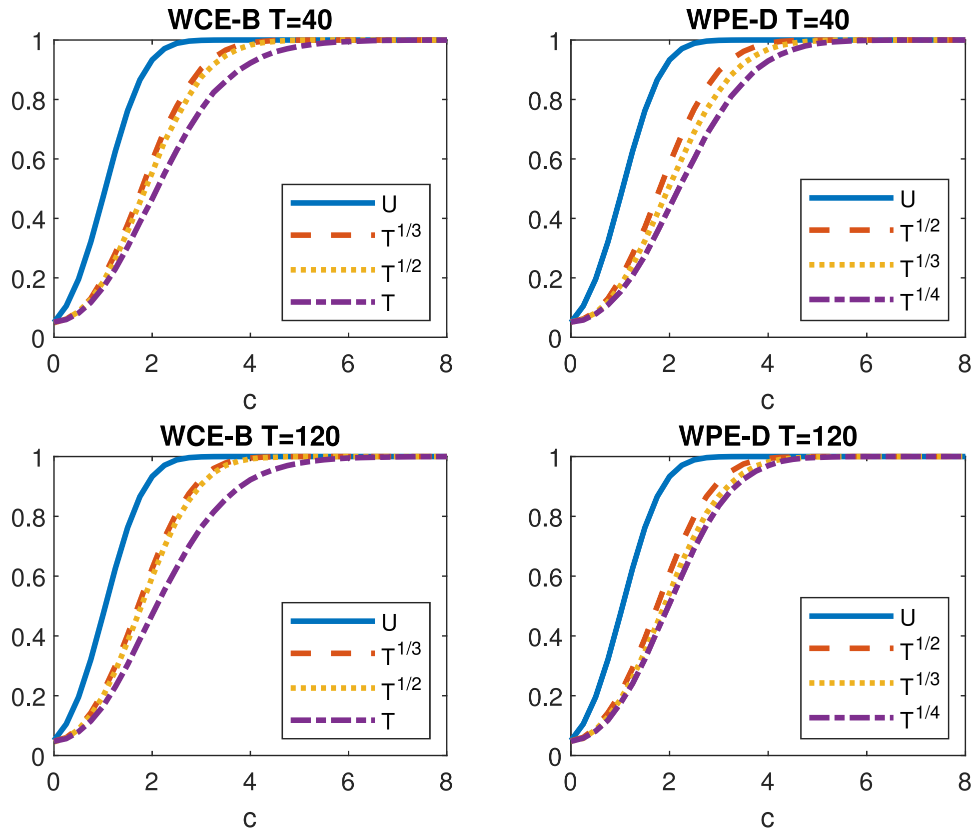
Figure 3.4: DM finite sample local power and quadratic loss function



Note: the Figure reports power performances of the Diebold and Mariano test with quadratic loss function in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

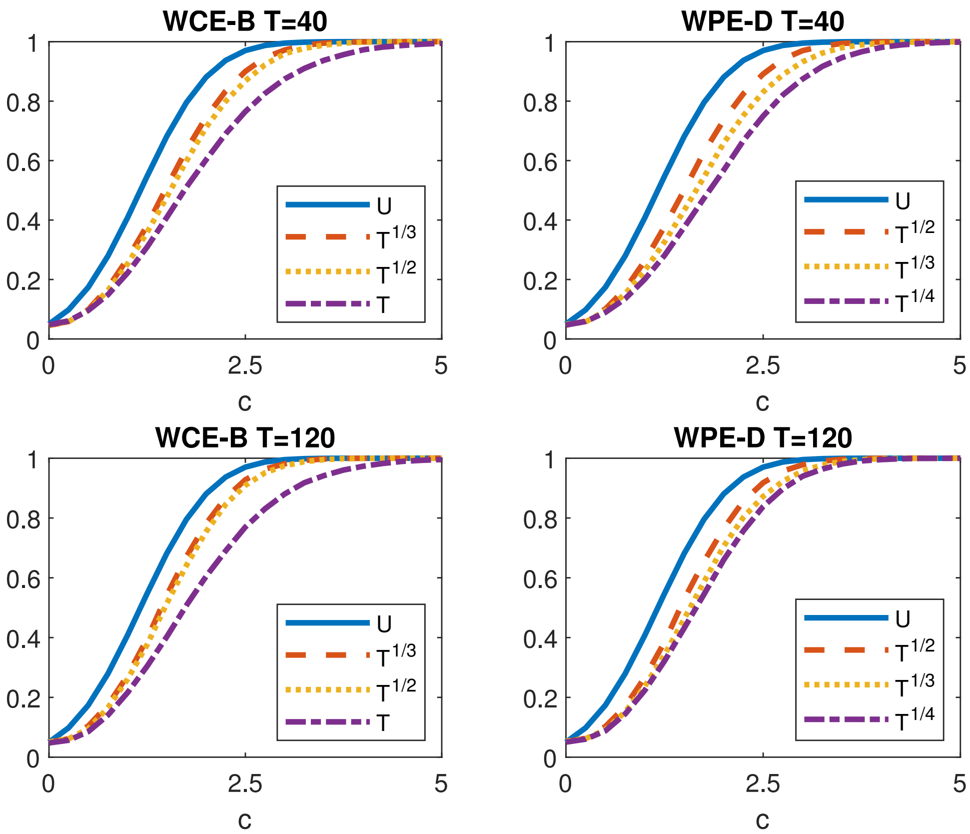
in the middle of the two former lines but close to the orange dashed one. Taking also into consideration results for the empirical size, the WCE bandwidth offering the best size-power performances is $M = \lfloor T^{1/2} \rfloor$ consistently for all loss functions as in Coroneo and Iacone (2020). The WPE bandwidth leading to the best power is $m = \lfloor T^{1/2} \rfloor$ (orange dashed line), while the one leading to the lowest power is $m = \lfloor T^{1/4} \rfloor$ (dot-dashed purple line). The yellow dotted line for powers of tests using bandwidth $m = \lfloor T^{1/3} \rfloor$ remains in the middle of the previous lines. However, considering empirical size and power together, the bandwidth $m = \lfloor T^{1/3} \rfloor$ should be preferred, in line with existing literature (Coroneo and Iacone, 2020; Lazarus, Lewis and Stock, 2019). Results remain consistent across loss functions. However, for the absolute loss, the Lin-Lin loss and in particular for the Linex loss with $\alpha = 0.5$, the power reaches 1 for small values of c. This exercise confirms that findings from Harvey, Leybourne and Whitehouse (2017) and Coroneo and Iacone (2020) extend to asymmetric loss functions and forecast accuracy evaluation can be carried out with this particular type of loss functions.

Figure 3.5: DM finite sample local power and absolute loss function



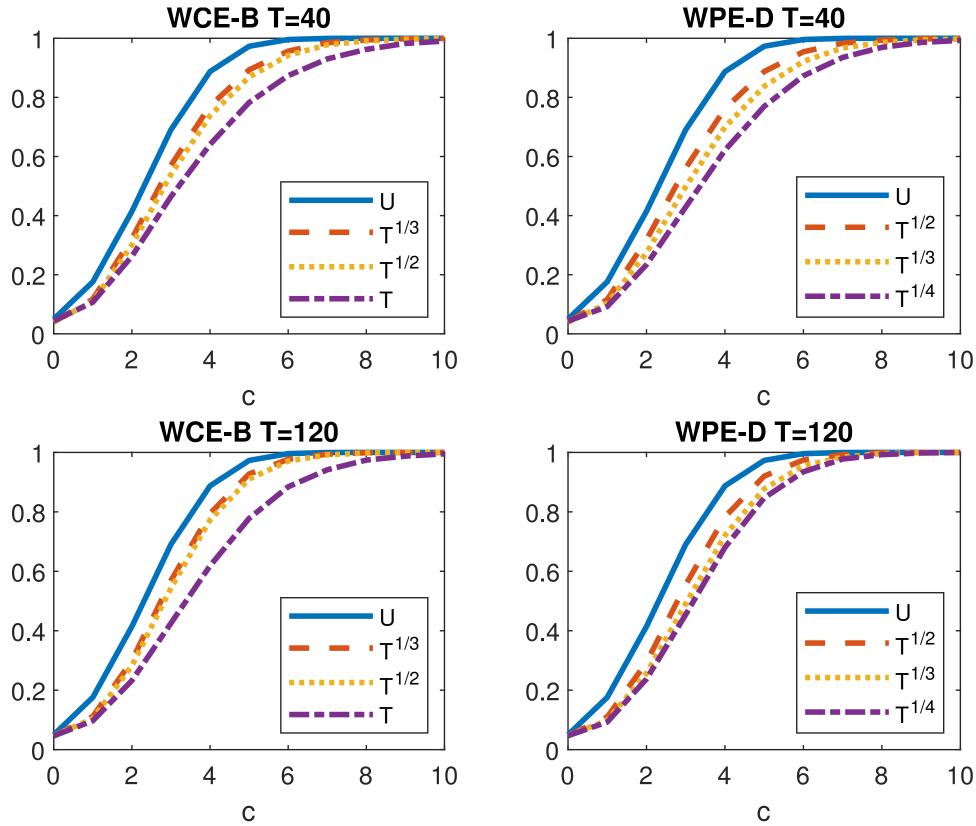
Note: the Figure reports power performances of the Diebold and Mariano test with absolute loss function in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.6: DM finite sample local power and Lin-Lin loss function $\alpha = 0.9$



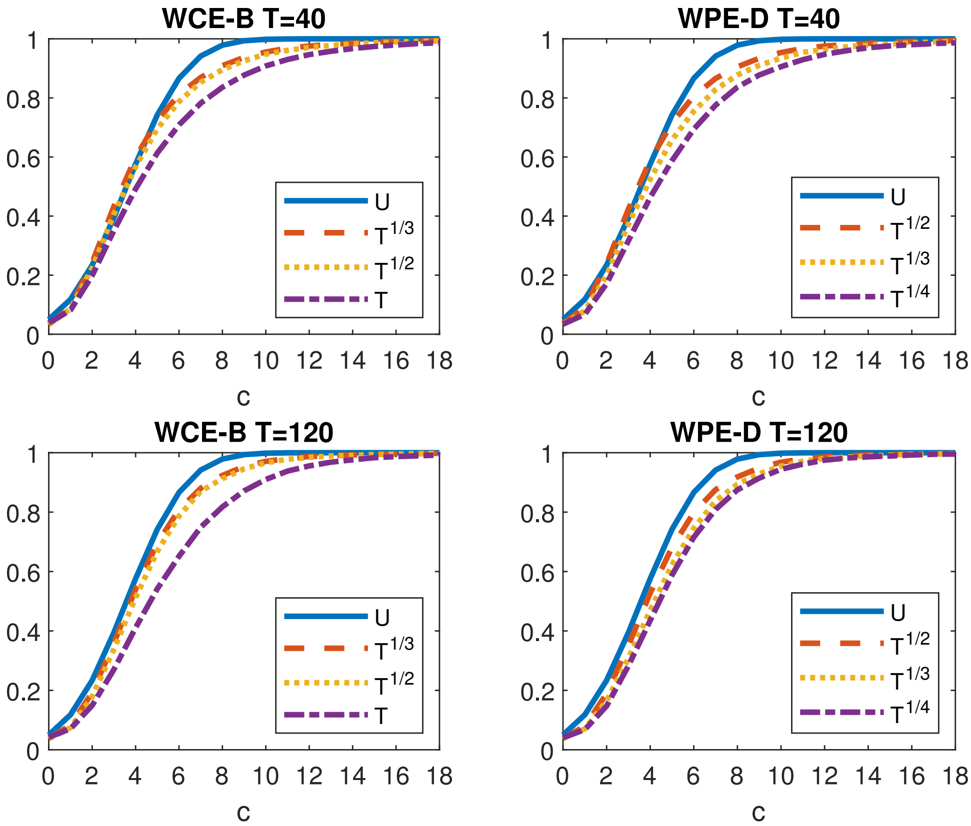
Note: the Figure reports power performances of the Diebold and Mariano test with Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.7: DM finite sample local power and squared Lin-Lin loss function, $\alpha = 0.9$



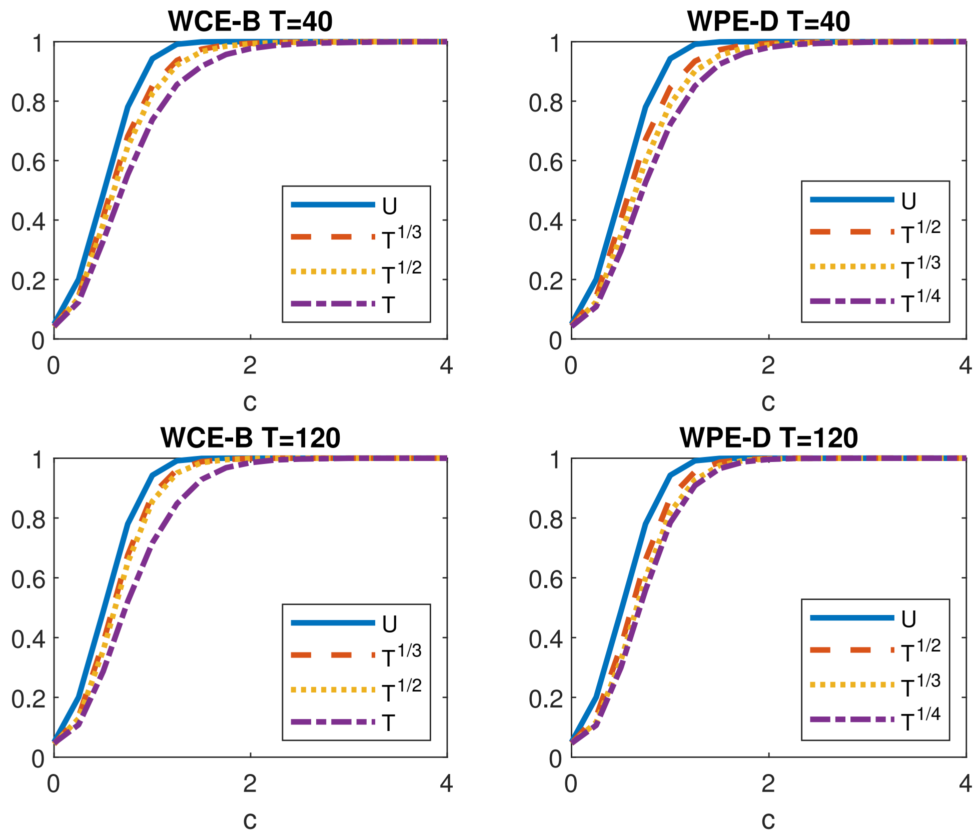
Note: the Figure reports power performances of the Diebold and Mariano test with squared Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.8: DM finite sample local power and Linex loss function, $\alpha = 1$



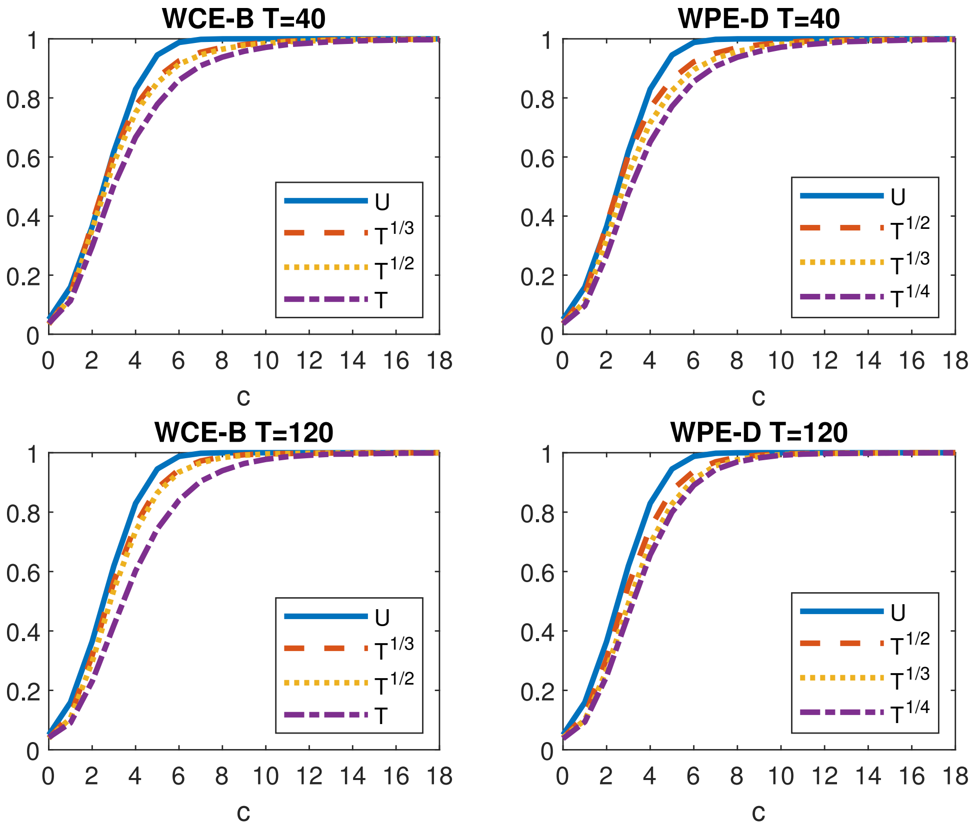
Note: the Figure reports power performances of the Diebold and Mariano test with Linex loss function, asymmetry parameter $\alpha = 1$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.9: DM finite sample local power and Linex loss function, $\alpha = 0.5$



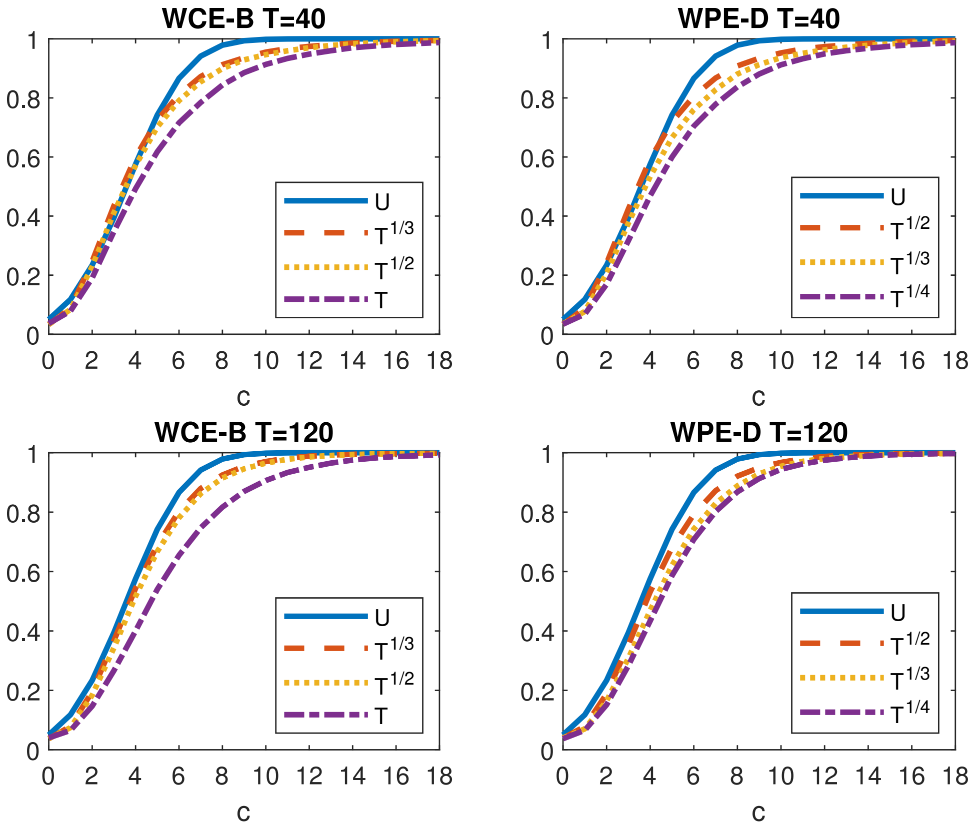
Note: the Figure reports power performances of the Diebold and Mariano test with Linex loss function, asymmetry parameter $\alpha = 0.5$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.10: DM finite sample local power and Linex loss function, $\alpha = -0.9$



Note: the Figure reports power performances of the Diebold and Mariano test with Linex loss function, asymmetry parameter $\alpha = -0.9$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.11: DM finite sample local power and Linex loss function, $\alpha = -1$



Note: the Figure reports power performances of the Diebold and Mariano test with Linex loss function, asymmetry parameter $\alpha = -1$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used and the test statistic has standard normal limiting distribution. All other lines represent different bandwidths used in the estimation of the long run variance.

3.6 Empirical Results

The European Central Bank runs the Survey of Professional Forecasters to collect private sector's expectations on macroeconomic variables and to assess the credibility of the inflation target the bank sets. For a complete description of this survey, see Chapter 1.

In this Section, I test the accuracy of HICP inflation, unemployment rate and real GDP growth aggregate point forecasts in the ECB SPF surveys from 2002.Q1 until 2010.Q3 for a total of 35 observations obtained from the ECB website. Core inflation has been excluded from this analysis given the limited number of surveys available. The aggregate forecasts reported in the survey are obtained from the unweighted average of each respondent's forecasts which proved to be a good way to obtain more accurate forecasts from individual ones (Clemen, 1989; Aiolfi, Capistrán and Timmermann, 2011; Manski, 2011; Genre, Kenny, Meyler and Timmermann, 2013; Meyler, 2020). The horizons considered are rolling horizons one-year and two-years forecasts and long term forecasts (five years). The choice of the start and the end date of the sample relies on the availability of realisations in different vintages in the Real-time Database for the euro area available on the European Central Bank Statistical Data Warehouse; realised data about inflation spans from December 2002 to June 2015 (end of quarter months of the HICP annual growth), about unemployment, from November 2002 to May 2015 (middle of quarter months of the unemployment rate) and about real GDP growth, from 2002.Q3 to 2015.Q1. To carry out this empirical exercise, I consider several alternative series of historical realisations according to their vintage: the first-ever release of the realised target variable, four releases after the first, twenty releases after the first and latest release available at 01/02/2018.

Competing forecasts are obtained from benchmarks models described in Section 3.2: a random walk model, an IAR model estimated using a rolling window of the last 30 quarterly observations at vintage V, and a DAR model estimated using a rolling window of the last $30 - h$ quarterly observations at vintage V. Estimation and forecasting are performed in real-time, i.e. using only data and vintage V available to forecasters when they had to submit their forecast, in order to have comparable alternative forecasts.

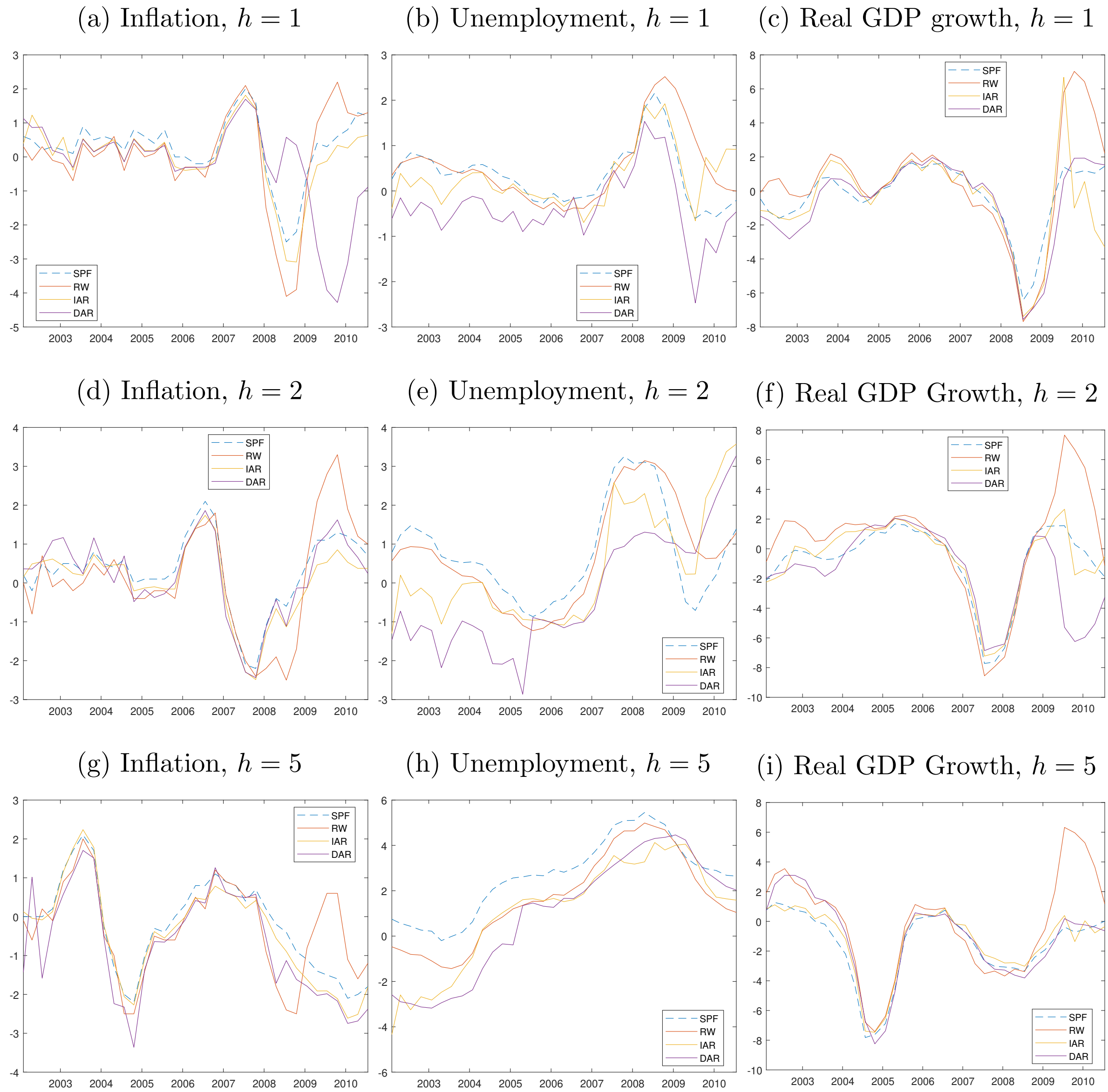
To test the null hypothesis of equal forecast accuracy of the ECB SPF and benchmark models forecasts, I employ the Diebold and Mariano test presented in Section 3.4 using the different loss functions described in Section 3.3: quadratic, absolute, Lin-Lin, squared Lin-Lin and Linex loss functions.

The long run variance involved in the test statistic is estimated using a Weighted Covariance Estimator with Bartlett kernel (WCE-B bandwidth $M = \lfloor T^{1/2} \rfloor$) and a Weighted Periodogram Estimator with Daniell kernel (WPE-D bandwidth $m = \lfloor T^{1/3} \rfloor$). Bandwidths are chosen as advised in Coroneo and Iacone (2020) and taking into consideration results of my Monte Carlo simulation. Fixed-b asymptotics critical values used with the WCE-B variance estimator come from simulations in Kiefer and Vogelsang (2005) and fixed-m asymptotics critical values for WPE-D variance estimator are taken from a Student-t distribution as advised in Hualde and Iacone (2017). Revision of realised data does not affect the asymptotic distribution of DM type tests and asymptotic distribution results from previously cited papers remain valid according to Clark and McCracken (2009).

Figure 3.12 reports forecast errors for the last release of the three target variables. The impact of the global financial crisis is evident for all variables and especially for the real GDP growth as the largest errors appear in the 2008 - 2010 period. In general, it seems models obtain forecast errors similar to the SPF ones and close to zero before 2007 for one-year ahead forecasts. Forecast errors start to increase in magnitude after then. A similar pattern is observed for two-years ahead forecasts. For longer horizons, all models and SPF forecasts produce quite large errors of opposite signs that cancel out over the whole sample. This feature of forecast errors can potentially change the result of the test according to the loss function used.

Figures 3.13 - 3.30 report the Diebold and Mariano test using all the different loss functions and the four different releases of the realised variable. Blue stars represent the DM test statistic for the comparison of the random walk forecasts to the SPF forecasts.

Figure 3.12: Forecast errors for HICP inflation, unemployment rate and real GDP growth



Note: the plot reports forecast errors for all three forecast horizons h, expressed in years, for HICP inflation, unemployment rate and real GDP growth. The vintage of the target variable is taken at the current release available on 01/02/2018. The dotted blue line is the forecast error from the SPF, the red solid line is the forecast error from the random walk forecasts, the solid yellow line is the forecast error from the IAR model and the purple solid line is the forecast error from the DAR model.

Yellow dots and red crosses are the calculated DM test statistics for the comparison of the SPF forecasts to the IAR and DAR forecasts respectively. If the test statistics indicators fall in the positive section of the plot, the error associated with the benchmark forecast was bigger than the error associated with the SPF forecast and vice versa. The solid black, the dot dashed blue and the dashed red lines represent the 20%, 10% and

5% fixed-smoothing asymptotics two sides critical values respectively. WCE-B refers to the Weighted Covariance Estimator with Bartlett kernel for the long run variance with bandwidth $M = \lfloor T^{1/2} \rfloor$ and WPE-D refers to Weighted Periodogram Estimator with Daniell kernel for the long run variance with bandwidth $m = \lfloor T^{1/3} \rfloor$.

Figures 3.13 - 3.18 refer to HCPI inflation test results. Revision in the realised data is not generally affecting the outcome of the test, possibly because it is modest in this variable. The null hypothesis of equal forecast accuracy is not rejected for any loss function in one-year and two-years ahead tests. However, the DM test statistic is generally in the positive section of the plot indicating that SPF forecasts are generating, on average, smaller forecast errors. With asymmetric loss functions that penalise overprediction, there is a rejection at 20% significance in favour of the IAR benchmark, which is indicating that the IAR model is slightly underpredicting inflation with respect to the SPF forecasts. For long term forecasts, ECB SPF perform better than the DAR model under quadratic, absolute and Linex with $\alpha = -1$ in which negative forecast errors are more penalised, which suggests that DAR forecasts are on average bigger than the realised inflation.

Figures 3.19 - 3.24 report DM test statistic for the unemployment rate; for this variable, revision is greater than inflation and it makes the value of the test change noticeably. With quadratic, absolute and Linex loss with $\alpha = -1$ the outcome of the test for one-year ahead forecasts changes. For other horizons, results are less impacted and, the Lin-lin loss, the squared Lin-Lon loss and the Linex loss function with $\alpha = 1$ seem to be more robust to revision.

For short and medium term forecasts, there is no rejection of the null of equal forecast accuracy. However, the test statistic is usually in the positive part of the plot suggesting that SPF forecast errors are smaller than benchmarks forecast errors. This is especially true for the DAR model, which takes the test statistics in the 20% rejection region in some cases, for early releases. In particular, the DAR model seems to underestimate unemployment because the test based on early releases falls in the 10% rejection region in favour of SPF with a Linex loss with $\alpha = -1$.

For long term forecasts, instead, benchmarks seem more accurate than SPF. In the comparison with random walk forecasts, the test strongly rejects the null hypothesis for all loss functions except the Linex loss with $\alpha = -1$ for which the test statistics lay in the positive side of the plot in support of ECB SPF forecasts. Other benchmarks are also performing well compared to SPF.

Real GDP growth test results are reported in Figures 3.25 - 3.30. The effect of revision is negligible for short term and long term forecasts and moderate for medium term forecasts. Absolute, Lin-Lin and squared Lin-Lin loss functions are the most sensitive to revision. In particular the absolute loss and the Lin-Lin loss seem to favour SPF forecasts as data is being revised. For one and two-years ahead forecasts, there is evidence in favour of the predictive ability of the ECB SPF, in particular with asymmetric loss function that penalise positive forecast errors indicating that benchmark models tend to overpredict GDP growth. In the long term, predictive ability decreases and test statistics show support for benchmark forecasts except with the Lin-Lin loss function, which favours the ECB SPF over the random walk.

In general, results are robust to the loss function used and to the revision of the realised variable used in the evaluation. Unemployment seems to be the variable most affected by revision. In fact, it is the variable with most revision of all three considered. ECB SPF seem to provide reliable long term forecasts for inflation, however, for unemployment and real GDP growth, benchmark forecasts and in particular IAR forecasts seem to outperform SPF. For short and medium term forecasts, the DM test does not reject the null hypothesis of equal forecast accuracy but the random walk model show smaller forecast errors than other benchmarks confirming findings of Atkeson and Ohanian (2001) and Balcilar, Gupta, Majumdar and Miller (2015) that random walk is hard to beat.

Existing literature about the evaluation of ECB SPF is scarce because of the small sample available. Bowles, Friz, Genre, Kenny, Meyler and Rautanen (2011) evaluate real GDP growth and unemployment over the period 1999.Q1 - 2008.Q4 without using a specific forecast accuracy test like the DM one but my findings confirm their results. They focus on forecast errors and statistics like the Mean Squared Error and Mean Absolute Error. Their findings suggest ECB SPF are generally superior to naive benchmarks however, forecast errors are quite persistent. They also find that the performance of the aggregate forecasts replicates the one of individual forecasts because of the information available or not available to single respondents is the same.

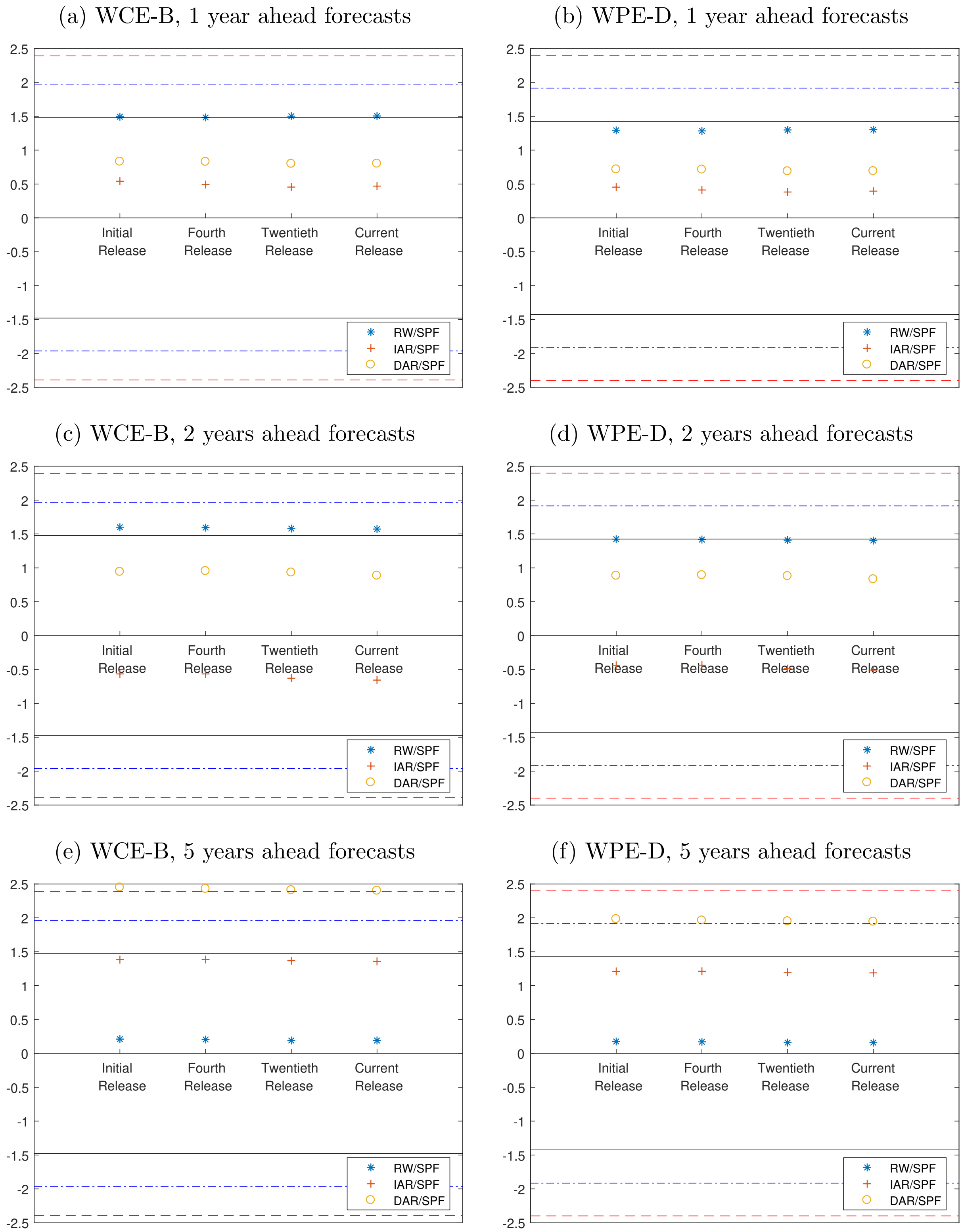
Coroneo and Iacone (2020) also perform a real-time evaluation of the ECB SPF forecasts over the period 2006.Q1 - 2016.Q4 using the same procedure of this work but not taking into account asymmetric loss functions or different vintages of the realisation of the target variable. They find limited evidence that ECB SPF forecasts are better than the ones from a random walk model.

Several authors evaluated the performance of US SPF, instead. Stark (2010) evaluates the US SPF forecasts from 1985.Q1 to 2007.Q4 using root mean square forecast errors from several different vintages of the realised series. They find that US SPF are good forecasts and they always outperform all benchmark models. Revision has no effect on unemployment and very small on inflation while it has a strong effect on real GDP. For the latter variable, US SPF forecasts become more inaccurate as new revisions are released. Results are consistent across different benchmark models.

Also Coroneo and Iacone (2020) assess the predictive accuracy of the US SPF forecasts over the period 1987.Q1 - 2016.Q4 against a random walk benchmark. They find that, especially for short horizons, the predictive ability of SPF is strong but it worsen as the horizon increases confirming findings of D'Agostino, Giannone and Surico (2006) and Demetrescu, Hanck and Kruse (2018).

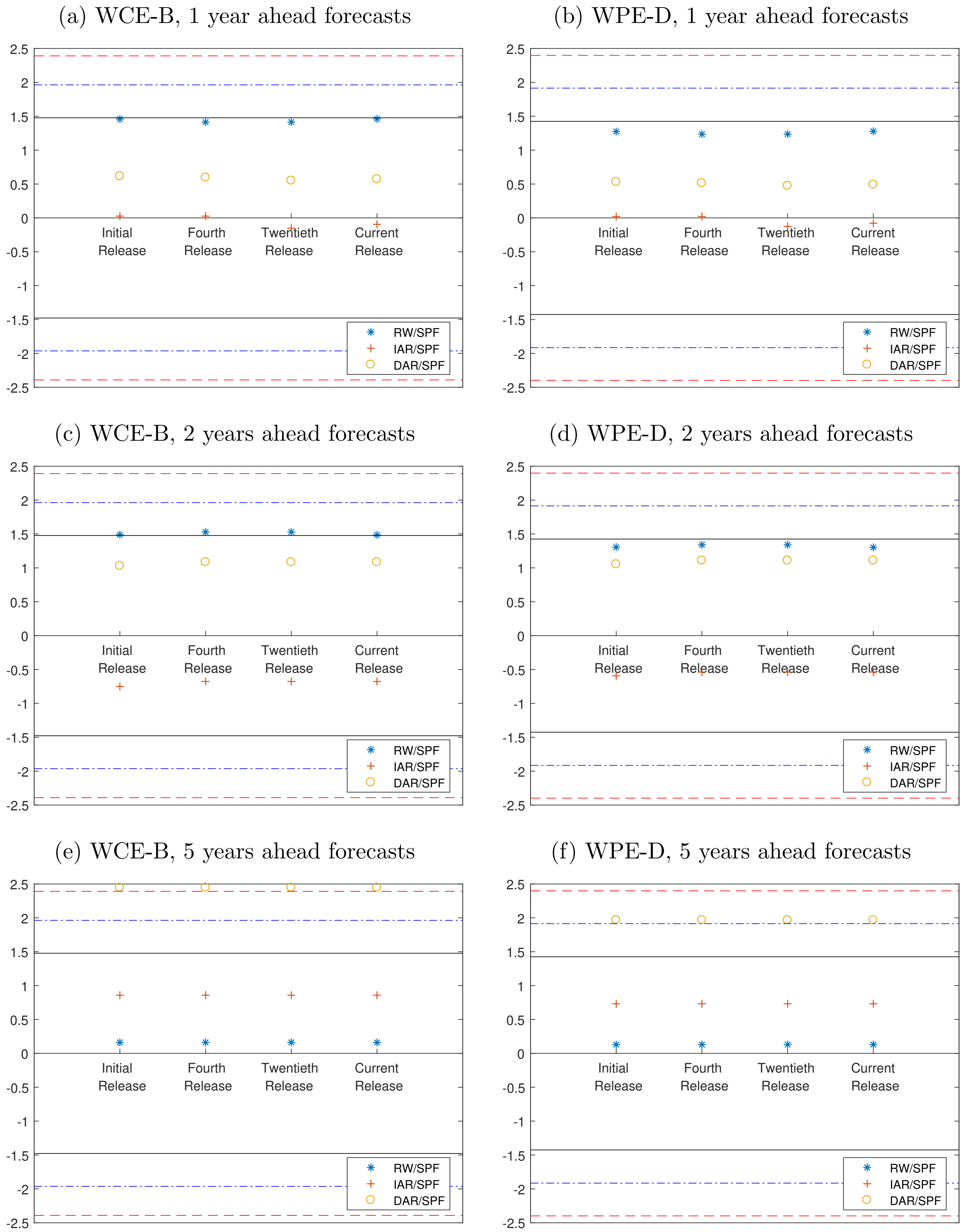
Given this evidence, US SPF appears to provide forecasts as accurate as ECB SPF but, in this comparison it is important to take into account that horizons for US SPF are expressed in quarters and the longest horizon is four quarters ahead i.e. one year ahead. For ECB SPF instead, the shortest horizon is one year and the longest is five years. Considering this difference, it seems that ECB SPF forecasts remain accurate for longer horizons than US SPF despite their more recent inception.

Figure 3.13: DM test statistic for inflation and quadratic loss function



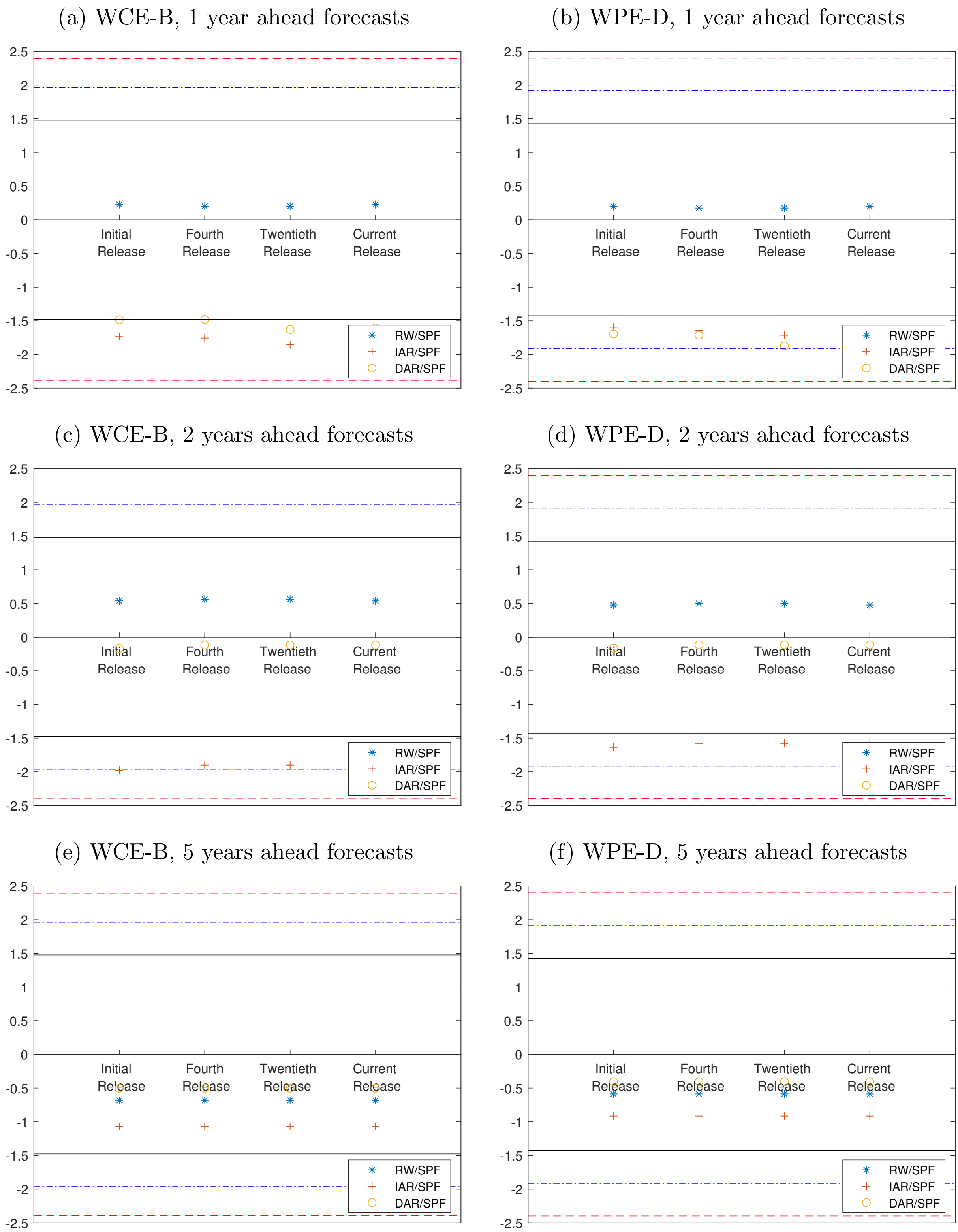
Note: plots report DM test statistic for inflation and quadratic loss function for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.14: DM test statistic for inflation and absolute loss function



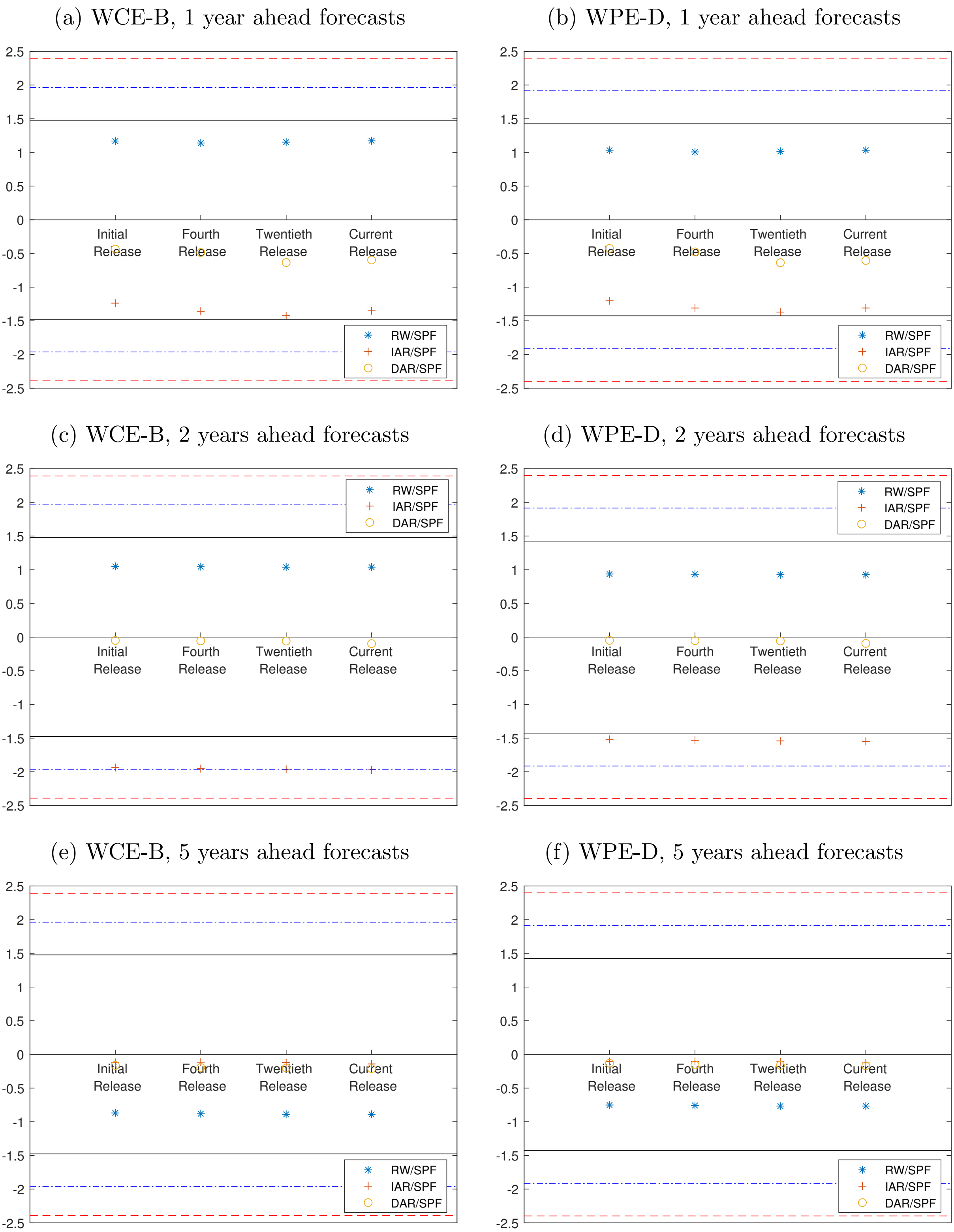
Note: plots report DM test statistic for inflation and absolute loss function for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.15: DM test statistic for inflation and Lin-Lin loss function



Note: plots report DM test statistic for inflation and Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

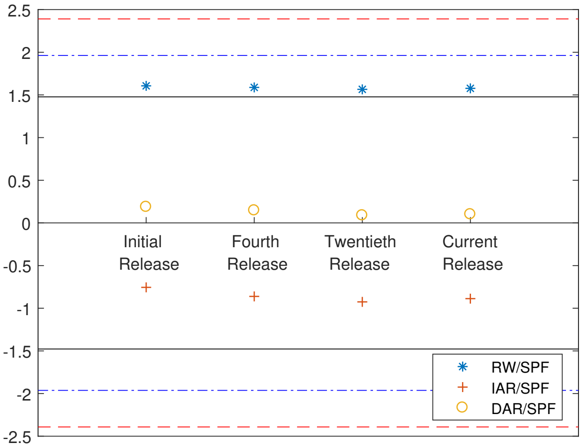
Figure 3.16: DM test statistic for inflation and squared Lin-Lin loss function



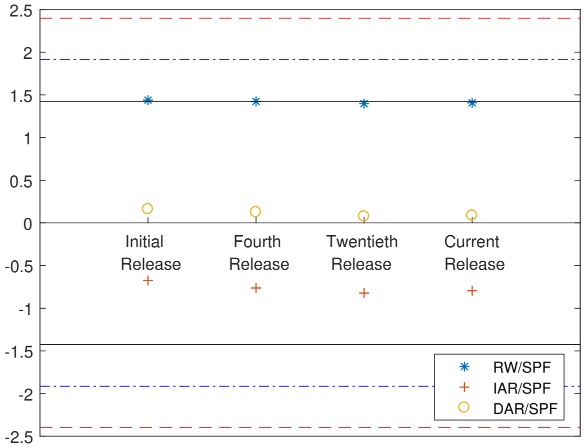
Note: plots report DM test statistic for inflation and squared Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.17: DM test statistic for inflation and Linex loss function with $\alpha = 1$

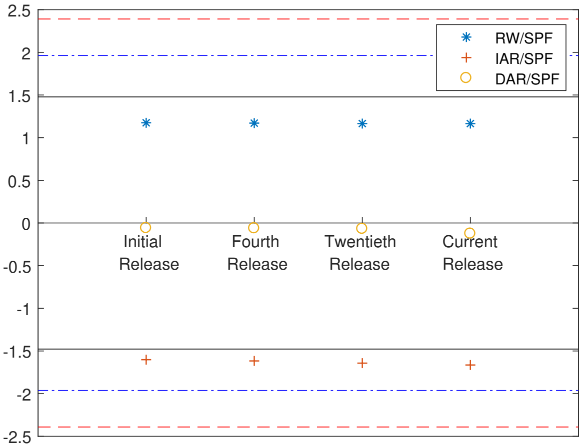
(a) WCE-B, 1 year ahead forecasts



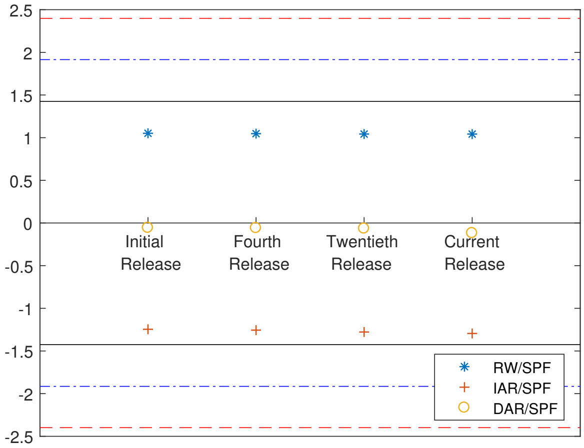
(b) WPE-D, 1 year ahead forecasts



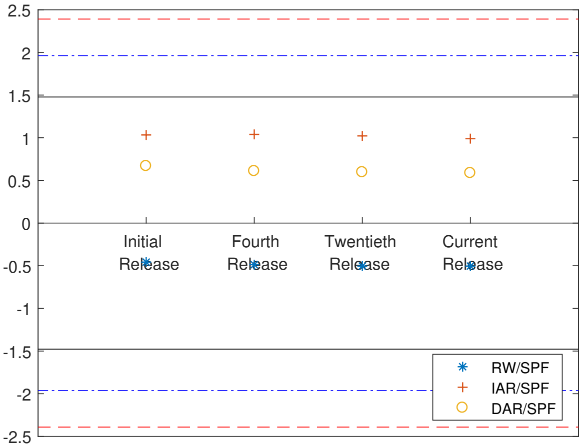
(c) WCE-B, 2 years ahead forecasts



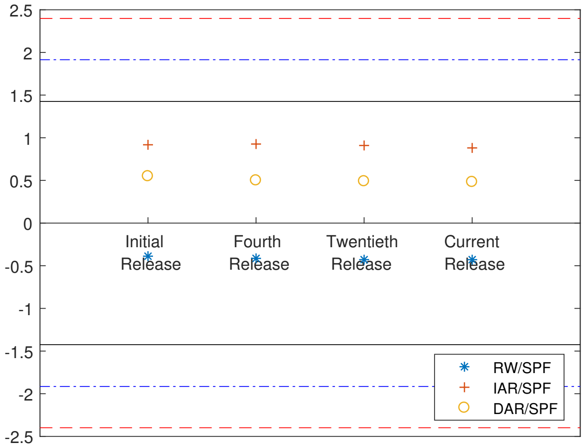
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts



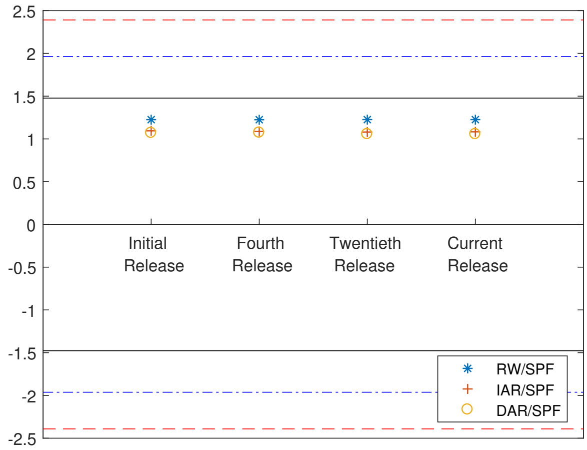
(f) WPE-D, 5 years ahead forecasts



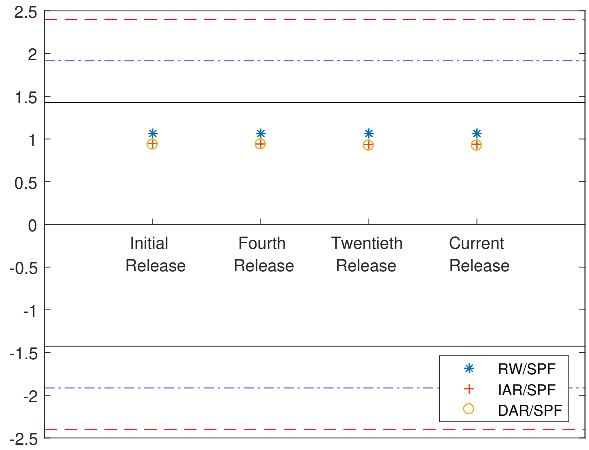
Note: plots report DM test statistic for inflation and Linex loss function with asymmetry parameter $\alpha = 1$ for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.18: DM test statistic for inflation and Linex loss function with $\alpha = -1$

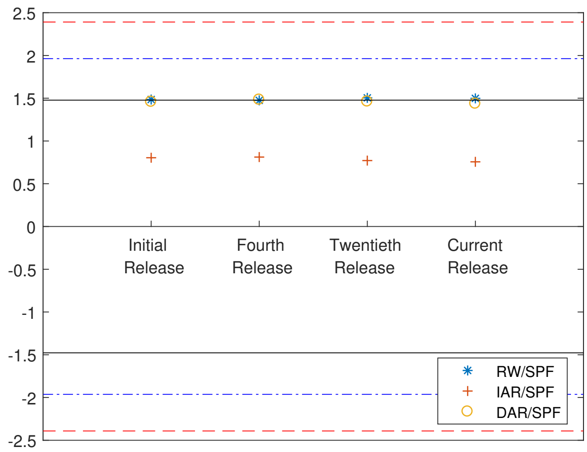
(a) WCE-B, 1 year ahead forecasts



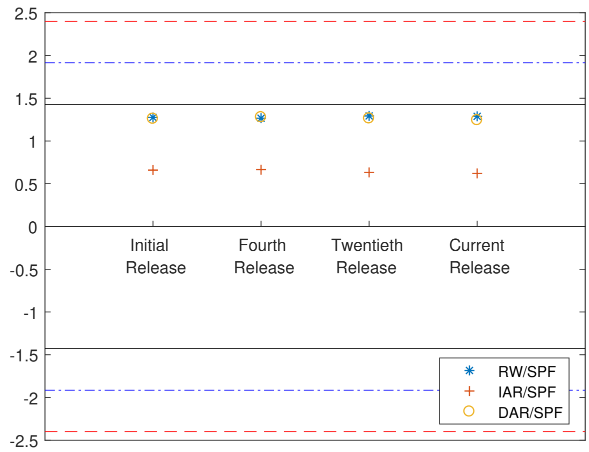
(b) WPE-D, 1 year ahead forecasts



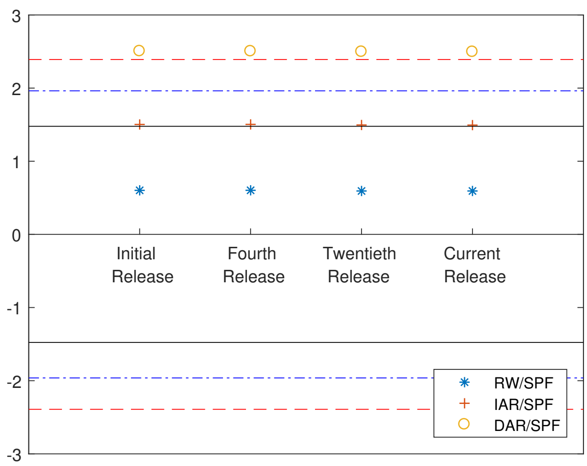
(c) WCE-B, 2 years ahead forecasts



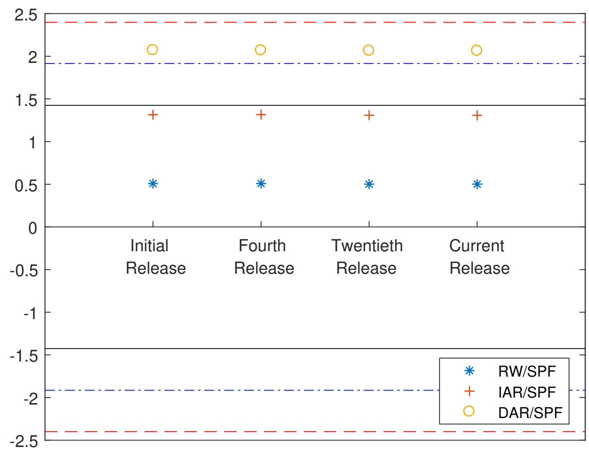
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts

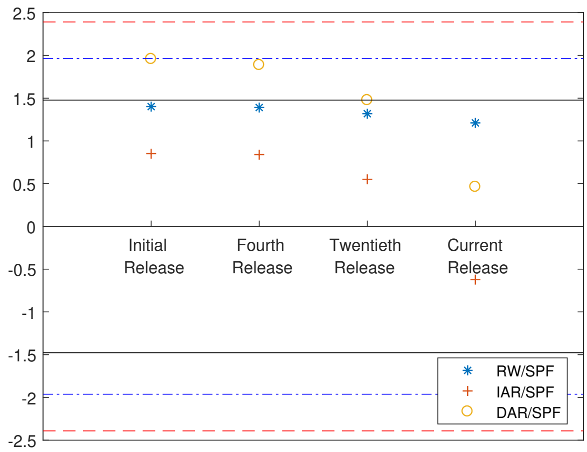


(f) WPE-D, 5 years ahead forecasts

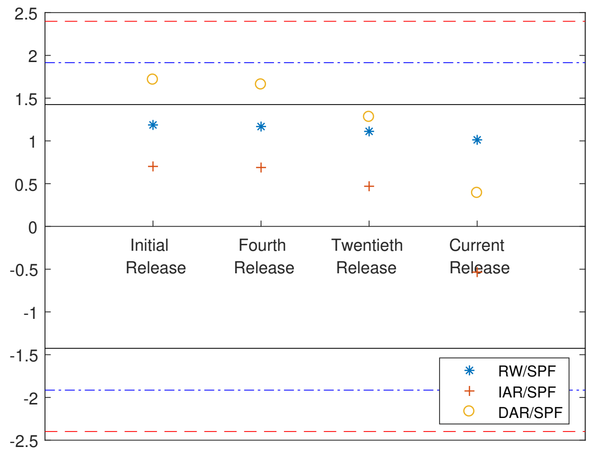


Note: plots report DM test statistic for inflation and Linex loss function with asymmetry parameter $\alpha = -1$ for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

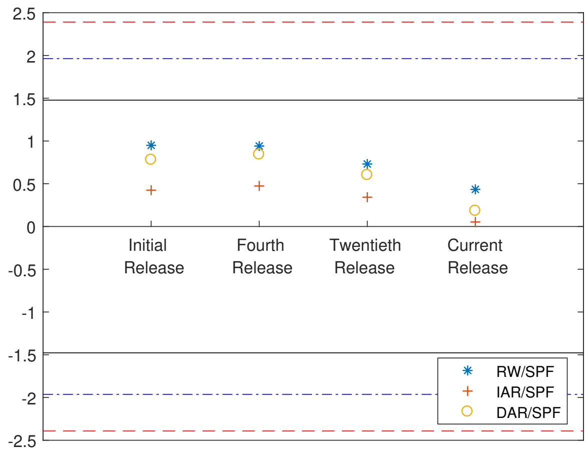
Figure 3.19: DM test statistic for unemployment and quadratic loss function



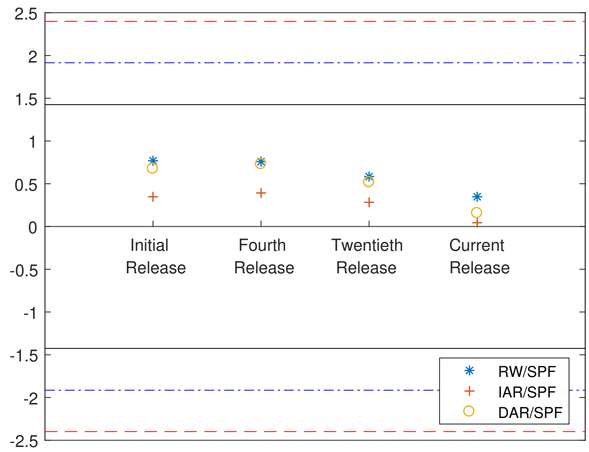
(a) WCE-B, 1 year ahead forecasts



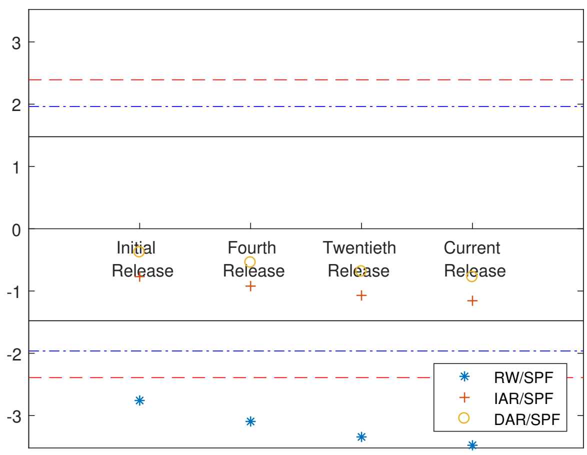
(b) WPE-D, 1 year ahead forecasts



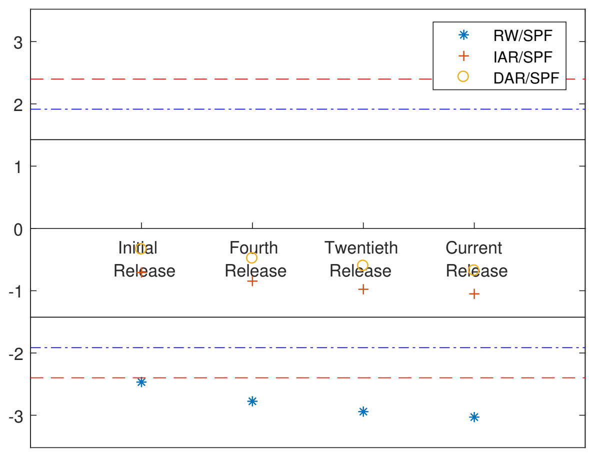
(c) WCE-B, 2 years ahead forecasts



(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts

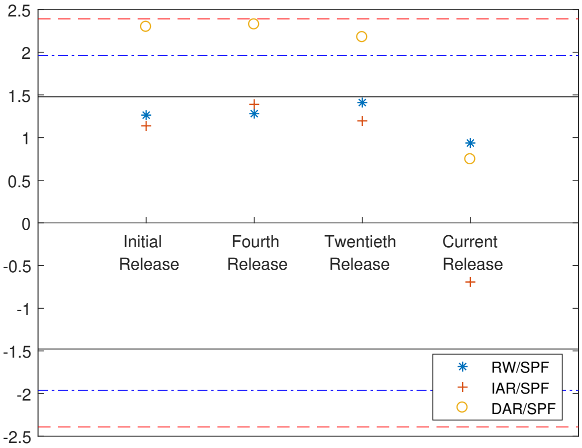


(f) WPE-D, 5 years ahead forecasts

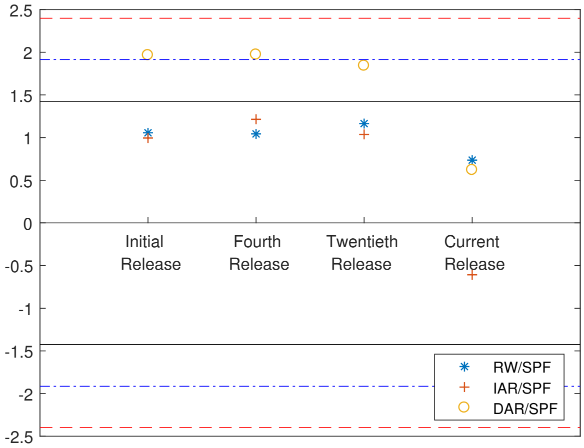
Note: plots report DM test statistic for unemployment and quadratic loss function for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.20: DM test statistic for unemployment and absolute loss function

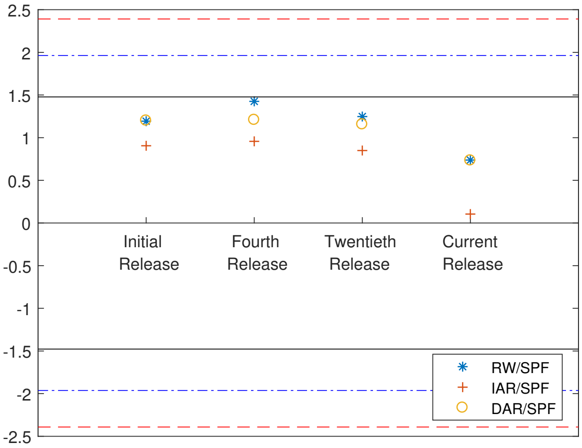
(a) WCE-B, 1 year ahead forecasts



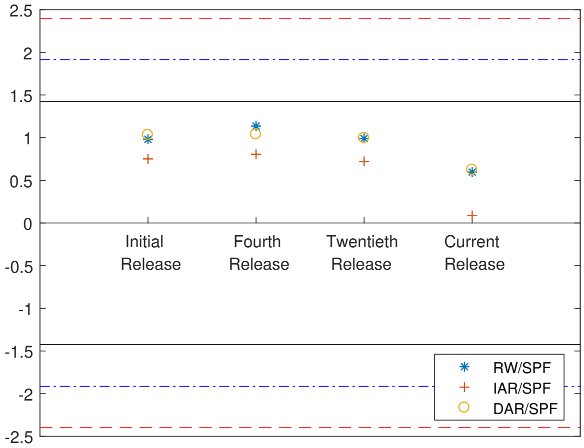
(b) WPE-D, 1 year ahead forecasts



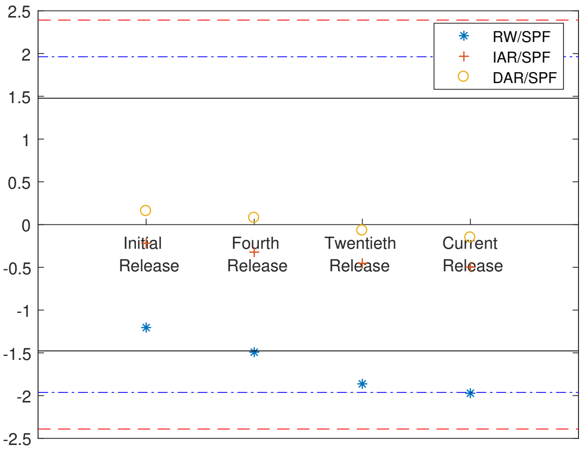
(c) WCE-B, 2 years ahead forecasts



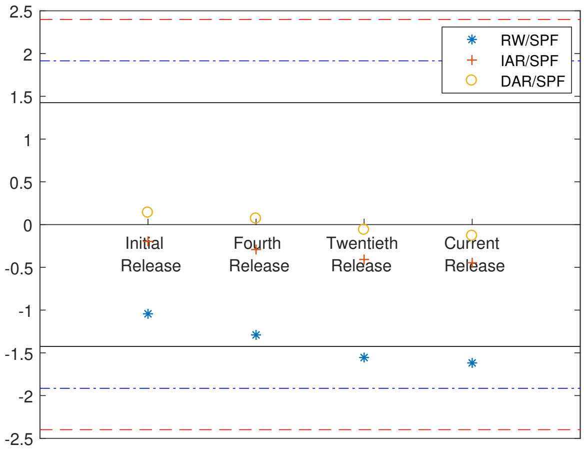
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts



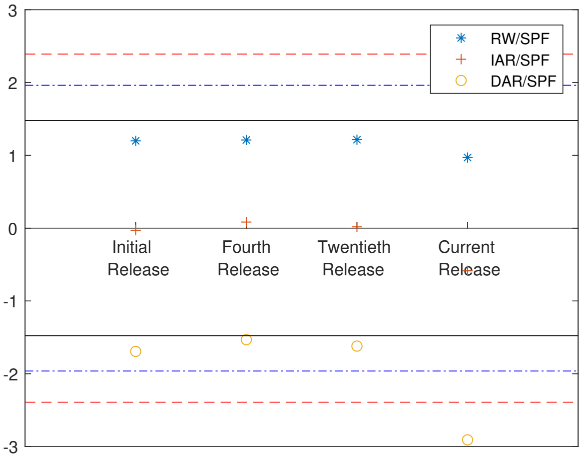
(f) WPE-D, 5 years ahead forecasts



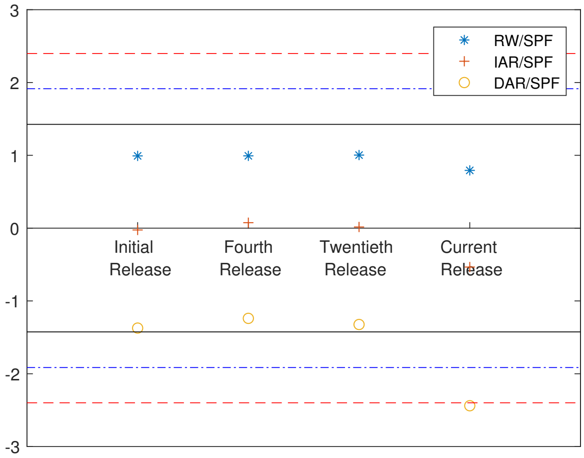
Note: plots report DM test statistic for unemployment and absolute loss function for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.21: DM test statistic for unemployment and Lin-Lin loss function

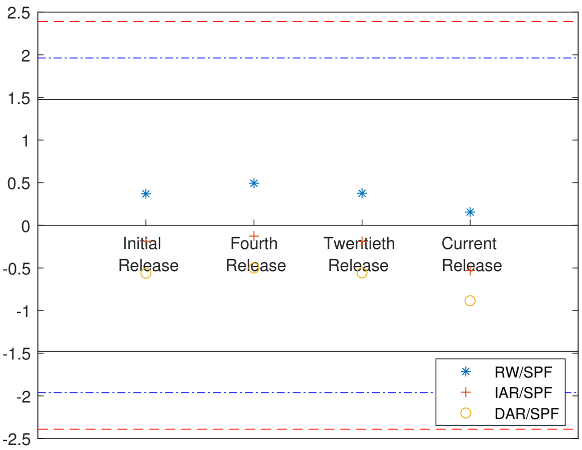
(a) WCE-B, 1 year ahead forecasts



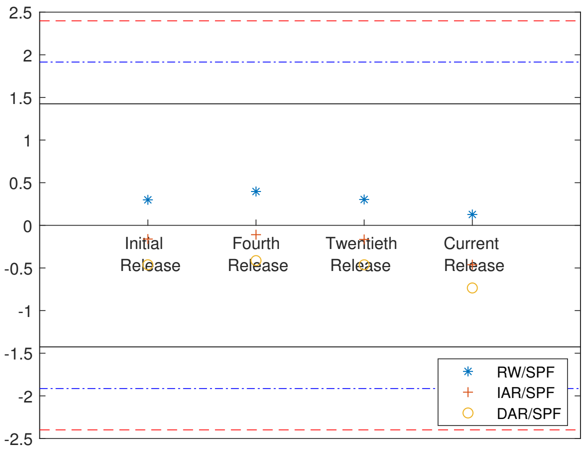
(b) WPE-D, 1 year ahead forecasts



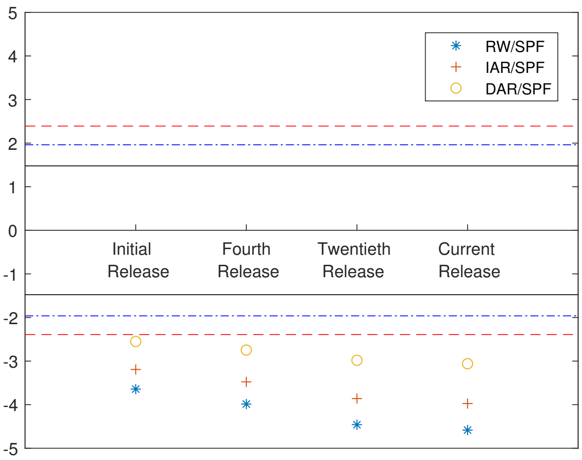
(c) WCE-B, 2 years ahead forecasts



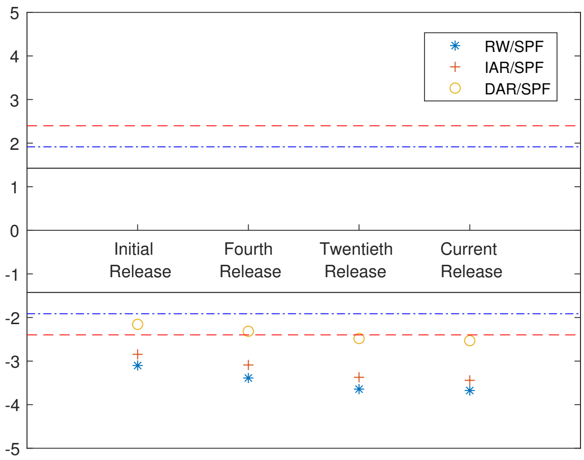
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts



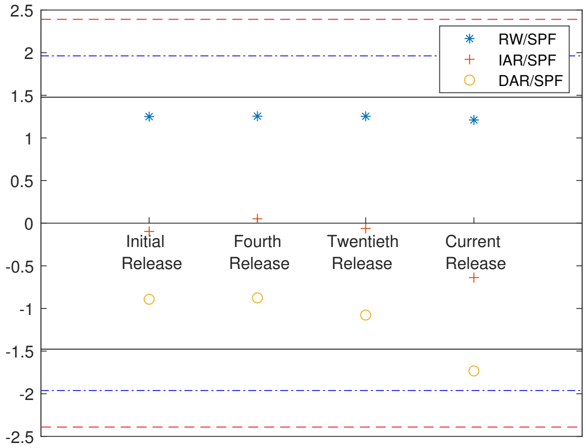
(f) WPE-D, 5 years ahead forecasts



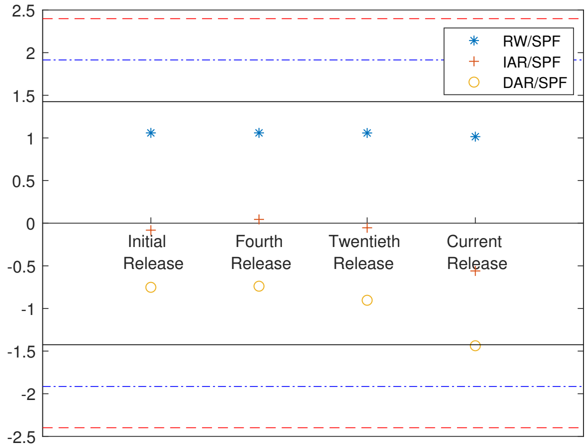
Note: plots report DM test statistic for unemployment and lin-lin loss function, asymmetry parameter $\alpha = 0.9$, for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.22: DM test statistic for unemployment and squared Lin-Lin loss function

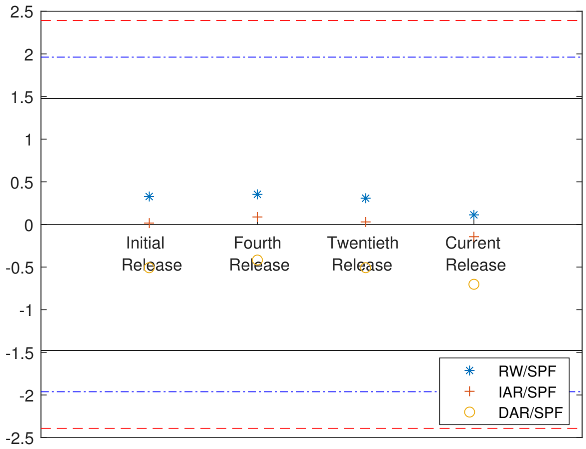
(a) WCE-B, 1 year ahead forecasts



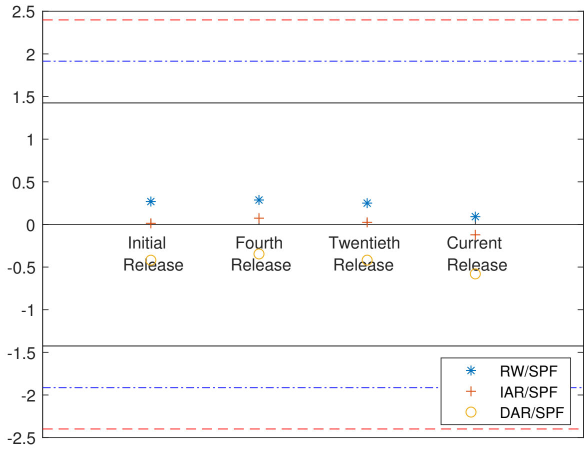
(b) WPE-D, 1 year ahead forecasts



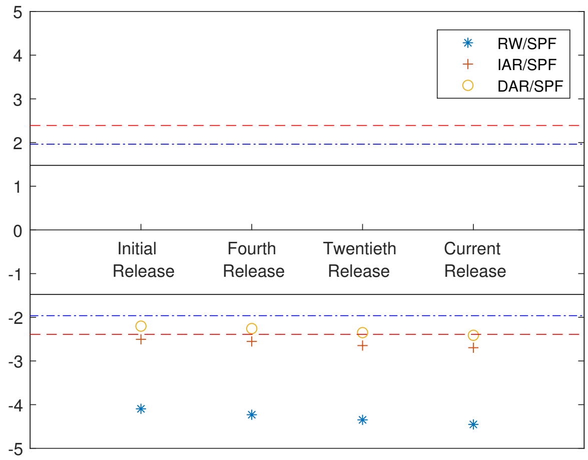
(c) WCE-B, 2 years ahead forecasts



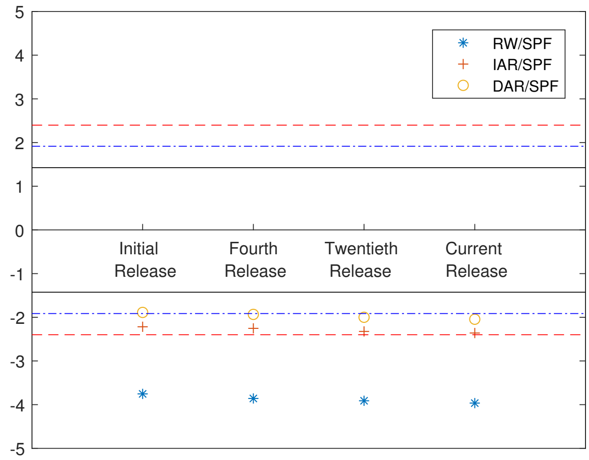
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts



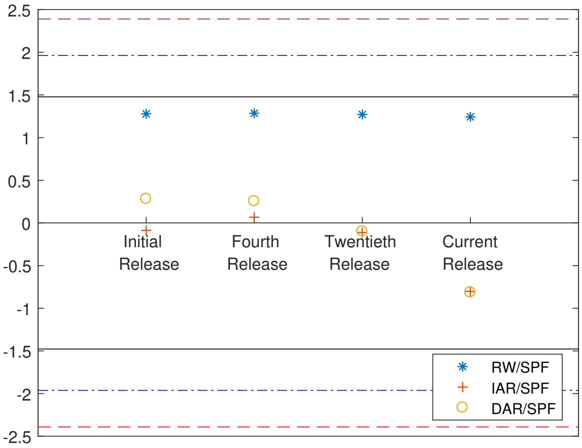
(f) WPE-D, 5 years ahead forecasts



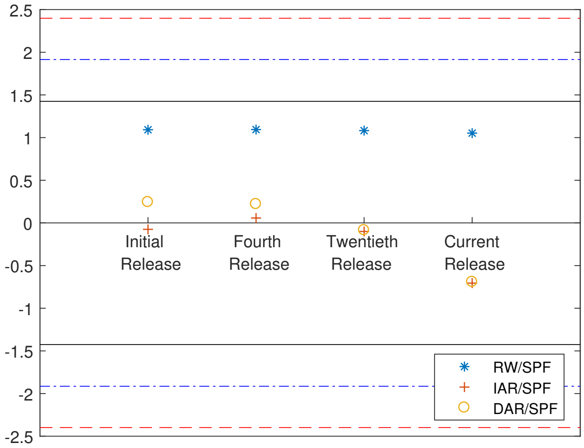
Note: plots report DM test statistic for unemployment and squared Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.23: DM test statistic for unemployment and Linex loss function with $\alpha = 1$

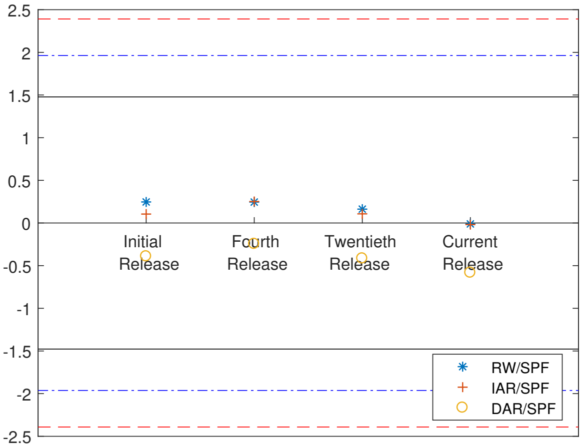
(a) WCE-B, 1 year ahead forecasts



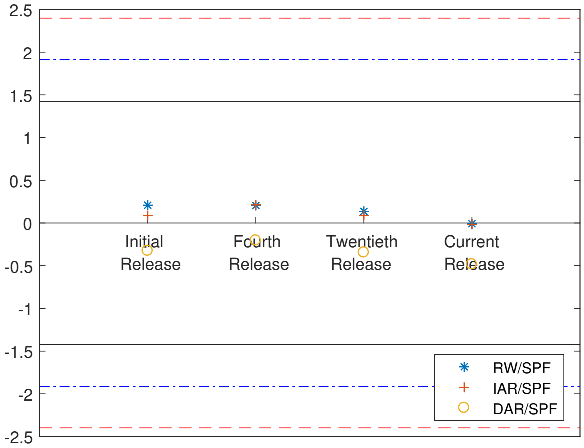
(b) WPE-D, 1 year ahead forecasts



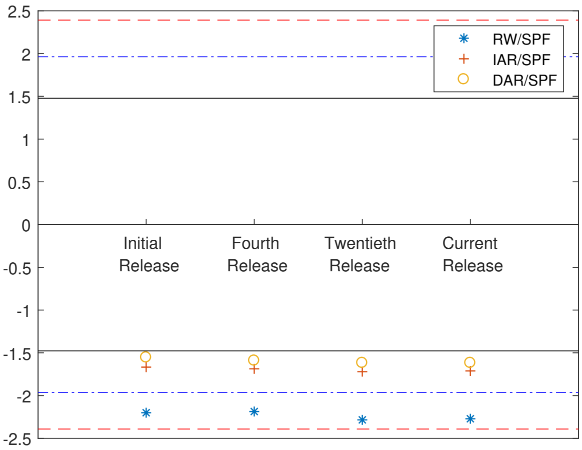
(c) WCE-B, 2 years ahead forecasts



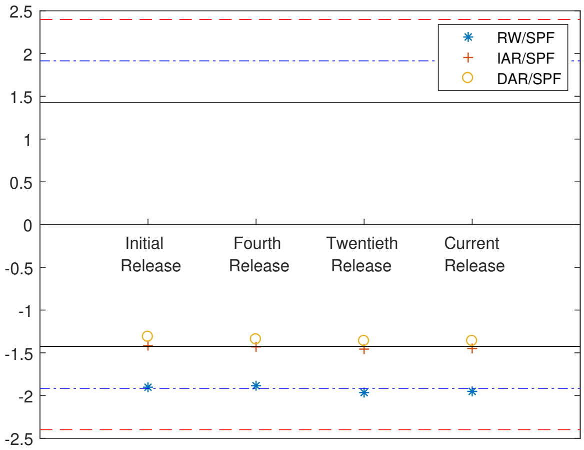
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts

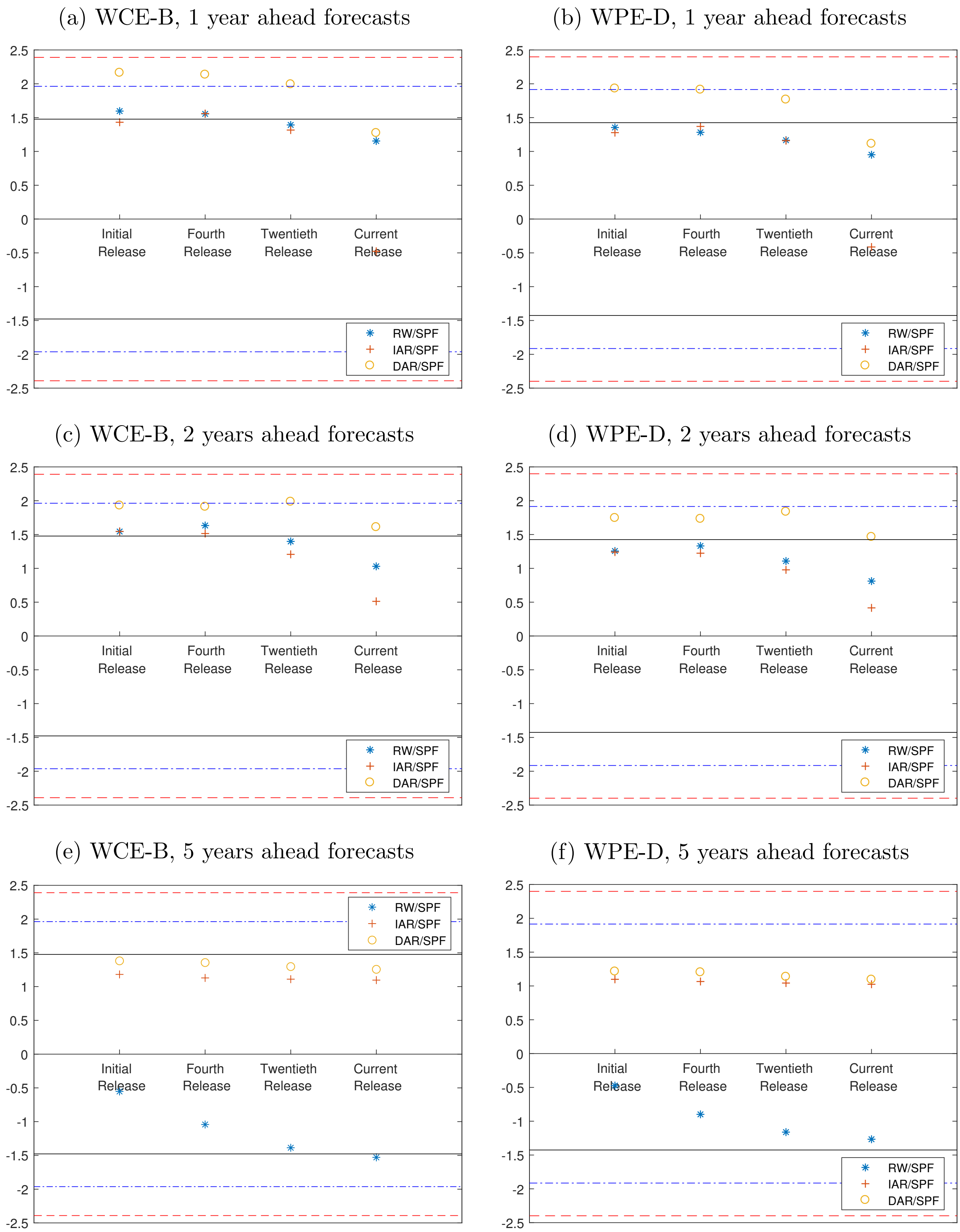


(f) WPE-D, 5 years ahead forecasts



Note: plots report DM test statistic for unemployment and Linex loss function with asymmetry parameter $\alpha = 1$ for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

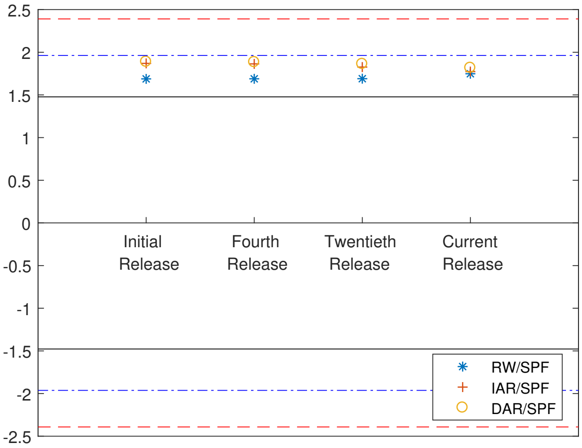
Figure 3.24: DM test statistic for unemployment and Linex loss function with $\alpha = -1$



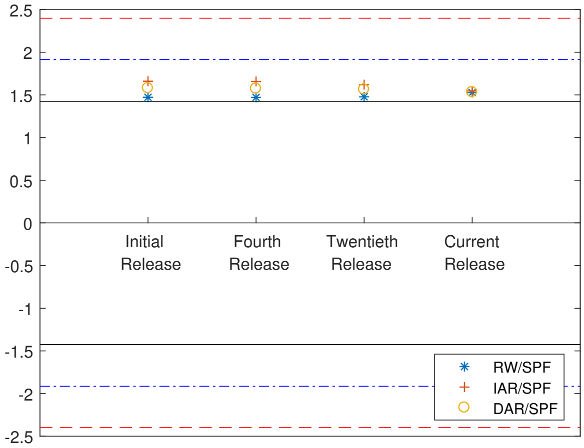
Note: plots report DM test statistic for unemployment and Linex loss function with asymmetry parameter $\alpha = -1$ for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.25: DM test statistic for real GDP growth and quadratic loss function

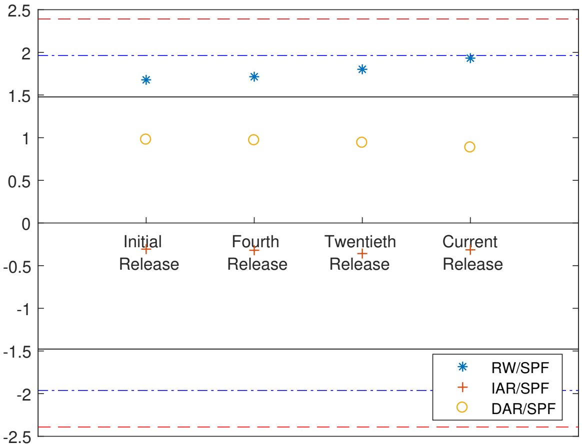
(a) WCE-B, 1 year ahead forecasts



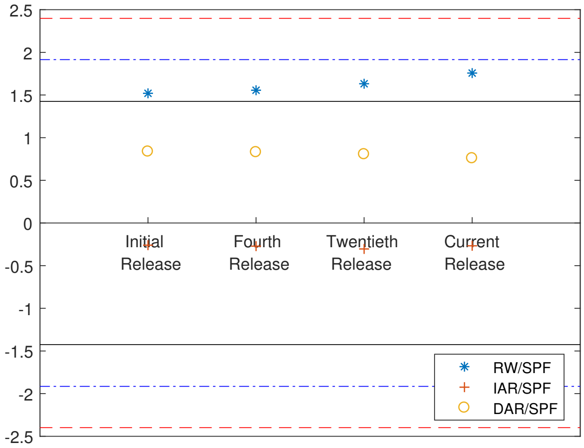
(b) WPE-D, 1 year ahead forecasts



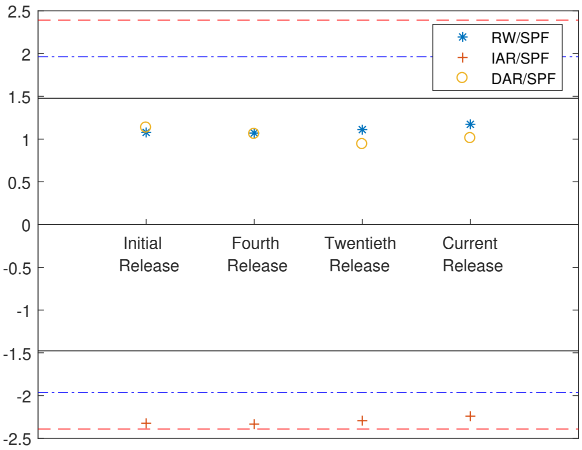
(c) WCE-B, 2 years ahead forecasts



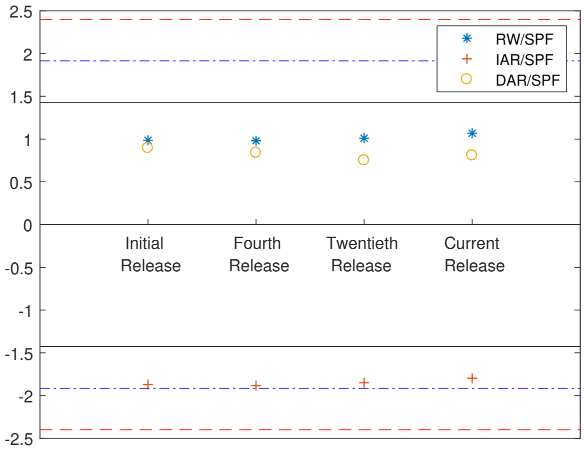
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts



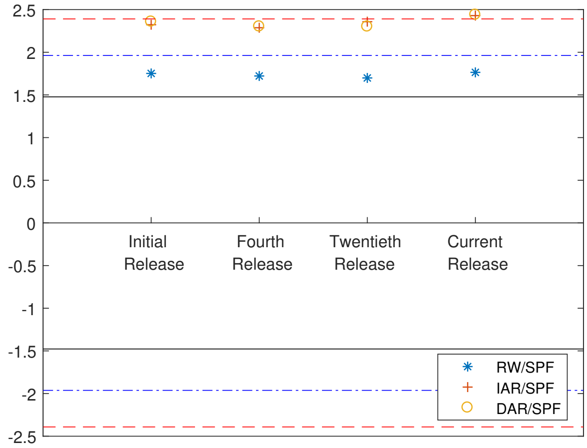
(f) WPE-D, 5 years ahead forecasts



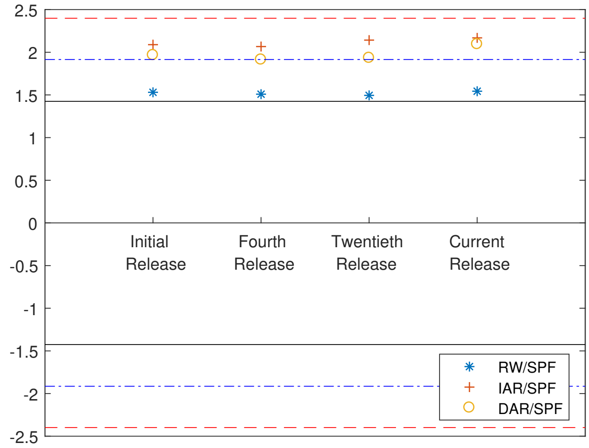
Note: plots report DM test statistic for real GDP growth and quadratic loss function for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.26: DM test statistic for real GDP growth and absolute loss function

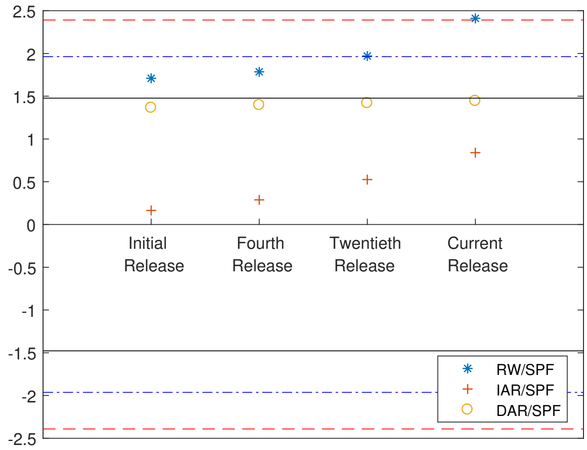
(a) WCE-B, 1 year ahead forecasts



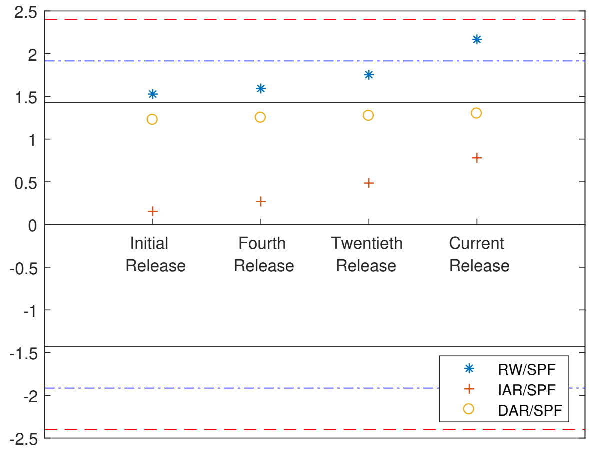
(b) WPE-D, 1 year ahead forecasts



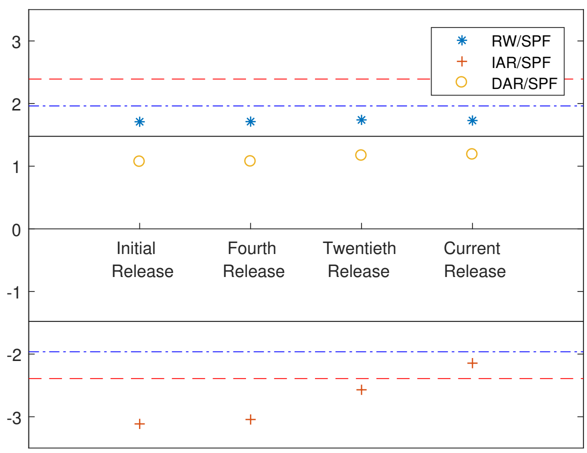
(c) WCE-B, 2 years ahead forecasts



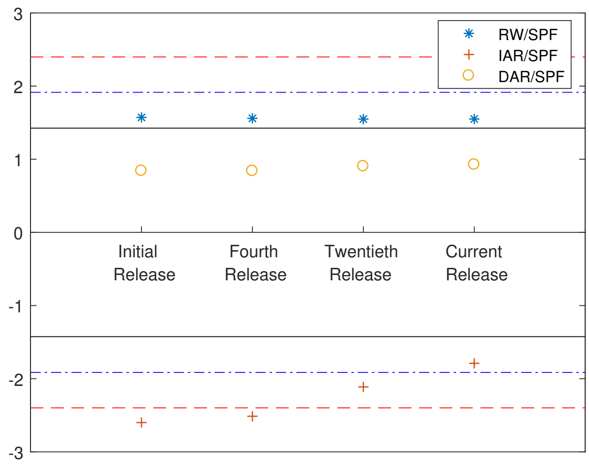
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts

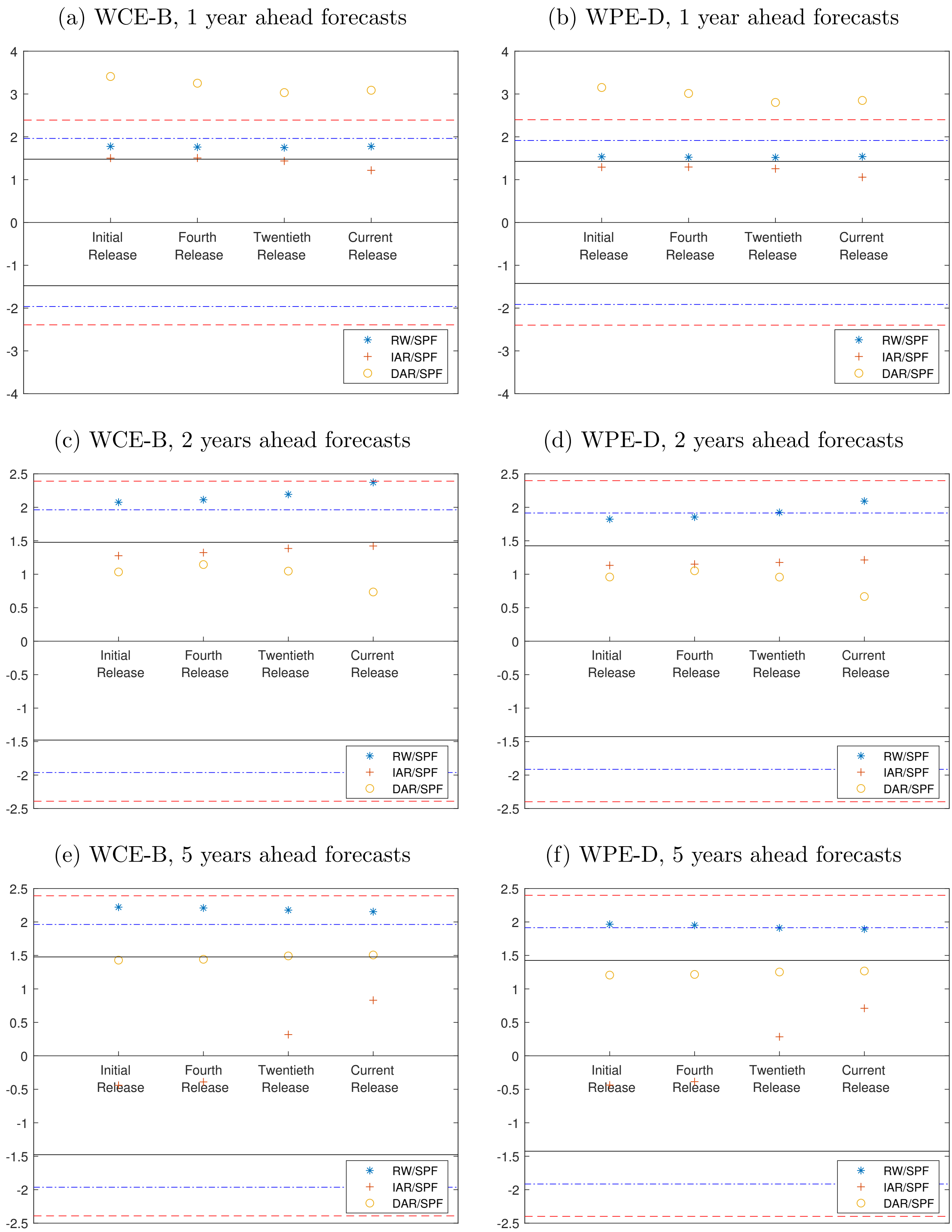


(f) WPE-D, 5 years ahead forecasts



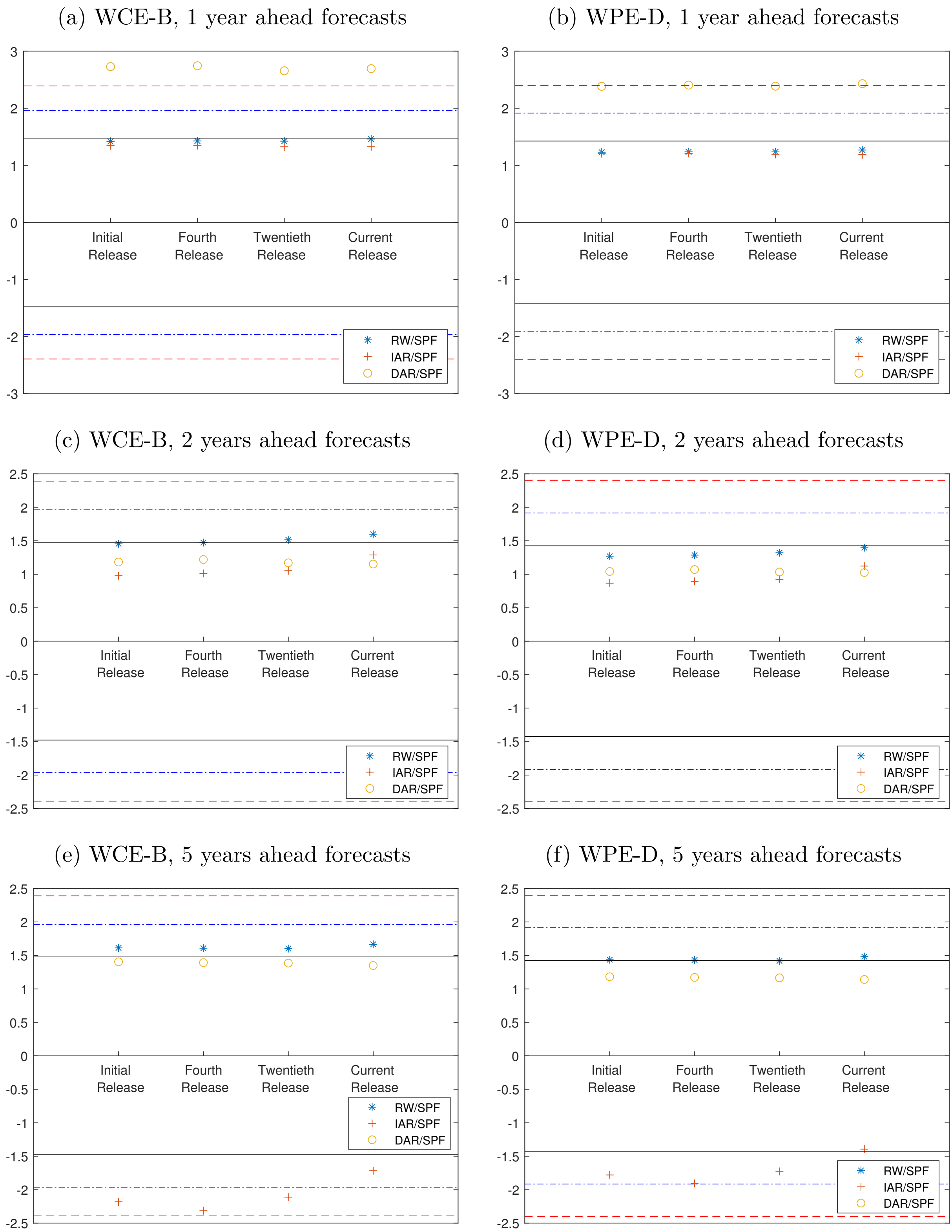
Note: plots report DM test statistic for real GDP growth and absolute loss function for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.27: DM test statistic for real GDP growth and Lin-Lin loss function



Note: plots report DM test statistic for real GDP growth and lin-lin loss function, asymmetry parameter $\alpha = 0.9$, for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

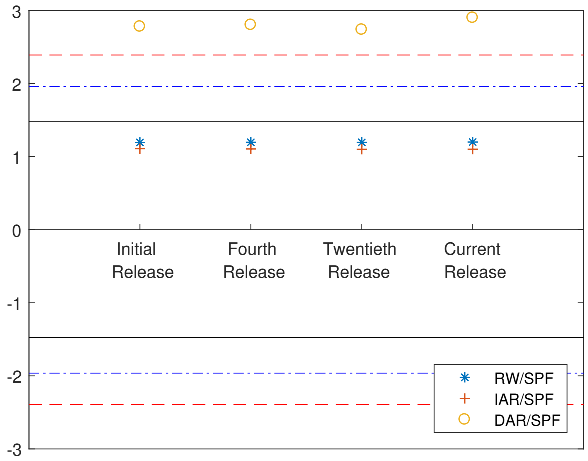
Figure 3.28: DM test statistic for real GDP growth and squared Lin-Lin loss function



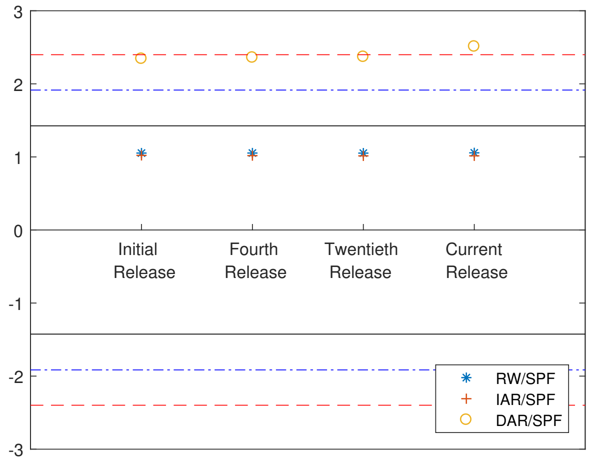
Note: plots report DM test statistic for real GDP growth and squared Lin-Lin loss function, asymmetry parameter $\alpha = 0.9$, for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.29: DM test statistic for real GDP growth and Linex loss function with $\alpha = 1$

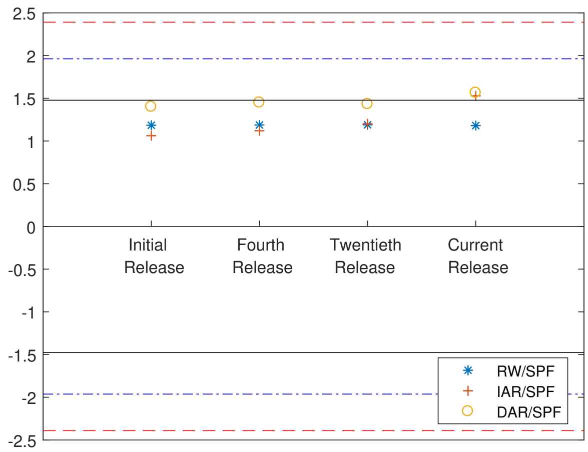
(a) WCE-B, 1 year ahead forecasts



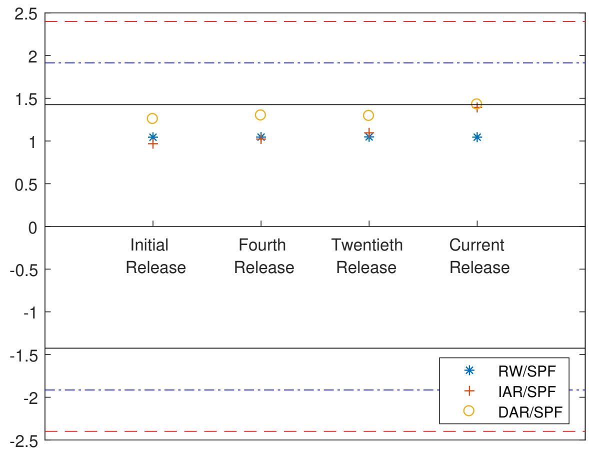
(b) WPE-D, 1 year ahead forecasts



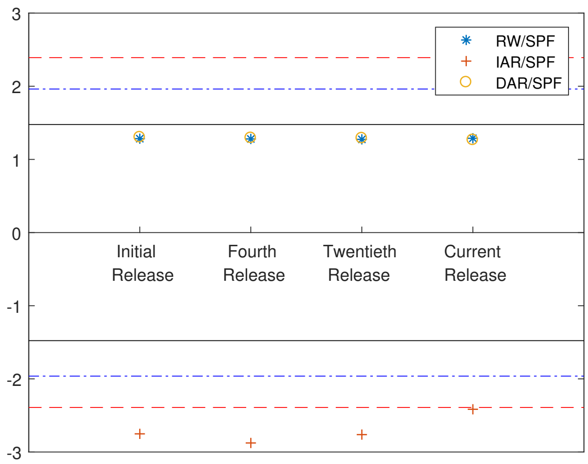
(c) WCE-B, 2 years ahead forecasts



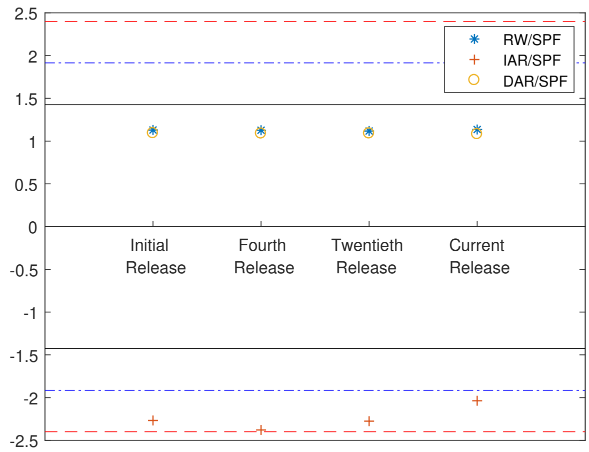
(d) WPE-D, 2 years ahead forecasts



(e) WCE-B, 5 years ahead forecasts

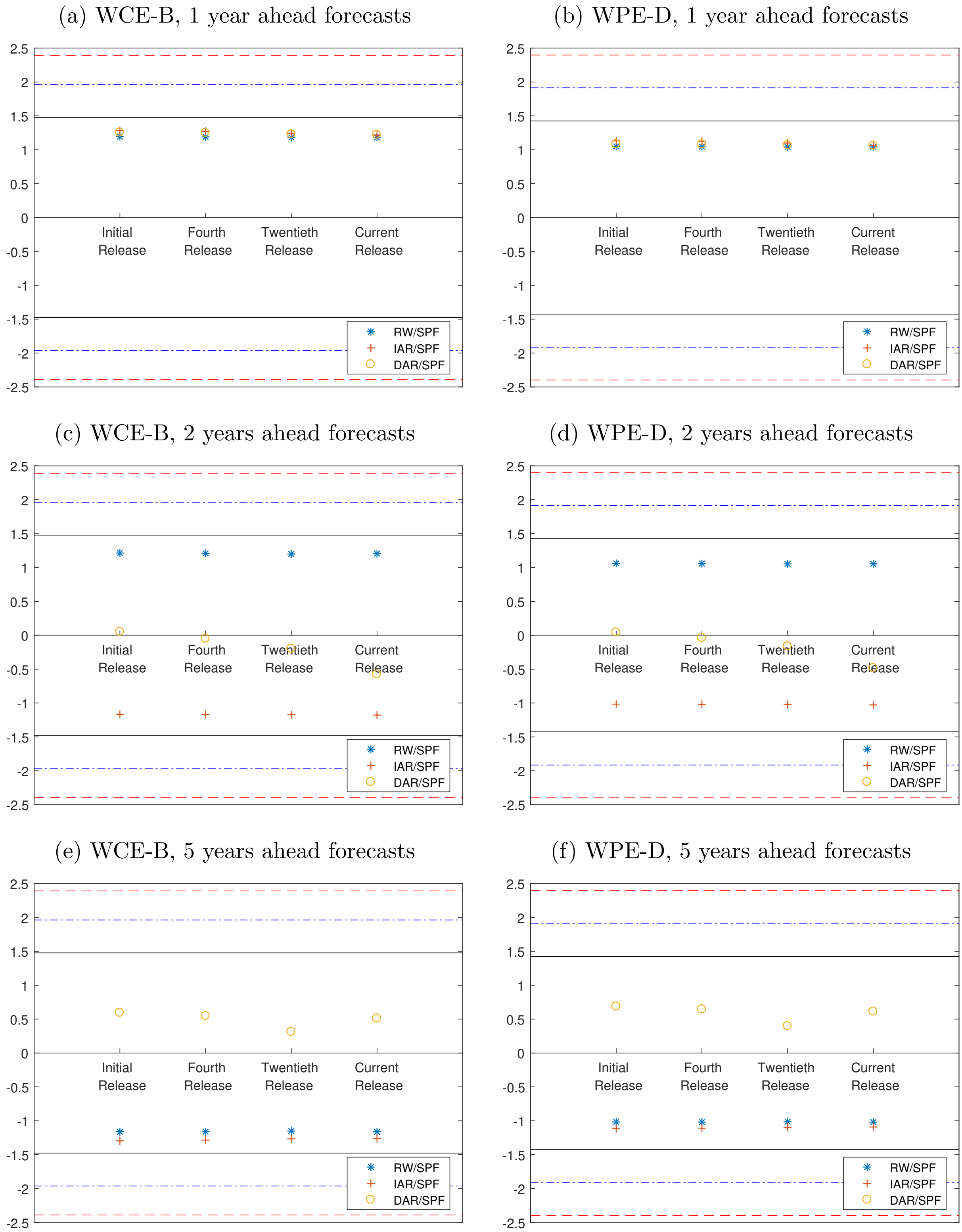


(f) WPE-D, 5 years ahead forecasts



Note: plots report DM test statistic for real GDP growth and Linex loss function with asymmetry parameter $\alpha = 1$ for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

Figure 3.30: DM test statistic for real GDP growth and Linex loss function with $\alpha = -1$



Note: plots report DM test statistic for real GDP growth and Linex loss function with asymmetry parameter $\alpha = -1$ for the sample 2002.Q1 - 2010.Q3 ($T = 35$). Blue stars indicate DM tests for SPF and random walk forecasts, red crosses indicate DM tests for SPF and IAR forecasts and yellow dots indicate DM tests for SPF and DAR forecasts. Lines indicate two side critical values taken from a non standard distribution in the case of WCE with fixed-b asymptotics (red dashed: 5%, 2.3911; blue dash-dotted: 10%, 1.9626; black solid: 20%, 1.4774) and from a Student-t distribution with $2m$ degrees of freedom in the case of WPE with fixed-m asymptotics (red dashed: 5%, 2.3986; blue dash-dotted: 10%, 1.9147; black solid: 20%, 1.4253). WCE-B and WPE-D bandwidths are $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ respectively.

3.7 Conclusions

In this Chapter, I perform a fully real-time evaluation of ECB SPF forecasts about HICP inflation, unemployment rate and real GDP growth using the Diebold and Mariano test for equal forecast accuracy with symmetric and asymmetric loss functions. Benchmark forecasts are taken from three simple models which should be easy to beat: random walk without drift, indirect autoregressive and direct autoregressive. I consider different vintages for the realisations of the target variables: first release, four releases after the first, twenty releases after the first and the latest release available. The sample available is small so, to account for small sample bias of this type of test, I use fixed-b and fixed-m asymptotics critical values, which proved to deliver good sized tests even in small samples in the existing literature and in my Monte Carlo exercise. Results show that ECB SPF outperform benchmark models in some cases especially for short horizons, while their advantage decreases as the horizon extends. In this light, random walk and IAR model seem to be good at predicting long term unemployment and real GDP growth respectively. However, in comparison to existing results about US SPF forecasts, ECB SPF seems to provide more reliable medium and long term forecasts.

3.A Appendix: Alternative DM Test Monte Carlo Results

This Appendix presents results of the Monte Carlo simulation with the modified Diebold and Mariano test statistics proposed by Harvey, Leybourne and Newbold (1997).

Harvey, Leybourne and Newbold (1997) suggest to modify the Diebold and Mariano test altering the divisor on the variance term. The modification follows from noting that the Diebold and Mariano test does not use degrees of freedom to adjust variances. The authors also suggest that the t-distribution with $T-1$ degrees of freedom should be used to construct critical values. This advice is not based on theoretical considerations but serves to increase the critical values for tests that appear oversized in Monte Carlo experiments.

The modified test statistic for a h-period forecast horizon is

$$DM_{HLN} = \left[\frac{T+1-2h+T^{-1}h(h-1)}{T}\right]^{1/2} \times DM, \tag{3.17}$$

where DM is the original test statistic by Diebold and Mariano (1995) and T is the sample size.

Harvey, Leybourne and Newbold (1998) provide Monte Carlo evidence that their corrections reduce size distortion under a quadratic loss function, this Appendix complete their size and power study using asymmetric loss functions.

The setting of this Monte Carlo simulation remains the one described in Section 3.5.

Results for empirical size of the modified test are reported in Tables 3.12 - 3.16, q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel and critical values form a t_{T-1} distribution are used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator. Results are similar to the ones of the original DM test. However, the DM_{HLN} appears slightly undersized when fixed-smoothing asymptotics are used for asymmetric loss functions, which it is undesirable for my empirical application.

Power performances are reported in Figures 3.31 - 3.38 and are in line with those of the original test.

Overall, considering size and power results under asymmetric loss functions, the original test paired with fixed-smoothing asymptotics has a better behaviour and for this reason, I have decided to use it in the empirical exercise.

Table 3.9: Empirical size of DM_{HLN} with quadratic loss function

Standard Asymptotics

T = 40

	WCE				WPE			
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.057	0.074	0.094	0.328	0.100	0.076	0.059	0.059
2	0.066	0.075	0.090	0.328	0.088	0.067	0.055	0.071
3	0.085	0.078	0.089	0.331	0.089	0.068	0.058	0.087
4	0.101	0.077	0.083	0.314	0.075	0.057	0.059	0.099
5	0.117	0.071	0.077	0.303	0.069	0.051	0.052	0.100

T = 120

q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.050	0.064	0.073	0.330	0.086	0.074	0.056	0.059
2	0.058	0.074	0.083	0.327	0.088	0.078	0.059	0.073
3	0.056	0.071	0.077	0.330	0.080	0.072	0.056	0.077
4	0.062	0.080	0.079	0.334	0.085	0.074	0.060	0.095
5	0.067	0.083	0.077	0.329	0.080	0.069	0.057	0.105

Fixed-smoothing Asymptotics

T = 40

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.047	0.044	0.047	0.039	0.038	0.041	0.054
2	0.044	0.038	0.041	0.032	0.031	0.038	0.064
3	0.049	0.044	0.043	0.034	0.032	0.041	0.079
4	0.045	0.035	0.035	0.027	0.027	0.038	0.090
5	0.039	0.030	0.035	0.024	0.021	0.034	0.090

T = 120

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.050	0.046	0.045	0.043	0.043	0.045	0.055
2	0.060	0.051	0.048	0.045	0.045	0.048	0.069
3	0.058	0.044	0.045	0.037	0.038	0.043	0.072
4	0.065	0.049	0.047	0.042	0.041	0.048	0.090
5	0.067	0.046	0.044	0.037	0.038	0.045	0.100

Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under a quadratic loss function with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.10: Empirical size of DM_{HLN} with absolute loss function

Standard Asymptotics

$T = 40$

q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.059	0.076	0.095	0.326	0.103	0.080	0.064	0.061
2	0.071	0.078	0.095	0.325	0.099	0.075	0.060	0.073
3	0.096	0.087	0.100	0.322	0.098	0.079	0.067	0.096
4	0.112	0.087	0.095	0.315	0.087	0.069	0.068	0.108
5	0.131	0.086	0.090	0.297	0.084	0.065	0.066	0.108

$T = 120$

q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.053	0.065	0.075	0.322	0.086	0.077	0.058	0.057
2	0.059	0.074	0.083	0.331	0.092	0.082	0.062	0.071
3	0.060	0.075	0.079	0.325	0.082	0.076	0.058	0.079
4	0.068	0.082	0.082	0.331	0.089	0.079	0.064	0.096
5	0.068	0.084	0.081	0.320	0.082	0.073	0.063	0.103

Fixed-smoothing Asymptotics

$T = 40$

	WCE			**WPE**			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.049	0.046	0.049	0.046	0.043	0.046	0.055
2	0.052	0.047	0.046	0.039	0.039	0.045	0.067
3	0.057	0.051	0.049	0.041	0.041	0.050	0.086
4	0.057	0.047	0.047	0.035	0.035	0.051	0.098
5	0.052	0.042	0.042	0.032	0.032	0.048	0.099

$T = 120$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.052	0.046	0.046	0.046	0.046	0.047	0.055
2	0.062	0.052	0.050	0.049	0.048	0.050	0.067
3	0.060	0.047	0.045	0.039	0.041	0.046	0.074
4	0.068	0.054	0.050	0.046	0.046	0.053	0.091
5	0.070	0.053	0.047	0.042	0.042	0.051	0.098

Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under an absolute loss function with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.11: Empirical size of DM_{HLN} with Lin-Lin loss function $\alpha = 0.9$

Standard Asymptotics

$T = 40$

q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.055	0.082	0.099	0.334	0.101	0.077	0.063	0.067
2	0.061	0.083	0.094	0.327	0.085	0.068	0.064	0.100
3	0.075	0.092	0.101	0.336	0.087	0.068	0.073	0.132
4	0.083	0.092	0.089	0.328	0.072	0.057	0.075	0.146
5	0.095	0.094	0.083	0.316	0.066	0.053	0.079	0.149

$T = 120$

q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.049	0.066	0.076	0.325	0.086	0.075	0.055	0.060
2	0.055	0.078	0.082	0.329	0.088	0.077	0.061	0.085
3	0.056	0.082	0.082	0.331	0.082	0.073	0.062	0.104
4	0.059	0.089	0.080	0.324	0.080	0.068	0.064	0.124
5	0.059	0.099	0.082	0.323	0.078	0.069	0.066	0.143

Fixed-smoothing Asymptotics

$T = 40$

	WCE			**WPE**			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.049	0.046	0.049	0.039	0.040	0.044	0.060
2	0.053	0.043	0.042	0.031	0.033	0.044	0.091
3	0.061	0.043	0.044	0.033	0.029	0.052	0.124
4	0.057	0.037	0.039	0.024	0.023	0.056	0.135
5	0.057	0.036	0.037	0.024	0.023	0.057	0.139

$T = 120$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.052	0.045	0.045	0.045	0.044	0.044	0.056
2	0.064	0.052	0.051	0.041	0.045	0.049	0.080
3	0.070	0.052	0.048	0.042	0.041	0.050	0.098
4	0.074	0.051	0.048	0.036	0.040	0.052	0.119
5	0.084	0.050	0.043	0.037	0.037	0.052	0.137

Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under a Lin-Lin loss function, $\alpha = 0.9$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.12: Empirical size of DM_{HLN} with squared Lin-Lin loss function $\alpha = 0.9$

Standard Asymptotics

	WCE				WPE			
				$T = 40$				
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.045	0.070	0.089	0.341	0.090	0.067	0.050	0.057
2	0.049	0.072	0.080	0.329	0.073	0.055	0.049	0.086
3	0.061	0.077	0.083	0.340	0.074	0.055	0.056	0.117
4	0.069	0.078	0.079	0.331	0.064	0.048	0.061	0.131
5	0.084	0.078	0.073	0.317	0.056	0.044	0.066	0.137

				$T = 120$				
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.046	0.062	0.071	0.326	0.082	0.072	0.053	0.058
2	0.049	0.074	0.077	0.339	0.079	0.068	0.055	0.077
3	0.052	0.080	0.079	0.336	0.081	0.071	0.058	0.099
4	0.051	0.084	0.076	0.331	0.074	0.065	0.058	0.120
5	0.055	0.089	0.074	0.329	0.070	0.061	0.057	0.136

Fixed-smoothing Asymptotics

	WCE			WPE			
				$T = 40$			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.040	0.036	0.040	0.032	0.030	0.035	0.049
2	0.041	0.033	0.034	0.024	0.023	0.031	0.078
3	0.046	0.036	0.038	0.025	0.023	0.038	0.108
4	0.046	0.030	0.035	0.022	0.019	0.041	0.121
5	0.047	0.027	0.029	0.018	0.017	0.046	0.124

				$T = 120$			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.049	0.042	0.044	0.042	0.043	0.041	0.055
2	0.058	0.045	0.045	0.040	0.040	0.043	0.073
3	0.065	0.049	0.044	0.041	0.040	0.045	0.094
4	0.067	0.045	0.042	0.031	0.034	0.044	0.114
5	0.072	0.044	0.037	0.031	0.031	0.046	0.130

Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under a squared Lin-Lin loss function, $\alpha = 0.9$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.13: Empirical size of the DM_{HLN} test with Linex loss function $\alpha = 1$

Standard Asymptotics

T = 40

	WCE				WPE			
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.045	0.053	0.072	0.340	0.075	0.056	0.039	0.037
2	0.052	0.049	0.064	0.330	0.063	0.048	0.035	0.047
3	0.068	0.050	0.062	0.331	0.062	0.042	0.034	0.061
4	0.085	0.049	0.057	0.321	0.053	0.038	0.033	0.068
5	0.100	0.045	0.051	0.300	0.048	0.034	0.032	0.069

T = 120

q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.044	0.052	0.064	0.334	0.073	0.063	0.046	0.048
2	0.044	0.058	0.066	0.341	0.071	0.060	0.045	0.055
3	0.045	0.057	0.062	0.342	0.064	0.055	0.044	0.067
4	0.046	0.059	0.058	0.337	0.060	0.053	0.042	0.078
5	0.049	0.056	0.054	0.332	0.057	0.048	0.040	0.082

Fixed-smoothing Asymptotics

T = 40

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.029	0.028	0.033	0.027	0.024	0.026	0.031
2	0.027	0.024	0.028	0.021	0.021	0.023	0.040
3	0.026	0.021	0.027	0.020	0.017	0.022	0.054
4	0.023	0.018	0.023	0.016	0.014	0.020	0.060
5	0.022	0.018	0.021	0.015	0.012	0.021	0.062

T = 120

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.040	0.035	0.036	0.034	0.033	0.034	0.045
2	0.043	0.036	0.039	0.033	0.030	0.033	0.050
3	0.044	0.032	0.034	0.027	0.027	0.031	0.061
4	0.046	0.031	0.032	0.025	0.023	0.030	0.072
5	0.042	0.028	0.029	0.022	0.022	0.029	0.077

Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under a Linex function, $\alpha = 1$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.14: Empirical size of the DM_{HLN} test with Linex loss function $\alpha = 0.5$

Standard Asymptotics

$T = 40$

q	WCE DM_{HLN}	WCE $\lfloor T^{1/3} \rfloor$	WCE $\lfloor T^{1/2} \rfloor$	WCE T	WPE $\lfloor T^{1/4} \rfloor$	WPE $\lfloor T^{1/3} \rfloor$	WPE $\lfloor T^{1/2} \rfloor$	WPE $\lfloor T^{2/3} \rfloor$
1	0.059	0.067	0.085	0.332	0.091	0.069	0.051	0.051
2	0.066	0.061	0.081	0.325	0.080	0.060	0.046	0.062
3	0.085	0.069	0.082	0.330	0.076	0.058	0.050	0.080
4	0.100	0.066	0.077	0.317	0.067	0.052	0.051	0.089
5	0.115	0.061	0.067	0.300	0.062	0.046	0.045	0.090

$T = 120$

q	WCE DM_{HLN}	WCE $\lfloor T^{1/3} \rfloor$	WCE $\lfloor T^{1/2} \rfloor$	WCE T	WPE $\lfloor T^{1/4} \rfloor$	WPE $\lfloor T^{1/3} \rfloor$	WPE $\lfloor T^{1/2} \rfloor$	WPE $\lfloor T^{2/3} \rfloor$
1	0.049	0.060	0.069	0.325	0.081	0.071	0.054	0.053
2	0.053	0.068	0.074	0.334	0.082	0.070	0.056	0.065
3	0.055	0.068	0.073	0.334	0.073	0.066	0.053	0.075
4	0.059	0.073	0.073	0.333	0.074	0.067	0.054	0.090
5	0.060	0.073	0.066	0.327	0.068	0.061	0.051	0.098

Fixed-smoothing Asymptotics

$T = 40$

q	WCE $\lfloor T^{1/3} \rfloor$	WCE $\lfloor T^{1/2} \rfloor$	WCE T	WPE $\lfloor T^{1/4} \rfloor$	WPE $\lfloor T^{1/3} \rfloor$	WPE $\lfloor T^{1/2} \rfloor$	WPE $\lfloor T^{2/3} \rfloor$
1	0.039	0.039	0.042	0.035	0.032	0.036	0.045
2	0.038	0.034	0.035	0.029	0.029	0.033	0.055
3	0.040	0.034	0.035	0.028	0.026	0.034	0.073
4	0.038	0.030	0.033	0.023	0.021	0.032	0.081
5	0.034	0.027	0.032	0.020	0.019	0.031	0.081

$T = 120$

q	WCE $\lfloor T^{1/3} \rfloor$	WCE $\lfloor T^{1/2} \rfloor$	WCE T	WPE $\lfloor T^{1/4} \rfloor$	WPE $\lfloor T^{1/3} \rfloor$	WPE $\lfloor T^{1/2} \rfloor$	WPE $\lfloor T^{2/3} \rfloor$
1	0.048	0.044	0.042	0.042	0.040	0.042	0.050
2	0.054	0.044	0.045	0.038	0.039	0.044	0.061
3	0.054	0.041	0.041	0.035	0.034	0.040	0.071
4	0.059	0.042	0.043	0.035	0.034	0.043	0.085
5	0.056	0.039	0.038	0.032	0.030	0.038	0.092

Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under a Linex loss function, $\alpha = 0.5$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.15: Empirical size of the DM_{HLN} test with Linex loss function $\alpha = -0.9$

Standard Asymptotics

		WCE			WPE			
				$T = 40$				
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.047	0.054	0.073	0.334	0.076	0.058	0.040	0.038
2	0.057	0.054	0.070	0.330	0.068	0.049	0.039	0.053
3	0.073	0.054	0.068	0.325	0.065	0.049	0.037	0.067
4	0.085	0.051	0.061	0.313	0.054	0.039	0.037	0.073
5	0.101	0.048	0.054	0.300	0.049	0.035	0.035	0.071
				$T = 120$				
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.045	0.056	0.067	0.345	0.080	0.069	0.051	0.048
2	0.047	0.062	0.069	0.342	0.078	0.065	0.047	0.062
3	0.049	0.060	0.064	0.334	0.069	0.061	0.044	0.065
4	0.053	0.065	0.066	0.345	0.069	0.058	0.045	0.084
5	0.058	0.070	0.065	0.344	0.064	0.056	0.047	0.095

Fixed-smoothing Asymptotics

	WCE			WPE			
			$T = 40$				
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.030	0.028	0.031	0.027	0.024	0.027	0.034
2	0.030	0.027	0.031	0.023	0.021	0.026	0.046
3	0.030	0.027	0.029	0.024	0.023	0.025	0.058
4	0.029	0.022	0.027	0.019	0.017	0.026	0.065
5	0.025	0.018	0.025	0.018	0.014	0.023	0.062
			$T = 120$				
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.043	0.037	0.041	0.035	0.037	0.039	0.044
2	0.045	0.038	0.041	0.035	0.036	0.035	0.057
3	0.047	0.036	0.035	0.028	0.032	0.034	0.060
4	0.050	0.035	0.037	0.030	0.028	0.034	0.079
5	0.053	0.035	0.036	0.027	0.028	0.035	0.088

Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under a Linex loss function, $\alpha = -0.9$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Table 3.16: Empirical size of the DM_{HLN} test with Linex loss function $\alpha = -1$

Standard Asymptotics

	WCE				WPE			
				$T = 40$				
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.044	0.052	0.069	0.333	0.071	0.055	0.039	0.037
2	0.054	0.050	0.067	0.330	0.066	0.047	0.036	0.050
3	0.070	0.051	0.065	0.327	0.063	0.047	0.034	0.062
4	0.081	0.048	0.057	0.310	0.051	0.037	0.035	0.068
5	0.096	0.046	0.050	0.299	0.047	0.033	0.032	0.066

				$T = 120$				
q	DM_{HLN}	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.043	0.054	0.064	0.345	0.076	0.066	0.049	0.045
2	0.045	0.060	0.068	0.344	0.076	0.063	0.046	0.061
3	0.047	0.058	0.062	0.339	0.068	0.060	0.042	0.064
4	0.051	0.064	0.064	0.346	0.066	0.054	0.043	0.082
5	0.055	0.066	0.062	0.348	0.062	0.052	0.045	0.092

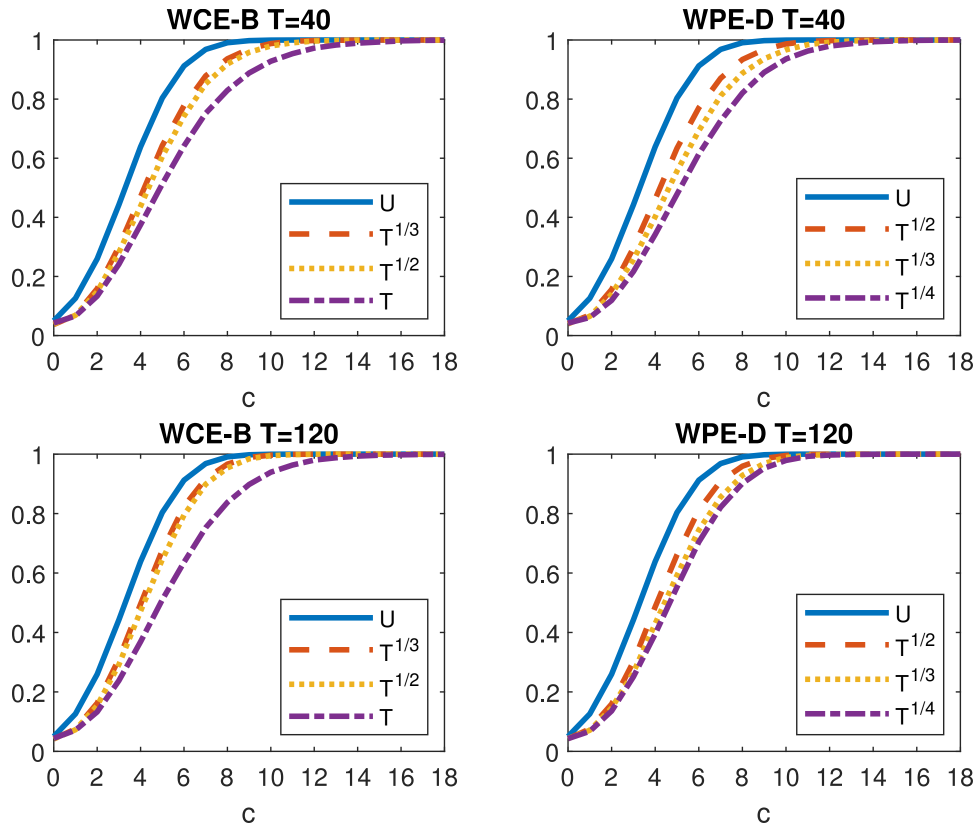
Fixed-smoothing Asymptotics

	WCE			WPE			
				$T = 40$			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.028	0.027	0.029	0.024	0.022	0.025	0.031
2	0.027	0.025	0.028	0.022	0.019	0.023	0.042
3	0.028	0.024	0.028	0.023	0.021	0.023	0.054
4	0.026	0.020	0.025	0.017	0.016	0.024	0.061
5	0.022	0.017	0.023	0.016	0.013	0.021	0.058

				$T = 120$			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.041	0.036	0.039	0.034	0.035	0.038	0.041
2	0.045	0.036	0.039	0.033	0.034	0.034	0.055
3	0.044	0.035	0.034	0.028	0.031	0.032	0.058
4	0.048	0.033	0.035	0.027	0.027	0.031	0.076
5	0.051	0.032	0.035	0.025	0.027	0.033	0.085

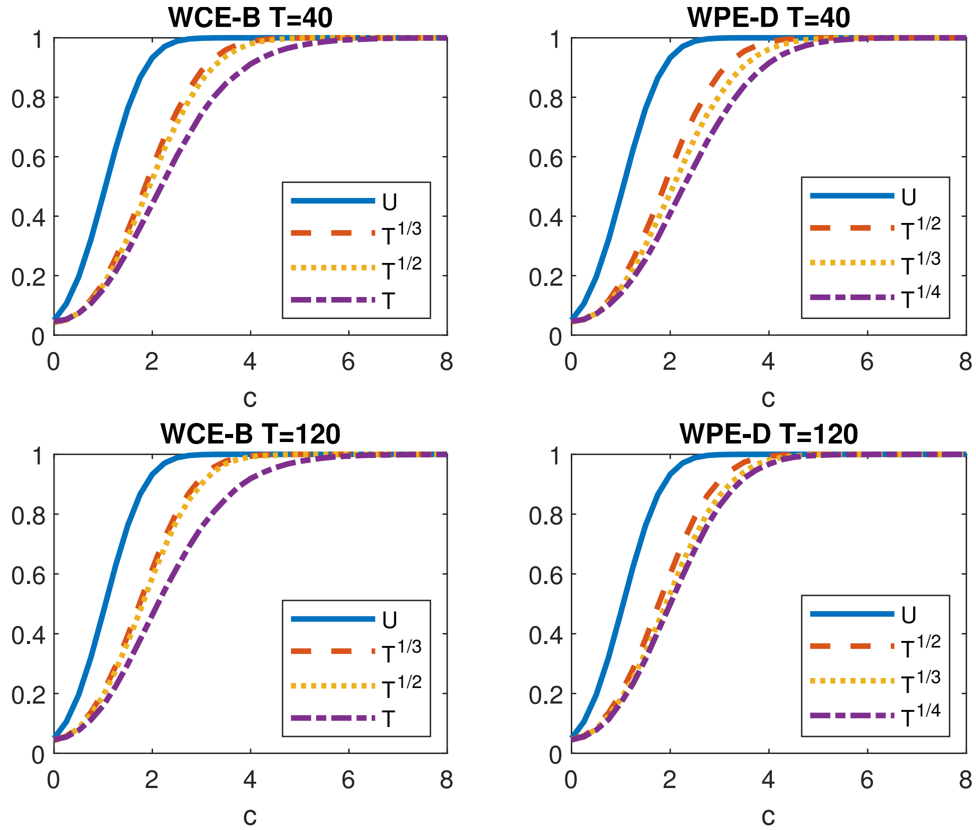
Note: the Table reports the empirical size of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) under a Linex loss function, $\alpha = -1$, with standard and fixed-smoothing asymptotics. The theoretical size is 5%. q indicates the level of serial correlation of forecast errors, the higher the q the higher the serial correlation. T is the sample size. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance except for the column DM_{HLN} where the rectangular kernel is used; WPE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Powers of T are the bandwidths used to calculate the long run variance estimator.

Figure 3.31: DM_{HLN} finite sample local power and quadratic loss function



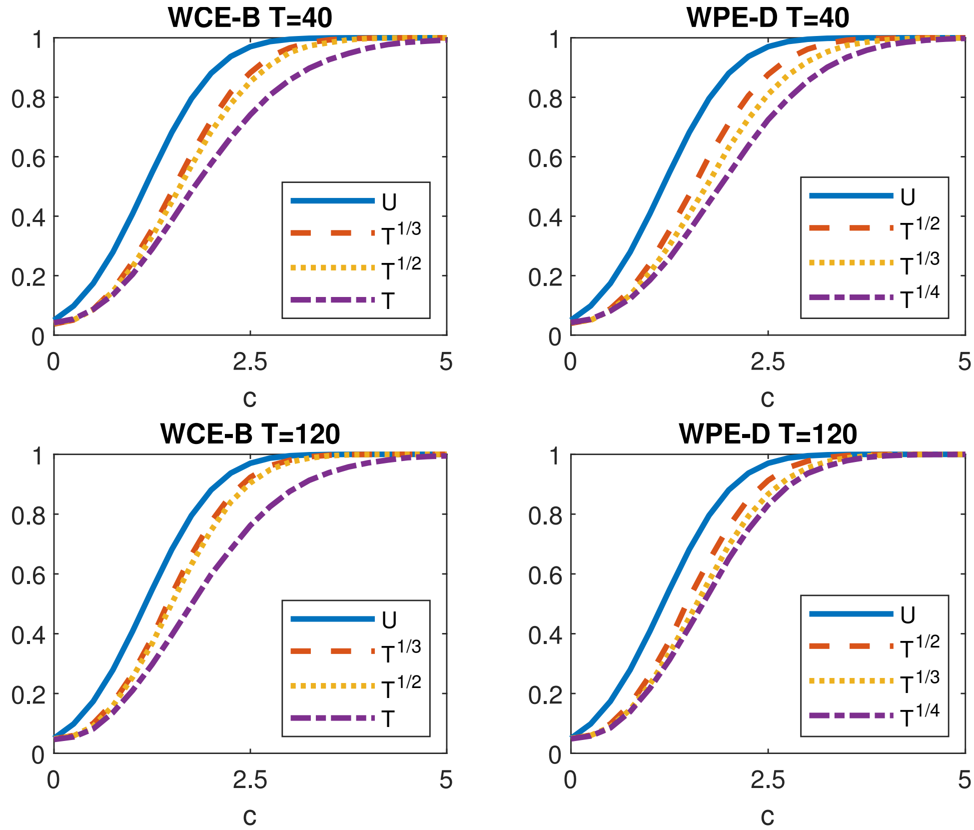
Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with quadratic loss function in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.32: DM_{HLN} finite sample local power and absolute loss function



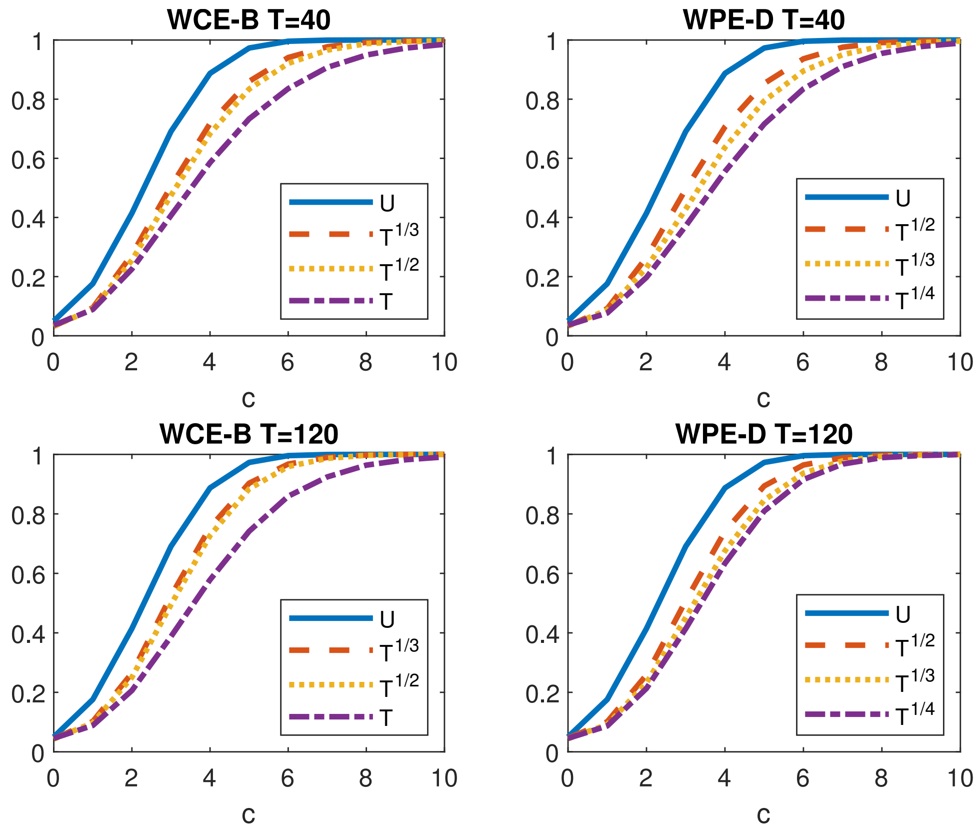
Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with absolute loss function in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.33: DM_{HLN} finite sample local power and Lin-Lin loss function



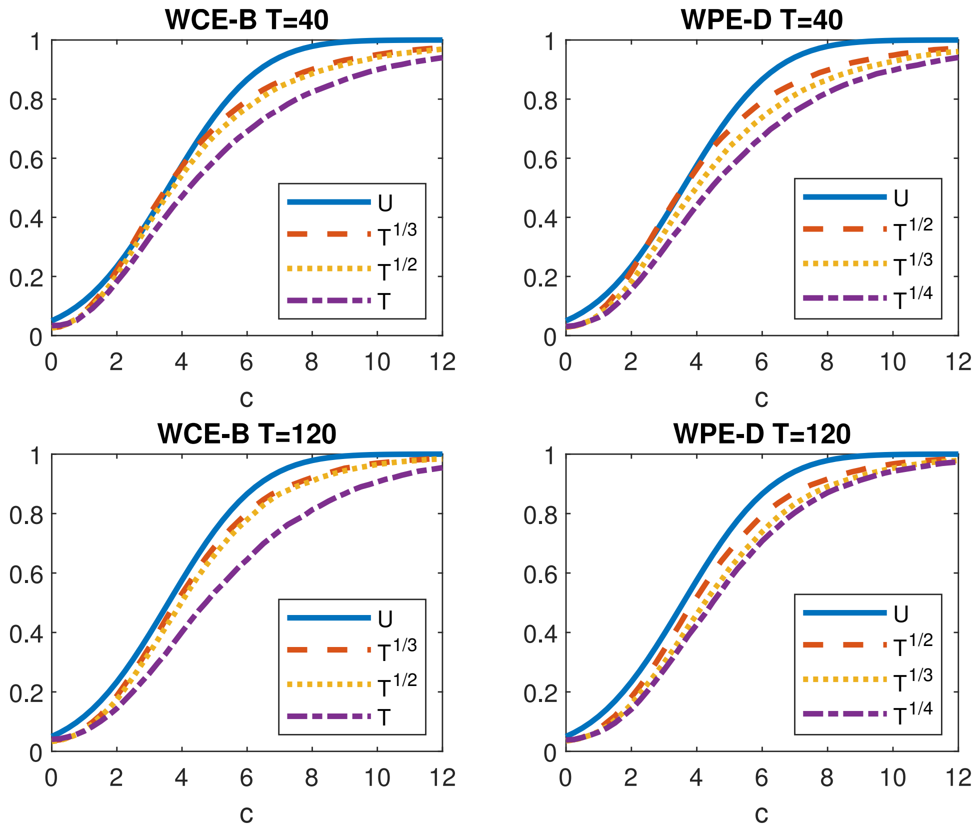
Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with Lin-Lin loss function, $\alpha = 0.9$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.34: DM_{HLN} finite sample local power and squared Lin-Lin loss function



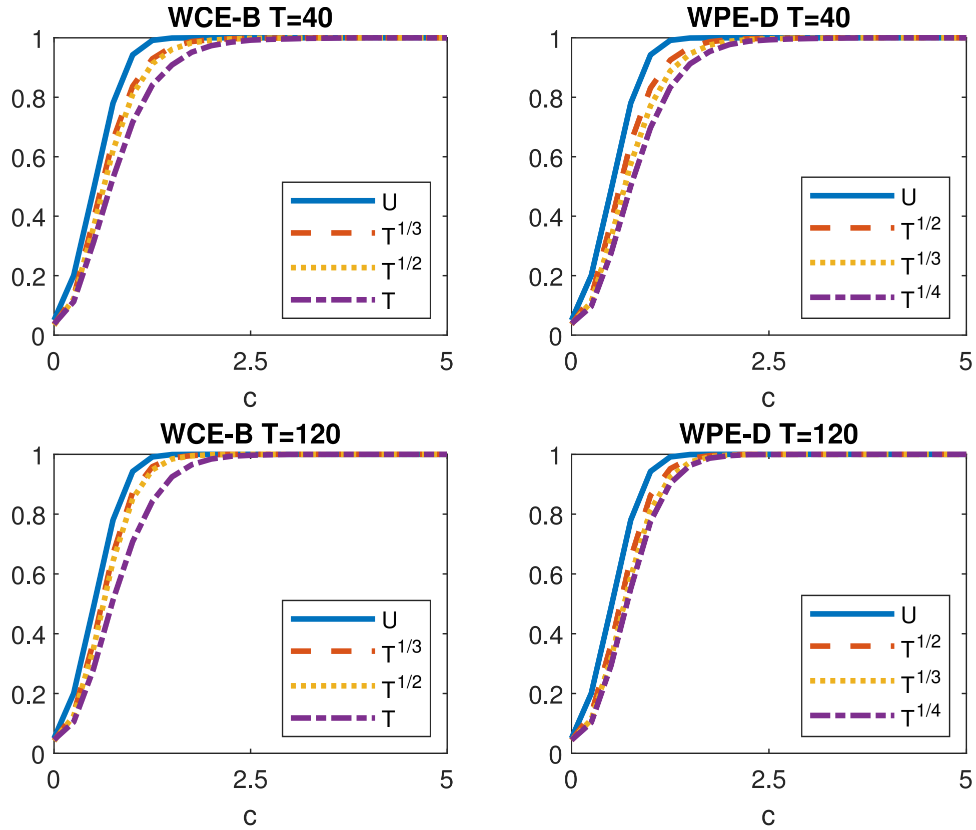
Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with squared Lin-Lin loss function, $\alpha = 0.9$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.35: DM_{HLN} finite sample local power and Linex $\alpha = 1$ loss function



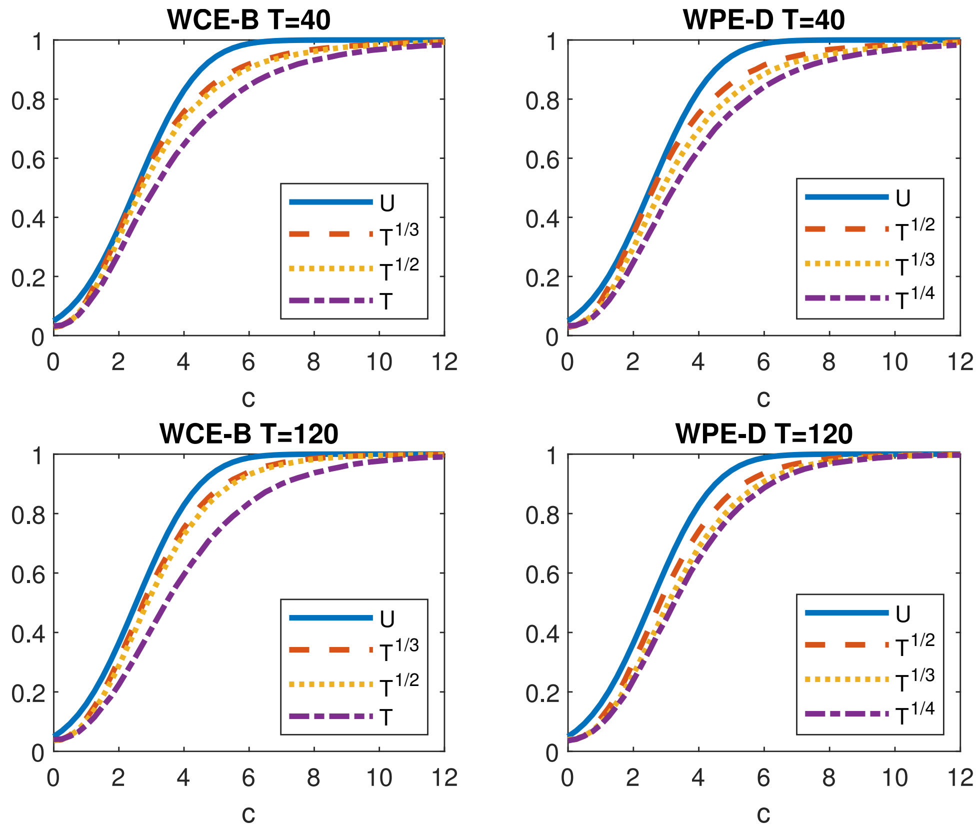
Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with Linex loss function, $\alpha = 1$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.36: DM_{HLN} finite sample local power and Linex $\alpha = 0.5$ loss function



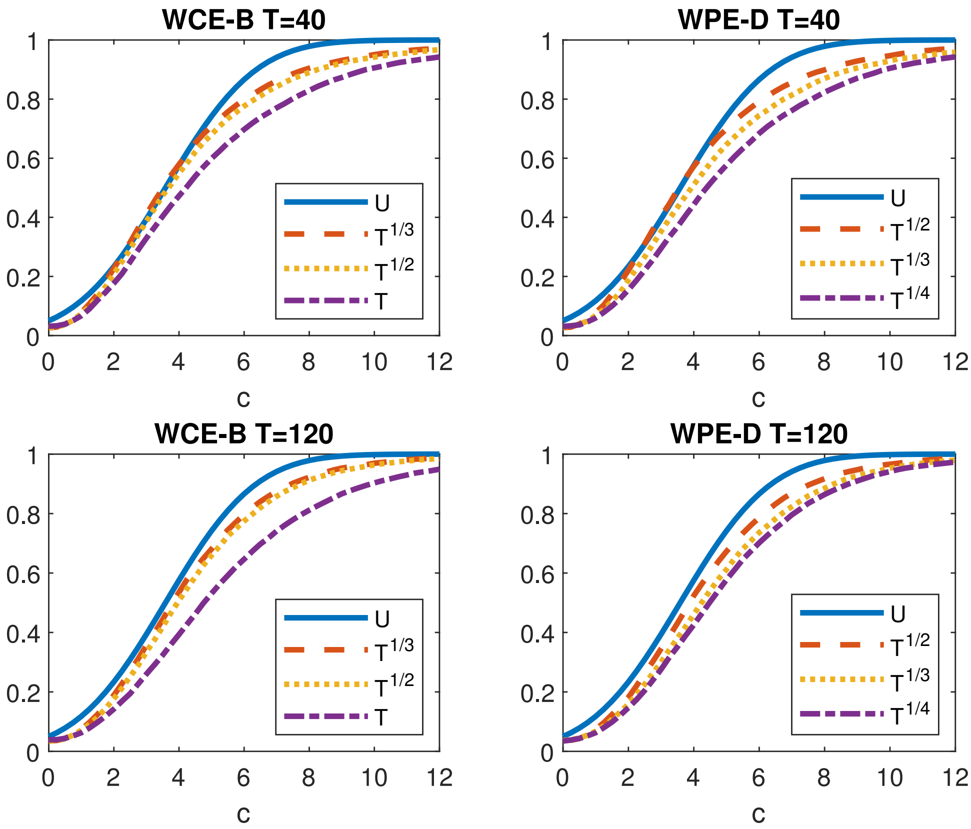
Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with Linex loss function, $\alpha = 0.5$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.37: DM_{HLN} finite sample local power and Linex $\alpha = -0.9$ loss function



Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with Linex loss function, $\alpha = -0.9$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Figure 3.38: DM_{HLN} finite sample local power and Linex $\alpha = -1$ loss function



Note: the Figure reports power performances of the alternative Diebold and Mariano test by Harvey, Leybourne and Newbold (1997) with Linex loss function, $\alpha = -1$, in samples of size $T = 40, 120$. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. The blue solid line U refers to the infeasible case in which the unknown variance is used. All other lines represent different bandwidths used in the estimation of the long run variance.

Chapter 4

The Accuracy of Density Aggregate Forecasts of the ECB SPF

ABSTRACT: I perform a real-time density forecast evaluation of the European Survey of Professional Forecasters (ECB SPF) using the Diebold and Mariano test for equal forecast accuracy with Quadratic Probability Score and Ranked Probability Score functions. As the sample size for ECB SPF is small and this affects size performances of the Diebold and Mariano test, I use fixed-smoothing asymptotics that proved to alleviate size distortion in small samples in other evaluation settings. A Monte Carlo exercise shows that fixed-smoothing asymptotics deliver correctly sized tests also in this occasion. Empirical results show that unemployment and real GDP growth ECB SPF forecasts perform better than benchmarks especially for short horizons while HICP inflation forecasts do not outperform benchmarks as inflation expectations are close to the ECB target.

Keywords: Diebold and Mariano test, long run variance estimation, fixed-smoothing asymptotics, Heteroscedasticity Autocorrelation Robust (HAR) inference, SPF, Real-time forecast evaluation, Hypothesis testing, density forecasts

JEL Classification: C12, C32, C51, C53, E17

This Chapter is adapted from the working paper '*A Real-time Density Forecast Evaluation of the ECB Survey of Professional Forecasters*' with Dr Laura Coroneo (University of York) and Dr Fabrizio Iacone (Università degli Studi di Milano and University of York).

4.1 Introduction

Survey forecasts play a crucial role in the economy as they are used by several economic agents such as central banks, financial institutions and investors to plan decision and monetary policy.

Surveys usually provide several types of forecasts: point forecasts are the expected value in the future of the target variable. They are the easiest to interpret and they are widely available. However, the information they provide is limited. For instance, there is no description of forecast uncertainty which gives information about the confidence of the forecaster about their prediction while its importance is well recognised in the literature as noted by Garratt, Lee, Pesaran and Shin (2003), Abel, Rich, Song and Tracy (2016) and Müller and Watson (2016) among others.
Interval forecasts (Chatfield, 1993; Christoffersen, 1998) provide a range in which the true outcome may fall and in this way they provide agents with slightly more information than point forecasts.
Density forecasts, instead, associate to every possible outcome of the target variable a probability and, for this reason, they are more informative than other types of forecasts. Tay and Wallis (2000) maintain '*a density forecast of the realization of a random variable at some future time is an estimate of the probability distribution of the possible future values of that variable*'. In practice, density forecasts give a wider understanding of the uncertainty associated with the prediction and their benefits have been recognised in the literature since the 80s (Fair, 1980; Dawid, 1984).
Many survey forecasts, nowadays, collect density forecasts in a systematic manner. The longest running density forecasts survey of macroeconomic variables series was set up by The Business and Economic Statistics Section of the American Statistical Association and the National Bureau of Economic Research in 1968 and then taken over by the Federal Reserve Bank of Philadelphia (FED) that named the survey 'Survey of Professional Forecasters' (US SPF). The FED provides a series of intervals, or bins, and asks participants to report the probability the future value of the target variable might fall in. In this way,

density forecasts are reported as histograms. In the UK, the Bank of England started reporting density forecasts of inflation in 1996 and now runs a quarterly survey called 'Survey of External Forecasters'. Few years later, the European Central Bank started its Survey of Professional Forecasters (ECB SPF) providing every quarter density forecasts about inflation, unemployment rate and real GDP growth for the euro area.

Considering the importance of survey density forecasts, there is the need for methods to check their quality and reliability. In this light, a body of the literature discusses absolute density forecast evaluation which refers to the correct specification of the density and it is implemented using a Probability Integral Transform (PIT) by Rosemblatt (1952). Dawid (1984), Diebold, Tay and Wallis (1997) and Diebold, Gunther and Tay (1998) are among the first to perform absolute evaluation and since them other works became available. Clements and Smith (2000) and Rossi and Sekhposyan (2014) use PIT on US density forecasts of output growth and unemployment, Clements (2004) and Boero, Smith and Wallis (2008) test Bank of England Survey of External Forecasters for unbiasedness and efficiency, Rossi and Sekhposyan (2013) test the correct specification of density forecasts extending the Corradi and Swanson (2006) setting. More recently, Clements (2018) performs a detailed absolute evaluation of US SPF and then compare the same survey forecasts to a set of competing unconditional density forecasts. The latter method is referred to as relative density forecast evaluation which concerns the assessment of a forecast with respect to another one of the same target variable. The standard approach for relative evaluation involves the Kullback-Leibler Information Criterion (KLIC) as in Bao, Lee and Saltoglu (2004), Mitchell and Hall (2005), Bao, Lee and Saltoglu (2007), Diks, Panchenko and Van Dijk (2011) and Clements (2018) among others. If the KLIC is used as a loss function in a Diebold and Mariano test, this type of evaluation is the one proposed by Amisano and Giacomini (2007), Gneiting and Ranjan (2011) and Diks, Panchenko and Van Dijk (2011), which exploits the Diebold and Mariano (1995) framework but rather than comparing Mean Squared Errors of the two forecasts, it compare their weighted logarithmic scores (Good, 1952). Using a logarithmic score implies that the density forecasts to compare are provided as a continuous distribution and the probability is always bigger than zero but this is not always the case. For example, Mitchell and Hall (2005) take National Institute Economic Review inflation density forecasts, which are reported as histograms with zero probability associated to some bins, and derive continuous density forecasts assuming normality to perform evaluation. However, the true distribution is unknown and assuming normality to obtain a viable density function may not be appropriate. In this concern, Boero, Smith and Wallis (2011) present a survey of loss functions that can be used to evaluate density forecasts and suggest the Quadratic Probability Score (QPS) by Brier (1950) and the Ranked Probability Score (RPS) by Epstein (1969) as suitable alternatives. Existing literature includes Lopez (2001), which employs the QPS in a Diebold and Mariano test for predictive accuracy of volatility density forecasts while Kenny, Kostka and Masera (2014) evaluate individual density forecasts of real GDP growth and inflation from the ECB SPF with a Diebold and Mariano test with QPS and RPS as loss functions.

The Diebold and Mariano framework is simple and the test statistic is easy to compute; these features make it widely used also for density forecast evaluation. However, as recognised by Diebold and Mariano (1995) themselves and by Clark (1999) among others, it suffers from small sample size distortion. Correctly sized test can be delivered using fixed-smoothing asymptotics for the limit distribution of the expected loss differential of the two competing forecasts. For point forecasts, Coroneo and Iacone (2020) suggest fixed-b asymptotics by Kiefer and Vogelsang (2005) in which the limiting distribution is non standard and it is derived taking into account the kernel function used in the estimate of the long run variance and the bandwidth to sample ratio b. Alternatively, they suggest using fixed-m asymptotics as in Sun (2013), Hualde and Iacone (2015a) and Hualde and Iacone (2017) in which the limiting distribution is a t-Student and the long run variance estimate is based on a Weighted Periodogram Estimate with Daniell kernel and bandwidth parameter m constant as the sample size increases. Simulations in these works show that tests are correctly sized when a quadratic loss function and an absolute function are used. To assess whether the Diebold and Mariano test with QPS and RPS remains correctly sized, I perform a Monte Carlo exercise of the test paired with these two loss functions for density forecast evaluation using a small sample. Results show that the test is correctly sized with a correct choice of the bandwidth and so it can be used to evaluate density forecasts in cases where survey rounds available are limited as in the ECB SPF case.

Previous works such as Kenny, Kostka and Masera (2014) and Krüger (2017), disregarded the sample size distortion and performed density forecast evaluation of ECB SPF with a Diebold and Mariano test presenting potentially spurious results. In this work, after observing promising Monte Carlo results, I employ fixed-smoothing asymptotics to evaluate rolling horizon one-year and two-years aggregate density forecasts of inflation, real GDP growth and unemployment rate from the ECB SPF. With this approach, there is no need to transform histograms forecasts of the surveys in continuous probability distribution assuming a particular distributional form as true and the test is correctly sized thanks to fixed-smoothing asymptotics. I compare ECB SPF forecasts to competing forecasts from simple models that should be easily beaten by professional forecasters: Uniform distribution assigning the same probability to all bins, a Gaussian random walk based on the assumption that the target variable follows a random walk without drift and a naive forecasts taken from the previous round of SPF forecasts. All forecasts are produced in real-time, that is using the same information available to professional forecasters at each survey deadline. Results show the superiority of ECB SPF forecasts for short term forecasts of unemployment and real GDP growth while the null hypothesis of equal forecast accuracy cannot be rejected for inflation no matter the forecast horizon indicating that expectations are anchored to the inflation target set by the ECB. Results are robust to the loss function used and the vintage of the realised target variable.

The remainder of this Chapter proceed as follows. The next Section describes benchmark forecasts used to assess the accuracy of ECB SPF density forecasts. The third Section presents the loss functions used. Section four describes the Monte Carlo exercise and its results. Section five illustrate the empirical exercise of ECB SPF density forecast evaluation and Section six concludes.

4.2 Benchmark Forecasts

This Section describes competing forecast models used as benchmarks to ECB SPF forecasts in this evaluation study. Benchmark forecasts should be simple enough to be easily beaten by ECB SPF forecasts and they are constructed to be fully real-time in the sense that only information available to professional forecasters up to the deadline for responding to each survey round is used. In a coherent way, all benchmarks report density forecasts as histograms assigning to each possible interval of outcomes a probability. These benchmarks are taken from the works of Kenny, Kostka and Masera (2014) and Krüger (2017) with some alterations to be fully real-time.

4.2.1 Uniform

This benchmark is based on a uniform distribution with constant probability between two thresholds identified using the historical realisations of the target variable at the vintage available at the survey deadline. The two thresholds are found from the maximum and minimum of the target variable observed among the latest 20 quarterly observations at the latest vintage available at each survey deadline. The constant probability assigned to the function between the thresholds is calculated dividing 1 by the number of bins of the considered survey round which fall between the thresholds. The probability density function is defined as

$$f_U(y_{t+h}) = \begin{cases} \dfrac{1}{k_t} & a^V \le y_{t+h} \le b^V \\ 0 & otherwise, \end{cases} \tag{4.1}$$

where $a^V = \min(y^V_t, y^V_{t-1}, y^V_{t-2}, \ldots, y^V_{t-19},)$, $b^V = \max(y^V_t, y^V_{t-1}, y^V_{t-2}, \ldots, y^V_{t-19},)$, k_t is the number of bins between a^V and b^V, V is the vintage of past observations available at the deadline of each survey round and h is the forecast horizon.

With this benchmark, the one-year ahead forecast coincides with the two-years ahead forecast and the probability assigned to each bin is the same for all bins falling between the minimum and maximum thresholds detected from past observations and zero for all the remaining bins.

4.2.2 Gaussian Random Walk

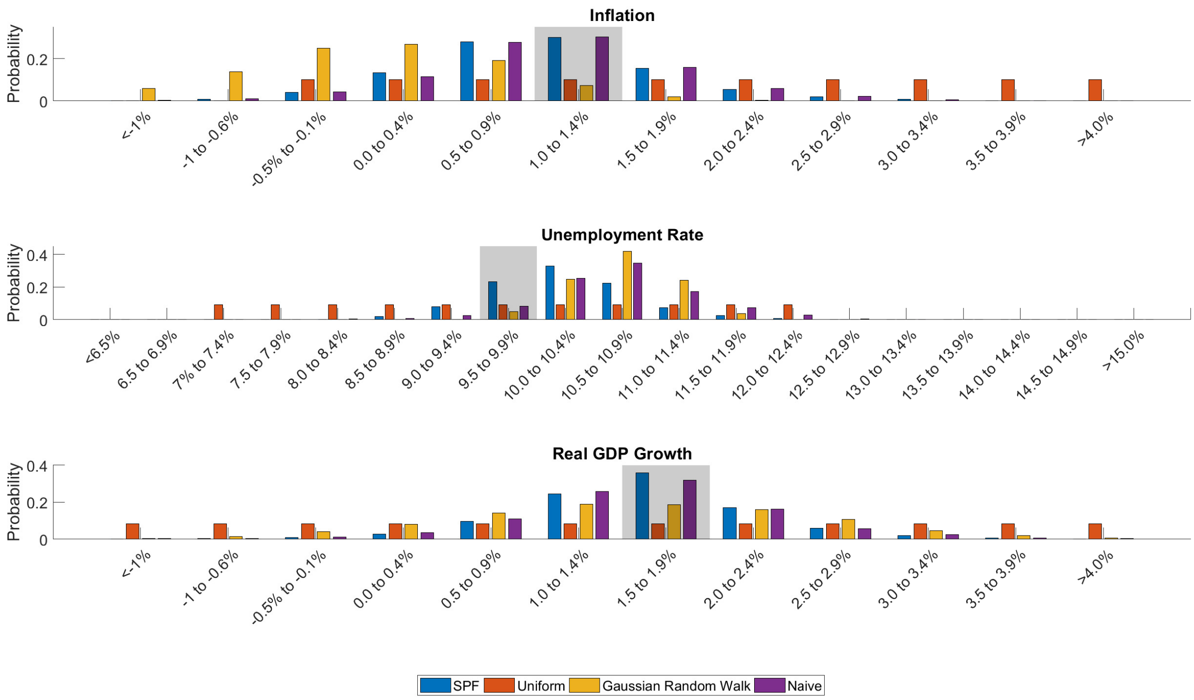
Forecasts of this benchmark are drawn from a normal distribution $N(\mu_{t+h}, \sigma^2_{t+h})$ with parameters calculated using the latest 20 past observations $y^V_t, \ldots, y^V_{t-19}$ of the target variable at quarterly frequency at the vintage V available at each survey deadline under the assumption that every target variable follows a random walk without drift. $\mu_{t+h} = y^V_t$ is the conditional expectation of the random walk and $\sigma^2_{t+h} = h \times (19)^{-1} \sum_{i=t-19}^{t} (y^V_i - y^V_{i-1})^2$ is the conditional variance, h is the forecast horizon in quarters.

4.2.3 Naive

For this benchmark, no distributional assumption is made. The forecast for the current survey is assumed to be the same as the previous aggregate density survey forecast for the same forecast horizon. In the case of different bins available from a survey round to the following, forecasts are slightly modified to accommodate the new bins structure: if in the new survey round there are more bins than in the previous, the probability of the last bin in the previous round is equally split across the additional bins available in the new round while, if there are less bins in the current survey round than the previous round, the probabilities of extreme bins are summed up and placed in the only available bin.

Figure 4.1 illustrates competing forecasts for each of the target variables for every benchmark and the ECB SPF: the blue histogram is the SPF forecast, the red histogram is the Uniform benchmark forecast, the yellow histogram is the Gaussian random walk forecast and the purple histogram is the naive benchmark forecast. The shaded area indicates the interval in which the realised at the latest revision falls.

Figure 4.1: Competing one-year ahead forecasts for the 2016.Q1 survey round



Note: competing one-year ahead forecasts for the 2016.Q1 survey round for inflation (top plot), unemployment rate (middle plot) and real GDP growth (bottom plot). The blue histograms are the ECB SPF forecasts, the red histograms denote the uniform benchmark forecasts, the yellow histograms refer to the Gaussian random walk forecasts and the purple histograms denote the naive benchmark forecasts. The shaded areas indicate the intervals selected by the realised outcomes at the current release.

4.3 Density Forecasts Evaluation

The evaluation of density forecasts is usually performed adopting a logarithmic score by Good (1952) as in the setting of Amisano and Giacomini (2007), Gneiting and Ranjan (2011) and Diks, Panchenko and Van Dijk (2011). In the case of survey forecasts in which probabilities associated with every possible outcome are reported as discrete densities and sometimes zero probability is assigned to a specific outcome, the logarithmic score is not ideal as the logarithm of zero is not defined. Boero, Smith and Wallis (2011) suggests alternative loss functions which may be more appropriate to histogram forecasts such as the Quadratic Probability Score (QPS) by Brier (1950),

$$QPS^V = \sum_{k=1}^{K} (f_{t+h}^k - x_{t+h}^{k,V})^2, \tag{4.2}$$

where K is the number of bins or intervals of the histogram for every survey round, f^k_{t+h} is the probability assigned to the outcome of the k-th bin for the period $t+h$ forecast and $x^{k,V}_{t+h}$ is a binary random variable that takes the value of 1 if the period $t+h$ outcome at vintage V falls in bin k and zero otherwise. In this way, $x^{k,V}_{t+h}$ represents the realised value of the target variable which is subject to revision and this is incorporated in $x^{k,V}_{t+h}$ itself which can change according to the vintage of the realisation of the target variable used in the evaluation. This loss function penalises severely any probability assigned to events that do not occur.

The Ranked Probability Score (RPS) by Epstein (1969), instead, is based on the cumulative distribution functions of the density forecasts F^k_{t+h} and of the binary variable $X^{k,V}_{t+h}$. It considers the overall tendency of the forecast probability density function and it tends to penalise less severely density forecasts which assign relatively larger probabilities to outcomes that are close to the true outcome.

$$RPS^V = \sum_{k=1}^{K}(F^k_{t+h} - X^{k,V}_{t+h})^2 \tag{4.3}$$

Reflecting this sensitivity to distance, the RPS gives some reward to a density forecast that has a near miss while the QPS will not distinguish between two competing forecasts in this way. According to Gneiting and Raftery (2007), RPS has the desirable property of being proper in the sense that encourages the forecasters to quote their true belief and reward higher moments feature of the density forecast.

QPS and RPS can be used as loss functions in a DM test presented in Section 3.4 of Chapter 2 to test the null hypothesis of equal forecast accuracy in density forecasts reported as histograms. The process to estimate the long run variance is still valid and fixed-smoothing asymptotics are effective to address the sample size distortion problem of DM tests.

4.4 Monte Carlo Study for Size and Power

Existing literature investigated how fixed-smoothing asymptotics can alleviate size distortion in Diebold and Mariano type tests under several types of loss functions. To check whether the size improvements for this test still hold in the case of QPS and RPS loss functions, I perform a Monte Carlo exercise for the empirical size and power with a small set of simulated forecasts and the presence of serial correlation.

To replicate the setting of the empirical exercise with SPF forecasts reported as histograms and showing serial correlation, the DGP should come from a discrete distribution and with a good degree of dependence. A possible way to generate dependence in a discrete distribution setting is the following: the DGP series x_t starts from taking a random integer number from a discrete Uniform distribution $\mathrm{U}(1,k)$, where k is an odd integer indicating the number of possible outcomes for x_t. Every other following observations of x_t after the first is set as follows:

$$x_t = \begin{cases} U(1, \lfloor k/2 \rfloor + 1) & \text{if } 1 \leq x_{t-1} \leq \lfloor k/2 \rfloor, \\ U(1, k) & \text{if } x_{t-1} = \lfloor k/2 \rfloor + 1, \\ U(\lfloor k/2 \rfloor + 1, k) & \text{if } \lfloor k/2 \rfloor + 1 < x_{t-1} \leq k. \end{cases}$$

After q iterations, the DGP resets and x_t is selected again randomly from a discrete Uniform distribution $\mathrm{U}(1,k)$ and the following observations are set accordingly to the preceding realisation of x_t as before; the same process repeats until the end of the sample, resetting every q iterations. Hence, when $q > 1$ the process is dependent and the dependence increases with q. To see how dependence is generated take, for example, $k = 5$ and $q = 2$. x_t is generated as follows: when $t = 1$, $P(x_1 = k) = 1/5$ for $k = 1, \ldots, 5$. For $t = 2$, $P(x_2 = k | x_1 = 1 \text{ or } 2) = 1/3$ for $k = 1, \ldots, 3$; $P(x_2 = k | x_1 = 3) = 1/5$ for $k = 1, \ldots, 5$; $P(x_2 = k | x_1 = 4 \text{ or } 5) = 1/3$ for $k = 3, \ldots, 5$. When $t = 3$ the DGP resets and $P(x_3 = k) = 1/5$ for $k = 1, \ldots, 5$, and so on.

A forecast for x_t, f_1, is taken from a discrete Uniform distribution $U(1, \lfloor k/2 \rfloor + 1)$ and the competing forecast f_2 is taken from a discrete Uniform distribution $U(\lfloor k/2 \rfloor + 1, k)$. In this way, the null hypothesis of equal forecast accuracy is always satisfied because both forecasts are on average equally wrong under H_0. To evaluate power performances, the competing forecasts f_2 are taken from a discrete Uniform distribution $U(\lfloor k/2 \rfloor + 1 + c, k + c)$ so that f_2 is shifted by $c \in [0, 12]$, the bigger the c, the more wrong the competing forecast is. The sample sizes of $T = \{30, 60\}$ are considered to match those available in the empirical exercise and I repeat the experiment for $10,000$ replications.

Tables 4.1 - 4.4 report empirical size of the test when critical values from both standard asymptotics and fixed-smoothing asymptotics are used. In this simulation, q takes integer values between 1 and 5 and k is set to 21 and 5. In columns WCE, the long run variance estimate is derived using (2.13) with a Bartlett kernel bandwidths $M = \lfloor T^{1/3} \rfloor$, $M = \lfloor T^{1/2} \rfloor$ and $M = T$ except for the colum DM in which the long run variance is obtained using (2.11). In WPE columns, the estimator used is (2.15) with a Daniell kernel bandwidths $m = \lfloor T^{1/4} \rfloor$, $m = \lfloor T^{1/3} \rfloor$, $m = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{2/3} \rfloor$. With both sample sizes, standard asymptotics cannot provide correctly sized tests also when $q = 1$ and the performance worsen as the dependence, q, increases. Fixed-smoothing asymptotics instead, produce correctly sized tests for WCE bandwidth $M = \lfloor T^{1/2} \rfloor$ and WPE bandwidth $m = \lfloor T^{1/3} \rfloor$ for both QPS and RPS loss functions confirming results in the literature for point forecasts (Harvey, Leybourne and Whitehouse, 2017; Coroneo and Iacone, 2020). However, in the sample of size $T = 30$, tests are oversized also with fixed-smoothing asymptotics for $q \geq 3$ still, the size distortion is reduced in comparison to standard asymptotics. In the same sample, as q increases, the WPE-D bandwidth $m = \lfloor T^{2/3} \rfloor$ fails to provide acceptably sized tests because the long run variance estimator uses 9 Fourier frequencies over a sample of only 30 observations. This may cause the estimate of σ^2 to be subject to too much bias.

Power performances are reported in Figures 4.2 and 4.3 using only bandwidths that showed good size performances, namely WCE-B $M = \lfloor T^{1/2} \rfloor$ and WPE-D $m = \lfloor T^{1/3} \rfloor$,

Table 4.1: Empirical size of the DM test with standard asymptotics, $T = 60$

Quadratic Probability Score

$k = 21$

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.073	0.066	0.086	0.333	0.121	0.100	0.070	0.062
2	0.086	0.097	0.106	0.339	0.128	0.105	0.080	0.091
3	0.095	0.121	0.112	0.355	0.123	0.101	0.088	0.135
4	0.112	0.157	0.127	0.353	0.127	0.109	0.105	0.196
5	0.124	0.190	0.139	0.349	0.132	0.112	0.128	0.238

$k = 5$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.071	0.066	0.087	0.330	0.122	0.099	0.068	0.058
2	0.085	0.091	0.104	0.340	0.125	0.105	0.080	0.083
3	0.099	0.110	0.108	0.348	0.124	0.101	0.084	0.115
4	0.114	0.137	0.120	0.349	0.129	0.108	0.097	0.152
5	0.124	0.140	0.117	0.345	0.124	0.102	0.099	0.161

Ranked Probability Score

$k = 21$

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.073	0.067	0.084	0.331	0.117	0.094	0.071	0.061
2	0.088	0.095	0.104	0.340	0.125	0.103	0.083	0.088
3	0.096	0.116	0.113	0.350	0.125	0.099	0.088	0.131
4	0.112	0.156	0.127	0.352	0.128	0.110	0.104	0.190
5	0.124	0.183	0.136	0.346	0.130	0.111	0.124	0.224

$k = 5$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.074	0.067	0.087	0.329	0.119	0.097	0.071	0.060
2	0.084	0.091	0.103	0.341	0.125	0.104	0.079	0.084
3	0.097	0.113	0.110	0.347	0.124	0.098	0.086	0.117
4	0.116	0.135	0.121	0.346	0.133	0.110	0.096	0.149
5	0.125	0.139	0.119	0.341	0.123	0.101	0.098	0.159

Note: the Table reports the empirical size of the Diebold and Mariano test with standard asymptotics and sample size $T = 60$. The theoretical size is 5%. k is an odd number indicating the number of possible realisations of the target variable. q indicates the number of periods after which the DGP resets. The higher the q, the higher the level of dependence in the process. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance (except for column DM, where the rectangular kernel is used); WCE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance.

Table 4.2: Empirical size of the DM test with fixed-smoothing asymptotics, $T = 60$

Quadratic Probability Score

$k = 21$

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.045	0.045	0.050	0.050	0.054	0.051	0.052
2	0.071	0.056	0.057	0.055	0.058	0.059	0.080
3	0.093	0.064	0.058	0.053	0.052	0.065	0.123
4	0.124	0.076	0.065	0.055	0.058	0.079	0.178
5	0.155	0.085	0.073	0.060	0.062	0.101	0.217

$k = 5$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.043	0.043	0.047	0.052	0.053	0.050	0.049
2	0.065	0.055	0.057	0.053	0.057	0.056	0.073
3	0.083	0.061	0.057	0.052	0.053	0.063	0.104
4	0.101	0.070	0.059	0.054	0.058	0.073	0.137
5	0.109	0.069	0.062	0.053	0.057	0.075	0.145

Ranked Probability Score

$k = 21$

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.045	0.045	0.049	0.049	0.053	0.051	0.050
2	0.069	0.058	0.057	0.054	0.056	0.058	0.076
3	0.090	0.061	0.059	0.054	0.051	0.063	0.116
4	0.122	0.074	0.065	0.058	0.058	0.080	0.173
5	0.149	0.083	0.070	0.058	0.059	0.099	0.208

$k = 5$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.045	0.044	0.048	0.050	0.052	0.050	0.052
2	0.064	0.056	0.057	0.053	0.054	0.058	0.073
3	0.082	0.059	0.056	0.051	0.052	0.062	0.106
4	0.102	0.070	0.062	0.054	0.057	0.071	0.134
5	0.108	0.068	0.062	0.052	0.058	0.076	0.143

Note: the Table reports the empirical size of the Diebold and Mariano test with fixed-smoothing asymptotics and sample size $T = 60$. The theoretical size is 5%. k is an odd number indicating the number of possible realisations of the target variable. q indicates the number of periods after which the DGP resets. The higher the q, the higher the level of dependence in the process. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WCE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance.

Table 4.3: Empirical size of the DM test with standard asymptotics, $T = 30$

Quadratic Probability Score

$k = 21$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.102	0.083	0.106	0.326	0.124	0.100	0.079	0.065
2	0.128	0.117	0.130	0.346	0.131	0.106	0.096	0.119
3	0.159	0.147	0.151	0.356	0.137	0.118	0.123	0.182
4	0.205	0.188	0.169	0.374	0.143	0.131	0.156	0.256
5	0.211	0.230	0.203	0.371	0.161	0.169	0.204	0.276

$k = 5$

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.101	0.083	0.105	0.339	0.123	0.100	0.078	0.064
2	0.130	0.111	0.126	0.350	0.128	0.107	0.092	0.107
3	0.160	0.139	0.143	0.357	0.134	0.115	0.112	0.158
4	0.205	0.162	0.157	0.367	0.136	0.124	0.134	0.188
5	0.227	0.180	0.167	0.361	0.136	0.132	0.151	0.213

Ranked Probability Score

$k = 21$

	WCE				WPE			
q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.098	0.080	0.102	0.335	0.119	0.099	0.077	0.062
2	0.123	0.114	0.126	0.348	0.132	0.106	0.091	0.110
3	0.157	0.149	0.150	0.357	0.138	0.116	0.120	0.181
4	0.202	0.186	0.171	0.378	0.146	0.134	0.158	0.236
5	0.210	0.222	0.197	0.367	0.148	0.157	0.201	0.269

$k = 5$

q	DM	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.100	0.083	0.104	0.335	0.120	0.099	0.079	0.064
2	0.126	0.112	0.124	0.351	0.127	0.107	0.091	0.107
3	0.160	0.138	0.142	0.355	0.133	0.115	0.111	0.156
4	0.204	0.160	0.156	0.361	0.138	0.124	0.132	0.186
5	0.229	0.177	0.167	0.362	0.137	0.130	0.152	0.210

Note: the Table reports the empirical size of the Diebold and Mariano test with standard asymptotics and sample size $T = 30$. The theoretical size is 5%. k is an odd number indicating the number of possible realisations of the target variable. q indicates the number of periods after which the DGP resets. The higher the q, the higher the level of dependence in the process. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance (except for column DM, where the rectangular kernel is used); WCE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance.

Table 4.4: Empirical size of the DM test with fixed-smoothing asymptotics, $T = 30$

Quadratic Probability Score

$k = 21$

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.041	0.040	0.048	0.050	0.050	0.051	0.051
2	0.070	0.063	0.057	0.055	0.061	0.064	0.101
3	0.104	0.082	0.075	0.064	0.066	0.097	0.157
4	0.131	0.099	0.084	0.060	0.086	0.125	0.226
5	0.179	0.121	0.084	0.053	0.075	0.175	0.254

$k = 5$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.042	0.042	0.047	0.053	0.051	0.049	0.049
2	0.062	0.058	0.056	0.055	0.058	0.062	0.085
3	0.083	0.069	0.068	0.057	0.060	0.078	0.133
4	0.101	0.079	0.074	0.061	0.070	0.095	0.161
5	0.120	0.090	0.081	0.063	0.077	0.114	0.187

Ranked Probability Score

$k = 21$

	WCE			WPE			
q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.040	0.040	0.045	0.047	0.049	0.051	0.049
2	0.066	0.060	0.058	0.053	0.058	0.063	0.091
3	0.094	0.075	0.071	0.060	0.066	0.087	0.154
4	0.128	0.096	0.080	0.059	0.081	0.121	0.208
5	0.166	0.112	0.084	0.058	0.089	0.164	0.243

$k = 5$

q	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	T	$\lfloor T^{1/4} \rfloor$	$\lfloor T^{1/3} \rfloor$	$\lfloor T^{1/2} \rfloor$	$\lfloor T^{2/3} \rfloor$
1	0.042	0.041	0.046	0.051	0.052	0.052	0.050
2	0.062	0.057	0.056	0.054	0.058	0.061	0.086
3	0.082	0.069	0.065	0.057	0.060	0.076	0.130
4	0.102	0.079	0.075	0.061	0.071	0.097	0.157
5	0.117	0.090	0.079	0.063	0.077	0.113	0.183

Note: the Table reports the empirical size of the Diebold and Mariano test with fixed-smoothing asymptotics and sample size $T = 30$. The theoretical size is 5%. k is an odd number indicating the number of possible realisations of the target variable. q indicates the number of periods after which the DGP resets. The higher the q, the higher the level of dependence in the process. WCE refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WCE refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance.

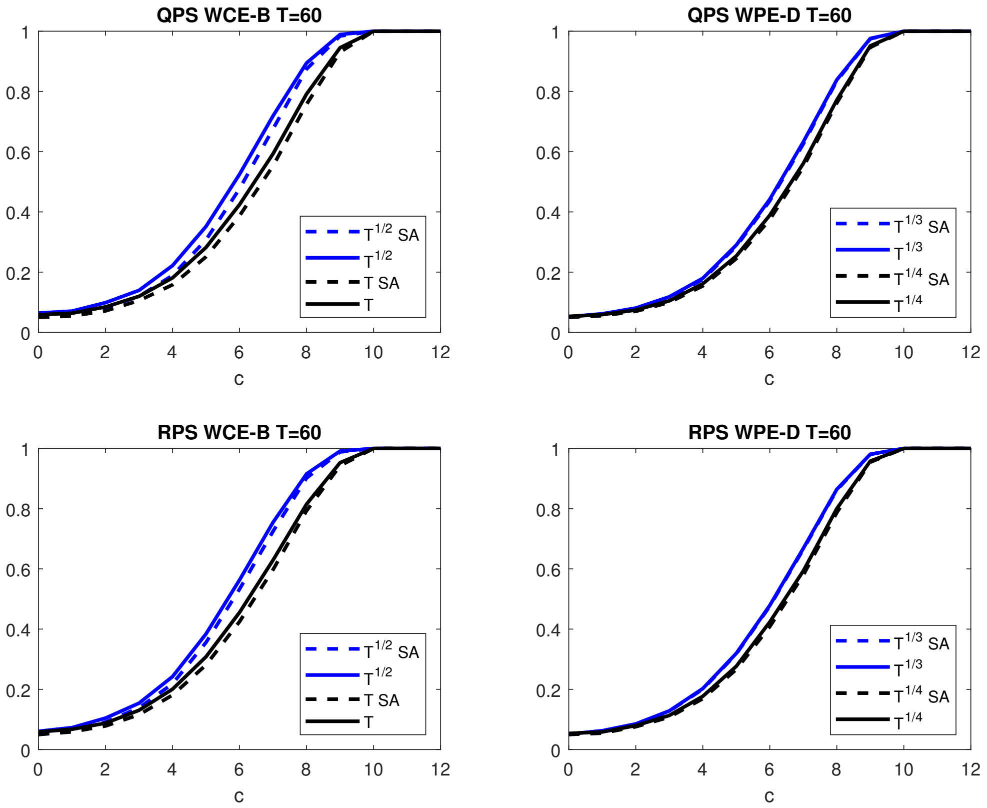
and critical values from fixed-smoothing asymptotics. k is set to 21 and q is 3. Bandwidths reported as solid blue lines are those cited before, that showed the best performances in size and the dashed lines depict size-adjusted power. In all cases the empirical power is a good approximation of the size adjusted power, again offering support to the assumption that fixed-smoothing asymptotics are a valuable instrument for inference. However, when the WCE estimate is used, the feasible power curves deviate from the size-adjusted ones, especially for the sample of only 30 elements. In line with the existing literature, smaller M (or larger m) give better power performances and $M = \lfloor T^{1/2} \rfloor$ and $m = \lfloor T^{1/3} \rfloor$ confirm to be the most suitable bandwidths for the sample sizes considered here. Power performances are not influenced by the type of loss function used and there is no reason to prefer QPS or RPS on the basis of size or power.

4.5 Empirical Results

For this empirical exercise, I use aggregate density forecasts of HICP inflation, unemployment rate and real GDP growth from the ECB SPF described in Chapter 1. Core inflation was excluded from this exercise because its sample is still too limited. I evaluate surveys between 2001.Q1 and 2016.Q2 ($T = 62$). To check the effect of the 2008 financial crises on forecaster, I also split the sample in two sub-samples of 31 observations each: 2001.Q1 - 2008.Q3 and 2008.Q4 - 2016.Q2. The sample of the realised series is taken from the Real-time database by Giannone, Henry, Lalik and Modugno (2012) and starts in December 2001 to March 2018 for HICP, from November 2001 to February 2018 for the unemployment rate and from 2001.Q3 to 2017.Q4 for real GDP growth. The choice of the start and end date of the evaluation period was driven by the data available in the real-time database.

ECB SPF density forecasts are compared to three simple benchmarks described in Section 4.2 that should be easy to beat by the ECB SPF: Uniform, Gaussian random walk and naive. The comparison is done using the Diebold and Mariano test with two different

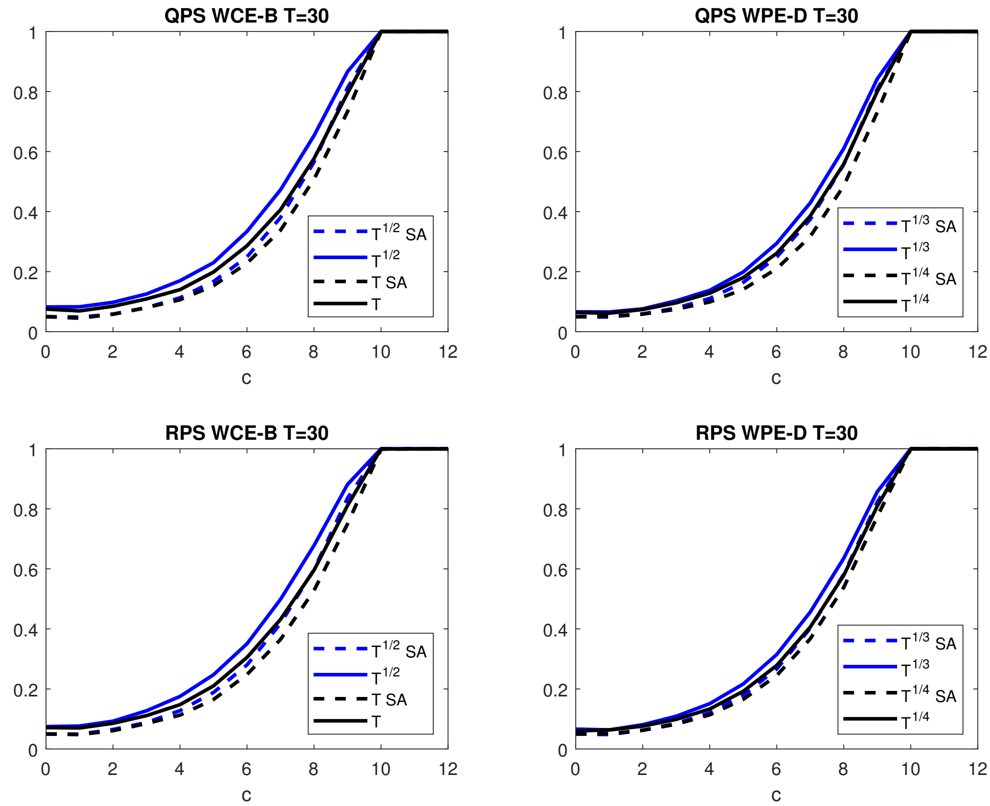
Figure 4.2: Finite sample local power, $T = 60$



Note: the Figure reports power performances of the Diebold-Mariano test with Quadratic Probability Score (QPS) and Ranked Probability Score (RPS) in a sample of size $T = 60$. The dashed lines refer to power performances using size-adjusted critical values while solid lines use fixed-smoothing asymptotics. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Data generating process with the number of possible realisations of the target variable $k = 21$ and the number of periods after which the DGP resets $q = 3$.

loss functions: Quadratic Probability Score and Ranked Probability Score which are largely employed in density forecast evaluation when forecasts are reported as histograms. Considering the results of the Monte Carlo exercise, the long run variance estimator is calculated using WCE bandwidth $M = \lfloor T^{1/2} \rfloor$ with the Bartlett kernel and with bandwidth $m = \lfloor T^{1/3} \rfloor$ for the WPE case with Daniell kernel. To perform a fully real-time evaluation, I use three different vintages of the realised variable: first release, four releases after the first and the latest available release on 30/01/2019. I also take revision into account when constructing benchmark forecasts such that only information and vintage available to forecasters are used to produce competing forecasts. Although information

Figure 4.3: Finite sample local power, $T = 30$



Note: the Figure reports power performances of the Diebold and Mariano test with Quadratic Probability Score (QPS) and Ranked Probability Score (RPS) in a sample of size $T = 30$. The dashed lines refer to power performances using size-adjusted critical values while solid lines use fixed-smoothing asymptotics. The parameter c indicates the distance from the null hypothesis. WCE-B refers to the test statistic with Weighted Covariance Estimate with Bartlett kernel for the long run variance; WPE-D refers to the test statistic with Weighted Periodogram Estimate with Daniell kernel for the long run variance. Data generating process with the number of possible realisations of the target variable $k = 21$ and the number of periods after which the DGP resets $q = 3$.

available to forecasters is the one reported in Table 1.1 of Chapter 1, in the case of inflation, at the survey deadline 2007.Q1, the latest realisation available to forecasters was January 2007 instead of December 2006. In my exercise, I use the exact information forecasters had available at the survey deadline as my aim is to perform a fully real-time exercise. In the case of unemployment, there is more revision and forecasters may not use newly available information because they are aware it is not reliable. In this case, I try both keeping and ignoring the additional information to obtain benchmark forecasts and results, available upon request, do not change. In this regard, surveys affected by this phenomenon are 2007.Q1 for which forecasters had one more realisation and 2004.Q2, 2008.Q4 and 2009.Q3 for which forecasters had one realisation less. For real GDP growth, the latest information available is the one expected but it has already been revised once except in the case of survey 2002.Q2 and I always use the latest revision available at the survey deadline.

DM test statistics values are reported in Tables 4.5 - 4.13. A negative value of the test indicates that the benchmark is performing better on average than the ECB SPF forecast while a positive value indicates the ECB SPF is predicting better than the benchmark, on average. Rejections from standard asymptotics critical values are indicated shading the appropriate cell; and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics critical values are reported using the superscript '*' such that ** and * indicate, respectively, two-sided significance at the 5% and 10% level using fixed-smoothing asymptotics from the appropriate distribution according to fixed-b or fixed-m case. The top panel of each Table uses the Quadratic Probability Score and the bottom panel uses the Ranked Probability Score.

In the case of HICP, there is no evidence ECB SPF forecasts outperform benchmarks. While this behaviour still appears for the first part of the sample, in the second part, rejections in favour of the ECB SPF forecasts are more frequent, although rather weak, especially against the naive benchmark. This evidence is coherent with de Vincent-Humphreys, Dimitrova, Falck and Henkel (2019) and the empirical works of Łyziak and Paloviita (2017) and Grishchenko, Mouabbi and Renne (2019) showing that when inflation is more loosely anchored to the target, like in the second part of the sample, professional forecasters take this into account and change their forecasting practice making it more judgement based.

For the unemployment rate, short term ECB SPF forecasts appear to always outperform benchmarks in all three samples. However, the predictive ability decreases as the release of the realised series is updated. SPF forecasts superiority is more marked in the second half of the sample in which, also, test statistics with RPS usually appear bigger than test statistics with QPS loss indicating that SPF forecasts often near-miss the true realization of the unemployment rate.

Testing real GDP growth gives similar results to those obtained for unemployment, ECB SPF forecast are superior in the short term, especially in the second half of the sample. However, the naive benchmark still proves to be quite easy to beat. In the second sub-sample, the ECB SPF significantly outperforms the uniform and Gaussian random walk forecasts.

The vintage of the realised series has no effect on HICP forecast evaluation while there is a minor effect on the unemployment rate and real GDP forecasts: using the first release, there is stronger evidence that ECB SPF outperform benchmarks. In general, RPS seems to be less sensitive to revision than QPS. Using RPS in the second sub-sample of inflation leads to rejections of the DM in favour of the ECB SPF and this indicates that professional forecasters placed probability in bins close to the true outcome only near-missing it.

The uniform benchmark seems the most difficult to beat especially for two-years ahead forecasts, in this case, the test statistic is often negative indicating that, on average, the benchmark was outperforming professional forecasters. On the other hand, professional forecasters always outperform the naive benchmark indicating that they actively update forecasts as new information becomes available.

Using fixed-smoothing asymptotics leads to less frequent rejections especially in sub-samples where tests are only performed on 31 rounds and the size distortion problem with standard asymptotics is exacerbated.

The results presented here are in line with those in Kenny, Kostka and Masera (2014), however, they are more robust due to the use of fixed-smoothing asymptotics. In addition, I could use a larger sample and analyse a sub-sample after the global financial crisis to confirm the change in forecasting practice which is consistent to the analysis in Łyziak and Paloviita (2017) and Grishchenko, Mouabbi and Renne (2019).

Table 4.5: DM test for the HICP inflation. Full sample Q1.2001 - Q2.2016, $T = 62$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
	First release					
WCE-DM	-0.55	-1.04	-0.10	-0.08	0.41	1.48
WCE-B	-0.53	-1.04	-0.11	-0.08	0.46	0.85
WPE-D	-0.57	-1.00	-0.11	-0.07	0.46	0.66
	Fifth release					
WCE-DM	-0.31	-0.87	0.30	-0.11	0.44	1.76
WCE-B	-0.30	-0.86	0.31	-0.11	0.48	1.40
WPE-D	-0.33	-0.82	0.25	-0.09	0.43	1.21
	Current release					
WCE-DM	-0.60	-1.06	0.06	0.02	0.48	0.90
WCE-B	-0.58	-1.06	0.06	0.02	0.54	0.62
WPE-D	-0.60	-1.02	0.05	0.01	0.56	0.48

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
	First release					
WCE-DM	1.05	0.90	1.26	0.07	1.41	1.48
WCE-B	0.92	0.80	1.53	0.07	1.41	1.27
WPE-D	0.87	0.77	1.31	0.07	1.31	1.06
	Fifth release					
WCE-DM	1.17	0.89	1.40	0.08	1.49	1.51
WCE-B	1.07	0.79	1.64	0.09	1.49	1.29
WPE-D	1.04	0.76	1.36	0.08	1.36	1.08
	Current release					
WCE-DM	0.99	0.94	1.32	0.10	1.40	1.45
WCE-B	0.86	0.83	1.55	0.11	1.40	1.24
WPE-D	0.81	0.79	1.31	0.10	1.30	1.03

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the inflation rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the full sample Q1.2001 - Q2.2016 ($T = 62$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.6: DM test for the HICP inflation. Sub-sample Q1.2001 - Q3.2008, $T = 31$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
	First release					
WCE-DM	-0.03	-0.46	-1.30	0.13	0.20	0.20
WCE-B	-0.03	-0.47	-1.33	0.14	0.22	0.19
WPE-D	-0.02	-0.47	-1.20	0.13	0.18	0.19
	Fifth release					
WCE-DM	0.14	-0.29	-1.35	0.08	0.24	0.87
WCE-B	0.12	-0.29	-1.38	0.09	0.26	0.83
WPE-D	0.11	-0.29	-1.24	0.08	0.22	0.75
	Current release					
WCE-DM	-0.03	-0.46	-1.30	0.26	0.32	-0.05
WCE-B	-0.03	-0.47	-1.33	0.28	0.36	-0.04
WPE-D	-0.02	-0.47	-1.20	0.25	0.30	-0.05

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
	First release					
WCE-DM	-0.93	1.07	-1.47	-0.68	0.47	0.15
WCE-B	-0.87	1.11	-1.45	-0.75	0.53	0.14
WPE-D	-0.76	1.10	-1.27	-0.65	0.45	0.13
	Fifth release					
WCE-DM	-0.79	1.06	-1.30	-0.72	0.63	0.22
WCE-B	-0.74	1.10	-1.30	-0.79	0.70	0.20
WPE-D	-0.66	1.09	-1.12	-0.70	0.59	0.19
	Current release					
WCE-DM	-0.93	1.07	-1.47	-0.56	0.45	0.12
WCE-B	-0.87	1.11	-1.45	-0.62	0.51	0.11
WPE-D	-0.76	1.10	-1.27	-0.54	0.43	0.10

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the inflation rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the sub-sample Q1.2001 - Q3.2008 ($T = 31$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.7: DM test for the HICP inflation. Sub-sample Q4.2008 - Q2.2016, $T = 31$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	-0.77	-0.94	1.08	-0.27	0.39	1.73
WCE-B	-0.78	-0.95	1.56	-0.29	0.41	1.17
WPE-D	-0.64	-0.79	2.03*	-0.23	0.34	1.18
Fifth release						
WCE-DM	-0.61	-0.85	1.42	-0.27	0.39	1.73
WCE-B	-0.62	-0.85	1.71	-0.29	0.41	1.17
WPE-D	-0.50	-0.70	1.66	-0.23	0.34	1.18
Current release						
WCE-DM	-0.86	-0.97	1.47	-0.27	0.39	1.73
WCE-B	-0.87	-0.98	1.78	-0.29	0.41	1.17
WPE-D	-0.70	-0.81	2.04*	-0.23	0.34	1.18

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	2.08	0.45	1.92	0.58	1.49	1.79
WCE-B	1.86	0.42	2.28*	0.63	1.57	1.68
WPE-D	1.66	0.34	2.31*	0.53	1.38	1.57
Fifth release						
WCE-DM	2.12	0.43	2.02	0.62	1.49	1.79
WCE-B	1.95	0.41	2.27*	0.67	1.57	1.68
WPE-D	1.73	0.33	2.24*	0.56	1.38	1.57
Current release						
WCE-DM	2.00	0.50	2.00	0.56	1.49	1.79
WCE-B	1.78	0.47	2.31*	0.61	1.57	1.68
WPE-D	1.59	0.38	2.33*	0.51	1.38	1.57*

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the inflation rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the sub-sample Q4.2008 - Q2.2016 ($T = 31$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.8: DM test for the unemployment rate. Full sample Q1.2001 - Q2.2016, $T = 62$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
			First release			
WCE-DM	2.64	2.36	3.34	-0.41	1.88	3.11
WCE-B	2.75**	2.06*	3.95**	-0.52	1.80	3.26**
WPE-D	3.04**	1.94*	4.54**	-0.68	1.43	3.62**
			Fifth release			
WCE-DM	1.93	2.51	3.62	-0.56	1.74	3.20
WCE-B	1.91	2.25*	4.46**	-0.70	1.66	3.35**
WPE-D	2.23*	1.91*	4.18**	-0.90	1.29	3.54**
			Current release			
WCE-DM	1.69	2.67	3.32	-1.33	1.85	1.88
WCE-B	1.68	2.47**	3.80**	-1.60	1.82	2.22*
WPE-D	1.55	2.40*	3.50**	-1.77	1.57	5.07**

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
			First release			
WCE-DM	4.47	2.17	3.50	0.46	1.79	2.46
WCE-B	3.78**	2.05*	4.16**	0.51	1.76	2.88**
WPE-D	2.98**	2.17*	4.96**	0.51	1.44	4.29**
			Fifth release			
WCE-DM	4.08	2.30	3.40	0.40	1.88	2.34
WCE-B	3.43**	2.17*	4.34**	0.43	1.82	2.79**
WPE-D	2.76**	2.25*	5.38**	0.43	1.46	4.22**
			Current release			
WCE-DM	2.97	2.14	3.24	-0.16	1.58	2.19
WCE-B	2.52**	1.97*	3.68**	-0.17	1.55	2.59**
WPE-D	2.08*	1.84	4.19**	-0.17	1.29	3.76**

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the unemployment rate against the uniform, the Gaussian random walk and the naive benchmark forecasts on the full sample Q1.2001 - Q2.2016 ($T = 62$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.9: DM test for the unemployment rate. Sub-sample Q1.2001 - Q3.2008, $T = 31$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
			First release			
WCE-DM	1.95	1.50	1.39	-0.28	1.12	1.95
WCE-B	1.99	1.38	1.58	-0.32	1.17	2.06*
WPE-D	1.68	1.22	1.50	-0.25	1.11	1.83
			Fifth release			
WCE-DM	1.76	1.96	2.25	-0.17	0.99	2.48
WCE-B	1.72	1.88	2.48**	-0.20	1.03	2.57**
WPE-D	1.39	1.64	2.14*	-0.15	0.95	2.27*
			Current release			
WCE-DM	1.37	1.82	2.20	-1.31	0.89	0.48
WCE-B	1.40	1.74	2.73**	-1.47	0.94	0.55
WPE-D	1.32	1.28	2.68**	-1.19	0.88	0.50

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
			First release			
WCE-DM	3.43	1.78	2.73	0.02	1.16	1.61
WCE-B	3.28**	1.68	2.57**	0.02	1.22	1.92
WPE-D	2.78**	1.54	2.56**	0.01	1.13	1.68
			Fifth release			
WCE-DM	2.93	2.13	3.10	0.03	1.16	1.55
WCE-B	2.80**	2.10*	3.25**	0.03	1.21	1.86
WPE-D	2.41*	1.74	2.68**	0.02	1.11	1.62
			Current release			
WCE-DM	1.20	1.72	3.49	-0.94	0.37	1.06
WCE-B	1.16	1.77	3.74**	-1.02	0.39	1.29
WPE-D	1.03	1.33	3.19**	-0.88	0.37	1.27

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the unemployment rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the sub-sample Q1.2001 - Q3.2008 ($T = 31$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.10: DM test for the unemployment rate. Sub-sample Q4.2008 - Q2.2016, $T = 31$

Quadratic Probability Score

	1 year ahead Uniform	GRW	Naive	2 years ahead Uniform	GRW	Naive
			First release			
WCE-DM	2.39	1.83	3.79	-0.35	1.74	2.72
WCE-B	2.77**	1.69	4.51**	-0.39	1.79	2.51**
WPE-D	2.22*	1.44	4.51**	-0.30	1.59	2.12*
			Fifth release			
WCE-DM	1.16	1.58	3.09	-0.86	1.71	2.02
WCE-B	1.32	1.47	3.51**	-0.97	1.75	1.96
WPE-D	1.07	1.24	3.49**	-0.73	1.54	1.60
			Current release			
WCE-DM	1.19	1.99	2.69	-0.55	1.98	2.96
WCE-B	1.35	1.82	2.75**	-0.62	2.09*	2.87**
WPE-D	1.06	1.56	2.43*	-0.47	1.82	2.30*

Ranked Probability Score

	1 year ahead Uniform	GRW	Naive	2 years ahead Uniform	GRW	Naive
			First release			
WCE-DM	3.33	1.77	2.97	0.71	1.46	2.08
WCE-B	2.93**	1.71	3.67**	0.77	1.53	2.19*
WPE-D	2.53**	1.41	3.24**	0.63	1.35	1.76
			Fifth release			
WCE-DM	3.16	1.77	2.59	0.59	1.57	1.96
WCE-B	2.76**	1.71	3.16**	0.65	1.63	2.09*
WPE-D	2.41*	1.40	2.71**	0.53	1.45	1.68
			Current release			
WCE-DM	3.46	1.90	2.48	0.77	1.67	2.04
WCE-B	3.11**	1.85	2.75**	0.84	1.76	2.18*
WPE-D	2.67**	1.55	2.22*	0.68	1.55	1.75

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the unemployment rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the sub-sample Q4.2008 - Q2.2016 ($T = 31$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). ■ and □ indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.11: DM test for the real GDP growth. Full sample Q1.2001 - Q2.2016, $T = 62$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	1.34	2.33	1.79	-0.50	1.22	0.90
WCE-B	1.40	2.27*	1.87	-0.50	1.23	0.98
WPE-D	1.15	1.78	1.74	-0.38	0.96	1.45
Fifth release						
WCE-DM	1.05	2.07	1.52	-0.47	1.26	1.08
WCE-B	1.04	1.95*	1.53	-0.48	1.29	1.19
WPE-D	0.86	1.53	1.39	-0.37	1.01	1.77
Current release						
WCE-DM	0.22	1.59	2.14	-0.83	1.10	0.96
WCE-B	0.25	1.62	2.35**	-0.90	1.19	1.13
WPE-D	0.21	1.28	2.30*	-0.72	0.97	1.96*

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	2.67	2.57	2.74	0.01	1.61	1.37
WCE-B	2.38**	2.18*	2.46**	0.01	1.80	1.67
WPE-D	2.02*	1.75	2.21*	0.01	1.53	1.74
Fifth release						
WCE-DM	2.49	2.56	2.65	0.04	1.68	1.37
WCE-B	2.18*	2.18*	2.38**	0.05	1.89	1.68
WPE-D	1.82	1.75	2.13*	0.04	1.61	1.73
Current release						
WCE-DM	1.94	2.55	2.90	-0.13	1.96	1.02
WCE-B	1.77	2.21*	2.63**	-0.15	2.30**	1.39
WPE-D	1.55	1.77	2.34*	-0.15	2.03*	1.48

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the real GDP growth rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the full sample Q1.2001 - Q2.2016 ($T = 62$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). ■ and ■ indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.12: DM test for the real GDP growth. Sub-sample Q1.2001 - Q3.2008, $T = 31$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	-0.30	0.19	0.74	-2.10	-1.13	2.45
WCE-B	-0.33	0.22	0.98	-2.35*	-1.22	1.91
WPE-D	-0.32	0.22	0.88	-2.40*	-1.18	1.81
Fifth release						
WCE-DM	-0.79	-0.13	0.40	-1.95	-0.83	2.42
WCE-B	-0.89	-0.15	0.55	-2.24*	-0.96	1.98
WPE-D	-0.81	-0.14	0.55	-2.29*	-0.94	1.92
Current release						
WCE-DM	-0.99	-0.06	0.93	-1.70	-0.34	1.76
WCE-B	-1.09	-0.07	1.34	-1.91	-0.49	1.89
WPE-D	-0.91	-0.06	1.19	-1.83	-0.36	1.86

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	0.58	1.27	2.08	-1.31	0.52	1.14
WCE-B	0.58	1.26	2.56**	-1.51	0.57	1.20
WPE-D	0.63	1.02	1.93	-1.46	0.51	0.95
Fifth release						
WCE-DM	0.20	1.29	2.02	-1.25	0.69	1.01
WCE-B	0.20	1.30	2.48**	-1.44	0.77	1.07
WPE-D	0.22	1.05	1.88	-1.36	0.70	0.84
Current release						
WCE-DM	-0.06	1.28	2.19	-1.23	1.35	0.88
WCE-B	-0.06	1.33	2.66**	-1.47	1.56	0.93
WPE-D	-0.06	1.08	2.06*	-1.41	1.39	0.74

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the real GDP growth rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the sub-sample Q1.2001 - Q3.2008 ($T = 31$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

Table 4.13: DM test for the real GDP growth. Sub-sample Q4.2008 - Q2.2016, $T = 31$

Quadratic Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	3.02	3.72	1.71	0.99	2.25	-0.08
WCE-B	2.89**	3.58**	1.67	1.04	2.49**	-0.09
WPE-D	2.61**	3.19**	1.63	0.94	1.96*	-0.08
Fifth release						
WCE-DM	3.03	3.31	1.61	0.95	2.19	0.17
WCE-B	2.86**	3.24**	1.56	1.00	2.40*	0.20
WPE-D	2.54**	2.87**	1.52	0.90	1.91	0.19
Current release						
WCE-DM	1.53	2.29	2.06	0.52	1.71	0.03
WCE-B	1.76	2.40*	2.12*	0.56	1.90	0.04
WPE-D	1.59	2.11*	1.95*	0.46	1.43	0.04

Ranked Probability Score

	1 year ahead			2 years ahead		
	Uniform	GRW	Naive	Uniform	GRW	Naive
First release						
WCE-DM	3.20	2.49	2.06	0.94	1.66	0.63
WCE-B	3.13**	2.33*	1.90	1.03	1.88	0.69
WPE-D	2.80**	2.07*	1.70	0.89	1.53	0.57
Fifth release						
WCE-DM	3.25	2.45	1.99	0.89	1.66	0.74
WCE-B	3.16**	2.30*	1.84	0.97	1.88	0.81
WPE-D	2.84**	2.05*	1.65	0.84	1.54	0.70
Current release						
WCE-DM	2.47	2.44	2.22	0.75	1.74	0.49
WCE-B	2.47**	2.30*	2.09*	0.84	1.99	0.56
WPE-D	2.21*	2.04*	1.86	0.69	1.60	0.49

Note: the Table reports DM test statistic values for one-year and two-year ahead ECB SPF density forecasts for the real GDP growth rate against the uniform, the Gaussian random walk and the naive benchmark forecasts for the sub-sample Q4.2008 - Q2.2016 ($T = 31$). A negative sign implies that benchmarks perform better than the ECB SPF. Long run variances are estimated using WCE with rectangular kernel and Diebold and Mariano (1995) bandwidth (WCE-DM), WCE with the Bartlett kernel and bandwidth $M = \lfloor T^{1/2} \rfloor$ (WCE-B) and WPE with Daniell kernel and bandwidth $m = \lfloor T^{1/3} \rfloor$ (WPE-D). and indicate, respectively, two-sided significance at the 5% and 10% level using standard asymptotics. Rejections using fixed-smoothing asymptotics are reported using ** and * to indicate, respectively, two-sided significance at the 5% and 10% level.

4.6 Conclusions

I perform ECB SPF aggregate density forecasts evaluation using the Diebold and Mariano test for equal forecast accuracy with fixed-smoothing asymptotics. As I consider forecasts reported as histogram with some bins associated with zero probability, I adopt loss functions that are well behaved in this case: the Quadratic Probability Score and the Ranked Probability Score. A Monte Carlo exercise shows that improvement in size performance and good power of the test still holds with these loss functions. Competing forecasts are obtained from a uniform distribution, a normal distribution and using the previous survey round forecast. In general, ECB SPF forecasts show a good predictive ability especially for unemployment rate and real GDP growth forecasts. For HICP inflation, there is no evidence from the sample that the ECB SPF forecasts are better than simple benchmark forecasts. For all variables there is a general improvement in predictive ability after 2008 supporting the evidence of a change in the forecasting practice after the financial crisis. The RPS function seems to be less sensitive to revision of the target variable than the QPS and highlights near-misses of professional forecasters in the second sub-sample.

Chapter 5

Conclusions

This thesis investigates the quality of the European Central Bank Survey of Professional Forecasters (ECB SPF), first by testing the rationality under a general loss function of point aggregate forecasts using the test proposed by Patton and Timmermann (2007) and then by testing the accuracy of point forecasts with the Diebold and Mariano (1995) test with asymmetric loss functions. The last Chapter deepens the study of forecast accuracy employing the Diebold and Mariano (1995) test and appropriate loss functions for density forecasts collected in histogram format as the ECB SPF aggregate density forecasts are.

All the tests mentioned above suffer from severe size distortion with standard asymptotics in small samples, such as the one available for ECB SPF. However, existing literature suggests using fixed-smoothing asymptotics to obtain correctly sized tests. Building on this, I use fixed-smoothing asymptotics in all my empirical exercises to obtain correctly sized tests after confirming with a series of Monte Carlo exercises that existing literature results hold in my setting.

Chapter 2 shows that ECB SPF point forecasts over the period 2002.Q1 - 2016.Q4 are rational under a general loss function but also under a symmetric MSE loss function. This result holds also for inflation forecasts despite the fact that the ECB has an asymmetric inflation target. A plausible explanation for this result could be the lack of trust professional forecasters have in the ECB to maintain inflation below the 2% target.

Chapter 3 examines the ECB SFP aggregate point forecasts with the Diebold and Mariano (1995) test for equal forecast accuracy paired with asymmetric loss functions. The need to consider asymmetric loss functions in forecast evaluation comes from the fact that a growing literature highlights that central banks use an asymmetric loss function and have asymmetric inflation targets. On the professional forecasters side, Weber (1994) suggests they minimise an asymmetric loss function because they care about their reputation and can adjust their future forecasts learning from past forecast errors. The assumption of different costs associated with negative and positive forecast errors is plausible in real life for example because firms may face different costs to upscale rather than downscale production. SPF forecasts are compared to forecasts from simple benchmarks that should be easy to beat by expert forecasters. Results show that ECB SPF outperform benchmark models in some cases especially for short horizons and seem more reliable than the Philadelphia Federal Reserve Survey of Professional Forecasters.

Chapter 4 tests the accuracy of ECB SPF density forecasts adapting the Diebold and Mariano (1995) test for density forecasts reported as histograms. This involves using appropriate loss functions, such as the Quadratic Probability Score and Ranked Probability Score. To verify size and power performances of the test with these loss functions and fixed-m asymptotics, I perform a Monte Carlo exercise before the actual empirical exercise. Density forecasts are compared to three simple benchmarks and ECB SPF forecasts show a good predictive ability especially for unemployment rate and real GDP growth forecasts. For HICP inflation, there is no evidence from the sample that the ECB SPF forecasts are better than simple benchmark forecasts.

Considering all empirical results, ECB SPF forecasts are rational and accurate. These forecasts could be reliably used to inform policy and business decisions. However, these qualities should be monitored continuously as the panel of forecasters changes and the quality of forecasts could change over time especially during uncertain periods such as the COVID pandemic. On a similar note, the assessment of individual forecasts should also be considered as methods explored in this thesis can also be applied on each participant's forecasts. Sequential Diebold and Mariano tests could also be employed to produce an aggregate forecast only from a selection of particularly well-performing forecasters that beats the consensus forecasts obtained from the simple average of respondents' forecasts.

An appealing development of the work presented in this thesis would be to apply fixed-smoothing asymptotics to other tests based on the Diebold and Mariano framework, such as the one suggested by Breitung and Knüppel (2018) to test the maximum forecast horizon at which forecasts become uninformative. This test could be applied to ECB SPF forecasts using fixed-m asymptotics to reduce small sample size distortion. In particular, the analysis of five-years forecasts collected in the survey using this test could clarify the usefulness and justify the collection of long-term forecasts. However, given the current results that forecasts tend to become less accurate as the forecast horizon increases, I would expect long-term forecasts not to be informative.

Another interesting extension would be to adapt techniques for the evaluation of density forecasts presented in Chapter 4 to accommodate an asymmetric loss functions along the same lines of Chapter 3 about point forecasts.

References

Abel, J., Rich, R., Song, J., and Tracy, J. (2016). The measurement and behavior of uncertainty: evidence from the ECB Survey of Professional Forecasters. *Journal of Applied Econometrics*, *31*(3), 533–550.

Aguiar, A. and Martins, M. M. (2008). Testing for asymmetries in the preferences of the Euro-area monetary policymaker. *Applied Economics*, *40*(13), 1651–1667.

Ahn, Y. B. and Tsuchiya, Y. (2019). Asymmetric loss of macroeconomic forecasts in South Asia: Evidence from the SPF survey of India, Indonesia, and Singapore. *Emerging Markets Finance and Trade*, *0*(0), 1–19.

Aiolfi, M., Capistrán, C., and Timmermann, A. (2011). Forecast combinations. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting* chapter 11, (pp. 355–388). Oxford: Oxford University Press.

Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, *25*(2), 177–190.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817–858.

Artis, M. and Marcellino, M. (2001). Fiscal forecasting: The track record of the IMF, OECD and EC. *The Econometrics Journal*, *4*(1), 20–36.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis. Quarterly Review-Federal Reserve Bank of Minneapolis*, *25*(1), 2.

Auffhammer, M. (2007). The rationality of EIA forecasts under symmetric and asymmetric loss. *Resource and Energy Economics*, *29*(2), 102–121.

Balcilar, M., Gupta, R., Majumdar, A., and Miller, S. M. (2015). Was the recent downturn in US real GDP predictable? *Applied Economics*, *47*(28), 2985–3007.

Bao, Y., Lee, T.-H., and Saltoglu, B. (2004). A test for density forecast comparison with applications to risk management. Technical report.

Bao, Y., Lee, T.-H., and Saltoglu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, *26*(3), 203–225.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, *37*(1/2), 1–16.

Bergbauer, S., Hernborg, N., Jamet, J.-F., Persson, E., and Schölermann, H. (2020). Citizens' attitudes towards the ECB, the euro and Economic and Monetary Union. Technical report, European Central Bank.

Bhansali, R. J. (2002). Multi-step forecasting. In *A Companion to Economic Forecasting* chapter 9, (pp. 206–221). Wiley.

Boero, G., Smith, J., and Wallis, K. F. (2008). Evaluating a three-dimensional panel of point forecasts: the Bank of England Survey of External Forecasters. *International Journal of Forecasting*, *24*(3), 354–367.

Boero, G., Smith, J., and Wallis, K. F. (2011). Scoring rules and survey density forecasts. *International Journal of Forecasting*, *27*(2), 379–393.

Bonham, C. and Cohen, R. (1995). Testing the rationality of price forecasts: Comment. *The American Economic Review*, *85*(1), 284–289.

Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., and Rautanen, T. (2007). The ECB Survey of Professional Forecasters (SPF)-a review after eight years' experience. *ECB Occasional Paper*, (50).

Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A., and Rautanen, T. (2011). An evaluation of the growth and unemployment forecasts in the ECB survey of professional forecasters. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, *2010*(2), 1–28.

Breitung, J. and Knüppel, M. (2018). How far can we forecast? Statistical tests of the predictive content. Technical report, Deutsche Bundesbank.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3.

Brillinger, D. R. (1975). *Time series: data analysis and theory*. SIAM.

Capistrán, C. (2008). Bias in Federal Reserve inflation forecasts: Is the Federal Reserve irrational or just cautious? *Journal of Monetary Economics*, *55*(8), 1415–1427.

Capistrán, C. and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit and Banking*, *41*(2-3), 365–396.

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

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business and Economic Statistics*, *11*(2), 121–135.

Chong, Y. Y. and Hendry, D. F. (1986). Econometric evaluation of linear macroeconomic models. *The Review of Economic Studies*, *53*(4), 671–690.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*(4), 841–862.

Christoffersen, P. F. and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, *13*(6), 808–817.

Clark, T. E. (1999). Finite-sample properties of tests for equal forecast accuracy. *Journal of Forecasting*, *18*(7), 489–504.

Clark, T. E. and McCracken, M. W. (2009). Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics*, *27*(4), 441–454.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, *5*(4), 559–583.

Clements, M. P. (2004). Evaluating the Bank of England density forecasts of inflation. *The Economic Journal*, *114*(498), 844–866.

Clements, M. P. (2018). Are macroeconomic density forecasts informative? *International Journal of Forecasting*, *34*(2), 181–198.

Clements, M. P. (2019). *Macroeconomic Survey Expectations*. Springer.

Clements, M. P. and Smith, J. (2000). Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting*, *19*(4), 255–276.

Coroneo, L. and Iacone, F. (2020). Comparing predictive accuracy in small samples using fixed-smoothing asymptotics. *Journal of Applied Econometrics*, *35*(4), 391–409.

Coroneo, L., Iacone, F., and Profumo, F. (2019). A real-time density forecast evaluation of the ECB Survey of Professional Forecasters. Technical report, Working paper, University of York.

Corradi, V. and Swanson, N. R. (2006). Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics*, *133*(2), 779–806.

Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, *105*(1), 111–130.

D'Agostino, A., Giannone, D., and Surico, P. (2006). (Un) predictability and macroeconomic stability. *ECB Working Paper Series*, (605).

Daniell, P. J. (1946). Discussion on symposium on autocorrelation in time series. *Supplement to Journal of the Royal Statistical Society*, *8*, 88–90.

Dawid, A. P. (1984). Present position and potential developments: Some personal views statistical theory the prequential approach. *Journal of the Royal Statistical Society: Series A (General)*, *147*(2), 278–290.

de Vincent-Humphreys, R., Dimitrova, I., Falck, E., and Henkel, L. (2019). Twenty years of the ECB Survey of Professional Forecasters. *Economic Bulletin Articles*, *1*.

Demetrescu, M., Hanck, C., and Kruse, R. (2018). Robust inference under time-varying volatility: A real-time evaluation of professional forecasters. Technical report.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

Diebold, F. X. and Lopez, J. A. (1996). Forecast evaluation and combination. *Handbook of Statistics*, *14*, 241–268.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–262.

Diebold, F. X. and Rudebusch, G. D. (1991). Forecasting output with the composite leading index: A real-time analysis. *Journal of the American Statistical Association*, *86*(415), 603–610.

Diebold, F. X., Tay, A. S., and Wallis, K. F. (1997). Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters. NBER Working Papers 6228, National Bureau of Economic Research, Inc.

Diks, C., Panchenko, V., and Van Dijk, D. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, *163*(2), 215–230.

Ehrbeck, T. and Waldmann, R. (1996). Why are professional forecasters biased? Agency versus behavioral explanations. *The Quarterly Journal of Economics*, *111*(1), 21–40.

El-Shagi, M. (2019). Rationality tests in the presence of instabilities in finite samples. *Economic Modelling*, *79*, 242–246.

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

Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, *6*(1), 122–157.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, *8*(6), 985–987.

Fair, R. C. (1980). Estimating the expected predictive accuracy of econometric models. *International Economic Review*, *21*(2), 355–378.

Fair, R. C. and Shiller, R. J. (1989). The informational content of ex ante forecasts. *The Review of Economics and Statistics*, *71*(2), 325–331.

Fair, R. C. and Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *The American Economic Review*, *80*(3), 375–389.

Fama, E. F. (1975). Short-term interest rates as predictors of inflation. *The American Economic Review*, *65*(3), 269–282.

Fourçans, A. and Vranceanu, R. (2007). The ECB monetary policy: choices and challenges. *Journal of policy Modeling*, *29*(2), 181–194.

Garcia, J. A. (2003). An introduction to the ECB's Survey of Professional Forecasters. *ECB Occasional Paper*, (8).

Garcia, J. A. and Manzanares, A. (2007). Reporting biases and survey results: evidence from european professional forecasters. *ECB Working Paper Series*, (836).

Garratt, A., Lee, K., Pesaran, M. H., and Shin, Y. (2003). Forecast uncertainties in macroeconomic modeling: An application to the uk economy. *Journal of the American Statistical Association*, *98*(464), 829–838.

Genre, V., Kenny, G., Meyler, A., and Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, *29*(1), 108–121.

Geraats, P. M. (2008). ECB credibility and transparency. *Economic Papers*, *330*, 1–33.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Giannone, D., Henry, J., Lalik, M., and Modugno, M. (2012). An area-wide real-time database for the Euro area. *Review of Economics and Statistics*, *94*(4), 1000–1013.

Giordani, P. and Söderlind, P. (2006). Is there evidence of pessimism and doubt in subjective distributions? implications for the equity premium puzzle. *Journal of Economic Dynamics and Control*, *30*(6), 1027–1043.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business and Economic Statistics*, *29*(3), 411–422.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *14*(1), 107–114.

Granger, C. W. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review*, *1*(2), 161–173.

Granger, C. W. J. and Newbold, P. (1986). *Forecasting economic time series*. Academic Press.

Grishchenko, O., Mouabbi, S., and Renne, J.-P. (2019). Measuring inflation anchoring and uncertainty: A U.S. and Euro Area comparison. *Journal of Money, Credit and Banking*, *51*(5), 1053–1096.

Gros, D. and Roth, F. (2010). The financial crisis and citizen trust in the European Central Bank. Technical Report 334, Centre for European Policy Studies, Brussels.

Grothe, M. and Meyler, A. (2015). Inflation forecasts: Are market-based and survey-based measures informative? Technical report, European Central Bank.

Hannan, E. J. (1970). *Multiple time series*, volume 38. John Wiley and Sons.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, *13*(2), 281–291.

Harvey, D. I., Leybourne, S. J., and Whitehouse, E. J. (2017). Forecast evaluation tests and negative long-run variance estimates in small samples. *International Journal of Forecasting*, *33*(4), 833–847.

Harvey, D. S., Leybourne, S. J., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics*, *16*(2), 254–259.

Howarth, D. and Loedel, P. (2003). *The European Central Bank: The New European Leviathan?* Springer.

Hualde, J. and Iacone, F. (2015a). Autocorrelation robust inference using the Daniell kernel with fixed bandwidth. Technical Report 15/14, The University of York.

Hualde, J. and Iacone, F. (2015b). Small-b and fixed-b asymptotics for weighted covariance estimation in fractional cointegration. *Journal of Time Series Analysis*, *36*(4), 528–540.

Hualde, J. and Iacone, F. (2017). Fixed bandwidth asymptotics for the studentized mean of fractionally integrated processes. *Economics Letters*, *150*, 39–43.

Jansson, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica*, *72*(3), 937–946.

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

Kenny, G., Kostka, T., and Masera, F. (2014). How informative are the subjective density forecasts of macroeconomists? *Journal of Forecasting*, *33*(3), 163–185.

Kiefer, N. M. and Vogelsang, T. J. (2002a). Heteroskedasticity–autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, *70*(5), 2093–2095.

Kiefer, N. M. and Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, *18*(6), 1350–1366.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, *21*(6), 1130–1164.

Koopmans, L. H. (1995). *The spectral analysis of time series*. Elsevier.

Krüger, F. (2017). Survey-based forecast distributions for euro area growth and inflation: ensembles versus histograms. *Empirical Economics*, *53*(1), 235–246.

Laster, D., Bennett, P., and Geoum, I. S. (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics*, *114*(1), 293–318.

Lazarus, E., Lewis, D. J., and Stock, J. H. (2019). The size-power tradeoff in har inference. Technical report.

Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). HAR inference: Recommendations for practice. *Journal of Business and Economic Statistics*, *36*(4), 541–559.

Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models. *Journal of Forecasting*, *20*(2), 87–109.

Łyziak, T. and Paloviita, M. (2017). Anchoring of inflation expectations in the euro area: Recent evidence based on survey data. *European Journal of Political Economy*, *46*, 52 – 73.

Mankiw, N. G. and Shapiro, M. D. (1986). News or noise? An analysis of GNP revisions. Technical report, National Bureau of Economic Research Cambridge, Massachusetts, USA.

Manski, C. F. (2011). Interpreting and combining heterogeneous survey forecasts. In M. P. Clements and D. F. Hendry (Eds.), *Oxford handbook of economic forecasting* chapter 16, (pp. 457–472). Oxford: Oxford University Press.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, *135*(1-2), 499–526.

Meyler, A. (2020). Forecast performance in the ECB SPF: ability or chance? Technical report, European Central Bank.

Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). NBER.

Mitchell, J. and Hall, S. G. (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR fan charts of inflation. *Oxford Bulletin of Economics and Statistics*, *67*, 995–1033.

Mitchell, J. and Wallis, K. F. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, *26*(6), 1023–1040.

Müller, U. K. (2014). HAC corrections for strongly autocorrelated time series. *Journal of Business and Economic Statistics*, *32*(3), 311–322.

Müller, U. K. and Watson, M. W. (2016). Measuring uncertainty about long-run predictions. *Review of Economic Studies*, *83*(4), 1711–1740.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*(3), 315–335.

Naghi, A. A. (2015). A forecast rationality test that allows for loss function asymmetries. Technical report, Working paper, University of Warwick.

Neave, H. R. (1970). An improved formula for the asymptotic variance of spectrum estimates. *Ann. Math. Statist.*, *41*(1), 70–77.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, *55*(4), 819–847.

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

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, *61*(4), 631–653.

Nobay, R. A. and Peel, D. A. (2003). Optimal discretionary monetary policy in a model of asymmetric central bank preferences. *The Economic Journal*, *113*(489), 657–665.

Patton, A. J. (2019). Comparing possibly misspecified forecasts. *Journal of Business and Economic Statistics*, *0*(0), 1–23.

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

Patton, A. J. and Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business and Economic Statistics*, *30*(1), 1–17.

Pesaran, M. H. and Weale, M. (2006). Survey expectations. In *Handbook of Economic Forecasting*, volume 1 (pp. 715–776). Elsevier.

Pierdzioch, C., Reid, M. B., and Gupta, R. (2016). Forecasting the South African inflation rate: On asymmetric loss and forecast rationality. *Economic Systems*, *40*(1), 82–92.

Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics*. Number 04; QA280, P7. New York: Academic Press.

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

Rosemblatt, M. (1952). Remarks on multivariate transformations. *Annals of Mathematical Statistics*, *23*, 470–472.

Rossi, B. and Sekhposyan, T. (2013). Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics*, *177*(2), 199–212.

Rossi, B. and Sekhposyan, T. (2014). Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, *30*(3), 662–682.

Rossi, B. and Sekhposyan, T. (2016). Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts. *Journal of Applied Econometrics*, *31*(3), 507–532.

Ruge-Murcia, F. J. (2000). Uncovering financial markets' beliefs about inflation targets. *Journal of Applied Econometrics*, *15*(5), 483–512.

Ruge-Murcia, F. J. (2003). Does the Barro–Gordon model explain the behavior of US inflation? A reexamination of the empirical evidence. *Journal of monetary economics*, *50*(6), 1375–1390.

Schorfheide, F. (2005). VAR forecasting under misspecification. *Journal of Econometrics*, *128*(1), 99–136.

Sims, C. A. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity*, *2002*(2), 1–40.

Stark, T. (2010). Realistic evaluation of real-time forecasts in the survey of professional forecasters. *Federal Reserve Bank of Philadelphia Research Rap, Special Report*, *1*.

Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *The Econometrics Journal*, *16*(1), 1–26.

Sun, Y. (2014). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, *178*, 659 – 677.

Sun, Y., Phillips, P. C., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, *76*(1), 175–194.

Tay, A. S. and Wallis, K. F. (2000). Density forecasting: a survey. *Journal of Forecasting*, *19*(4), 235–254.

Theil, H. (1958). *Economic forecasts and policy*. North-Holland.

Ulu, Y. (2015). Rationality of inflation–output forecasts of MMS survey: international evidence. *Applied Economics*, *47*(12), 1187–1198.

Varian, H. R. (1975). A bayesian approach to real estate assessment. *Studies in Bayesian Econometric and Statistics in honor of Leonard J. Savage*, *81*(394), 195–208.

Wang, Y. and Lee, T.-H. (2014). Asymmetric loss in the Greenbook and the Survey of Professional Forecasters. *International Journal of Forecasting*, *30*(2), 235–245.

Weber, C. S. and Forschner, B. (2014). ECB: Independence at risk? *Intereconomics*, *49*(1), 45–50.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, *115*(2), 228.

Weiss, A. A. (1996). Estimating time series models using the relevant cost function. *Journal of Applied Econometrics*, *11*(5), 539–560.

Wilson, E. B. (1934). The periodogram of American business activity. *The Quarterly Journal of Economics*, *48*(3), 375–417.

Wooldridge, J. M. and White, H. (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric theory*, *4*(2), 210–230.

Wu, W. B. and Shao, X. (2006). Invariance principles for fractionally integrated nonlinear processes. In *Recent developments in nonparametric inference and probability* (pp. 20–30). Institute of Mathematical Statistics.

Zarnowitz, V. (1985). Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics*, *3*(4), 293–311.

Zhang, X. and Shao, X. (2013). Fixed-smoothing asymptotics for time series. *The Annals of Statistics*, *41*(3), 1329–1349.