An Investigation into the Uncertainty Revision Process of Professional Forecasters

Michael P. Clements
Robert Rich
Joseph Tracy*

August 13, 2024

Abstract: Following Manzan (2021), this paper examines how professional forecasters revise their fixed-event uncertainty (variance) forecasts and tests the Bayesian learning prediction that variance forecasts should decrease as the horizon shortens. We show that Manzan's (2021) use of first moment "efficiency" tests are not applicable to study revisions of variance forecasts. Instead, we employ monotonicity tests developed by Patton and Timmermann (2012) in the first application of these tests to second moments of survey expectations. We find strong evidence that the variance forecasts are consistent with the Bayesian learning prediction of declining monotonicity.

Keywords: Variance forecasts, survey expectations, Bayesian learning

JEL Codes: C53, E17, E37.

---

* Clements: ICMA Centre, Henley Business School, University of Reading; Rich: Federal Reserve Bank of Cleveland; Tracy: Daniels School of Business, Purdue University and American Enterprise Institute. The authors thank Sebastiano Manzan and Andrew Patton for helpful comments. We are also grateful to Andrew Patton for the use of computer programs available on his website. Alexander Cline and Kristoph Naggert provided excellent research assistance. The views expressed in the paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Cleveland or the Federal Reserve System. Address for correspondence: Research Department, P.O. Box 6387, Federal Reserve Bank of Cleveland, Cleveland, OH 44101-1387. Phone: 216-579-2928. Email: Robert.rich@clev.frb.org

I. Introduction

The study of expectations and the process underlying their formation remains a topic of considerable interest and importance. The recently published *Handbook of Economic Expectations* (2023) speaks to the extensive literature exploring this topic across many dimensions. One area of focus has been the statistical properties of survey expectations. Some studies have tested if reported expectations are unbiased predictors while others have compared their accuracy to that of model-based forecasts. There has also been significant interest in determining whether survey expectations possess certain optimality properties. Patton and Timmermann (2012, p.6) describe the common properties, as well as some of the less common properties, by extending the popular weak and strong forecast efficiency tests of Mincer and Zarnowitz (1969). Finally, Patton and Timmerman (2012) and others have looked at how expectations are updated over time and if the revisions fully incorporate available information.

While most analyses of survey expectations have examined *point* forecasts because of their greater availability, studies have also examined *density* forecasts. An attractive feature of density forecasts is that they provide a basis to derive estimates of uncertainty. Density-based estimates of uncertainty are often regarded as both theoretically and empirically superior to alternative approaches such as using disagreement (see, e.g., Zarnowitz and Lambros (1987), Manski (2011), and Rich and Tracy (2021b)) or using model-based estimates (Giordani and Söderlind (2003)).[1]

In contrast to survey-based point forecasts, there have been few attempts to conduct formal tests for the optimality and efficiency of the forecast revision process for survey-based uncertainty measures. Manzan (2021) is commendable as a notable exception, both testing whether forecasts of precision (uncertainty) conform to the Bayesian Learning Model (BLM) of expectations formation, and modelling some of the determinants of the revisions. Bayesian learning implies that the precision of the posterior forecast is equal to the precision of the prior forecast plus the precision of the signal. If we interpret the first of two forecasts of the same target variable as the prior forecast and its precision as the prior precision, then the subsequent forecast which is based on a larger information set becomes the posterior and should consequently have a higher precision. Because precision is just the inverted variance, Bayesian learning implies that the shorter-horizon variance forecast should be no larger than the prior (i.e., the longer-horizon) variance forecast.

[1] We do not review the arguments here as to the importance of macroeconomic uncertainty for the macroeconomy or policy: see e.g., Bloom (2009) and Haddow et al. (2013).

This paper revisits Manzan's (2021) analysis and approach to testing the BLM prediction that variance forecasts should decline monotonically as the survey date draws nearer to the target date. Manzan uses the univariate Optimal Revision Regression (ORR) approach of Patton and Timmermann (2012) which extends the earlier Mincer and Zarnowitz (1969) test of weak efficiency. Applying this testing procedure to survey-based forecasts of precision, Manzan concludes that professional forecasters tend to revise their forecasts in a manner that violates the prediction of Bayesian learning. However, we show that Mincer-Zarnowitz (MZ) type regressions should not be used to investigate the monotonicity of variance (or precision) forecasts, and, furthermore, that a rejection using an MZ test does not indicate "non-optimality" or inefficiency. There appear to be two sources of confusion. The first is the belief that the MZ tests (and the ORR extension) for the efficient use of information for mean (first-moment) forecasts carry over directly to variance (second-moment) forecasts. The second is the belief that tests of efficiency can be used to test for (non-) monotonicity.

While the MZ and ORR frameworks should not be used to test the BLM prediction, the non-parametric tests of monotonicity of Patton and Timmermann (2012) provide a valid alternative approach. Accordingly, we apply the monotonicity tests of mean squared forecasts and mean squared forecast revisions to the variance forecasts. An attractive feature of this approach is that it does not require observed realizations of the target variable. To the best of our knowledge, our study provides the first application of the Patton-Timmermann monotonicity tests to second moments of survey expectations.

Following Manzan, we examine the U.S. Survey of Professional Forecasters (US-SPF) and the European Central Bank Survey of Professional Forecasters (ECB-SPF) and focus on fixed-event density forecasts of growth, inflation, and unemployment. We construct respondent-level variance forecasts using three approaches: the probability-based measure of Manzan (2021), the interquartile range (IQR), and the entropy-based measure proposed by Krüger and Pavlova (2024). In contrast to Manzan (2021), we find strong evidence that the variance forecasts are consistent with the implication of forecast rationality of declining monotonicity. Specifically, formal tests at the individual level overwhelmingly fail to reject the predictions of a decreasing pattern in the variance forecasts and their revisions as the forecast horizon declines.

While our sample period, measures of second-moment forecasts, and panel composition are not identical to those in Manzan (2021), we also find that the application of the ORR tests to our data generate results that broadly match those of Manzan. Hence, we are reasonably confident that the different conclusions regarding forecast optimality reflect the alternative testing procedures. It is important to note, though, that Manzan (2021) also presents several interesting results concerning the determinants of second-moment forecasts and their revisions, and our critique of his tests of rationality do not apply to these findings.

This paper is organized as follows. Section II reviews the regression frameworks of Mincer-Zarnowitz (1969) and the multiple-horizon extension of Patton-Timmermann (2012). We show why these testing procedures are problematic for evaluating the properties of variance forecasts and then discuss how the monotonicity tests in Patton and Timmermann (2012) – mean squared forecasts and mean squared forecast revisions – can be used to evaluate the rationality of the variance forecasts. Section III describes the density forecasts from the US-SPF and the ECB-SPF, the construction of the variance forecasts, and the participation criterion used to select respondents. Section IV presents the empirical findings. Section V provides concluding remarks.

II. Evaluating the Moments of Fixed-event Forecasts

The Mincer and Zarnowitz (1969) regression approach (and the related Optimal Revision Regression extension of Patton and Timmermann (2012)) has been widely used to evaluate the properties of point forecasts. Consider fixed-event forecasts that involve a sequence of forecasts of the same event or target variable that are made at different dates. The MZ and ORR approaches test the condition that the forecast revision - the difference between two forecasts of the same target - is not systematically related to information available at the time the earlier forecast was made (see, e.g., Nordhaus (1987)).

The development of surveys fielding fixed-event density forecasts has allowed studies investigating the updating process of forecasters to expand their scope beyond first moments to include second moments of subjective distribution forecasts. Because forecast rationality is expected to hold across all features of predictive behavior, it might be reasonable to assume that conventional tests for forecast efficiency of first moments can also be used to analyze second moments of density forecasts. For example, Manzan (2021) adapts the ORR framework to evaluate variance forecasts. However, as we explain in more detail below, there are reasons why the approach is not directly applicable in this situation.

Intuitively, there is a fundamental difference in the properties of revisions to first and second moments of expectations. Mean forecast revisions can be of either sign, with the ORR framework testing whether movements in the series are unpredictable as the forecast horizon shortens. The design of the ORR framework, however, is inconsistent with variance forecast revisions that are predicted to decrease (or not to increase) as the forecast horizon shortens. Because the implications of forecast rationality for the predictive content of available information are not the same across first and second moments of the data, the evaluation of variance forecast revisions requires an alternative approach. Accordingly, we propose using the Patton and Timmermann (2012) tests of monotonicity.

2.1 The Mincer-Zarnowitz Testing Approach

The simplest Mincer-Zarnowitz test provides a useful starting point for the analysis. The MZ test considers whether forecasters efficiently make use of the information available at the time of the forecast, which must of course include the forecast itself. Hence, the null hypothesis of forecast efficiency is $H_0: \delta_0 = 0 \cap \delta_1 = 1$ in:[2]

$$y_t = \delta_0 + \delta_1 y_{t|t-h_1} + \mu_t \tag{1}$$

where y_t is the target variable, $y_{t|t-h_1}$ is the forecast of y_t made h_1 periods previously, and the observations range over *t* for a given h_1. Nordhaus (1987) refers to this as a weak-efficiency test. Strong-efficiency tests are more stringent and test the orthogonality of the forecast error and variables known at $t-h_1$. This test can be achieved by adding $g'_{t-h_1}\theta$ to equation (1), where g'_{t-h_1} is a vector of variables known at time $t-h_1$. The null hypothesis of rationality is then $H_0: \delta_0 = 0 \cap \delta_1 = 1 \cap \theta = 0$.

Patton and Timmermann (2012) note that a short-horizon forecast can be written as:

$$y_{t|t-h_1} \equiv y_{t|t-h_H} + d_{t|h_1,h_2} + \ldots + d_{t|h_{H-1},h_H} \equiv y_{t|t-h_H} + \sum_{i=1}^{H-1} d_{t|h_i,h_{i+1}} \tag{2}$$

where $h_1 < h_2 < \ldots < h_H$, h_H is the longest-horizon forecast of y_t, and $d_{t|h_i,h_{i+1}} = y_{t|t-h_i} - y_{t|t-h_{i+1}}$. That is, the short-horizon forecast is the longest-horizon fixed-event forecast plus the series of revisions between adjacent horizon forecasts from the longest horizon up to the shortest horizon.

[2] See Clements (2022) for further discussion.

The Optimal Revision Regression (ORR) test substitutes equation (2) into equation (1) and allows for a free coefficient on each component of $y_{t|t-h_1}$. We can then estimate the following regression:

$$y_t = \delta_0 + \delta_H y_{t|t-h_H} + \sum_{i=1}^{H-1} \delta_i d_{t|h_i,h_{i+1}} + \mu_t, \tag{3}$$

and test the null hypothesis $H_0 : \delta_0 = 0 \cap \delta_1 = \delta_2 = \ldots = \delta_H = 1$. Under the null hypothesis, equation (3) becomes $y_t = y_{t|t-h_1} + \mu_t$ which predicts that the error for the short-horizon forecast $y_{t|t-h_1}$ is uncorrelated with all prior forecasts of y_t, where the prior forecasts were formed using smaller information sets.

As explained by Patton and Timmermann (2012, p.6), the ORR specification tests whether 'agents optimally and consistently revise their forecasts at the interim points between the longest and shortest-forecast horizons and also that the long-run forecast is unbiased. Hence, it generalizes the conventional single horizon MZ regression . . .'. An attractive feature of the ORR testing procedure is that it can be applied to many of the fixed-event forecasts found in various surveys (see, e.g., Nordhaus (1987) and Clements (1995, 1997)), including those of professional forecasters such as the US-SPF and the ECB-SPF.

The ORR regression framework allows one of the forecasts (the shortest) to be substituted for the actual outcome in equation (3). This solves the difficulty of finding a proxy for y_t when needed, as is the case for the variance forecasts.[3] The interpretation of the regression with y_t replaced by $y_{t|t-h_1}$ changes in that the null hypothesis under equation (3) could be satisfied by forecasts that are unrelated to the (unobserved) actuals. This was famously described by Nordhaus (1987, p. 673): 'A baboon could generate a series of weakly efficient forecasts by simply wiring himself to a random-number generator, but such a series of forecasts would be completely useless.'

2.2 ORR Tests and Survey-based Variance Forecasts

Manzan (2021) investigates the revision process of professional forecasters but differs from previous studies by examining forecasts of precision derived from fixed-event density forecasts. As

[3] This may also be desirable when the choice of data vintage to use as realized values for point forecasts is unclear.

part of his analysis, Manzan focuses on the prediction of Bayesian learning that the precision of density forecasts should not decline as the forecast horizon shortens.

For his evaluation of forecast precisions (or their inverse, variances), Manzan (2021, p.17) uses equation (3) with y_t replaced by the shortest-horizon forecast because the true values are not observed. Let $\psi_{q,t}$ denote the precision for quarter *q* of survey year *t*, and $\psi_{q-1,t}$ denote the precision in the previous quarter. Manzan estimates individual and pooled versions of:

$$\psi_{4,t} = \alpha + \beta_1 \psi_{1,t} + \sum_{q=2}^{3} \beta_q \phi_{q,t} + \varepsilon_t \tag{4}$$

where $\phi_{q,t}$ denote the revisions, $\phi_{q,t} = \psi_{q,t} - \psi_{q-1,t}$. He notes that the BLM predicts that $\psi_{q,t} \geq \psi_{q-1,t}$, (for *q* = 2, 3, 4), but reports tests of the null hypothesis $H_0: \alpha = 0 \cap \beta_1 = \ldots = \beta_3 = 1$, directly corresponding to the Patton and Timmermann (2012) null hypothesis of rationality for point forecasts. However, under this null hypothesis equation (4) is $\psi_{4,t} = \psi_{3,t} + \varepsilon_t$. Because ε_t is required to be zero mean, this is inconsistent with the property $\psi_{4,t} \geq \psi_{3.t}$ except when it holds as an equality. Therefore, it would initially seem more sensible to relax the restriction that $\alpha = 0$ and instead simply to consider the null hypothesis $H_0: \beta_1 = \ldots = \beta_3 = 1$.

However, there are more fundamental problems with the application of the ORR framework in the current setting. Consider equation (1), where we replace the actual value by a short-horizon forecast and, for the remainder of the analysis, elect to work with variances defined as $V_{q,t} = \psi_{q,t}^{-1}$:[4]

$$V_{4,t} = \delta_0 + \delta_1 V_{3,t} + \mu_t \tag{5}$$

For a simple data generating process capable of producing time-varying variance forecasts, such as an AR(1) process with an ARCH(1) error, we show in Appendix 1 that $\delta_1 < 1$ for optimal variance forecasts. Consequently, the MZ approach adopted by Manzan (2021) cannot be applied to test for the optimality of variance forecasts because the data will reject $\delta_1 = 1$ in equation (5) even if the

[4] Our decision to work with variances rather than precisions is discussed further in Section III.

variance forecasts are optimal and decline monotonically as the forecast horizon shortens.[5] While one might consider developing an approach based on testing the null hypothesis $H_0 : \delta_1 \leq 1$ against the alternative hypothesis $H_1 : \delta_1 > 1$, this appears to be closely related to the tests of monotonicity of Patton and Timmermann (2012) and we prefer that approach.

2.3 Tests of Monotonicity

According to the Bayesian learning model a requirement for optimality is that the variance forecasts should decline monotonically as the forecast horizon shortens. Hence, in the context of the US-SPF and ECB-SPF forecasts of inflation and GDP growth in the current year relative to the previous year, the variances should steadily decline between the first and fourth quarters of the year. The same pattern should also hold for the ECB-SPF forecasts of the survey-year unemployment rate. As we now show, we can use the Patton and Timmermann (2012) tests of mean squared forecasts (MSF - their section 2.4) and of mean squared forecast revisions (MSFR - section 2.3) to evaluate monotonicity of the variance forecasts.

For the MSF test, Patton and Timmermann (2012, equation (7)) show that the expected squared optimal forecasts of a first moment should be non-decreasing as the horizon shortens. However, we need to modify the MSF testing procedure in our analysis for two reasons. First, the variance forecasts are non-negative and therefore do not require the application of a squaring operator.[6] Second, in contrast to the expected squared optimal forecasts of a first moment, the variance forecasts should be non-increasing as the horizon shortens. Consequently, the null hypothesis is $H_0 : \Delta^f \leq 0$ versus $H_1 : \Delta^f \nleq 0$, where Δ^f is the vector of population parameters, which in our case consists of the quarter-on-quarter forecast variances:

$$\Delta^f = \left[E\left(V_{4,t}\right) - E\left(V_{3,t}\right), E\left(V_{3,t}\right) - E\left(V_{2,t}\right), E\left(V_{2,t}\right) - E\left(V_{1,t}\right) \right]' \tag{6}$$

The test is based on the sample analog of Δ^f, $\hat{\Delta}^f$, which is given by:

---

[5] Our findings for the MZ and, by extension, ORR frameworks would argue that the strong-efficiency tests discussed in Section 2.1 are also problematic because, as we have seen, the additional restrictions are not valid when the data pertain to variance forecasts. Specifically, evidence of statistical significance for the parameter vector θ in the extended version of specification (1) would not indicate the inefficient use of information, but rather would identify determinants of the variance revision process.

[6] Nevertheless, we did conduct the MSF test applying the squaring operator as a robustness check and the results lead to nearly identical conclusions.

$$\hat{\Delta}^f = \left[\frac{1}{T}\sum_{t=1}^{T}\left(V_{4,t} - V_{3,t}\right), \frac{1}{T}\sum_{t=1}^{T}\left(V_{3,t} - V_{2,t}\right), \frac{1}{T}\sum_{t=1}^{T}\left(V_{2,t} - V_{1,t}\right)\right]' \tag{7}$$

For the MSFR test, let Δ^d be the difference in mean squared forecast revisions defined as:

$$\Delta^d = \left[E\left(V_{4,t} - V_{3,t}\right)^2 - E\left(V_{4,t} - V_{2,t}\right)^2, E\left(V_{4,t} - V_{2,t}\right)^2 - E\left(V_{4,t} - V_{1,t}\right)^2\right]' \tag{8}$$

Under forecast rationality, the expected squared revision between the Q2 variance forecast and the Q4 variance forecast should be no smaller than that between the Q3 variance forecast and the Q4 variance forecast.[7] In addition, the expected squared revision between the Q1 variance forecast and the Q4 variance forecast should be no smaller than that between the Q2 variance forecast and the Q4 variance forecast. Consequently, the null hypothesis is $H_0 : \Delta^d \leq 0$ versus $H_1 : \Delta^d \not\leq 0$. Following from equation (7), we can use the sample analog of $\Delta^d, \hat{\Delta}^d,$ as the basis for the MSFR test:

$$\hat{\Delta}^d = \left[\frac{1}{T}\sum_{t=1}^{T}\left[\left(V_{4,t} - V_{3,t}\right)^2 - \left(V_{4,t} - V_{2,t}\right)^2\right], \frac{1}{T}\sum_{t=1}^{T}\left[\left(V_{4,t} - V_{2,t}\right)^2 - E\left(V_{4,t} - V_{1,t}\right)^2\right]\right]' \tag{9}$$

As discussed by Patton and Timmermann (2012, pp. 7 and 8), an advantage of testing *fixed-event* forecasts is that covariance stationarity is not necessary for the bounds inequalities to hold for optimal forecasts. By holding the date of the target variable fixed and varying the horizon (as opposed to varying the target with a fixed horizon), the usual requirement of covariance stationarity is put to one side, thereby widening the applicability of the approach.

We view the formulations in (7) and (9) as valid tests of rationality (based on the Bayesian learning model), whereas the tests in Section 2.2 are not valid. We have not studied the small-sample properties of the tests of monotonicity in our context. Patton and Timmermann (2012, section 5) provide some Monte Carlo simulation evidence, but it is unclear how relevant this is to survey-based variance forecasts. Both the ORR tests and monotonicity tests are likely to have low power given the relatively small sample sizes available at the individual level, but, beyond that, it is more likely that the differences we observe between the outcomes of the ORR tests

[7] This follows directly from Manzan (2021, equation 4).

and monotonicity tests are due to the invalidity of the former rather than the small-sample properties of the tests.

III. Data and Variance Forecast Measures

The analysis examines data from the US-SPF and ECB-SPF. Both surveys collect point and density forecasts, where density forecasts are reported using histograms based on a set of intervals provided in the survey instrument. The surveys also include a mixture of fixed- and rolling-event (or fixed-horizon) forecasts, but we restrict our attention to the fixed-event density forecasts and focus on the revisions to the variance forecasts.

The US-SPF survey began in the fourth quarter of 1968 and was initially conducted by the American Statistical Association and the National Bureau of Economic Research. In 1990 the survey was taken over by the Federal Reserve Bank of Philadelphia.[8] The average number of panelists has been around 40 since the Philadelphia Federal Reserve has overseen the survey. The respondents are anonymous but have identifiers allowing the forecasts of a given respondent to be followed across surveys. The survey provides forecasts of U.S. macro variables and is conducted quarterly after the release of the advance report of the national income and product accounts from the Bureau of Economic Analysis (BEA). Since 1990:Q2, the deadlines for responses have been around the middle of each quarter—before the BEA's second report. The Philadelphia Federal Reserve posts quarterly reports on their website summarizing the survey results and the underlying data.

The ECB-SPF began in January 1999 and provides quarterly surveys of forecasts for the euro area. The survey's principal aim is to elicit expectations about inflation, real GDP growth, and unemployment. The survey is typically fielded in January, April, July, and October, with approximately 55 responses received on average per quarter. While there is a listing of the respondents' names and institutions, individual responses are anonymous and tracked across surveys by an assigned identification number. Like the US-SPF, there is a quarterly report summarizing the results of the survey and the data are available on the ECB's website.[9]

The US-SPF data are density forecasts for inflation and output from 1992:Q1 to 2019:Q4. For inflation, the survey asks for density forecasts of the annual rate of GDP deflator inflation in the year of the survey relative to the previous year, and of the next year relative to the current year. That is, of the percentage rate of change in the annual GDP deflator between years. The survey structure is the

[8] See Zarnowitz (1969) on the original objectives of the survey, and Croushore (1993) and Croushore and Stark (2019) on the revival of the survey by the Philadelphia Federal Reserve.
[9] For additional details about the ECB-SPF, see Garcia (2003) and Bowles et al. (2007).

same for output which reports density forecasts of real GDP growth. The ECB-SPF data are density forecasts for inflation, output, and unemployment from 1999:Q1-2019:Q4. The density forecasts of inflation and growth for the current calendar year and the next calendar year parallel those in the US-SPF, with inflation measured by the Harmonised Index of Consumer Prices. For unemployment, the current calendar year and the next calendar year refer to the average of monthly unemployment rates in the current year and the subsequent year of the survey, respectively.

The empirical analysis focuses on forecasts of the survey-year target variables, so that for a given year we have four fixed-event forecasts with approximate horizons of 4, 3, 2, and 1 quarters that are made, respectively, in the 1st, 2nd, 3rd and 4th quarters of the year. While the extension of the US-SPF and ECB-SPF target variables into the next calendar year can provide four additional fixed-event forecasts with approximate horizons of 8, 7, 6, and 5 quarters, we do not use these forecasts. This is because our participation criterion, discussed later in this section, requires respondents to provide a complete sequence of quarterly fixed-event forecasts of a target variable and including these additional forecasts would greatly reduce the sample size for the analysis.

As noted above, surveys that elicit density forecasts typically ask each respondent *i* to assign a sequence of probabilities, $p_i(k)$, to a set of $k = 1, \dots, K$ pre-specified outcome intervals, where $k=1$ and $k=K$ denote, respectively, the lowest and highest intervals.[10] To facilitate discussion at the moment, we assume that the number of intervals do not change over time, and therefore, we omit a time subscript for this term here.

There are various ways of calculating variances from histograms, as explained by Manzan (2021) and Clements et al. (2022). One method involves a distribution-free approach where the uncertainty implicit in a histogram is quantified using only the probabilities assigned to the different range of possible outcomes. An alternative method involves fitting a parametric distribution to the histograms (e.g., Engelberg et al. (2009)). Manzan (2021) adopts the first method and uses the following probabilities-based approach to derive the variance forecast measure for respondent *i*:

$$\sigma_{i,q,t}^2 = \sum_{k=1}^{K} \left[p_{i,q,t}(k) \left(\bar{x}(k) - \mu_{i,q,t} \right)^2 \right] - w^2/12 \qquad (10)$$

[10] We verify that probabilities sum to unity for each density forecast used in the analysis.

where $\mu_{i,q,t} = \sum_{k=1}^{K} p_{i,q,t}(k)\bar{x}(k)$ is the density mean, $\bar{x}(k)$ is the mid-point of the *k*th interval, w is the interval width, and $w^2/12$ represents the Sheppard's correction (Stuart and Ord (2010).[11]

For robustness, we also report results for two other variances constructed using a distribution-free approach: the interquartile range (IQR) and the entropy function of the Ranked Probability Score (ERPS) suggested by Krüger and Pavlova (2024). Following Abel et al. (2016) and Glas (2020), the individual IQR variance forecast measure for respondent *i* is given by:

$$V_{i,q,t}^{IQR} = p_{i,q,t}^{0.75} - p_{i,q,t}^{0.25} \tag{11}$$

where $p_{i,q,t}^{0.75}$ and $p_{i,q,t}^{0.25}$ denote the estimated 75th and 25th percentiles, respectively, of respondent *i*'s density forecast. As described in Krüger and Pavlova (2024), the individual ERPS variance forecast measure for respondent *i* is given by:

$$V_{i,q,t}^{ERPS} = \sum_{k=1}^{K} P_{i,q,t}(k)(1 - P_{i,q,t}(k)) \tag{12}$$

where $P_{i,q,t}(k) = \sum_{j=1}^{k} p_{i,q,t}(j)$ is the cumulative probability of the first *k* intervals.

Additional auxiliary assumptions are required for the calculation of the variance forecasts using the three methods we have described. One issue is that the density forecasts from both surveys contain open intervals on each end of the histogram that, if the respondent assigned a probability to either, must be closed to calculate the (mean and) variance. We follow the usual—although ad hoc—practice and assign twice the width of the interior closed intervals to the open intervals. The degree to which this assumption will impact any estimate depends on the amount of probability assigned to each open interval.[12]

A second issue that is relevant for the IQR measure concerns the location of the probability mass within a specified closed interval. As previously noted, Manzan (2021) assumes that the

[11] Equation (10) occasionally yielded a negative variance forecast. We set the value to zero in those cases.
[12] The IQR is typically more robust than a standard deviation/variance estimate to situations when respondents place probability in open intervals. Specifically, the IQR is unaffected unless the respondent places more than a 25 percent probability in an open interval. Consequently, the inclusion of the IQR-based forecast variance allows us to assess the extent to which the open interval boundary is a problem and the sensitivity of the results to our approach to address it.

probability mass is located at the mid-point of each interval which parallels the approach in some studies [Boero et al. (2008a), Rich and Tracy (2010), Kenny et al. (2015), and Poncela and Senra (2017)]. However, we will follow other studies [Zarnowitz and Lambros (1987), Abel et al. (2016), and Rich and Tracy (2021a)] and assume that the probability is distributed uniformly within each interval. Under this assumption, the lower and upper quartiles in equation (11) are calculated by linear interpolation.

A third issue is that the interval widths for the density forecasts occasionally change over time. A drawback of the ERPS is that it is not invariant to changes in in the interval width across surveys.[13] The issue of interval widths is not a concern for the ECB-SPF density forecasts because they remained constant at 0.5 percentage point for our sample period. However, this issue is relevant for the GDP-deflator inflation density forecasts for the US-SPF. Specifically, the interval width was one percentage point from 1992:Q1-2013:Q4 and then was reduced to half a percentage point in 2014:Q1. When we calculate ERPS for inflation for the post-2013 surveys, we first re-define the histogram to have interval widths of a percentage point by summing the probabilities for the adjacent half-percentage point histograms.

A final issue concerns the treatment of the 2009:Q1 ECB-SPF GDP growth density forecast. There is a well-known "piling-up" of probability mass in the lower open interval of this survey for GDP growth. This outcome resulted from the lower range of closed intervals in the survey instrument providing insufficient coverage for the overly pessimistic growth forecasts associated at the time with the global financial crisis. While the ECB-SPF subsequently expanded the lower range of closed intervals to address this concern, we view the 2009:Q1 survey as especially problematic and elected to drop it from the analysis which, because of our participation criterion, has the consequence of excluding all GDP growth density forecasts conducted in 2009.[14]

The empirical analysis also requires us to address the issue of the participation criterion. Manzan (2021) only includes forecasters that provide at least 30 consecutive predictions which results in 32 forecasters for the US-SPF and 35 forecasters for the ECB-SPF. Because the forecast revision process is central to our study, we only include data that allow us to view a complete path of

[13] For example, suppose the interval [1,2] has $p = 0.45$, and the interval [2,3] has $p = 0.55$. The resulting ERPS is 0.2475. If instead the interval widths were doubled, then [1,3] would have $p = 1$, and ERPS = 0.
[14] In contrast, Manzan (2021) proposes a method to solve the problem of probability assigned to open intervals that uses information from matched point forecasts. Consequently, Manzan includes these 2009 data in his analysis. See Manzan (2021) for details.

forecasts toward a target variable. Consequently, we restrict the sample to observations where respondents provide a complete sequence of four forecasts of the current target year.[15]

In deciding participation criterion, it is also important to recognize that composition effects for the surveys can impact the results. The US-SPF and ECB-SPF are unbalanced panels due both to the entry and exit of participants and to the fact that participants occasionally do not provide responses to all or part of the survey questionnaire. For robustness, we consider participants that provide a minimum of 7, 11, and 15 complete 4-quarter predictions. Unlike Manzan, we do not require the forecast sequence to be consecutive. However, the range of participants that we consider will span Manzan's set of forecasters. For the US-SPF, there are 6-7 participants under the most restrictive case (15 complete 4-quarter predictions) and 36-37 participants under the least restrictive case (7 complete 4-quarter predictions). For the ECB-SPF, there are 15-17 participants under the most restrictive case and 39-44 participants under the least restrictive case. By considering various participation criteria, we show that our results do not depend on a particular set of individuals and are a general feature of the survey forecasts.

Finally, Manzan (2021) considers precisions, whereas we analyze variances. As described above, our participation criterion is based on respondents providing a complete sequence of quarterly fixed-event forecasts of a target variable. While we view precisions and variances as an equivalent basis to investigate issues related to methodology and testing procedures, there are a few instances when the variance measures using Manzan (2021) and Krüger-Pavlova (2024) equal zero and generate a precision that is undefined. Because this outcome for precisions reduces the sample size, we choose to work with variances. In addition, we capture the essence of Manzan's (2021) results for the ORR tests using variances instead of precisions, as reported in the following section. Further, unreported results show that the monotonicity test results are not affected by the choice of variances versus precisions. Consequently, we conclude that this choice is not responsible for the different findings of the two studies.

IV. Empirical Results

As part of an exploratory analysis, we will examine the data to gain some initial insights into the behavior of the variance forecasts. While there is a large combination of measures and participation criterion, we found many similarities in the observed patterns which allow us to narrow our discussion to a few representative charts. Drawing upon Manzan (2021), Figures 1-4

[15] For example, a sequence of four forecasts of (say) the annual inflation rate in 1999 that would comprise histogram forecasts made in 1991:Q1, Q2, Q3, and Q4.

plot the average of a respondent's variance forecasts across the same quarterly survey round for the US-SPF and the ECB-SPF. The lines depict the IQR or ERPS measures for each forecaster in our panel meeting the participation criterion of a minimum of 7 complete 4-quarter predictions for the relevant target variable.

Taken together, the plots generally show a pattern of declining variance forecasts toward the end of the year. In addition, there is notable heterogeneity across forecasters in the level of uncertainty at all horizons, a finding previously documented by Manzan (2021). The lines do not show extensive crossings which suggests that this heterogeneity in the variance forecasts displays persistence. That is, individuals who tend to report relatively higher (lower) variance forecasts in surveys conducted in quarter 1 also tend to report relatively higher (lower) variance forecasts in surveys conducted in quarter 4. This finding is consistent with other studies that have documented persistent heterogeneity in various features of forecast behavior, although our evidence appears to be somewhat weaker than that reported by Manzan (2021).[16]

Looking at the US-SPF and ECB-SPF separately, forecasters appear to be more confident about predicting inflation relative to output. Comparing the US-SPF and the ECB-SPF, the IQR-based variance forecasts indicate that forecasters are more confident about predicting output and inflation in the euro area relative to the US variables. This contrasts with the ERPS-based measures that suggest predictability is more comparable.

Figures 1-4 also reveal instances when the variance forecasts of an individual remain relatively constant or show an actual increase as the horizon declines. While these features might be interpreted as *prima facie* evidence of a violation of declining monotonicity, visual inspection is not sufficient to make an informed evaluation. Instead, formal testing procedures that account for the presence of statistical uncertainty (i.e., "sampling" uncertainty) are required to draw reliable inferences about this property of the data.

Tables 1a-1b present the results from applying the ORR testing procedure to the US-SPF and ECB-SPF, respectively, using the probabilities-based variance forecasts ($\sigma^2_{i,q,t}$) as well as the IQR ($V^{IQR}_{i,q,t}$) and ERPS ($V^{ERPS}_{i,q,t}$) measures. Each panel reports the total number of respondents who provided density forecasts for the target variables based on the participation criteria of a minimum of 7, 11, and 15 complete 4-quarter sequences. We provide a count of the

[16] See D'Amico and Orphanides (2008), Patton and Timmermann (2010), Bruine de Bruin et al. (2011), Boero et al. (2008, 2015), Rich and Tracy (2021a, 2021b) and Clements (2014, 2022).

number of rejections of the relevant null hypothesis at the 10 percent level of significance. As shown, we consider two versions of the null hypothesis that each incorporates the joint restriction that the slope coefficients equal unity but differ in terms of the treatment of the constant term. We use a Wald test to evaluate the null hypothesis, where the test statistic is distributed as $\chi^2(4)$ or $\chi^2(3)$ depending on whether the additional restriction $\alpha = 0$ is imposed or relaxed, respectively.

The results in Tables 1a-1b show little ambiguity in terms of the outcomes associated with tests of the null hypothesis. While there are fewer rejections in almost all cases when we relax the restriction on the constant term, the rejection rates in both cases remain high. When translated into percentages, the rejection rates for the full set of restrictions range from 70-100 percent and broadly match those reported in Manzan (2021, Table 5). The high incidence of rejections of the null hypothesis is also remarkably robust across survey series, the choice of variance forecasts, and participation criterion. Viewed through the lens of Manzan's interpretation of his testing procedure, the overall evidence strongly suggests that most forecasters in the US-SPF and ECB-SPF update their density forecasts in a way that is inconsistent with Bayesian learning.

A critical question at the center of our analysis is the extent to which (mis)application of first-moment efficiency tests to variance forecasts may alter the conclusions about the behavior of forecasters. To answer this question, Tables 2a-2b present the results from applying the Patton and Timmermann (2012) testing procedure to the US-SPF and ECB-SPF, respectively.[17] The variance forecast measures and the participation criterion are the same as those in Tables 1a-1b. For optimal forecasts, the mean squared forecasts (MSF) should be non-increasing as the horizon shortens (see equation (6)). For the mean squared forecast revisions (MSFR), they should be non-increasing in the 'length' of the revision - see equation (8).

In sharp contrast to the results from the application of the ORR testing procedure, the findings in Tables 2a-2b show little evidence pointing to a violation of the decreasing variance forecast property. Looking across both survey series there appear to be more rejections for the variance forecasts of inflation as compared to GDP growth. The incidence of rejections using the IQR-based variance forecasts doesn't stand out compared to the other measures, suggesting that the results do not appear to be sensitive to the presence of probability in the open intervals.

[17] The MSF and MSFR tests were conducted using the Matlab code made available by Andrew Patton at http://public.econ.duke.edu/~ap172/code.html.

In addition, the MSFR tests largely result in more rejections compared to the MSF tests, but the associated rejection rates are still quite low. In particular, the highest rejection rate is 17 percent which occurs in a few instances for the US-SPF. For the ECB-SPF, the highest rejection rate is 14 percent and only occurs once. Another notable feature of the ECB-SPF is that there are fewer rejections of the null hypothesis of monotonicity. For example, we fail to reject the null hypothesis in all cases under the requirement of a minimum of 15 complete 4-quarter sequences. Moreover, this outcome is not a consequence of a small panel size, with the range of 16-17 participants over twice the size of the analogous panel appearing in the US-SPF.

Taken together, the findings in Tables 2a and 2b offer strong support for the predicted monotonicity properties of the variance forecasts across the declining forecast horizons. While we expressed concerns about the applicability and interpretation of Manzan's (2021) testing procedure for monotonicity, we were uncertain about the consequences for analyzing this feature of forecast behavior. We view the evidence from the MSF and MSFR tests as not only validating these concerns, but also underscoring their significance as the conclusions emerging from our study are essentially the opposite to those in Manzan (2021).

V. Conclusion

While there is an extensive literature focusing on the forecast revision process and the evaluation of its properties, studies have almost exclusively restricted their attention to first moments of the data. Motivated by the increased availability and interest in density forecasts, Manzan (2021) extends this line of research by considering second moments of survey expectations. As part of the analysis, he examines the fixed-event density forecasts of professional forecasters from the U.S. and the euro area and purportedly tests the prediction of Bayesian learning that precision increases as the forecast horizon shortens. Applying the ORR testing procedure of Patton and Timmermann (2012) to the survey data, Manzan finds the professional forecasters display a very high incidence of non-Bayesian behavior.

Our paper makes two contributions to tests of optimal forecast revision and the application to fixed-event density forecasts. First, we show that the popular Mincer and Zarnowitz (1969) approach to testing for forecast efficiency and the ORR extensions of Patton and Timmermann (2012) should not be applied to variance forecasts. The properties of revisions to variance forecasts and point forecasts are fundamentally different and do not lend themselves to being evaluated using similar testing procedures. Second, we show that the monotonicity tests of Patton and Timmermann (2012) can be slightly modified to test for the optimality of the variance forecasts. While the Patton-

Timmermann monotonicity tests have been previously used to evaluate the properties of point forecast series, our analysis represents the first application to second moments of survey expectations.

For the empirical analysis, we examine samples of professional forecasters in the U.S. and the euro area who report fixed event density forecasts. For robustness, we use three different approaches to construct variance forecasts and consider alternative participation criterion. The main result of the paper is that there is little evidence against the optimality of the variance forecasts, where this optimality finding is consistent with the prediction of Bayesian learning. In our context, this prediction is that the variance forecasts decline with the accumulation of information as the forecast origin nears its target. While the monotonicity tests are supportive of the optimality of the variance forecasts, we document that the use of first moment efficiency tests leads to a widespread erroneous rejection of their optimality.

This paper has principally focused on the appropriate way of testing revisions to forecast variances and has argued for the use of monotonicity tests for this purpose. We regard determining and modelling the key drivers of density forecast revisions as a separate topic, albeit an interesting one, for further research. Manzan (2021) has made some headway on this with his finding that large data surprises and the number of bins used by forecasters have predictive content for the quarterly revisions to the precision of their density forecasts. Drawing upon the work of Adrian et al. (2019), it might be interesting to investigate if financial conditions play a role in the updating of variance forecasts. Relatedly, the behavior and responsiveness of variance forecasts may differ across quantiles or phases of the business cycle. We leave these extensions for future work.

Figure 1: US-SPF IQR Measures

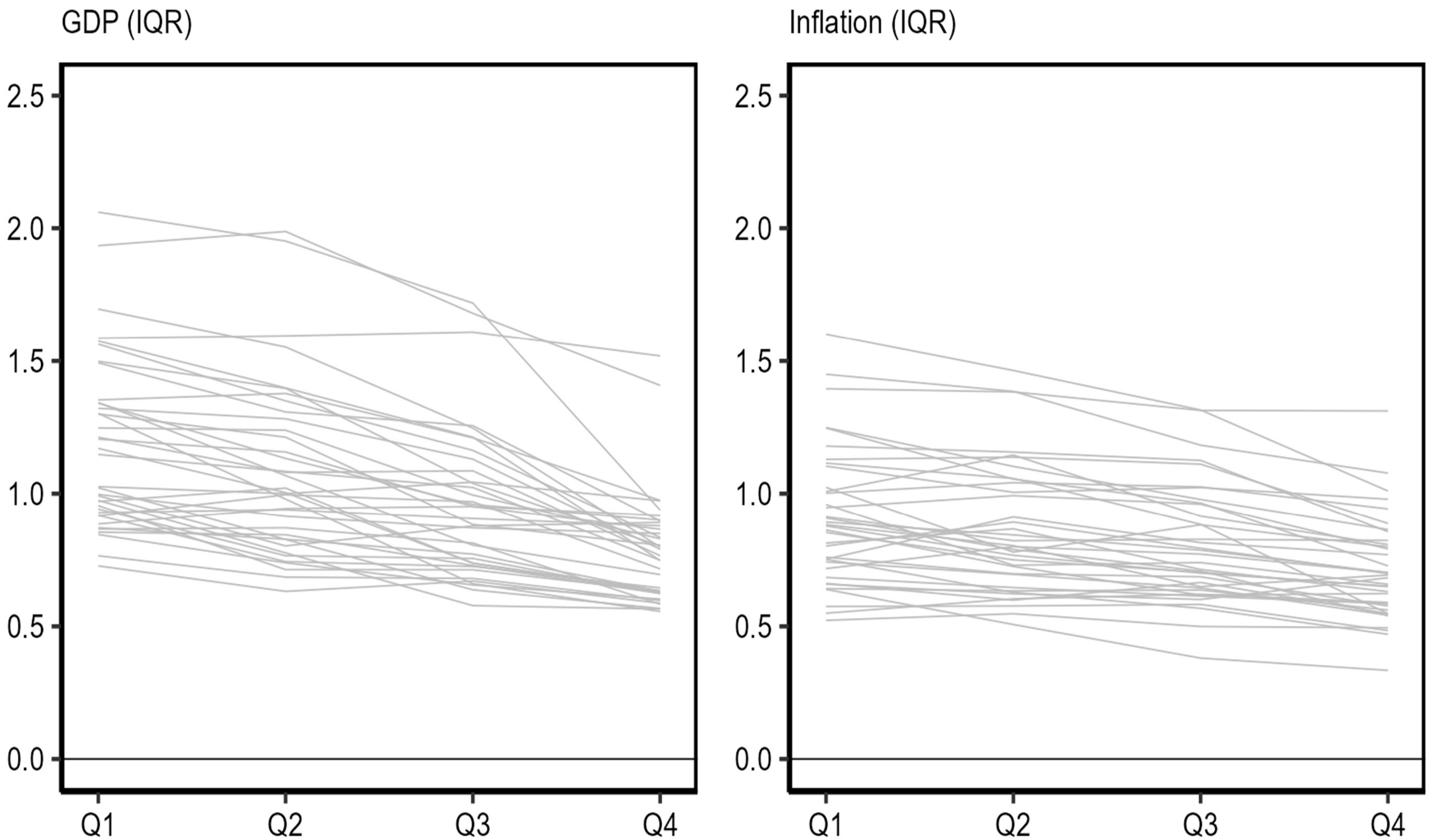

Figure 2: US-SPF ERPS Measures

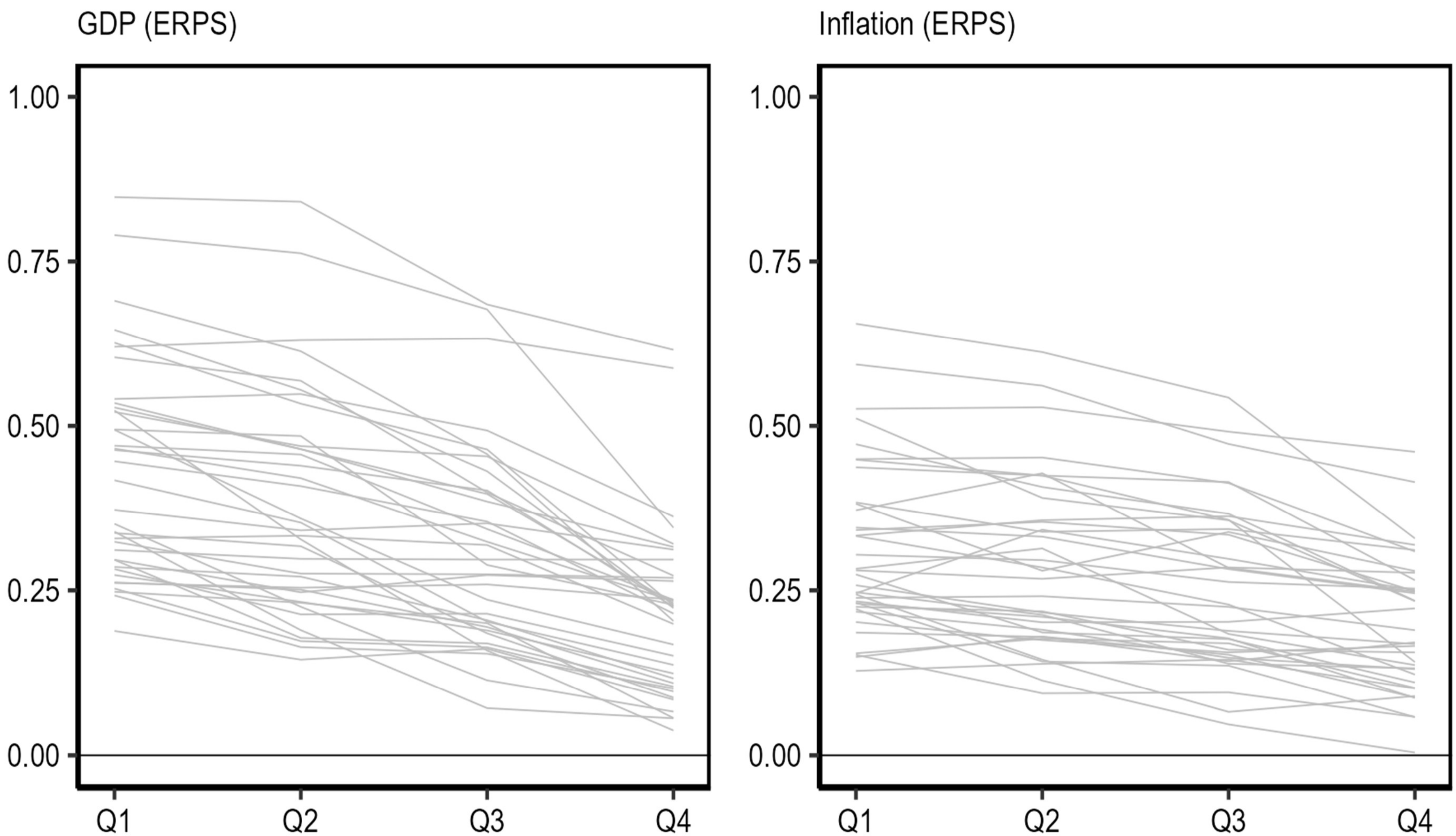

Figure 3: ECB-SPF IQR Measures

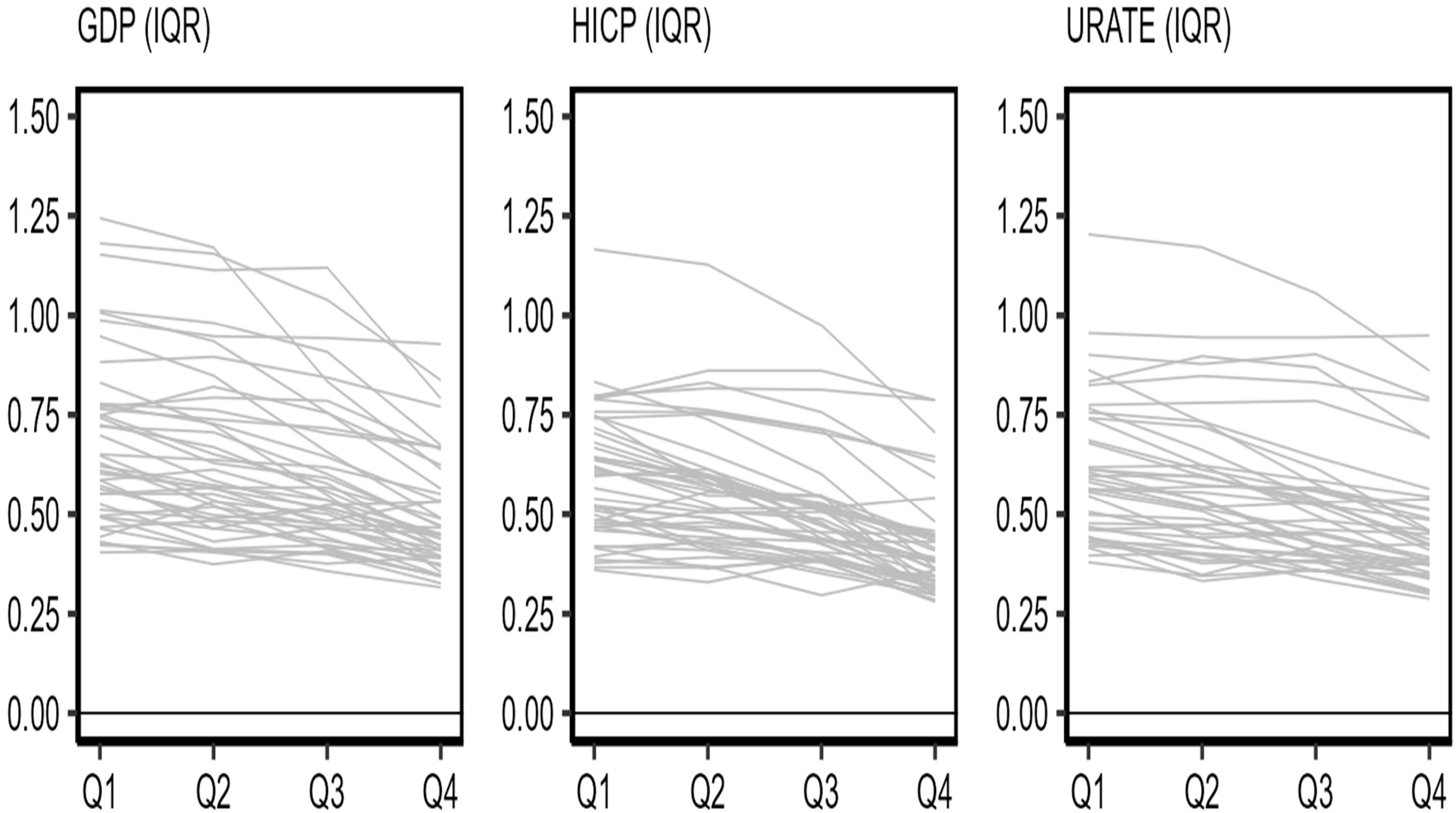

Figure 4: ECB-SPF ERPS Measures

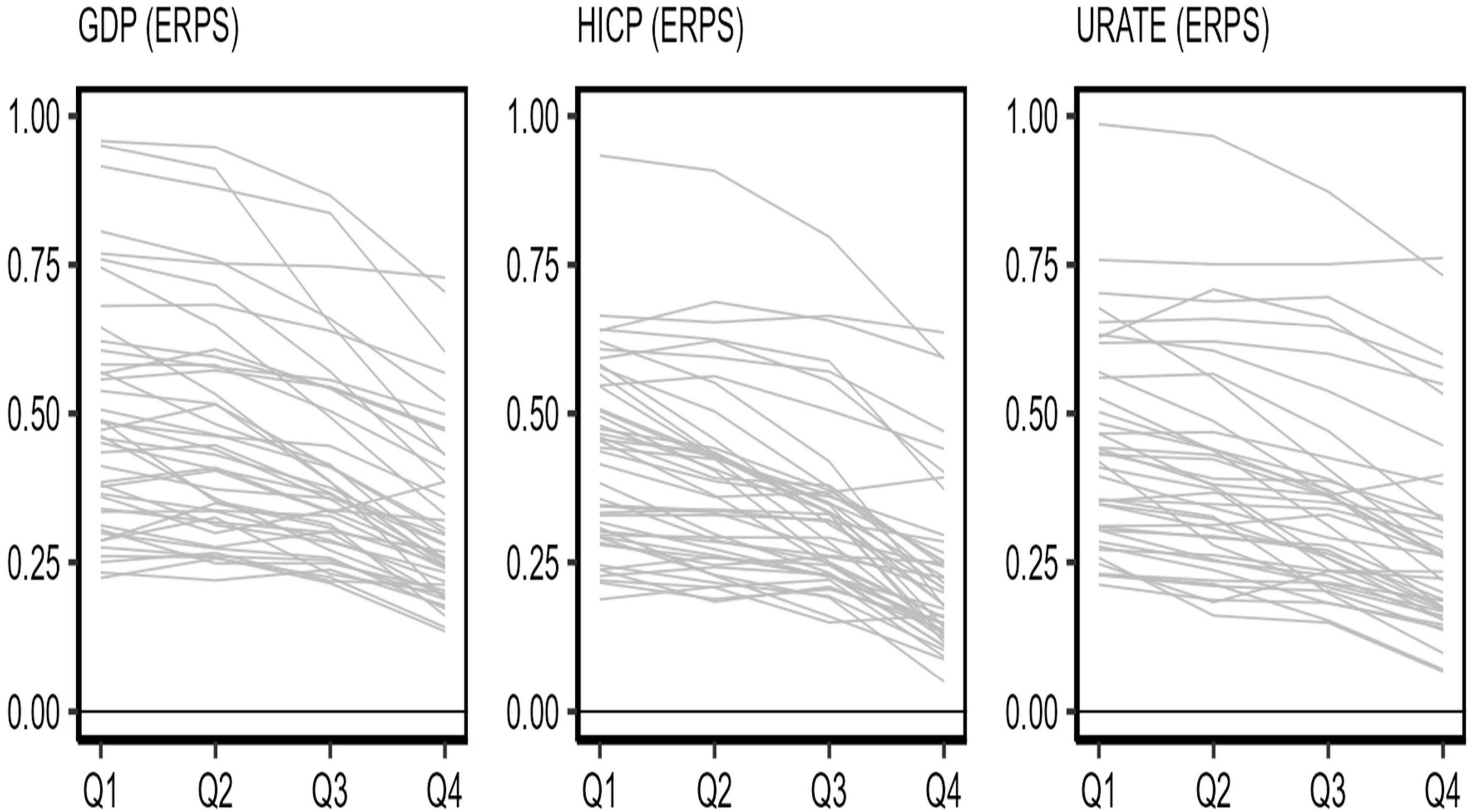

Table 1a – ORR Tests of US-SPF Variance Forecasts

$$V_{i,4,t} = \alpha_i + \beta_{i,1} V_{i,1,t} + \sum_{q=2}^{3} \beta_{i,q} (V_{i,q,t} - V_{i,q-1,t}) + \varepsilon_{i,4,t}$$

15 Complete 4-quarter Sequences											
GDP Growth						Inflation					
N=7						N=6					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
7	7	6	5	7	6	5	5	6	5	5	5

11 Complete 4-quarter Sequences											
GDP Growth						Inflation					
N=18						N=18					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
16	14	16	14	16	13	14	14	16	14	14	12

7 Complete 4-quarter Sequences											
GDP Growth						Inflation					
N=37						N=36					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
31	28	28	24	28	23	29	28	33	27	31	27

Values for $\chi^2(4)$ report rejections of $H_0 : \alpha = 0 \cap \beta_1 = \beta_2 = \beta_3 = 1$ at the 10% significance level.

Values for $\chi^2(3)$ report rejections of $H_0 : \beta_1 = \beta_2 = \beta_3 = 1$ at the 10% significance level.

Table 1b – ORR Tests of ECB-SPF Variance Forecasts

$$V_{i,4,t} = \alpha_i + \beta_{i,1} V_{i,1,t} + \sum_{q=2}^{3} \beta_{i,q} (V_{i,q,t} - V_{i,q-1,t}) + \varepsilon_{i,4,t}$$

15 Complete 4-quarter Sequences

GDP Growth						Inflation						Unemployment Rate					
N=15						N=17						N=16					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
12	9	13	11	13	9	14	13	15	14	17	13	15	14	15	11	14	8

11 Complete 4-quarter Sequences

GDP Growth						Inflation						Unemployment Rate					
N=24						N=27						N=28					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
20	16	21	18	20	14	20	20	25	22	26	20	23	22	25	20	25	16

7 Complete 4-quarter Sequences

GDP Growth						Inflation						Unemployment Rate					
N=42						N=45						N=39					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
37	31	39	32	38	27	37	37	41	35	43	35	33	32	33	28	34	24

Values for $\chi^2(4)$ report rejections of $H_0: \alpha = 0 \cap \beta_1 = \beta_2 = \beta_3 = 1$ at the 10% significance level.

Values for $\chi^2(3)$ report rejections of $H_0: \beta_1 = \beta_2 = \beta_3 = 1$ at the 10% significance level.

Table 2a – Patton-Timmermann Tests for Monotonicity of US-SPF Variance Forecasts

$$\text{MSFR: } \Delta_i^d = \left[E(V_{i,4,t} - V_{i,3,t})^2 - E(V_{i,4,t} - V_{i,2,t})^2, E(V_{i,4,t} - V_{i,2,t})^2 - E(V_{i,4,t} - V_{i,1,t})^2 \right]'$$

$$\text{MSF: } \Delta_i^f = \left[E\left(V_{i,4,t}\right) - E\left(V_{i,3,t}\right), E\left(V_{i,3,t}\right) - E\left(V_{i,2,t}\right), E\left(V_{i,2,t}\right) - E\left(V_{i,1,t}\right) \right]'$$

15 Complete 4-quarter Sequences											
GDP Growth						Inflation					
N=7						N=6					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
0	0	0	0	0	0	0	1	1	1	1	0

11 Complete 4-quarter Sequences											
GDP Growth						Inflation					
N=18						N=18					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
0	0	0	1	0	0	0	2	2	3	1	1

7 Complete 4-quarter Sequences											
GDP Growth						Inflation					
N=37						N=36					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
0	0	1	1	1	0	2	3	2	6	2	2

Values report rejections of $H_0 : \Delta^d \leq 0$ and $H_0 : \Delta^f \leq 0$ at the 10% significance level.

Table 2b – Patton-Timmermann Tests for Monotonicity of ECB-SPF Variance Forecasts

$$\text{MSFR: } \Delta_i^d = \left[E(V_{i,4,t} - V_{i,3,t})^2 - E(V_{i,4,t} - V_{i,2,t})^2, E(V_{i,4,t} - V_{i,2,t})^2 - E(V_{i,4,t} - V_{i,1,t})^2 \right]'$$

$$\text{MSF: } \Delta_i^f = \left[E\left(V_{i,4,t}\right) - E\left(V_{i,3,t}\right), E\left(V_{i,3,t}\right) - E\left(V_{i,2,t}\right), E\left(V_{i,2,t}\right) - E\left(V_{i,1,t}\right) \right]'$$

15 Complete 4-quarter Sequences

GDP Growth						Inflation						Unemployment Rate					
N=16						N=17						N=16					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

11 Complete 4-quarter Sequences

GDP Growth						Inflation						Unemployment Rate					
N=27						N=27						N=28					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
0	1	0	0	0	1	0	1	0	2	0	1	0	2	0	1	0	4

7 Complete 4-quarter Sequences

GDP Growth						Inflation						Unemployment Rate					
N=42						N=44						N=39					
$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$		$\sigma^2_{i,q,t}$		$V^{IQR}_{i,q,t}$		$V^{ERPS}_{i,q,t}$	
MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
0	1	0	1	0	1	0	2	0	4	0	1	1	4	0	1	0	4

Values report rejections of $H_0 : \Delta^d \leq 0$ and $H_0 : \Delta^f \leq 0$ at the 10% significance level.

Appendix 1

We illustrate the problem with the MZ test using a simple model that captures the essential features of variance forecast. We assume an AR(1) process for a stationary random variable y_t :

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{A1}$$

and assume $\varepsilon_t = \sqrt{h_t} z_t$, $z_t \sim N(0,1)$, and $h_t = \omega + \alpha \varepsilon_{t-1}^2$. The 1-step-ahead variance forecast of y_t, made at time t -1, is:

$$V_{t,1} \equiv Var(y_t \mid y_{t-1}) = Var(\varepsilon_t \mid y_{t-1}) = \omega + \alpha \varepsilon_{t-1}^2. \tag{A2}$$

From equation (A1),

$$y_t = \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \tag{A3}$$

and so:

$$\begin{aligned} V_{t,2} &\equiv Var(y_t \mid y_{t-2}) = \phi^2 Var(\varepsilon_{t-1} \mid y_{t-2}) + Var(\varepsilon_t \mid y_{t-2}) \\ &= \phi^2(\omega + \alpha \varepsilon_{t-2}^2) + \alpha^2 \varepsilon_{t-2}^2 + \alpha\omega + \omega \\ &= (\alpha\phi^2 + \alpha^2)\varepsilon_{t-2}^2 + \phi^2\omega + \alpha\omega + \omega \end{aligned} \tag{A4}$$

where the second line comes from $Var(\varepsilon_t \mid y_{t-2}) = h_{t|t-2}$, and $h_{t|t-2} - \sigma^2 = \alpha(h_{t-1|t-2} - \sigma^2)$, so $h_{t|t-2} = \alpha^2 \varepsilon_{t-2}^2 + \alpha\omega + \omega$, using $h_{t-1|t-2} = \alpha \varepsilon_{t-2}^2 + \omega$, where $\sigma^2 = \omega(1-\alpha)^{-1}$.

The MZ test regresses the shorter-horizon variance forecast $V_{t,1}$ on a constant and $V_{t,2}$ and tests the null hypothesis that the slope coefficient is unity. However, it can be shown that the population value of the slope coefficient is less than unity. Specifically, the slope coefficient of the MZ regression is given by:

$$\beta = \frac{Cov(V_{t,1}, V_{t,2})}{Var(V_{t,2})} \tag{A5}$$

The numerator is:

$$Cov(V_{t,1}, V_{t,2}) = \alpha^2(\phi^2 + \alpha) \times Cov(\varepsilon_{t-1}^2, \varepsilon_{t-2}^2) = \alpha^3(\phi^2 + \alpha) \times Var(\varepsilon_{t-2}^2) \qquad \text{(A6)}$$

using $\varepsilon_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \upsilon_t, \upsilon_t = \varepsilon_t^2 - h_t,$ and $Cov(\varepsilon_{t-1}^2, \varepsilon_{t-2}^2) = \alpha Var(\varepsilon_{t-2}^2).$

The denominator is:

$$Var(V_{t,2}) = \alpha^2(\phi^2 + \alpha)^2 Var(\varepsilon_{t-2}^2), \qquad \text{(A7)}$$

which results in:

$$\beta = \frac{\alpha^3(\phi^2 + \alpha)}{\alpha^2(\phi^2 + \alpha)^2} = \frac{\alpha}{\phi^2 + \alpha} \qquad \text{(A8)}$$

Hence, $\beta < 1$, because $\alpha > 0$, and $\phi^2 > 0$, since $-1 < \phi < 1$. This shows that the test of the null that $\beta = 1$ will be rejected (abstracting from small sample issues) for optimal variance forecasts.

References

Abel, Joshua, Robert Rich, Joseph Song, and Joseph Tracy. "The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters." *Journal of Applied Econometrics* 31 (May 2016): 533-550.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. "Vulnerable Growth." *American Economic Review* 109 (April 2019): 1263-1289.

Bachman, Rüdiger, Giorgio Topa, and Wilbert van der Klaauw. *Handbook of Economic Expectations*. Amsterdam, Elsevier, 2023.

Bloom, Nicholas. "The Impact of Uncertainty Shocks." *Econometrica* 77 (May 2009): 623-685.

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis. "Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters." *Economic Journal* 118 (July 2008): 1107-1127.

------, Jeremy Smith, and Kenneth F. Wallis. "The Measurement and Characteristics of Professional Forecasters' Uncertainty." *Journal of Applied Econometrics* 30 (December 2015): 1029-1046.

Bowles, Carlos, Roberta Friz, Veronique Genre, Geoff Kenny, Aidan Meyler, and Tuomas Rautanen. "The ECB Survey of Professional Forecasters (SPF): A Review After Eight Years' Experience." ECB Occasional Paper No 59. European Central Bank, April, 2007.

Bruine de Bruin, Wandi, Charles F. Manski, Giorgio Topa, and Wilbert van der Klaauw. "Measuring Consumer Uncertainty About Future Inflation." *Journal of Applied Econometrics* 26 (May 2011): 454-478.

Clements, Michael P. "Rationality and the Role of Judgement in Macroeconomic Forecasting." *Economic Journal* 105 (March 1995): 410-420.

------. "Evaluating the Rationality of Fixed-Event Forecasts." *Journal of Forecasting* 16 (July 1997): 225-239.

------. "Forecast Uncertainty - Ex Ante and Ex Post: US Inflation and Output Growth." *Journal of Business and Economic Statistics* 32 (April 2014): 206-216.

------. "Forecaster Efficiency and Disagreement: Evidence Using Individual-Level Survey Data." *Journal of Money, Credit and Banking* 54 (April 2022): 537-568.

------, Robert Rich, and Joseph Tracy. "Surveys of Professionals, Chapter 3." In *Handbook of Economic Expectations*, edited by Ruediger Bachmann, Giorgio Topa and Wilbert van der Klaauw, 71-106. Academic Press, Elsevier, 2022.

Croushore, Dean. "Introducing: The Survey of Professional Forecasters." *Federal Reserve Bank of Philadelphia Business Review*, December 1993, 3-15.

------, and Tom Stark. "Fifty Years of Professional Forecasters." *Economic Inquiry*, April 2019, 1-11.

D'Amico, Stefania, and Athanasios Orphanides. "Uncertainty and Disagreement in Economic Forecasting." Finance and Economics Discussion Series 2008-56. Federal Reserve Board, Washington DC, 2008.

Engelberg, Joseph, Charles F. Manski, and Jared Williams. "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters." *Journal of Business and Economic Statistics* 27 (January 2009): 30-41.

Garcia, Juan A. "An Introduction to the ECB's Survey of Professional Forecasters." ECB Occasional Paper No 8. European Central Bank, 2003.

Giordani, Paolo, and Paul Söderlind. "Inflation Forecast Uncertainty." *European Economic Review* 47 (December 2003): 1037-1059.

Glas, Alexander. "Five Dimensions of the Uncertainty-Disagreement Linkage." *International Journal of Forecasting* 36 (June 2020): 607-627.

Haddow, Abiail, Chris Hare, John Hooley, and Tamarah Shakir. "Macroeconomic Uncertainty: what is it, how can we measure it and why does it matter?" *Bank of England Quarterly Bulletin*, June 2013.

Kenny, Geoff, Thomas Kostka, and Federico Masera. "Can Macroeconomists Forecast Risk? Event-Based Evidence From the Euro-Area SPF." *International Journal of Central Banking* 11 (December 2015): 1-46.

Krüger, Fabian, and Lora Pavlova. "Quantifying Subjective Uncertainty in Survey Expectations." *International Journal of Forecasting* 40 (June 2024): 796-810.

Manski, Charles F. "Interpreting and Combining Heterogeneous Survey Forecasts." In *Oxford Handbook of Economic Forecasting*, edited by Michael P. Clements and D.F. Hendry, 457-472. Oxford, Oxford University Press, 2011.

Manzan, Sebastiano. "Are Professional Forecasters Bayesian?" *Journal of Economic Dynamics and Control* 123 (February 2021).

Mincer, Jacob A., and Victor Zarnowitz. "The Evaluation of Economic Forecasts." In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, edited by Jacob A. Mincer, 3-46. New York, National Bureau of Economic Research, 1969.

Nordhaus, William D. "Forecasting Efficiency: Concepts and Applications." *Review of Economics and Statistics* 69 (November 1987): 667-674.

Patton, Andrew J., and Allan Timmermann. "Why Do Forecasters Disagree? Lessons From the Term Structure of Cross-Sectional Dispersion." *Journal of Monetary Economics* 57 (October 2010): 803-820.

------, and Allan Timmermann. "Forecast Rationality Tests Based on Multi-Horizon Bounds." *Journal of Business and Economic Statistics* 30, no. 1 (2012): 36-40.

Poncela, Pilar, and Eva Senra. "Measuring Uncertainty and Assessing its Predictive Power in the Euro Area." *Empirical Economics* 53 (August 2017): 165-182.

Rich, Robert W., and Joseph Tracy. "The Relationships Among Expected Inflation, Disagreement, and Uncertainty: Evidence From Matched Point and Density Forecasts." *Review of Economics and Statistics* 92 (February 2010): 200-207.

------, and Joseph Tracy. "All Forecasters Are not the Same: Time-Varying Predictive Ability Across Forecast Environments." Working Paper WP 21-06. Federal Reserve Bank of Cleveland, February, 2021.

------, and Joseph Tracy. "A Closer Look at the Behavior of Uncertainty and Disagreement: Micro Evidence From the Euro Area." *Journal of Money, Credit, and Banking* 53 (February 2021): 233-253.

Stuart, Alan, and J Keith Ord. *Kendall's Advanced Theory of Statistics, Volume 1 Distribution Theory*. New York, John Wiley and Sons, 2010.

Zarnowitz, Victor. "The New ASA-NBER Survey of Forecasts by Economic Statisticians." *American Statistician* 23 (February 1969): 12-16.

------, and Louis A. Lambros. "Consensus and Uncertainty in Economic Predictions." *Journal of Political Economy* 95 (June 1987): 591-621.