

Clements, Michael P.; Rich, Robert W.; Tracy, Joseph S.

Working Paper

An Investigation into the Uncertainty Revision Process of Professional Forecasters

AEI Economics Working Paper, No. 2023-14

Provided in Cooperation with:

American Enterprise Institute (AEI), Washington, DC

Suggested Citation: Clements, Michael P.; Rich, Robert W.; Tracy, Joseph S. (2023) : An Investigation into the Uncertainty Revision Process of Professional Forecasters, AEI Economics Working Paper, No. 2023-14, American Enterprise Institute (AEI), Washington, DC

This Version is available at:

<https://hdl.handle.net/10419/280675>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



An Investigation into the Uncertainty Revision Process of Professional Forecasters

Michael P. Clements

ICMA Centre, Henley Business School, University of Reading

Robert Rich

Federal Reserve Bank of Cleveland

Joseph Tracy

American Enterprise Institute

AEI Economics Working Paper 2023-14

September 2023

An Investigation into the Uncertainty Revision Process of Professional Forecasters

Michael P. Clements
Robert Rich
Joseph Tracy*

September 20, 2023

Abstract: Following Manzan (2021), this paper examines how professional forecasters revise their uncertainty (variance) forecasts. We show that popular first moment “efficiency” tests are not applicable to study variance forecasts and instead employ monotonicity tests developed by Patton and Timmermann (2012). We find strong support for the Bayesian learning prediction of decreasing patterns in the variance of fixed-event density forecasts and their revisions as the forecast horizon declines. We explore the role of financial conditions indices in variance forecasts and document their predictive content for the revision process of US professional forecasters, although the evidence is weaker for euro area forecasters.

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: American Enterprise Institute. 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* (2022) speaks to the extensive literature exploring this topic across many dimensions. One area that has received attention focuses on evaluating the statistical properties of survey expectations. Within this area, some studies have tested if reported expectations are unbiased predictors or 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). Other studies, including Patton and Timmerman (2012), 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, there are also studies that have examined *density* forecasts. An attractive feature of density forecasts is that they provide a basis to derive estimates of uncertainty that are often regarded as both theoretically and empirically superior to alternative approaches such as using forecast disagreement (see, e.g., Zarnowitz and Lambros (1987), Manski (2011), and Rich and Tracy (2021b) on the lack of a theoretical backing), or using model-based estimates (Giordani and Söderlind (2003), for example, argue that such estimates will typically be slow to adapt, and may be less forward looking).¹ Within this strand of the survey expectations literature, there has been interest in comparing measures of *ex ante* forecast uncertainty with *ex post* or realized uncertainty, where the former is calculated from the histograms elicited by the surveys: see Clements (2014). The key finding is that agents are underconfident at within year horizons, at odds with the prevailing view in the literature on behavioral economics and finance (see, e.g., the surveys by Rabin (1998) and Hirshleifer (2001)).²

In contrast to survey-based point forecasts, there have been few attempts to conduct formal tests of optimality and the efficiency of the forecast revision process for survey-based uncertainty

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, Hare, Hooley and Shakir (2013).

² Knoppel and Schulte Frankenfeld (2019) examine inflation variance forecasts from the Bank of England, the Banco Central do Brasil, the Magyar Nemzeti Bank and the Sveriges Riksbank, and generally replicate the finding that forecasters are underconfident at short horizons, and overconfident at longer horizons.

measures. Manzan (2021) is a notable exception and reports that professional forecasters occasionally revise their fixed-event density forecasts in a manner that violates a prediction of Bayesian learning. Specifically, he finds evidence that measures of uncertainty derived from density forecasts (hereafter “variance forecasts”) do not always decline monotonically as the forecast date approaches the target date. 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 an enlarged information set becomes the posterior and should have a higher precision. Because precision is just an 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.

Drawing upon Manzan (2021), this paper studies how professional forecasters update their variance forecasts. Specifically, we examine the prediction of Bayesian learning that uncertainty should decline monotonically as the survey date nears the target date. The analysis examines fixed-event density forecasts of growth and inflation from the U.S. Survey of Professional Forecasters (US-SPF) and fixed-event density forecasts of growth, inflation, and unemployment from the European Central Bank Survey of Professional Forecasters (ECB-SPF). We construct variance forecasts at the individual forecaster level using two approaches: the interquartile range (IQR) of the density forecasts and an entropy-based measure recently proposed by Krüger and Pavlova (2020).

The important methodological contribution of this paper concerns the testing procedures for the variance forecasts. The term structure-based approach of Mincer and Zarnowitz (1969) has been widely adopted by studies to test whether forecasters make efficient use of past forecasts when revising first moment predictions. Manzan (2021) also makes use of Mincer-Zarnowitz (MZ) type regressions to investigate the monotonicity of the second moment of density forecasts. However, we demonstrate there is a misspecification in his testing procedure because MZ efficiency tests cannot be applied in a simple fashion to variance forecasts. Specifically, the application of MZ type regressions and associated parameter restrictions to the second moment of a density forecast is inconsistent with the null hypothesis of a monotonically decreasing pattern as the forecast horizon declines.

While the previous discussion argues that the empirical framework of Manzan (2021) is not valid to evaluate the monotonicity of variance forecasts, we can instead draw upon the tests of

monotonicity in Patton and Timmermann (2012) for this purpose. Specifically, we apply the tests of monotonicity of mean squared forecasts and mean squared forecast revisions to the variance forecasts. An attractive feature of the approach is that it does not require observed realizations of the target variable. To the best of our knowledge, our analysis represents the first application of the Patton-Timmermann rationality tests to second moments of survey expectations.

In contrast to the results of Manzan (2021), we find strong evidence that the density predictions of professional forecasters are consistent with the tenets of forecast rationality. Specifically, formal tests at the forecaster level fail to reject the predictions of a decreasing pattern in the variance forecasts and their revisions as the forecast horizon declines in almost all cases. Importantly, our testing procedure makes clear that visual inspection of the data is not sufficient for drawing inferences about the property of monotonicity. While our sample period, variance forecasts, and panel composition are not identical to those in Manzan (2021), we find that the application of MZ tests to our data yield results that broadly match those of Manzan. Hence, we can conclude that the alternative testing procedures in the two studies account for their different conclusions.

We then extend the analysis to investigate the process used by forecasters to form their variance forecasts. Specifically, we consider whether the revisions in variance forecasts are related to indices of current financial conditions for the U.S. and euro area. That this may be a useful way of modelling variance forecasts is motivated by the recent literature on 'growth-at-risk' (or GaR), and the finding that current financial conditions have predictive power for the lower quantiles of future real GDP growth (see Adrian, Boyarchenko and Giannone (2019)). We find that financial conditions have predictive content for the variance forecast revision process of U.S. professional forecasters, although the evidence is weaker for their counterparts in the euro area.

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. We then discuss how tests for monotonicity in Patton and Timmermann (2012) – for mean squared forecasts and mean squared forecast revisions – can be used to test 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 for the monotonicity tests and the predictive content of financial conditions in the variance forecast revision process. Section V contains concluding remarks.

II. Empirical Approach – Methods to Evaluate Variance Forecasts

The Mincer and Zarnowitz (1969) regression approach (and the related Optimal Revision Regression (ORR) extension of Patton and Timmermann (2012)) has been widely used to evaluate the properties of point forecasts. In the context of fixed-event forecasts, that is, when we have a sequence of forecasts of the same event or target variable made at a number of different origins, these approaches embody testing 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)). Manzan (2021) adapts this framework to evaluate variance forecasts, but, as we explain below, there are reasons why the approach is not directly applicable in this situation and argues instead for the use of the Patton and Timmermann (2012) tests of monotonicity.

2.1 The Mincer-Zarnowitz Testing Approach

The simplest Mincer-Zarnowitz (MZ) test considers whether forecasters efficiently make use of the information available at the time their forecast is made, which must of course include the forecast itself. Hence, the null hypothesis of forecast efficiency is that $H_0 : \delta_0 = 0$ and $\delta_1 = 1$ in:³

$$y_t = \delta_0 + \delta_1 y_{t-h_1} + \mu_t \quad (1)$$

where y_t is the target variable, y_{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 . This is referred to by Nordhaus (1987) 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 that $H_0 : \delta_0 = 0$ and $\delta_1 = 1$ and $\theta = 0$.

Patton and Timmermann (2012) note that one can write a short horizon forecast as:

³ See Clements (2022) for further discussion.

$$y_{t|t-h_1} \equiv y_{t|t-h_H} + d_{t|h_1, h_2} + \dots + d_{t|h_{H-1}, h_H} \quad (2)$$

where $h_1 < h_2 < \dots < h_H$, h_H is the longest horizon forecast of the target 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 revisions between the longest-horizon and the longest-horizon-less-one ($d_{t|h_{H-1}, h_H}$), and so on up to the revision between the shortest forecast horizon and shortest-plus-one

($d_{t|h_1, h_2}$). The Optimal Revision Regression (ORR) test substitutes for $y_{t|t-h_1} = y_{t|t-h_H} + \sum_{i=1}^{H-1} d_{t|h_i, h_{i+1}}$ in

equation (1) and allows for a free coefficient on each component of $y_{t|t-h_1}$. We can then estimate:

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

and test the null hypothesis $H_0: \delta_0 = 0$ and $\delta_1 = \delta_2 = \dots = \delta_H = 1$. Under the null hypothesis,

equation (3) becomes $y_t = y_{t|t-h_1} + \mu_t$ which shows that the error for the short-horizon forecast

$y_{t|t-h_1}$ is uncorrelated with all forecasts of y_t made at earlier times (and hence on smaller information sets). As explained by Patton and Timmermann (2012, p.6), the ORR 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 the U.S. Survey of Professional Forecasters (US-SPF) and the European Central Bank Survey of Professional Forecasters (ECB-SPF). Each survey provides a forecast of the survey-year rate of GDP growth and inflation, so that for a given year we have four fixed-event forecasts of approximate horizons of 4, 3, 2, and 1 quarters that are made, respectively, in the 1st, 2nd, 3rd and

4th quarters of the year. In addition, the ECB-SPF provides a forecast of the survey-year unemployment rate.⁴

The ORR regression approach allows one of the forecasts (the shortest) to be substituted for the actual in equation (3). This solves the difficulty of finding a proxy for \mathcal{Y}_t when needed, as is the case for the variance forecasts.⁵ Because there are multiple forecasts of the same target, this ensures a shorter horizon forecast is always available. The interpretation of the regression with \mathcal{Y}_t replaced by $\mathcal{Y}_{t|t-h_1}$ 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 MZ Tests and Survey-based Variance Forecasts

To evaluate forecast precisions (or their inverse, variances), Manzan (2021, p.17) uses equation (3) with \mathcal{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:

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

where $\phi_{q,t}$ are the revisions, $\phi_{q,t} = \psi_{q,t} - \psi_{q-1,t}$. He notes that the Bayesian Learning Model (BLM) suggests that $\psi_{q,t} \geq \psi_{q-1,t}$, (for $q = 2, 3, 4$), but reports tests of the null hypothesis $H_0: \alpha = 0$ and $\beta_q = 1, q = 1, 2, 3$, directly corresponding to the Patton and Timmermann (2012) null hypothesis of

⁴ The US-SPF and ECB-SPF also provide four additional fixed-event forecasts of approximate horizons of 8, 7, 6, and 5 quarters. As discussed in Section III, our participation criterion requires respondents to make 15 or more complete sequences of four quarterly forecasts of a variable for a current target year. Consequently, we do not use these additional forecasts because they would further restrict the number of available observations for the analysis.

⁵ This may also be desirable when the choice of data vintage to use as realized values for point forecasts is unclear.

rationality for point forecasts. However, equation (4) under this null hypothesis 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 seem more sensible not to require $\alpha = 0$. So the null hypothesis might instead be simply $H_0 : \beta_q = 1, q = 1, 2, 3$.

However, it is straightforward to show that there are more fundamental problems with the application of the MZ regression approach 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 $V_{q,t} = \psi_{q,t}^{-1}$.⁶

$$V_{4,t} = \delta_0 + \delta_1 V_{3,t} + \mu_t \quad (5)$$

For a simple data generating process capable of producing time-variant 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 directly applied to test for the optimality of variance forecasts because the associated parameter restriction $\delta_1 = 1$ does not hold. That is, the data will lead Manzan to reject $\delta_1 = 1$ in equation (5) even if the variance forecasts are optimal and decline monotonically as the forecast horizon shortens. While one might consider developing an approach based on testing the null hypothesis $H_0 : \delta \leq 1$ against the alternative hypothesis $H_1 : \delta > 1$, this appears to be closely related to the tests of monotonicity of Patton and Timmermann (2012), and we prefer to consider that approach instead.

⁶ We choose to work with variances because this is preferable for the uncertainty measure based on Krüger-Pavlova (2020). There are a few instances when the measure equals zero, resulting in the associated precision value being undefined. Similar to the point raised in footnote 4, our participation criterion in this situation would result in a smaller sample size. Because precisions and variances offer an equivalent basis to describe Manzan's (2021) methodology and testing procedure, the choice to cast our subsequent discussion in terms of variances reflects the benefit it provides for the empirical analysis compared to the use of precisions.

2.3 Tests of Monotonicity

According to the Bayesian learning model a simple requirement for optimality is that the variance forecasts should decline monotonically as the forecast horizon shortens. Hence, in the context of US-SPF and ECB-SPF quarterly surveys of inflation and GDP growth in the current year relative to the previous year, the histogram 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, eqn. (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 for two reasons. First, the variance forecasts are non-negative and therefore do not require the application of a squaring operator.⁷ Second, in contrast to the expected squared optimal forecasts of a first moment, the variance forecast should be non-increasing as the horizon shortens. Consequently, the null hypothesis is $H_0 : \Delta^f \leq 0$ versus $H_1 : \Delta^f \not\leq 0$, where Δ^f is the vector of population parameters, which in our case consists of the quarter-on-quarter variances:

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

The test is based on the sample analog of $\Delta^f, \hat{\Delta}^f$, which is given by:⁸

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

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

⁷ Nevertheless, we did conduct the MSF test applying the squaring operator as a robustness check and the results lead to nearly identical conclusions.

⁸ To calculate the distribution under the null hypothesis H_0 , we use the Matlab code made available by Andrew Patton at <http://public.econ.duke.edu/~ap172/code.html>.

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

Then the null hypothesis is $H_0 : \Delta^d \leq 0$ versus $H_1 : \Delta^d \not\leq 0$. For example, 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. (This follows directly from Manzan (2021, equation 4)). Following the approach for the MSF test, we can use the sample analog of $\Delta^d, \hat{\Delta}^d$, as the basis for the MSFR test.

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.

2.4 Variance Forecasts and Financial Conditions Indices

Our findings for the MZ regression suggest the “strong efficiency” test regression may also be problematic. As discussed in Section 2.1, the strong-efficiency test considers whether the forecast revision is systematically related to information known at the time the earlier forecast was made, while simultaneously restricting the constant term to zero, and the “slope” (the coefficient on the forecast, or the longer-horizon forecast) to one. But as we have seen, these restrictions are not valid when the data pertain to variance forecasts. For this reason, instead of attempting to test whether the variance forecast revisions are systematically related to earlier information using a strong-efficiency test, we attempt to shed light on the types of information that are likely to induce the survey respondents to revise their forecasts. A leading candidate for such a source of information, in the context of variance forecasts, is a financial conditions index (FCI). Manzan (2021) also considers potential determinants of forecaster behavior, by relating the time-variation in the Bayesian weight – the ratio of the prior to posterior precisions – to “surprises” in announcements (about the variable of interest), amongst other things.

Figure 1 and Figure 2 plot the FCIs for the U.S. and the euro area, respectively, used in

the analysis. The U.S. FCI is the Federal Reserve Bank of Chicago's National Financial Conditions Index which is a weighted average of 105 measures of financial activity each expressed as a standardized variable. The euro area FCI is the Goldman Sachs Euro Area FCI which is calculated as a weighted average of a policy rate, a long-term risk-free bond yield, a corporate credit spread, an equity price variable, a trade-weighted exchange rate, and a sovereign credit spread. The weights mirror the effects of the financial variables on real GDP growth in the models over a one-year horizon.⁹ The reference points for the U.S. FCI and euro area FCI are 0 and 100, respectively, and indicate average financial conditions, with values above (below) the average historically associated with tighter (looser)-than-average financial conditions.

We consider the following regression for our investigation into the role of FCIs in the variance forecast revision process:

$$V_{4,t} - V_{q,t} = \alpha + \delta_q (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t \quad (9)$$

where $FCI_{t,q}$ is the average value of the FCI during quarter q in year t up to the time the survey is conducted in that quarter. We estimate this regression for $q = 3$ and $q = 2$. We can also estimate equation (9) for Q3 forecasts instead of Q4 forecasts by replacing $V_{4,t}$ with $V_{3,t}$, replacing $(FCI_{t,4} - FCI_{t,q})$ by $(FCI_{t,3} - FCI_{t,q})$, and setting $q = 2$ and $q = 1$. Because positive (negative) values of the change in the FCI are associated with tighter-than-average (looser-than-average) financial conditions, our interest is whether there is a positive relationship between changes in financial conditions and the variance forecast revisions. Accordingly, the null hypothesis is $H_0 : \delta_q \leq 0$ versus $H_1 : \delta_q > 0$.

For the U.S., the results of Adams, Adrian, Boyarchenko and Giannone (2021) suggest that *tight* financial conditions help predict inflation uncertainty, as well as *downside* growth risk.

⁹ The Federal Reserve Bank of Chicago's National Financial Conditions Index is available at <https://www.chicagofed.org/research/data/nfci/>. Additional details on the Goldman Sachs Euro Area FCI are available in this report: <https://www.gspublishing.com/content/research/en/reports/2017/04/20/c10f888f-4faa-4ffc-b4c2-518cf5ffffe3.html>.

Adrian *et al.* (2019) show the value of an FCI for forecasting the lower quantiles of real GDP growth, i.e., 'growth-at-risk' (see also, Amburgey and McCracken (2023)), suggesting a non-linear relationship between an FCI and real GDP growth, as well as between an FCI and inflation. Given that we have relatively few observations per individual forecaster to capture the potential non-linearity, we adopt a simple approach. We regress the revision to the variance forecast on a variable which is the change in the FCI if financial conditions are tightening, or zero otherwise: a loosening is not restricted to have an equal effect (of the opposite sign)¹⁰

To capture an effect from tightening only, we modify the specification in equation (9) and estimate:

$$V_{4,t} - V_{q,t} = \alpha + \delta_q \text{Tight}_{t,4} \times (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t \quad (10)$$

where $\text{Tight}_{t,4} = 1$ if $FCI_{t,4} - FCI_{t,q} > 0$ and $\text{Tight}_{t,4} = 0$ otherwise. When equation (10) is specified for V_3 , then we use $\text{Tight}_{t,3} = 1$ if $FCI_{t,3} - FCI_{t,q} > 0$ and $\text{Tight}_{t,3} = 0$ otherwise. The null hypothesis $H_0 : \delta_q \leq 0$ versus $H_1 : \delta_q > 0$ remains the same as in equation (9).

III. Data and Variance Forecast Measures

The analysis examines data from the U.S. Survey of Professional Forecasters (US-SPF) and the European Central Bank Survey of Professional Forecasters (ECB-SPF). 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, but from 1990 to the present day has been run by the Federal Reserve Bank of Philadelphia.¹¹ The average number of panelists has been around 40 since the Philadelphia Fed has overseen the survey. The respondents are anonymous but have identifiers allowing the forecasts of a given individual to be followed across surveys. The survey of U.S. macro variables is conducted each quarter after the release of the advance report of the national income and product accounts (NIPA) from the Bureau of Economic Analysis (BEA). Since 1990:Q2, the

¹⁰ This should also capture the impact of downside risk for real GDP growth, to the extent that larger negative values of the lower quantiles will in practice likely inflate the variance of the distribution.

¹¹ 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 Federal Reserve Bank of Philadelphia.

deadlines for responses have been around the middle of each quarter—before the BEA's second report. The Federal Reserve Bank of Philadelphia's posts quarterly reports summarizing the survey results and the underlying data on their website.

The ECB-SPF began in January 1999 and provides a survey of forecasts for the euro area. The principal aim of the survey is to elicit expectations about inflation, real GDP growth, and unemployment. The survey is typically fielded in February, May, August, and November, with a little under 60 respondents on average per survey.¹² 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.¹³

The US-SPF and ECB-SPF both elicit 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 attention to the fixed-event density forecasts because of our interest in the revisions to the variance forecasts.

The analysis of the US-SPF examines 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.¹⁴ The survey structure is the same for output which reports density forecasts of real GDP growth. For the ECB-SPF, we examine 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, except that the inflation measure is the Harmonised Index of Consumer Prices (HICP). 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.

¹² The actual average of 58 panelists per survey is based on the history of the ECB-SPF, although there has been some decline in participation over time.

¹³ For additional details about the ECB-SPF, see Garcia (2003) and Bowles et al. (2007).

¹⁴ This results in the availability of eight “fixed-event” histogram forecasts of annual inflation in 2016, say, compared to 2015. As noted in footnote 4 and discussed later in this section, we restrict our attention to just the four quarterly forecasts of a variable due to considerations of sample size based on our participation criterion.

There are various ways of calculating variances from histograms, as explained by Clements, Rich and Tracy (2022). We report results for variances calculated as the interquartile range (IQR) and the entropy function of the Ranked Probability Score (RPS), namely ERPS, as suggested by Krüger and Pavlova (2020). 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 the lowest and highest bins, respectively.¹⁵ To facilitate discussion at the moment, we assume that the number of bins and the number of respondents do not change over time, and therefore, we omit a time subscript for these terms here.

Following Abel et al. (2016) and Glas (2020), we define the individual IQR measure as:

$$p_{i,t+h|t}^{IQR} = p_{i,t+h|t}^{0.75} - p_{i,t+h|t}^{0.25} \quad (11)$$

where $p_{i,t+h|t}^{0.75}$ and $p_{i,t+h|t}^{0.25}$ denote the 75th and 25th percentiles, respectively, of respondent i 's density forecast.

The expected ranked probability score (ERPS) for respondent i is given by

$$\sum_{k=1}^K P_{i,k} (1 - P_{i,k}), \text{ where } p_i(j) \text{ is the probability assigned to the } j\text{th bin and } P_{i,k} = \sum_{j=1}^k p_i(j) \text{ is the}$$

cumulative probability of the first k bins. The advantage of ERPS is that it quantifies the uncertainty implicit in a histogram using only the probabilities assigned to different ranges, whereas other popular methods, such as fitting parametric distributions (e.g., Engelberg, Manski and Williams (2009)) require additional assumptions, such as the choice of distribution, and how to proceed when there are too few non-zero probability bins to fit the distribution.

Some additional auxiliary assumptions are required for the calculation of the variance forecasts. 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 typical—although ad hoc—practice and assign

¹⁵ The probabilities are assumed to sum to unity.

twice the width of the interior closed intervals to the open intervals. The degree to which this will impact any estimate depends on the amount of assigned probability to each open interval.¹⁶

A second issue that is relevant for the IQR measure concerns the location of the probability mass within a specified closed interval. Some studies [Boero, Smith, and Wallis (2008a), Rich and Tracy (2010), Kenny, Kostka, and Masera (2015), and Poncela and Senra (2017)] assume that the probability mass is located at the midpoint of each bin. We will follow other studies [Zarnowitz and Lambros (1987), Abel et al. (2016), and Rich and Tracy (2021a)] and instead assume that the probability is distributed uniformly within each interval. Under this assumption, the lower and upper quartiles in equation (9) are calculated by linear interpolation.

A third issue is that the widths of the bins used to define the histogram for the density forecasts occasionally change over time. A drawback of the ERPS is that it is not invariant to changes in the bin width across surveys. To see the lack of invariance, 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 bin widths were doubled, then $[1,3]$ would have $p = 1$, and $\text{ERPS} = 0$. The issue of bin widths is not a concern for the ECB-SPF density forecasts because they remained constant at 0.5 percentage point for our sample period from 1999:Q1 to 2019:Q4. However, this issue is relevant for the GDP-deflator inflation density forecasts for the US-SPF for our sample period from 1992:Q4 to 2019:Q4. Specifically, the bin 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 bin widths of a percentage point by summing the probabilities for the adjacent half-percentage point histograms.

A final issue concerns composition effects for the surveys. The US-SPF and ECB-SPF involve panels of forecasters whose participation changes as some respondents leave and new respondents are added. The panels are unbalanced due both to entry and exit and to the fact that there are occasions when respondents do not provide responses to all or part of the survey questionnaire. Because the forecast revision process is central to our study, we only include data that allow us to view a complete path of forecasts toward a target variable. Consequently, we restrict the sample to observations where respondents provide a complete sequence of four forecasts of the

¹⁶ The IQR is typically more robust than a standard deviation/variance estimate to situations when respondents place probability in open intervals. The IQR is unaffected unless the respondent places more than a 25% probability in an open interval.

current target year. 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. In addition, we required a participant to have at least 15 complete 4-quarter sequences. For the US-SPF, these criteria result in the inclusion of 6 participants for the inflation density forecasts and 7 participants for the output density forecasts. For the ECB-SPF, there are 17 participants for the inflation density forecasts and 16 participants for the density forecasts for output and unemployment.

IV. Empirical Results

Drawing upon Manzan (2021), Figure 3 and Figure 4 plot the quarterly time average of the variance forecasts for the US-SPF and the ECB-SPF, respectively. Each line is obtained from applying either the IQR or ERPS measures to the density forecasts for each forecaster in our panel. Taken together, the plots generally show a pattern of declining variance forecasts toward the end of the year. It is interesting to note that despite differences in the panels of forecasters and methodologies, our variance forecasts and the precisions in Manzan (2012) display several similar features. For example, there is notable heterogeneity across forecasters in the level of uncertainty at all horizons. Moreover, there is limited evidence of the lines crossing which suggests that there is persistent heterogeneity in the variance forecasts. 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 evidence is consistent with other studies that have documented persistent heterogeneity in various features of forecast behavior.¹⁷

Further inspection of Figure 3 and Figure 4 reveals that there are also instances when the variance forecasts of an individual remain relatively constant or show an actual increase as the horizon declines. While these patterns might be interpreted as *prima facie* evidence of a violation of declining monotonicity, visual inspection is not sufficient to draw reliable inferences and conclusions about this property of the data. Instead, formal testing procedures are required to make an informed evaluation of the patterns.

¹⁷ See D’Amico and Orphanides (2008), Patton and Timmermann (2010), Bruine de Bruin et al. (2011), Boero, Smith and Wallis (2008, 2015), Rich and Tracy (2021a, 2021b) and Clements (2014, 2022).

Table 1 presents the results from applying the monotonicity testing procedure of Manzan (2021) and Patton and Timmermann (2012) to the ERPS- and IQR-based variance forecasts. We initially report the total number of respondents who filed an appropriate number of density forecasts for the target variables for the ECB-SPF and the US-SPF. We then provide a count of the number of rejections of the relevant null hypothesis for each testing procedure at the 10% level of significance. As shown, we consider two versions of the null hypothesis for the monotonicity testing procedure of Manzan (2021) that each incorporate the joint restriction that the slope coefficients equal unity but differ in their 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 the upper panel of Table 1 are based on Manzan's (2021) testing procedure and show with little ambiguity that most forecasters in the ECB-SPF and US-SPF update their density forecasts in a way that is inconsistent with Bayesian learning. 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). While there are fewer rejections in almost all cases when we relax the restriction on the constant term, the rejection rates remain extremely high.

The bottom panel of Table 1 reports the results from the Patton-Timmermann (2012) tests for monotonicity. For optimal forecasts, the mean squared forecasts, MSF, should be non-increasing as the horizon shortens (see equation (6)). As shown, we fail to reject the null hypothesis in all cases for the ECB-SPF density forecasts. For the US-SPF, we reject the null hypothesis for one of the six density forecasts for each variance forecast measure. Next, we turn to the results for the mean squared forecast revisions (MSFR) which should be non-increasing in the 'length' of the revision - see equation (8). Matching the results for the MSFs, we fail to reject the null hypothesis in all cases for the ECB-SPF density forecasts. For the US-SPF, we only reject the null hypothesis for one respondent using the IQR measure of the variance forecast.

The evidence from the MSF and MSFR tests offer strong support for the predicted monotonicity properties of the variance forecasts across forecast horizons and are in sharp contrast with the conclusions of Manzan (2021). As previously discussed, we acknowledge that our sample period, variance forecasts, and panel composition are not identical to those in Manzan

(2021). However, we also expressed concerns about the interpretation and applicability of Manzan's (2021) testing procedure for monotonicity and view the findings in Table 1 as largely validating these concerns as well as the adoption and inferences drawn from our alternative testing procedure.

Tables 2 and 3 report the findings for the regressions of the variance forecast revisions on the FCIs for the ECB-SPF and US-SPF, respectively. We estimate individual regressions for each survey respondent who again met the participation criterion. We report the number of individuals for whom we reject the null hypothesis that the coefficient on FCI is zero against the alternative of a positive value at a 10% significance level. Each table reports the results for the symmetric specification and for the asymmetric specification, separately.

For the ECB-SPF, for each variable and forecast horizon, we reject the null for no more than a quarter of the 16 or 17 individual forecasters. The results are similar for the two measures of the variance forecast and are broadly unchanged whether we allow for non-linearity in the response of the variance revision to the change in the FCI. However, the results for the US-SPF in Table 3 provide much more evidence of a role for the FCI for variance forecast revisions. While there are fewer individuals (6 or 7), there are nevertheless several instances where we reject the null for nearly half or over half the time for the individual forecasters. It is also interesting that the results are much stronger for GDP growth than inflation, offering evidence that the predictive content of the FCI extends beyond first moments of GDP growth to include second moments. However, the results match the ECB to the extent that the asymmetric specification yields similar results to the symmetric, and the results are not very different depending on the variance forecast measure.

Finally, Table 4 reports the results of pooling across forecasters, and is in the same format as Tables 2 and 3, except the cells report whether we reject the null hypothesis for the specified regression (that is, for the given variance forecast revision.) Pooling should counter the potential low power of the tests in Tables 2 and 3 due to the relatively small samples available at the individual level but will be misleading if the assumption of a homogeneous slope is not valid. As shown, we essentially fail to reject the null in all cases for the ECB-SPF. For the US-SPF, however, there is now evidence of a stronger relationship for inflation where we reject the null in half the cases for the symmetric specification and for 40 percent of the cases for the asymmetric specification. As noted above, the difference in the results in Table 2 and Table 4

for GDP growth and inflation for the US-SPF would suggest a greater degree of homogeneity in the slope coefficients for inflation.

V. Conclusion

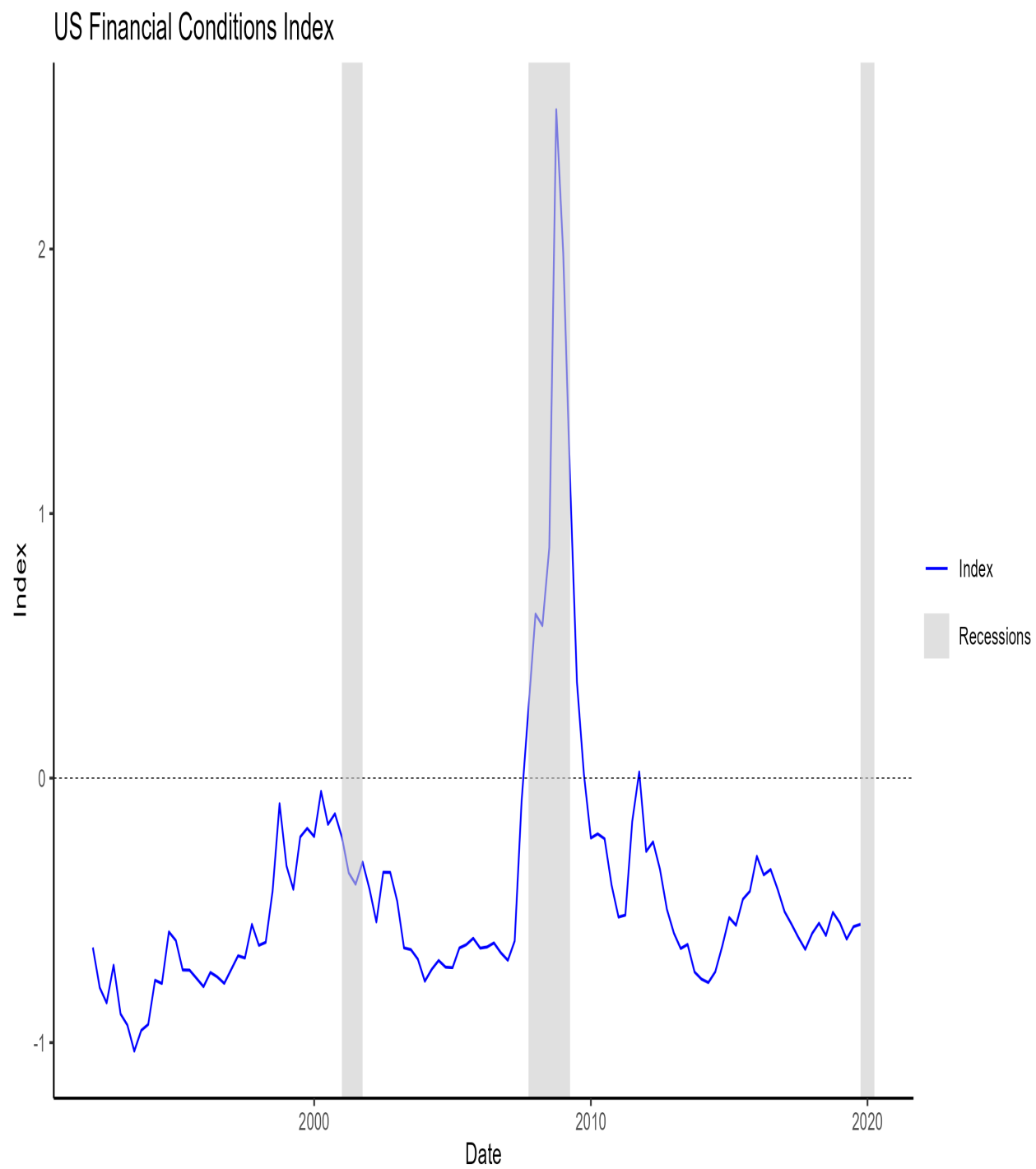
The main empirical takeaway of the paper is that there is little evidence against the optimality of the variance forecasts we examine. We consider samples of professional forecasters in the U.S. and the euro area. By optimality we mean that the variance forecasts decline as the length of the forecast horizon shortens, that is, as the fixed target draws near. This optimality finding is consistent with the Bayesian learning model, which in our context simply states that the variance forecasts decline as information accrues as the forecast origin nears the forecast target.

Our theoretical contribution is to show that the popular approach to testing for forecast efficiency of Mincer and Zarnowitz (1969), and the extensions of Patton and Timmermann (2012) to the Optimal Revision Regression, cannot be straightforwardly applied to variance forecasts. Variance forecasts are fundamentally different from point forecasts in this regard. However, the monotonicity tests of Patton and Timmermann (2012) are ideally suited to testing the optimality of the variance forecasts. We find that the monotonicity tests are supportive of the optimality of the variance forecasts, whereas the use of MZ tests leads to a widespread erroneous rejection of their optimality.

We also investigate a potential explanation of the magnitudes of the revisions to variance forecasts over time, namely, indices summarizing financial conditions. We find little evidence that financial conditions influence the revisions made by the ECB forecasters, but more evidence for the U.S. forecasters. We also consider whether there is a non-linear relationship between the revisions to the forecast variances and financial conditions, as suggested by the 'growth-at-risk' literature. We find little evidence that non-linearities matter, but caution that we have relatively small samples at the individual level and have not undertaken an extensive investigation of the different forms that non-linearity may take.

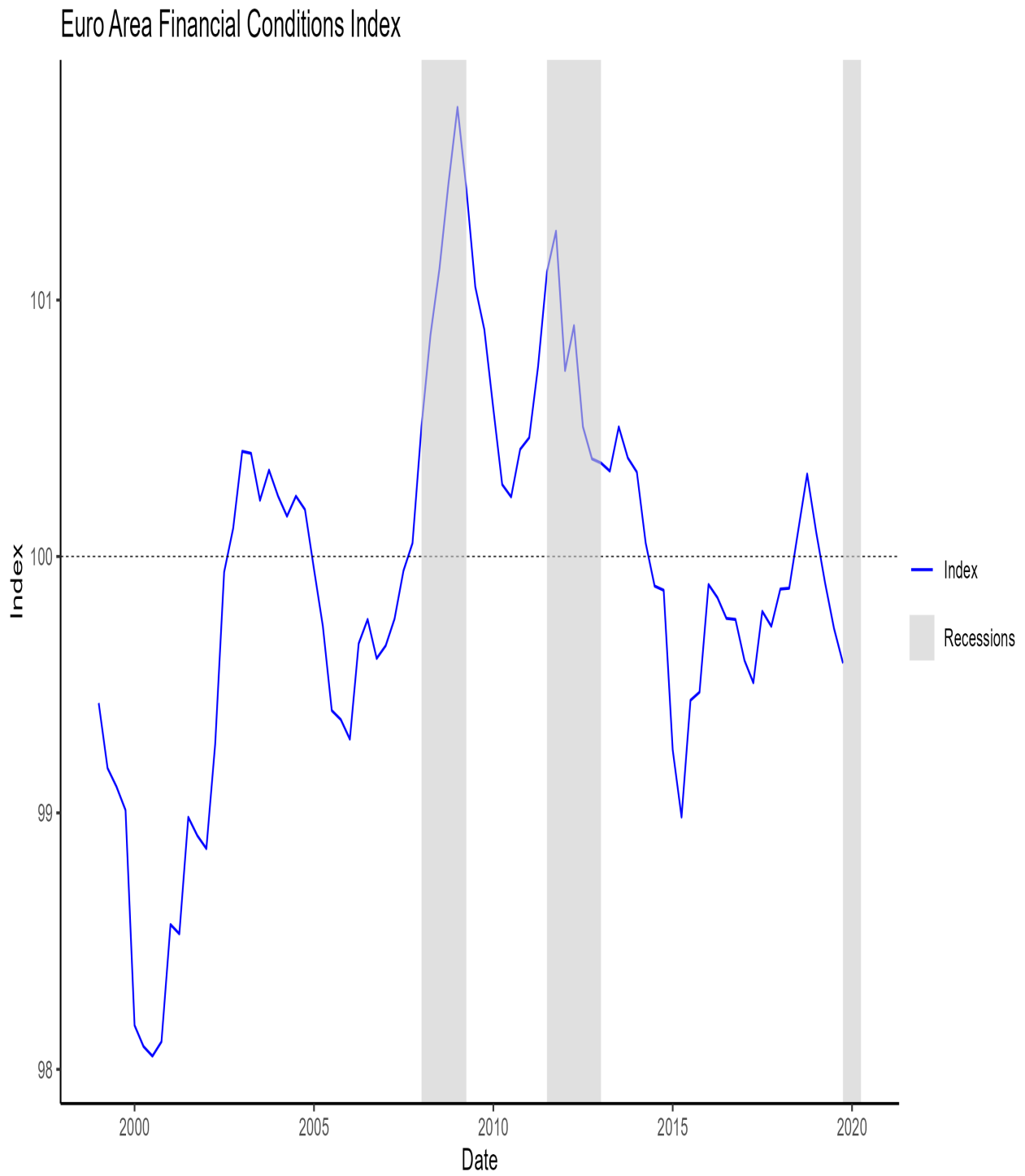
We regard determining and modelling the key drivers of density forecast revisions as an interesting topic for further research, including understanding the differences between the U.S. and ECB forecasters in terms of their responsiveness to financial conditions.

Figure 1



Source: Federal Reserve Bank of Chicago

Figure 2



Source: Goldman Sachs

Figure 3

US-SPF

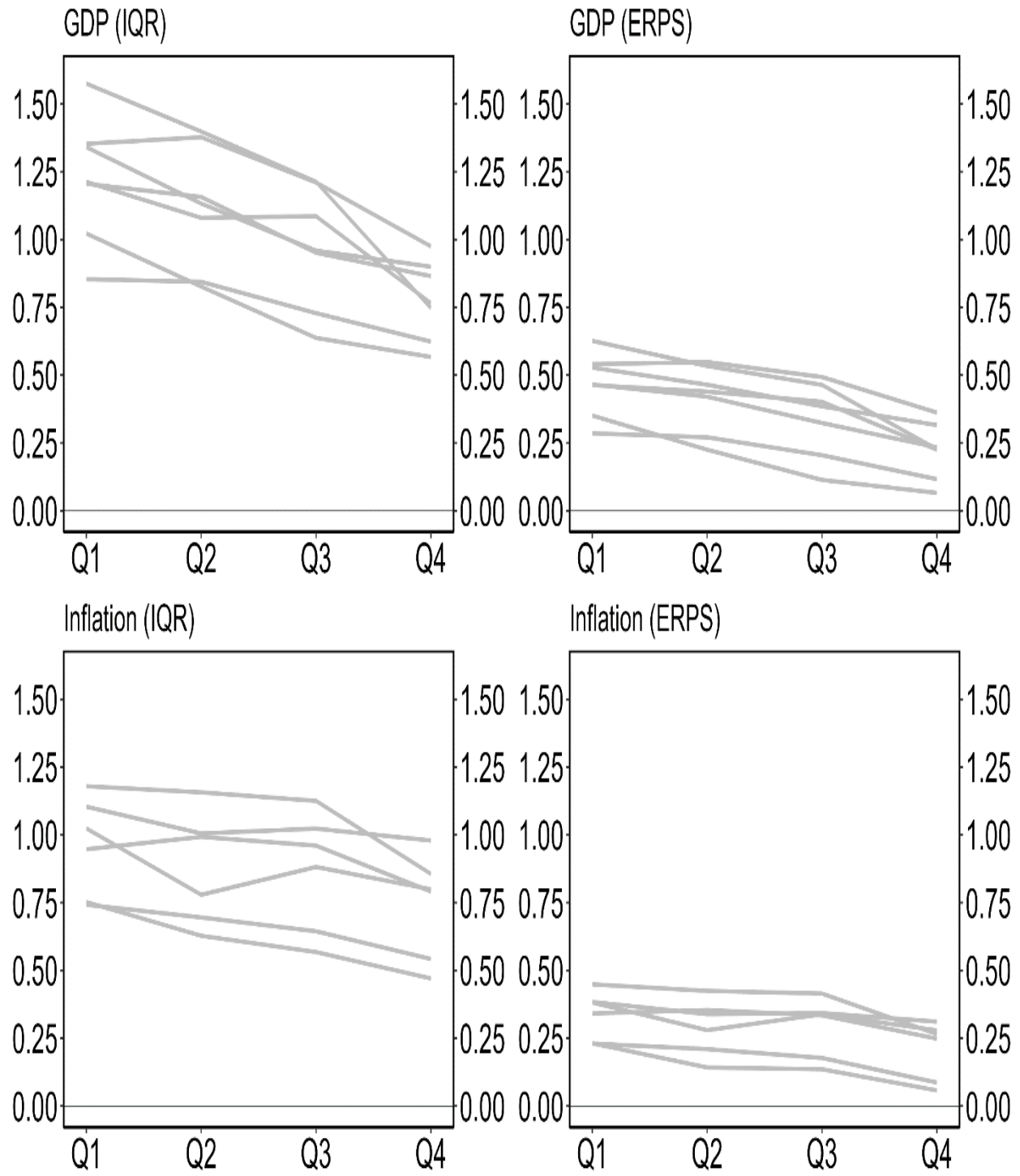


Figure 4

ECB-SPF

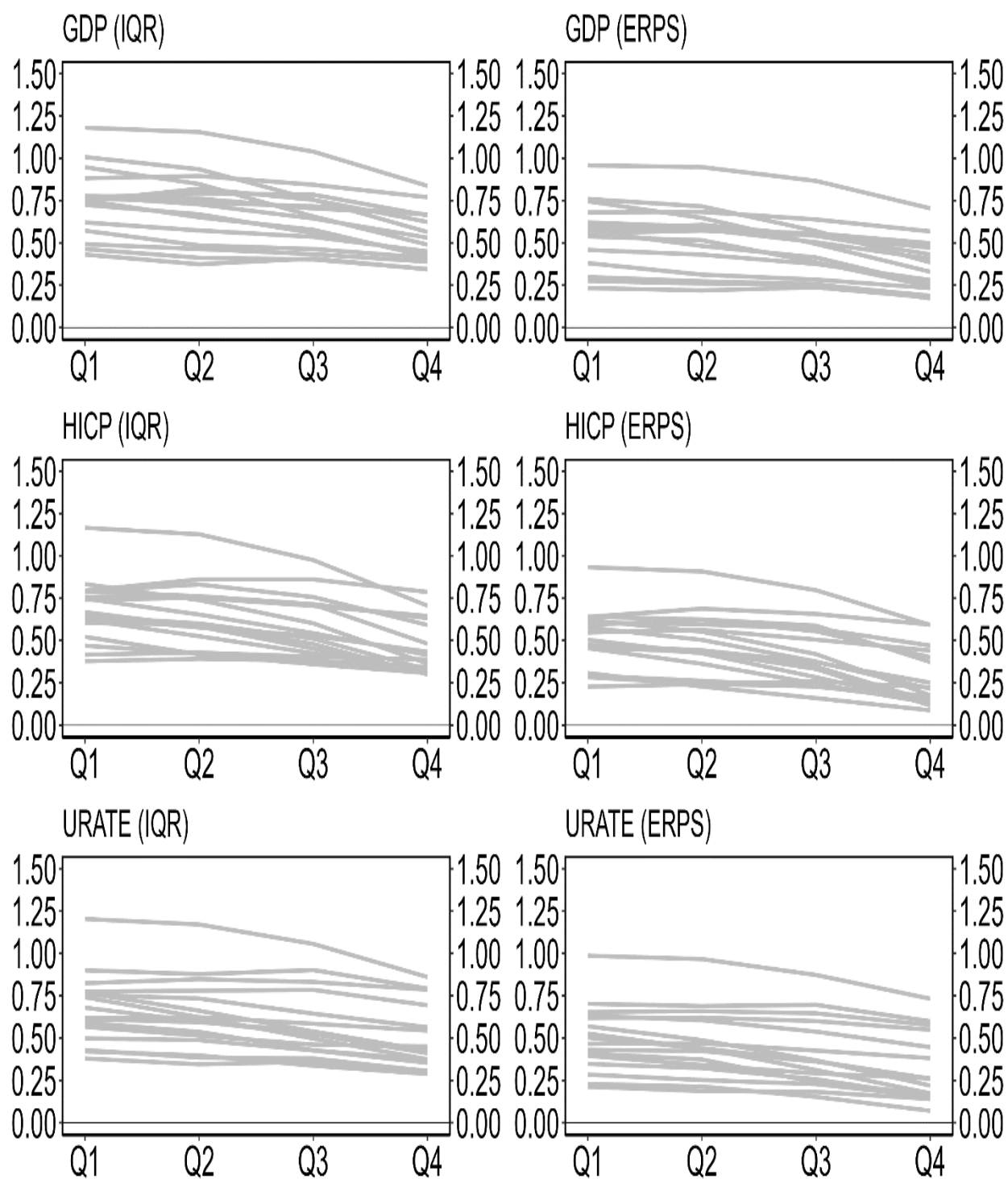


Table 1a – Manzan Tests for Monotonicity

$$V_{4,t} = \alpha + \beta_1 V_{1,t} + \sum_{q=2}^3 \beta_q (V_{q,t} - V_{q-1,t}) + \varepsilon_t$$

	GDP Growth				Inflation				Unemployment Rate			
ECB-SPF	N=16				N=17				N=16			
	ERPS		IQR		ERPS		IQR		ERPS		IQR	
	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
	14	12	15	13	17	13	15	14	14	8	15	11
US-SPF	N=7				N=6				---			
	ERPS		IQR		ERPS		IQR		ERPS		IQR	
	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$	$\chi^2(4)$	$\chi^2(3)$
	7	6	6	5	5	5	6	5	---		---	

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

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

Table 1b – Patton-Timmermann Tests for Monotonicity

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

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

	GDP Growth				Inflation				Unemployment Rate			
ECB-SPF	N=16				N=17				N=16			
	ERPS		IQR		ERPS		IQR		ERPS		IQR	
	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
	0	0	0	0	0	0	0	0	0	0	0	0
US-SPF	N=7				N=6				---			
	ERPS		IQR		ERPS		IQR		ERPS		IQR	
	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR	MSF	MSFR
	0	0	0	0	1	0	1	1	---	---	---	---

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

Table 2a – ECB-SPF Individual/Symmetric

$$V_{4,t} - V_{q,t} = \alpha + \delta_q (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth (N=16)		HICP Inflation (N=17)		Unemployment Rate (N=16)	
	ERPS	IQR	ERPS	IQR	ERPS	IQR
$V_4 - V_3$	2	2	2	3	2	1
$V_3 - V_2$	1	1	0	0	3	3
$V_2 - V_1$	2	1	2	2	0	2
$V_4 - V_2$	2	2	2	3	4	3
$V_3 - V_1$	1	2	2	2	1	4

Values report rejections of the null hypothesis $H_0 : \delta_q \leq 0$ at the 10% significance level.

Table 2b – ECB-SPF Individual/Asymmetric

$$V_{4,t} - V_{q,t} = \alpha + \delta_q Tight_{t,4} \times (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth (N=16)		HICP Inflation (N=17)		Unemployment Rate (N=16)	
	ERPS	IQR	ERPS	IQR	ERPS	IQR
$V_4 - V_3$	1	2	3	5	1	2
$V_3 - V_2$	1	1	1	1	4	2
$V_2 - V_1$	2	2	0	0	2	1
$V_4 - V_2$	1	2	1	1	3	3
$V_3 - V_1$	1	0	0	0	3	1

Values report rejections of the null hypothesis $H_0 : \delta_q \leq 0$ at the 10% significance level.

Table 3a – US-SPF Individual/Symmetric

$$V_{4,t} - V_{q,t} = \alpha + \delta_q (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth (N=7)		GDP Deflator Inflation (N=6)	
	ERPS	IQR	ERPS	IQR
$V_4 - V_3$	4	4	2	2
$V_3 - V_2$	3	3	2	2
$V_2 - V_1$	1	2	1	1
$V_4 - V_2$	3	3	3	2
$V_3 - V_1$	3	2	2	1

Values report rejections of the null hypothesis $H_0 : \delta_q \leq 0$ at the 10% significance level.

Table 3b – US-SPF Individual/Asymmetric

$$V_{4,t} - V_{q,t} = \alpha + \delta_q Tight_{t,4} \times (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth (N=7)		GDP Deflator Inflation (N=6)	
	ERPS	IQR	ERPS	IQR
$V_4 - V_3$	4	4	2	3
$V_3 - V_2$	3	2	2	2
$V_2 - V_1$	2	2	1	0
$V_4 - V_2$	3	3	3	3
$V_3 - V_1$	1	1	0	0

Values report rejections of the null hypothesis $H_0 : \delta_q \leq 0$ at the 10% significance level.

Table 4a – ECB-SPF Pooled Across Forecasters

$$V_{4,t} - V_{q,t} = \alpha + \delta_q (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth		HICP Inflation		Unemployment Rate	
	ERPS	IQR	ERPS	IQR	ERPS	IQR
$V_4 - V_3$						
$V_3 - V_2$					†	
$V_2 - V_1$						
$V_4 - V_2$						
$V_3 - V_1$						

† indicates rejection of the null hypothesis $H_0 : \delta_q \leq 0$ at the 10% significance level.

Table 4b – ECB-SPF Pooled Across Forecasters

$$V_{4,t} - V_{q,t} = \alpha + \delta_q Tight_{t,4} \times (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth		HICP Inflation		Unemployment Rate	
	ERPS	IQR	ERPS	IQR	ERPS	IQR
$V_4 - V_3$						
$V_3 - V_2$						
$V_2 - V_1$						
$V_4 - V_2$						
$V_3 - V_1$						

† indicates rejection of the null hypothesis $H_0 : \delta_q \leq 0$ at the 10% significance level.

Table 5a – US-SPF Pooled Across Forecasters

$$V_{4,t} - V_{q,t} = \alpha + \delta_q (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth		GDP Deflator Inflation	
	ERPS	IQR	ERPS	IQR
$V_4 - V_3$			†	†
$V_3 - V_2$	†		†	
$V_2 - V_1$	†	†		
$V_4 - V_2$			†	†
$V_3 - V_1$				

† indicates rejection of the null hypothesis $H_0 : \delta_q \leq 0$ at the 10% significance level.

Table 5b – US-SPF Pooled Across Forecasters

$$V_{4,t} - V_{q,t} = \alpha + \delta_q Tight_{t,4} \times (FCI_{t,4} - FCI_{t,q}) + \varepsilon_t$$

	GDP Growth		GDP Deflator Inflation	
	ERPS	IQR	ERPS	IQR
$V_4 - V_3$			†	
$V_3 - V_2$	†	†	†	†
$V_2 - V_1$	†			
$V_4 - V_2$			†	
$V_3 - V_1$				

† indicates rejection of the null hypothesis $H_0 : \delta_q \leq 0$ 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 :

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (\text{A1})$$

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:

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

From equation (A1),

$$y_t = \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \quad (\text{A3})$$

and so:

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

where the second line comes from $\text{Var}(\varepsilon_t | y_{t-2}) = h_{t|t-2}$, 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:

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

The numerator is:

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

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

The denominator is:

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

which results in:

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

Hence, $\beta < 1$, because $\alpha > 0$, and $\phi^2 > 0$, since $-1 < \phi < 1$.

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.
- Adams, Patrick A., Tobias Adrian, Nina Boyarchenko, and Domenico Giannone. "Forecasting Macroeconomic Risks." *International Journal of Forecasting* 37 (September 2021): 1173-1191.
- Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. "Vulnerable Growth." *American Economic Review* 109 (April 2019): 1263-1289.
- Amburgey, Aaron J., and Michael W. McCracken. "On the Real-Time Predictive Content of Financial Condition Indices for Growth." *Journal of Applied Econometrics* 38 (March 2023): 137-163.
- Bachmann, Ruediger, Giorgio Topa, and Wilbert Klaauw (2022). *Handbook of Economic Expectations*. Academic Press, Elsevier.
- 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 1, 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, Abigail, 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.
- Hirshleifer, David. "Investor Psychology and Asset Pricing." *Journal of Finance* 56 (August 2011): 1533-1597.
- 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.
- Knüppel, Malte, and Guido Schulte Frankenfeld. "Assessing the Uncertainty in Central Banks' Inflation Outlooks." *International Journal of Forecasting* 35 (December 2019): 1748-1769.
- Krüger, Fabian, and Lora Pavlova. "Quantifying Subjective Uncertainty in Survey Expectations." Working Paper #5. University of Berlin, March 1, 2020.
- 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.
- Rabin, Matthew. "Psychology and Economics." *Journal of Economic Literature* 36 (March 1998): 11-46.
- 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 2021a.
- , 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 2021b): 233-253.

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