Department of Economics

Group for Economic Analysis at Reading (GEAR)

Do Professional Forecasters' Phillips Curves Incorporate the Beliefs of Others?

by Michael P. Clements and Shixuan Wang

Discussion Paper No. 2023-05

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

Do Professional Forecasters' Phillips Curves Incorporate the Beliefs of Others?

Michael P. Clements
ICMA Centre, Henley Business School, University of Reading
and
Shixuan Wang*
Department of Economics, University of Reading

December 8, 2023

Abstract

We apply functional data analysis to survey expectations data, and show that functional principal component analysis combined with functional regression analysis is a fruitful way of capturing the effects of others' forecasts on a respondent's inflation forecasts. We estimate forward-looking Phillips curves on each respondent's inflation and unemployment rate forecasts, and show that for nearly a half of the respondents the forecasts of others are important. The functional principal components of the cross-sectional distributions of forecasts are shown to capture characteristics other than the mean or consensus forecast, and include forecaster disagreement.

Keywords: Inflation forecasting, Functional data analysis, Survey of Professional Forecasters, Forecast disagreement

JEL Classification: C53, E37

*E-mail address: m.p.clements@icmacentre.ac.uk (M. P. Clements); shixuan.wang@reading.ac.uk (S. Wang). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors report there are no competing interests to declare.

1 Introduction

In the recent literature there has been much interest in why survey respondents disagree with one another in terms of their expectations. One strand of the literature has asked whether forecasters disagree because they have different views (or models) about how the economy operates. In terms of inflation forecasting, for example, a set of individual forecasters may all have in mind a Phillips curve model whereby inflation responds to the level of "slack" in the economy, but might hold different views about the degree to which inflation responds.[1] This and related issues have been explored by a number of authors, including Casey (2020) and Clements (2023) *inter alia*, who fit Phillips curve-type models to the forecasts of survey respondents at the individual level, to determine whether the respondents' forecasts are consistent with "similar" Phillips curve models.

The Phillips curve was originally proposed by Phillips (1958) as an inverse relationship between UK wage inflation and unemployment, but following Calvo (1983) has emerged as a "micro-founded" relationship between inflation, expectations of future inflation, and a measure of slack: the so-called expectations-augmented Phillips curve.[2]

In this paper we investigate whether functional data analysis (FDA) can be used to include the 'forecasts of others' in individuals' Phillips curve models. This is motivated by the belief that forecasters are likely to be influenced by their fellow forecasters, and that FDA in principle should prove a useful tool. Given the relatively small numbers of forecasts typically available at the level of the individual forecaster, our baseline model is relatively simple: we estimate simple, constant-parameter linear Phillips curve models. Nevertheless, there is evidence that the actual Phillips curve (between actual inflation and a measure of slack in the economy, as opposed to individuals' forecasts of these quantities) may have flattened over time, and that the relationship may be non-linear (see, e.g., Hooper et al. (2019)). Indeed, for forecasts, Fendel et al. (2011) find the slope of the Phillips curve depends of the state of the business cycle. We explore possible dependence of the phase of the business cycle in what follows.

Whatever the Phillips curve model, we show that FDA can be used to allow for the possibility that an individual's forecasts might be influenced by the forecasts of others, and in particular, by the forecasts of other respondents which are known to the forecaster when they make their forecasts. We surmise that respondents might respond to different aspects of the forecasts of others to different degrees, and that this might itself constitute a source of forecaster disagreement.

[1] Other papers consider the theory-consistency of agents' expectations with a relationship between relative consumption growth and real exchange rate depreciations across countries (see Backus and Smith (1993)), or whether consumers' spending decisions are consistent with an Euler equation (Dräger and Nghiem (2021)), *inter alia*.

[2] See Gali and Gertler (1999) and Coibion et al. (2018) for a review of its historical development.

To make this concrete, consider Clements (2023), who estimates a Phillips curve model for each respondent j using j's *forecasts*, of the form:

$$E_{j,t}\pi_{t+h} = \beta_j E_{j,t}\pi_{t+h+1} + \alpha_j E_{j,t}u_{t+h} + e_{j,t}, \tag{1}$$

where $E_{j,t}\pi_{t+h}$ are j's h-step ahead forecasts of π_{t+h} made at time t, where π is the quarterly inflation rate, and $E_{j,t}u_{t+h}$ is the forecast of the unemployment rate (u). Thus, (1) is a relationship that would be expected to hold (subject to a disturbance term $e_{j,t}$) between individual j's forecasts of inflation (for periods $t+h$, and $t+h+1$) and the unemployment rate, all made at time t, if j's forecasts are consistent with a Phillips curve-type relationship.

In addition to the range of Phillips curve models that might guide expectations, there is a large literature suggesting that forecasters pay attention to the views of others for a variety of reasons (see, e.g., Clements (2018)). This is usually allowed for in the macro-forecasting literature by including the consensus or average of the (appropriately dated) cross-section of individuals' forecasts. The main aim of our paper is to show that there are alternatives to the consensus as ways of including others' forecasts, and that these alternatives may give rise to different findings, as well as yielding additional insights into macro-forecaster behavior.

The proposal is to extend (1) to allow for the forecasts of others in a flexible way. That is, we allow j's forecasts to be driven by an aggregator function of other individuals' forecasts (made at time $t-1$, so that these are known to j at time t). Because we do not know how professional forecasters use the information contained in the forecasts of others, we proceed by including the information in the entire distribution of all the individuals' time $t-1$ forecasts. Thus, in addition to including $E_{j,t}\pi_{t+h+1}$ on the right-hand-side (as suggested by the expectations-augmented Phillips curve applied to an individual's forecasts), we also include $\int \gamma_j(x)dF_{t+h+1|t-1}(x)$ in (1), which gives us:

$$E_{j,t}\pi_{t+h} = \beta_j E_{j,t}\pi_{t+h+1} + \int \gamma_j(x)dF_{t+h+1|t-1}(x) + \alpha_j E_{j,t}u_{t+h} + e_{j,t}, \tag{2}$$

where $F_{t+h+1|t-1}$ is the distribution of individuals' forecasts made at time $t-1$ for time $t+h+1$, and $\gamma_j(\cdot)$ is the coefficient function which serves as the flexible aggregator for the individual j. Notice that $\gamma(\cdot)$ is subscripted by j: $\gamma_j(\cdot)$. This allows individual j's forecasts to be influenced by the forecasts of others in a way which is peculiar to that individual.

The challenges to standard methods of model estimation and inference in (2) resulting from the inclusion of the distribution of individuals' forecasts can be overcome by the use of FDA. FDA deals with problems where one of the variables in the analysis can be naturally viewed as a smooth curve or function (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012). That is, FDA can be thought of as the statistical analysis of samples of curves, possibly combined with the quantities of standard analysis - vectors and scalars. FDA has witnessed rapid development over the last two decades. While the central ideas and methods of FDA have achieved a certain maturity, applications of FDA proceed at an accelerating speed. Economics and finance have benefited from the use of tools from FDA. Some of the prominent examples in economics and finance include: intraday price curves (Kokoszka et al., 2015), the term structure of interest rates (Bardsley et al., 2017; Horváth et al., 2022), forward curves of commodity futures (Horváth et al., 2020), and price signatures (Oomen, 2019). In our case, the distribution of individuals' forecasts can be characterized as a function. We will show how tools developed in FDA can be applied to make statistical inference on (2).

In terms of the use of FDA to extract information from the cross-section of forecasts, Meeks and Monti (2019) is similar to our study. They relate distributions of inflation expectations from surveys of professional forecasters to *actual* inflation. They apply functional principal component analysis (FPCA) to the distributions of inflation expectations and find that the first three principal components (PCs) can be interpreted as disagreement, skew, and "shape". They then estimate Phillips curve models of actual inflation with forward-looking components comprising these PCs, as a way of taking into account the individual-level heterogeneity. Their approach is founded within the literature that uses survey expectations as a source of expectations that is external to the model itself, obviating the need to instrument future values of variables.[3] Our study is different because we study how professional survey respondents' forecasts of inflation are affected by the forecasts of others. To the best of our knowledge, our study is the first to apply FDA to the analysis of the "theory consistency" of macro-expectations, where FDA is employed to capture potential interactions between forecasters.[4]

The plan of the rest of the paper is as follows. Section 2 describes the forecast data we use, and how and why the empirical cross-sectional inflation forecast distributions are transformed to cross-sectional log quantile densities (LQDs) of the inflation forecasts. Section 3 explains how we apply FDA to the estimation of the "Phillips curve models", of the form of (2), for each respondent j, and discusses the findings. Section 4 then explores the robustness of our findings to some of the modeling choices we have made, and to alternative formulations. Section 5 offers some concluding remarks, as well as some potential avenues for future research. We believe the methods we discuss and illustrate might be fruitfully used in many other contexts relating to survey expectations.

[3]See, for example, McCallum (1976) and Coibion et al. (2018), albeit that the prior literature simply used aggregate expectations.

[4]See Clements (2023) for a discussion of the literature on investigating whether macro forecasts are consistent with various macro-theories, including the Phillips curve.

2 Descriptive analysis

The forecast data is from the US Survey of Professional Forecasters (SPF), the longest-running survey of macro expectations of professionals.[5] The survey began in 1968:Q4, and has been held quarterly ever since. The individual-level data are made available on the website of the Federal Reserve Bank of Philadelphia. Extensive use of the SPF data has been made by academic researchers. For our purposes, it is important that the respondent to a given survey will know the responses made to the survey in the previous quarter, as well as the fact that individuals can be tracked through time even though they remain anonymous. We use the individual expectations from 1981:Q3 - 2022:Q2, inclusive, for:

- CPI Inflation Rate (CPI) Headline: annualized percentage points, seasonally adjusted, based on quarterly average index level.

- Civilian Unemployment Rate (UNEMP): percentage points, seasonally adjusted, quarterly average.

The cross-sectional distributions use the forecasts from all the respondents, some of whom make only a small number of returns, although we only estimate Phillips curve models for respondents who make more than a minimum number of forecasts.

2.1 Properties of the Forecasts

Table 1 presents the summary statistics of the SPF CPI and unemployment forecasts. Individuals give their forecasts of the current quarter - the quarter of the survey (these are denoted $h = 0$), of the next quarter ($h = 1$) and of the next 3 quarters. The table pools the forecasts across individuals and survey quarters. It is evident that the standard deviation of the CPI forecasts is greatest at $h = 0$, whereas the reverse is true for the unemployment rate forecasts.

2.2 Data transformation

We begin by plotting the cross-sectional distributions of the inflation forecasts for each quarter in the sample (1981:Q3 - 2022:Q2) for $h = 1$, where the density functions have been estimated using a kernel-type estimator: see the upper panel of Figure 1.

However, density functions satisfy two constraints - strict non-negativity, and integrating to one - which complicate their use in FDA.[6] The literature on FDA has devoted much

[5] Clements et al. (2022) provide a review of surveys of professional forecasters. On the U.S. SPF, see Croushore (1993) and Croushore and Stark (2001).

[6] Density functions can be viewed as elements of a Hilbert space, but they do not constitute a *linear*

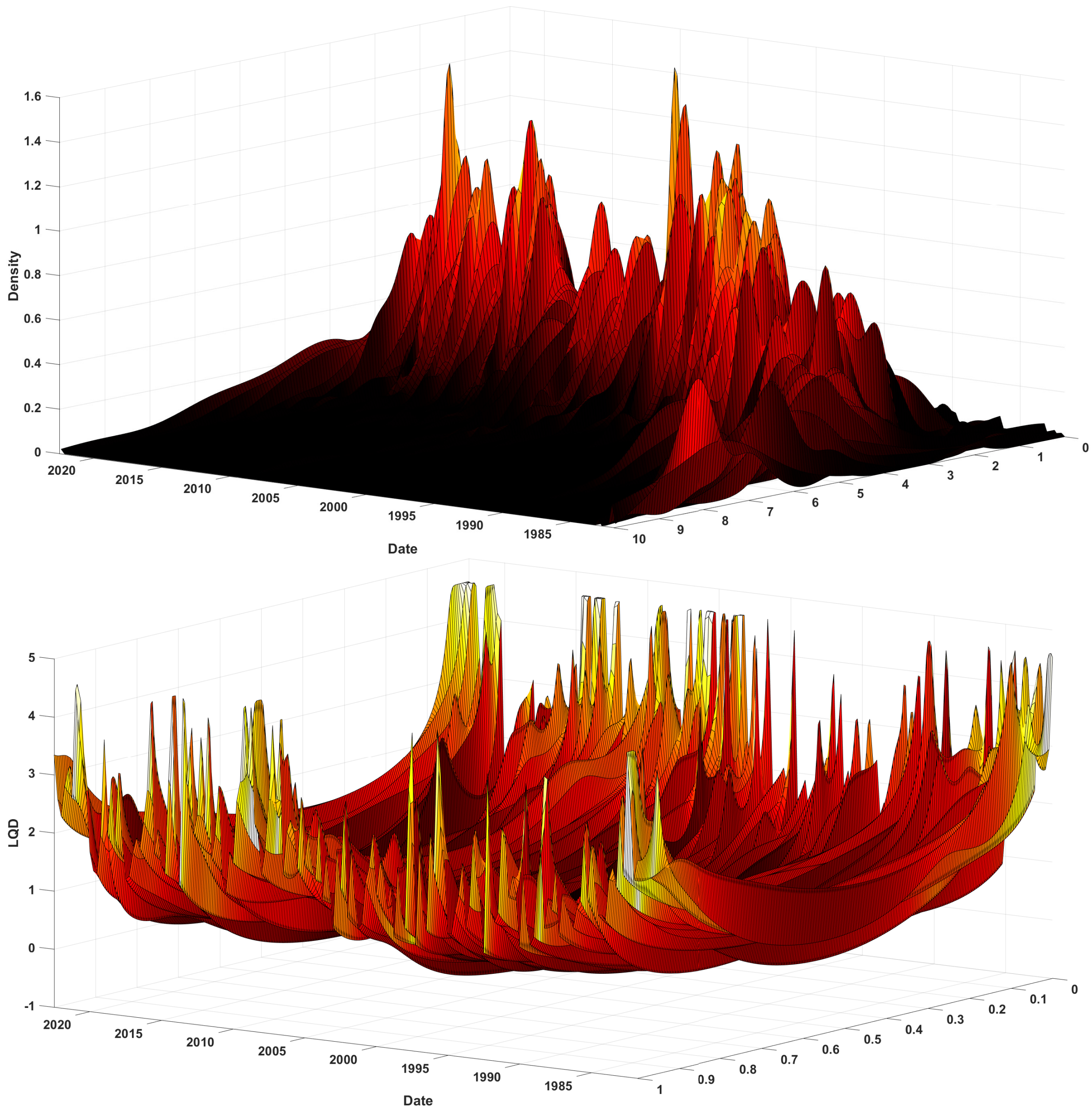

Figure 1: Upper Panel: Cross-sectional densities of inflation forecasts for $h = 1$; Lower Panel: Cross-sectional LQD of inflation forecasts for $h = 1$

Table 1: Summary statistics of CPI and unemployment forecasts

	CPI								
	Mean	SD	Skewness	Kurtosis	10% Q	25% Q	50% Q	75% Q	90% Q
$h = 0$	2.839	0.907	-0.202	3.875	1.822	2.346	2.855	3.388	3.848
$h = 1$	2.866	0.802	-0.306	4.492	2.014	2.488	2.893	3.298	3.725
$h = 2$	2.931	0.748	-0.223	4.480	2.156	2.596	2.943	3.324	3.723
$h = 3$	2.992	0.744	-0.072	4.582	2.218	2.639	2.998	3.367	3.797
$h = 4$	3.047	0.754	-0.011	4.509	2.252	2.685	3.051	3.422	3.859
	Unemployment								
	Mean	SD	Skewness	Kurtosis	10% Q	25% Q	50% Q	75% Q	90% Q
$h = 0$	6.217	0.142	0.211	3.925	6.064	6.142	6.212	6.294	6.376
$h = 1$	6.180	0.226	0.088	3.388	5.924	6.051	6.175	6.297	6.446
$h = 2$	6.125	0.287	0.066	3.470	5.801	5.947	6.123	6.283	6.463
$h = 3$	6.067	0.338	0.084	3.549	5.690	5.865	6.062	6.254	6.464
$h = 4$	6.009	0.383	0.112	3.599	5.585	5.781	6.000	6.225	6.456

effort to how best to deal with density curves. An influential paper is Petersen and Müller (2016). Their main idea is to suitably transform the densities (constrained curves) into unconstrained curves, so that one can embed the densities into the L^2 space. Then, standard methods of FDA, such as FPCA and functional regression (detailed below), can be applied to the transformed curves in the L^2 space. If all the densities share a common known support $\mathscr{T}$, the representations of the density curves themselves can be obtained by applying the inverse transform to the unconstrained curves in the L^2 space. Petersen and Müller (2016) recommend two choices for the functional transformation, the *log hazard transform* and the *log quantile density transformation*. The *log quantile density transformation* has become the more popular choice (and has been used in e.g., bioscience (Petersen et al., 2019) and engineering (Chen et al., 2019)), and we use this in our application.

Briefly, the *log quantile density transformation* (LQD) transformation works as follows. A cross-sectional distribution at time t is commonly represented in one of three ways: the cumulative distribution function (CDF) F_t, the probability density function (PDF) f_t, and the quantile function Q_t. The three representations have the following relationship

$$F_t(x) = \int_{-\infty}^{x} f_t(u)du, \quad -\infty < x < \infty,$$
$$Q_t(s) = F_t^{-1}(s) = \inf\{y \in (0,1) : F_t(y) \geq s\}, \quad s \in (0,1).$$

subspace. If the densities are treated as elements of L^2 space, then linear combinations of densities will be elements of L^2, but will not necessarily be densities. Thus, it is not appropriate to directly apply conventional FDA approaches to density curves.

A given distribution is uniquely characterized by any of the three functions, F_t, f_t, and Q_t, each of which may be used as a functional data object. Nevertheless, all three are subject to nonlinear constraints. Specifically, the CDF is constrained by $0 \leq F_t(x) \leq 1$ and $F_t'(x) \geq 0$; the PDF by the constraints $f_t > 0$; while the quantile function is the least constrained, requiring only that $Q_t'(x) \geq 0$. Unfortunately, the presence of the nonlinear constraints complicates the application of typical linear functional data methods, as noted above. To enable the application of FDA, we consider ways to characterize a distribution with fewer or no constraints. Jones (1992) proposes the so-called quantile density, $q_t = Q_t'$, which is only constrained to be nonnegative. The definition of the quantile density implies:

$$q_t(s) = Q_t'(s) = (F_t^{-1})'(s) = \frac{1}{F_t'(F_t^{-1}(s))} = \frac{1}{f_t(Q_t(s))}, \quad s \in (0,1).$$

Then, an unconstrained representation for a distribution is obtained by taking the log of the quantile density: this gives the LQD

$$X_t(s) = \log(q_t(s)) = -\log(f_t(Q_t(s))), \quad s \in (0,1). \tag{3}$$

One drawback of Petersen and Müller (2016) is that all densities are assumed to have the same support to enable the LQD in the L^2 space to be mapped back to the density space. However, this assumption is sometimes too restrictive, and fails to hold in some applications. In a follow-up study, Kokoszka et al. (2019) proposed a modified LQD transformation that allows for densities with different supports to be transformed to functions on a common domain. The only requirement for their modified LQD transformation is that the intersection of all the densities' supports contains a known point.

In our study, the densities of professional survey respondents' forecasts of inflation at different times t do not have the same support: the minimum and maximum of the inflation forecasts are not the same across quarters. Nevertheless, we can use the Petersen and Müller (2016) version of the LQD transformation, rather than the modified version. This is because we are interested in how the distribution of individuals' forecasts may affect a particular respondent, and we do not need to invert LQD back to the density space. Thus, the inversion step of Petersen and Müller (2016) is not needed in our study.[7]

Another consideration is whether or not to remove outliers. Technically, it becomes necessary to remove outliers when working with the LQD. This is because LQD can go to infinity if any $s \in (0,1)$ gives $f_t(Q_t(s)) = 0$. Keeping extreme outliers can bring this troublesome issue to the fore. For this reason, outliers at the 1% level in each quarter t for

[7] However, the Kokoszka et al. (2019) modified LQD transformation would not be applicable in any case because the intersection of all densities' supports does not contain a known point.

a given h are removed.[8]

In order to evaluate the stationarity of the series of LQD, we applied the functional KPSS test[9] developed in Horváth et al. (2014) separately for $h = 0, 1, 2, 3, 4$. The p-value of the null hypothesis of stationarity is as follows: 0.241 ($h = 0$), 0.451 ($h = 1$), 0.417 ($h = 2$), 0.258 ($h = 3$), and 0.140 ($h = 4$). The results suggest that it is reasonable to assume the series of LQD are stationary, which provides the justification for the functional regression model.

A visualization of the cross-sectional LQD of the inflation forecasts at each quarter of our sample period for $h = 1$ is provided in the lower panel of Figure 1. The LQD is typically U-shaped due to the negative sign in its construction (3). Most of the LQD curves are relatively smooth, but some occasionally have small wiggles. Importantly, we can observe that the LQD is always in the domain of $s \in (0, 1)$, which facilitates the functional data analysis. We stress that we work with the LQD, rather than the densities, in the rest of the study.

3 Functional data analysis

In this section we describe our application of FDA to the SPF data. There are two steps. The first is the use of FPCA to capture the main patterns in the variability of the distribution of professionals' forecasts. The second step uses functional linear regression to determine whether an individual's forecasts are affected by the forecasts of the other respondents, and which aspects of the distributions of the forecasts of the other respondents are relevant. One might suppose that the "consensus" view would be important - that is the mean of the distribution of the forecasts of the other respondents. Although we find a role for the mean, FDA suggests the findings are more nuanced.

We begin with the basic model. The different forecast horizons reported in the SPF allow Model (2) to be estimated for left-hand-side forecasts of $h = 0, 1, 2, 3$.[10] Model (2)

[8]We provide a robustness check of removing outliers at 5% in the Supplement.

[9]The specific settings we used are the same as Horváth et al. (2019); see their Appendix A for details.

[10]Technically, $h = -1$ could also be estimated. However, there is little cross-sectional variation in $E_{j,t}\pi_{t-1}$. This is because the advance estimate of the previous period has already been officially released, and is used by most respondents. Thus we do not include $h = -1$ in the analysis. Note that $h = 0$ means a "forecast" for the current quarter, which is typically termed as nowcasting.

can be written in full for each h as:

$$\begin{aligned}
E_{j,t}\pi_t &= \zeta_{j,0} + \beta_{j,0}E_{j,t}\pi_{t+1} + \int \gamma_{j,0}(s)X_{t+1|t-1}(s)ds + \alpha_{j,0}E_{j,t}u_t + e_{j,t,0},\\
E_{j,t}\pi_{t+1} &= \zeta_{j,1} + \beta_{j,1}E_{j,t}\pi_{t+2} + \int \gamma_{j,1}(s)X_{t+2|t-1}(s)ds + \alpha_{j,1}E_{j,t}u_{t+1} + e_{j,t,1},\\
E_{j,t}\pi_{t+2} &= \zeta_{j,2} + \beta_{j,2}E_{j,t}\pi_{t+3} + \int \gamma_{j,2}(s)X_{t+3|t-1}(s)ds + \alpha_{j,2}E_{j,t}u_{t+2} + e_{j,t,2},\\
E_{j,t}\pi_{t+3} &= \zeta_{j,3} + \beta_{j,3}E_{j,t}\pi_{t+4} + \int \gamma_{j,3}(s)X_{t+4|t-1}(s)ds + \alpha_{j,3}E_{j,t}u_{t+3} + e_{j,t,3}, \quad s \in (0,1),
\end{aligned} \tag{4}$$

where constant terms have been added, and $X_{t+h+1|t-1}$ is the LQD of the individual forecasts made at time $t-1$ of the value of inflation at time $t+h+1$. As written, the model parameters vary across respondent j and horizon h. Following Jain (2019) and Clements (2023), we assume the parameters are the same across h for a given respondent, giving rise to more precise parameter estimates. Doing so gives the "pooled" version of Model (2), written as:

$$E_{j,t}\pi_{t+h} = \zeta_j + \beta_j E_{j,t}\pi_{t+h+1} + \int \gamma_j(s)X_{t+h+1|t-1}(s)ds + \alpha_j E_{j,t}u_{t+h} + e_{j,t}, \text{ for } h = 0,1,2,3. \tag{5}$$

In the Supplement we show the results of allowing for horizon fixed effects. Generally doing so makes little differences, justifying their omission in what follows.

We could further restrict the heterogeneity in the parameters by assuming the "slope" parameters are the same across respondents in either (4) or (5) (and possibly using a fixed-effects panel data estimator). We prefer not to impose slope homogeneity across j, so do not use a panel estimator. Both Jain (2019) and Clements (2023) report considerable heterogeneity across j: Jain (2019) reports heterogeneity in respondents' beliefs regarding inflation persistence, and Clements (2023) considers the extent to which heterogeneity in individuals' beliefs in the Phillips curve reflects differing economic conditions when they were active survey participants.

The first step of the functional principal component regression applies FPCA to the $X_{t+h+1|t-1}$, to obtain empirical functional principal components (EFPC's) and their corresponding scores. The second step replaces the term in $X_{t+h+1|t-1}$ in (4) and (5) by the scores, recasting the scalar-on-function regression in a multiple regression framework.

3.1 Functional principal component analysis

Given that FPCA may not be widely known, and to make our study self-contained, we provide a brief summary in the context of our study.

At time t, all the respondents' forecasts are collected, and the corresponding LQD of the cross-sectional distribution is obtained, and denoted as $X_t(s), s \in (0,1)$. $X_t(s)$ is assumed to be square integrable with the condition $\mathbb{E}\int X_t(s)ds < \infty$, and thus it can regarded as a functional observation in the L^2 space. The sample $\{X_1(s), X_2(s), ..., X_T(s)\}$ is a sequence of functional observations, and can be viewed as the realizations of a random function $X(s)$. We can define the population mean and covariances of the random function $X(s)$ as:

$$\begin{aligned}\mu(s) &= \mathbb{E}X(s),\\ c(s,s') &= \mathbb{E}\left[(X(s)-\mu(s))\left(X(s')-\mu(s')\right)\right].\end{aligned}$$

Based on Mercer's lemma (Riesz et al., 1990), there exists an *orthonormal* sequence of continuous functions $v_k(s)$ in the sense that:

$$\int v_k(s)v_\ell(s)dt = \begin{cases} 0 & \text{if } k \neq \ell, \\ 1 & \text{if } k = \ell, \end{cases}$$

and a non-increasing sequence λ_k of positive numbers, such that:

$$c(s,s') = \sum_{k=1}^{\infty} \lambda_k v_k(s) v_k(s').$$

Typically, we obtain $v_k(s)$ as the eigenfunctions of the covariance function c, which are also called functional principal components (FPCs), and the λ_k are the eigenvalues of c, in non-increasing order $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. By the Karhunen-Loève expansion (Karhunen, 1947; Loève, 1960), one can decompose the random function $X(s)$ as:

$$X(s) = \mu(s) + \sum_{k=1}^{\infty} \xi_k v_k(s),$$

The random variables ξ_i are called the (principal component) *scores*, which are given by the projection of $(X(s) - \mu(s))$ in the direction of the k-th eigenfunction $v_k(s)$

$$\xi_k = \int (X(s) - \mu(s))\, v_k(s) ds.$$

It can be shown that

$$\mathbb{E}\xi_k = 0, \qquad \mathbb{E}\xi^2 = \lambda_k, \qquad \mathrm{Cov}(\xi_k, \xi_\ell) = 0, \text{ if } k \neq \ell,$$

and

$$\mathbb{E}\int (X(s) - \mu(s))^2\, ds = \sum_{k=1}^{\infty} \lambda_k.$$

Thus, it follows that λ_k is the variance of random function $X(s)$ in the direction of the k-th eigenfunction $v_k(s)$, and the sum of all λ_k is the total variance of $X(s)$.

In terms of estimation, the estimators of the population mean and covariance functions are the sample mean and covariance functions, defined by

$$\hat{\mu}(s) = \frac{1}{N}\sum_{t=1}^{T} X_t(s),$$
$$\hat{c}(s,s') = \frac{1}{N}\sum_{t=1}^{T} \left[\left(X_t(s) - \hat{\mu}(s)\right)\left(X_t(s') - \hat{\mu}(s')\right)\right].$$

The eigenfunctions of the sample covariance function $\hat{c}$ are the empirical functional principal components (EFPC's):

$$\hat{c}(s,s') = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{v}_k(s)\hat{v}_k(s'),$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq 0$ are the sample eigenvalues of $\hat{c}(s,s')$. Then via the Karhunen-Loève expansion, the functional observations $\{X_1(s), ..., X_T(s)\}$ can be decomposed as:

$$\begin{aligned} X_t(s) &= \hat{\mu}(s) + \sum_{k=1}^{\infty} \hat{\xi}_{k,t}\hat{v}_k(s), \\ &= \hat{\mu}(s) + \sum_{k=1}^{K} \hat{\xi}_{k,t}\hat{v}_k(s) + e_t(s), \qquad t = 1, 2, ..., T \end{aligned} \tag{6}$$

where $\hat{\xi}_{k,t} = \int \left(X_t(s) - \hat{\mu}(s)\right)\hat{v}_k(s)ds$ are the estimated (principal component) *scores* in the direction of the k-th eigenfunction for the t-th observation. When $K < T$ denotes the number of retained EFPCs, the $e_t(s)$ represents the truncation error function, and has a zero mean and finite variance.

Equation (6) illustrates the dimension reduction from infinity to a finite number K. This requires that the functional observation $X_t(s)$ can be well approximated from the first K EFPCs: that is, the information contained in functional observations $\{X_1(s), ..., X_T(s)\}$ can be effectively summarized by the finite-dimensional scores $\left\{\hat{\boldsymbol{\xi}}_1, ..., \hat{\boldsymbol{\xi}}_T\right\}$, where $\hat{\boldsymbol{\xi}}_t = \left(\hat{\xi}_{1,t}, ..., \hat{\xi}_{K,t}\right)^\top$. These scores have zero mean and variance $\hat{\lambda}_k$, and they are uncorrelated with each other.

A practical question is how to choose K. Several criteria have been proposed: eigenvalue ratio tests (Ahn and Horenstein, 2013), cross-validation (Ramsay and Silverman, 2005), Akaike's information criterion (Akaike, 1974), the bootstrap method (Hall and Vial, 2006), and the cumulative percentage of variance (CPV) explained (Horváth and Kokoszka, 2012).

CPV is commonly used (Kearney and Shang, 2020; Bouri et al., 2021; Shang and Kearney, 2022). The CPV explained by the first K EFPC's is computed by:

$$\mathrm{CPV}(K) = \frac{\sum_{k=1}^{K} \hat{\lambda}_k}{\sum_{k=1}^{T} \hat{\lambda}_k}.$$

We use the CPV method to select the number of K, and choose K such that at least 80% of the total variance is explained.

Empirical results of FPCA

Figure 2 shows the variance explained by each of the principal components, and the cumulative variance, for the pooled (over h) data. Three EFPC's (i.e. $K = 3$) are sufficient to reach the threshold of explaining 80% of the total variance.

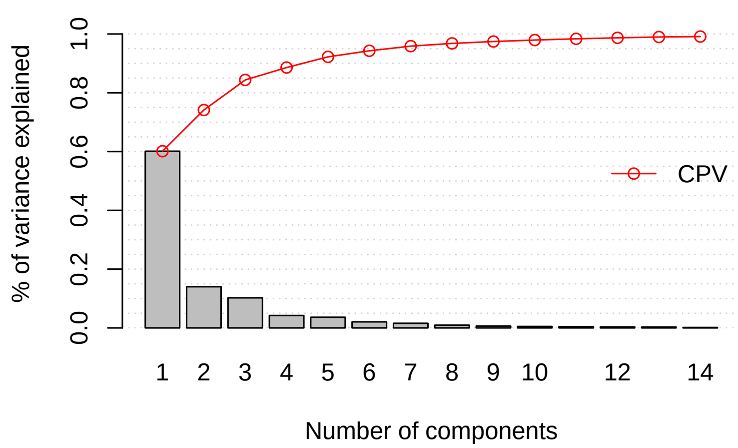

Figure 2: CPV explained by the principal components for the pooled data.

Figure 3 shows the EFPC's for the pooled (over h) cross-sectional distributions, and Figure 4 displays their corresponding scores for different horizons. The scores in Figure 4 are calculated with respect to the common EFPC's obtained from the pooled data.

A challenge often encountered in FDA is how to interpret the FPCA, and in particular, the EFPCs. (Similar issues arise of course in standard principal component analysis). The EFPCs are orthonormal to one another, but may not be readily interpretable (Kokoszka and Reimherr, 2017). The conventional way of interpreting the scores is to consider their correlation (linear or rank-based) with other known factors or quantities. A relatively high correlation is generally taken to suppose a tentative interpretation of the score in terms of that factor or quantity.

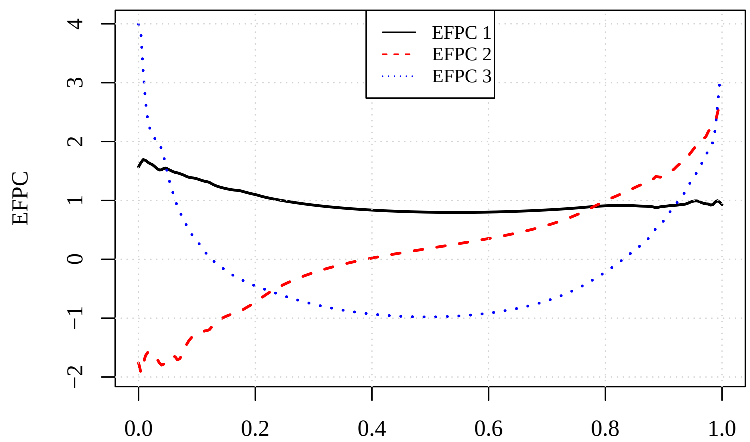

Figure 3: EFPC's for the pooled cross-sectional distributions.

Table 2 shows the estimates of Pearson's correlation and Spearman's rho between the first three estimated scores $\left(\hat{\xi}_{1,t}, \hat{\xi}_{2,t}, \hat{\xi}_{3,t}\right)$ and moments and descriptive statistics of the cross-sectional distributions[11], viz. the mean, median, standard deviation (SD), interquartile range (IQR), skewness, and kurtosis. As can be observed, the first score $\hat{\xi}_{1,t}$ has a very high Pearson's correlation (0.94) with the SD measure of disagreement, and a little lower correlation of 0.83 with the IQR measure. The correlation between the second score $\hat{\xi}_{2,t}$ and the skewness is relatively high at 0.80. The third score $\hat{\xi}_{2,t}$ is more moderately correlated with the kurtosis (0.60). Thus, we conjecture that the three scores are related to the disagreement, skewness, and kurtosis.

In contrast to Pearson's correlation which evaluates linear correlation, Spearman's rho evaluates monotonic relationships (whether linear or not). The Spearman's rho between $\hat{\xi}_{1,t}$ and SD is 0.96, between $\hat{\xi}_{2,t}$ and skewness is 0.86, and between $\hat{\xi}_{3,t}$ and kurtosis is 0.75. These are all higher than the corresponding linear correlations, and are indicative of nonlinear relationships between the scores and the moments (SD, skewness, and kurtosis). Figure 5 display the scatter plots between the scores and the moments, and the nonlinearity can be clearly observed. The evidence supports the conjecture that the three scores measure disagreement, skewness, and kurtosis in a *nonlinear* manner.

[11] We did not consider the *shape* factor used by Meeks and Monti (2019). This is because the calculation of the shape factor involves the pointwise time average distribution, which may not necessarily be in the density space, as described in Section 2.2.

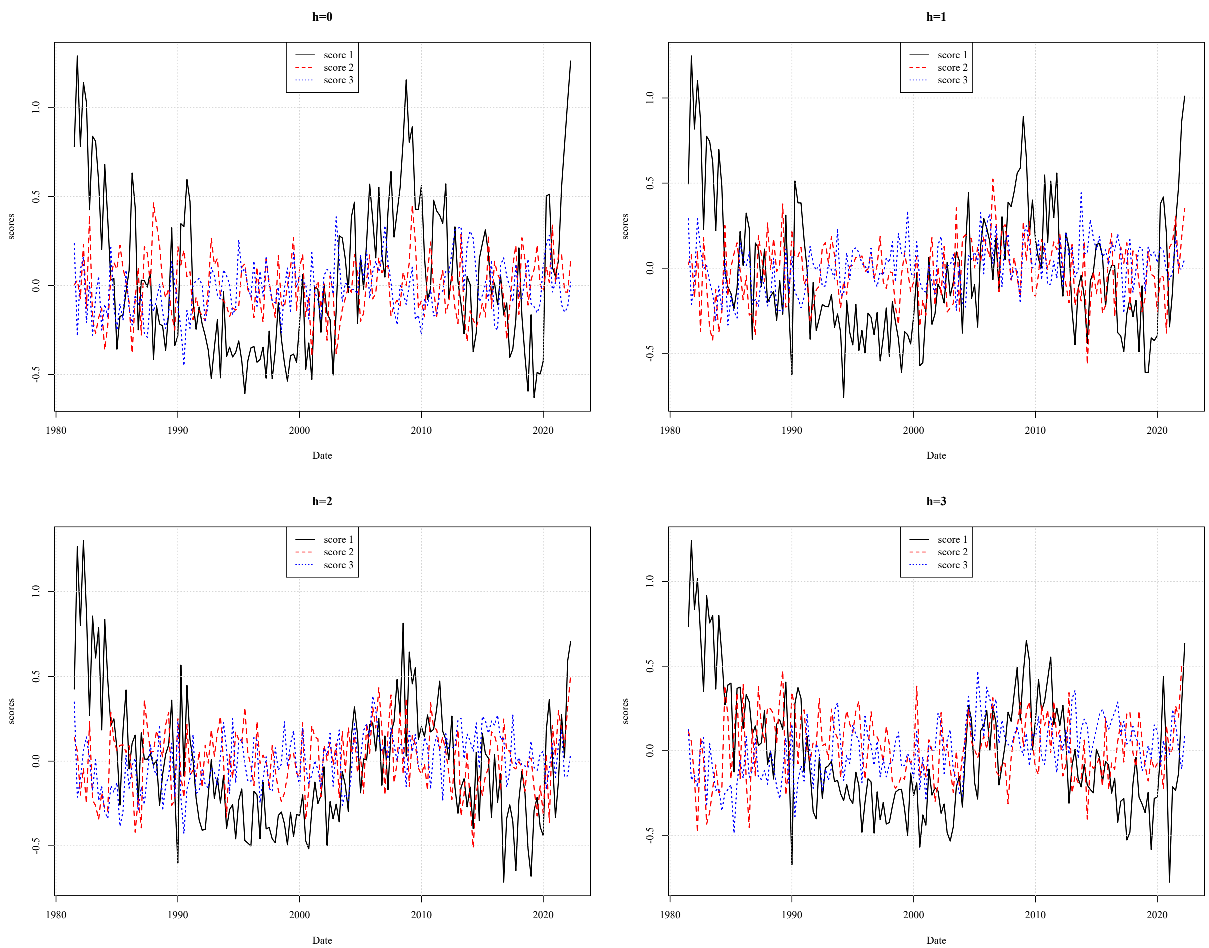

Figure 4: Scores for the pooled cross-sectional distributions.

3.2 Functional regression model

Model (5) admits a functional regression model in the category of "scalar-on-function", because one of the regressors is a "curve" and the response variable is scalar. Without imposing any regularity conditions, such a regression is infeasible due to the infinite dimension of the coefficient function $\gamma_j(\cdot)$. Reiss et al. (2017) provide a comprehensive review of the main methods for scalar-on-function regression. As they note, the main idea is to expand $\gamma_j(\cdot)$ in terms of a set of basis functions. The two types of basis functions are: 1) prior fixed bases, such as the B-spline basis and the Fourier basis; and 2) data-driven bases, often derived by FPCA or functional partial least squares. If the basis is from FPCA, it is known as functional principal component regression (FPCR), which is widely used in the literature (e.g. Meeks and Monti, 2019; Cao et al., 2020). We choose to use FPCR because of its simplicity and interpretability (as discussed in Section 3.1).

Table 2: Correlation between EFPC's and moments

	Pearson's Correlation			Spearman's rho		
	Score 1	Score 2	Score 3	Score 1	Score 2	Score 3
Mean	0.32	0.10	-0.32	0.08	0.15	-0.33
Median	0.35	0.02	-0.32	0.09	0.06	-0.33
SD	**0.94**	0.00	-0.05	**0.96**	0.09	0.01
IQR	0.83	0.07	-0.48	0.78	0.12	-0.49
Skewness	-0.29	**0.80**	-0.06	-0.26	**0.86**	-0.05
Kurtosis	0.30	-0.06	**0.60**	0.38	-0.05	**0.75**

Note: SD stands for standard deviation, and IQR stands for interquartile range.

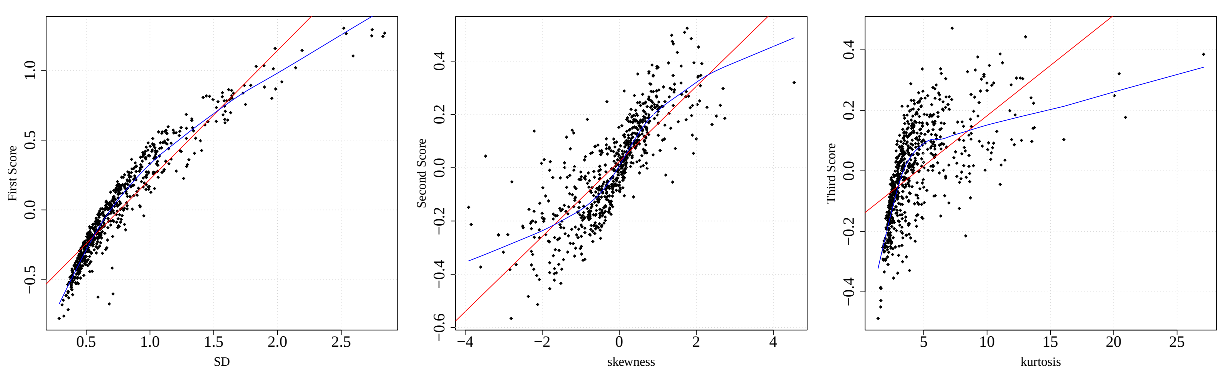

Figure 5: Evidence of nonlinearity between the scores and the moments.

Here we briefly summarize FPCR. According to (6), we can have the approximation:

$$X_t(s) \approx \hat{\mu}(s) + \sum_{k=1}^{K} \hat{\xi}_{k,t}\hat{v}_k(s), \qquad \text{where} \;\; \hat{\xi}_{k,t} = \int \left(X_t(s) - \hat{\mu}(s)\right)\hat{v}_k(s)ds.$$

Plugging this approximation into (5) gives:

$$\begin{aligned} E_{j,t}\pi_{t+h} &= \zeta_j + \beta_j E_{j,t}\pi_{t+h+1} + \int \gamma_j(s)\left(\hat{\mu}(s) + \sum_{k=1}^{K}\hat{\xi}_{k,t}\hat{v}_k(s)\right)ds + \alpha_j E_{j,t}u_{t+h} + e_{j,t}, \\ &= \delta_j + \beta_j E_{j,t}\pi_{t+h+1} + \sum_{k=1}^{K}\hat{\xi}_{k,t}\gamma_{j,k} + \alpha_j E_{j,t}u_{t+h} + e_{j,t} \end{aligned} \tag{7}$$

where:

$$\delta_j = \zeta_j + \int \gamma_j(s)\hat{\mu}(s)ds, \qquad \text{and} \;\; \gamma_{j,k} = \int \gamma_j(s)\hat{v}_k(s)ds,$$

and α_j and β_j are regarded as unknown (scalar) parameters. Now, the functional regression model (5) is recast as a multiple regression with the scores replacing the curve. Then define the $T \times (K+3)$ matrix:

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & E_{j,1}\pi_{1+h} & \hat{\xi}_{1,1} & \hat{\xi}_{2,1} & \cdots & \hat{\xi}_{K,1} & E_{j,1}u_{1+h} \\ 1 & E_{j,2}\pi_{2+h} & \hat{\xi}_{1,2} & \hat{\xi}_{2,2} & \cdots & \hat{\xi}_{K,1} & E_{j,2}u_{2+h} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & E_{j,T}\pi_{T+h} & \hat{\xi}_{1,T} & \hat{\xi}_{2,T} & \cdots & \hat{\xi}_{K,T} & E_{j,T}u_{T+h} \end{bmatrix},$$

so that the parameter vector $\boldsymbol{\theta} = [\delta_j, \beta_j, \gamma_{j,1}, ..., \gamma_{j,K}, \alpha_j]^\top$ can be estimated by least squares multiple regression of $\boldsymbol{Y} = \boldsymbol{\Omega\theta} + \boldsymbol{e}$. If we denote the estimated parameters as $\hat{\delta}_j, \hat{\beta}_j, \hat{\gamma}_{j,1}, ..., \hat{\gamma}_{j,K}, \hat{\alpha}_j$, then the estimates of the parameters in the functional regression model (5) can be obtained as:

$$\hat{\gamma}_j(s) = \sum_{k=1}^{K}\gamma_{j,k}\hat{v}_k(s), \quad \text{and} \quad \hat{\zeta}_j = \hat{\delta}_j - \sum_{k=1}^{K}\hat{\gamma}_{j,k}\int \hat{v}_k(s)\hat{\mu}(s)ds. \tag{8}$$

Empirical results of FPCR

There are in total 256 unique individual respondents in the U.S. SPF dataset for the period we consider.[12] In order to ensure sufficient forecasts to obtain reasonably precise estimates at an individual level, we select only the respondents who made 120 or more forecasts (for all $h = 0, 1, 2, 3$) which is at least 30 forecasts per h on average. This gave 64 respondents. Then we undertook FPCR for each respondent j, pooling over h for that j.

[12]The respondents are anonymous in the U.S. SPF, but each has a unique identifier, allowing an analysis of all the forecasts associated with a given identifier.

To summarize the results, Figure 6 provides a boxplot of the parameter estimates for Model (7) for each respondent, and Table 3 presents the summary statistics of parameter estimates across respondents. We report the cross-sectional mean, standard deviation, lower quartile, median, and upper quartile for each of the estimated parameters. We also report the proportion of respondents for whom we reject the null that the coefficient is zero, at 1%, 5%, and 10% levels. Additionally, we report summary statistics for the adjusted R^2's and number of observations[13].

We summarize our findings as follows. Firstly, as many as a nearly a half of all respondents (45%) appear to use the information in Score 1 when they produce their forecasts. Since Score 1 is a (nonlinear) measure of disagreement, it appears that disagreement at $t-1$ is an important determinant of around one half of the respondents forecasts made at time t (of $t+h$). Forecasters disagreement has often been regarded as a proxy for perceived (or actual) uncertainty about the future outlook, and may be playing a similar role in our analysis (see, e.g., Zarnowitz and Lambros (1987) and Haddow et al. (2013)). For the majority of agents, $\hat{\gamma}_{j,1} > 0$, suggesting that higher inflation disagreement at time $t-1$ results in an increase in j's forecast made at time t. Empirically this seems plausible because inflation rates and inflation variability have often been found to be positively correlated. The interpretation would be that if inflation disagreement last period ($t-1$) was high, the majority of respondents would increase their time t forecasts of inflation in $t+h$ by more than would be suggested by their own forecasts of inflation in $t+h+1$ and the unemployment rate in $t+h$.

Secondly, a modest percentage of all respondents (19% at a 5% significance test level) employ Score 2 in the cross-sectional distribution of all forecasts at $t-1$. Recall that Score 2 nonlinearly measures the skewness in the distribution. Thus, we find evidence that some respondents take skewness into consideration when they generate their forecasts. However, the effect of a change in skewness is less amenable to a behavioural interpretation. The parameter estimates of Score 2 are positive for most agents, albeit not statistically significantly different from zero. In Section 4 we attempt to shed further light on the effects of the skew, by allowing the effects of positive and negative skew to differ.

Thirdly, a small portion of respondents (8% at the 5% level) incorporate the information in Score 3 for their forecasts. As discussed, Score 3 is nonlinearly related to the kurtosis of the distribution.

As described above, FDA provides a number of insights into how respondents' forecasts are influenced by those of others. An overall test of the significance of the FPCA scores in the FPCRs is one way of assessing the importance of applying FDA to the analysis of macro survey expectations. We calculated an F-test of $\gamma_{j,1} = \gamma_{j,2} = \gamma_{j,3} = 0$ for each respondent.

[13]Since we pool all h, the number of observations for one respondent is required to be at least 120.

We rejected the null for nearly half of the respondents at the 5% level. Hence we find that nearly one half of the respondents use information contained in the distribution of other individuals' forecasts.

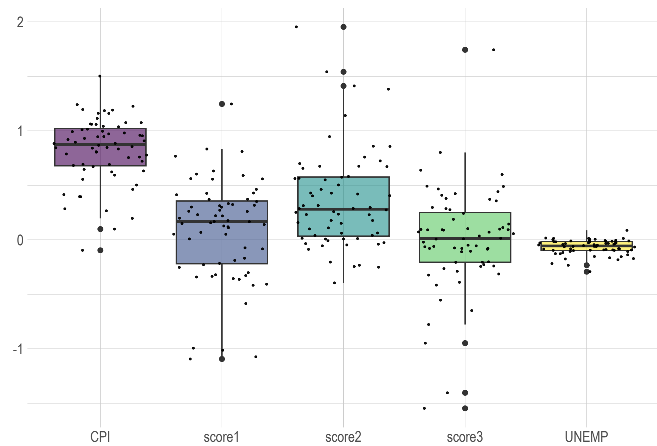

Figure 6: Boxplot of the parameter estimates

Table 3: Summary of parameter estimates of Model (7) across Respondents

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.758	0.749	0.333	0.816	1.073	46.9%	60.9%	75.0%
CPI	0.825	0.300	0.679	0.875	1.021	95.3%	96.9%	96.9%
Score 1	0.092	0.462	-0.220	0.167	0.356	28.1%	46.9%	56.3%
Score 2	0.364	0.454	0.033	0.280	0.576	6.3%	18.8%	28.1%
Score 3	0.000	0.475	-0.206	0.012	0.250	4.7%	7.8%	10.9%
UNEMP	-0.068	0.069	-0.098	-0.057	-0.016	32.8%	39.1%	45.3%
F-test on Scores 1-3						25.0%	45.3%	53.1%
Adj. R^2	46.8%	25.0%	28.5%	46.9%	63.7%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), for each respondent j. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

Characteristics of Respondents

Of interest is whether we are able to identify the characteristics of respondents which are correlated with the tendency to draw on information contained in (the distribution of) the forecasts of others. To investigate this question, we firstly select five characteristics which might potentially be relevant. They include measures of various aspects of the strength of the Phillips curve "imprint" on the forecasts.[14] That is, the extent to which an individual's inflation and unemployment rate forecasts reveal a belief that higher "slack" in the economy puts downward pressure on inflation. The characteristics are:

- $\hat{\alpha}_j$: Respondent j's estimated parameter of the unemployment rate in Model (5),
- $\text{sign}(\hat{\alpha}_j)$: the sign of $\hat{\alpha}_j$,
- $\hat{\beta}_j$: Respondent j's estimated parameter of the inflation rate in Model (5),
- Period_j: the proportion of Respondent j's forecasts made in the earlier two decades (1981:Q3 – 1999:Q4) of the sample period,
- $\tilde{\sigma}_j$: an uncertainty measure of Respondent j's forecasts across $h = 0, 1, 2, 3$.

Hence the first three characteristics are based on the individual's estimates of Model (5), and capture how respondent j believes future inflation is related to the unemployment rate. Following Clements (2023), we include an indicator (Period_j) of whether the respondent was primarily active earlier or later in the period. Higher values of Period_j indicate that the respondent made a larger proportion of their forecasts in the earlier period (1981:Q3 – 1999:Q4). A value of zero indicates that an individual was only active in the 21st century (2000:Q1 – 2022:Q3). We wish to allow for the possibility that the inflation process might have changed over the near 40 year period we consider, with a concomitant change in forecaster behavior. Finally, the fifth characteristic is a measure of the uncertainty of Respondent j's forecasts, calculated by $\tilde{\sigma}_j = \frac{1}{4}\sum_{h=0}^{3}\sqrt{\text{Var}(E_{j,t}\pi_{t+h})}$. This measures the volatility of Respondent j's forecasts.[15]

Table 4 shows summary statistics (the mean, standard deviation, and the number of respondents) of the five respondent-specific characteristics separately for the "significant" and the "insignificant" groups, according to each of Score 1, Score 2, and Score 3 (and jointly). That is, for Score 1 we separate the respondents into the group for whom we reject

[14] Recall that the respondents are anonymous, precluding analysing the effects of *personal* characteristics.

[15] An alternative would be to use the respondents' histogram forecasts to generate measures of perceived uncertainty, as described by Clements et al (2022). Complications would arise in doing so, in particular because the histograms are fixed-event and target the annual year-on-year growth rates, in contrast to the fixed-horizon quarterly point forecasts, but see Ganics et al. (2023) on deriving series of fixed-horizon forecasts from fixed-event forecasts. This is potentially interesting because one might conjecture that an individual's perceptions of uncertainty will vary over time and that more weight would be accorded to the forecasts of others at times of high perceived uncertainty. We do not pursue this here, but we consider whether reliance on others varies across the business cycle (with expansions/contractions being correlated with low/high uncertainty), and in the Supplement we check whether our findings are robust to the inclusion of VIX, as a general measure of uncertainty.

the null that Score 1 is statistically insignificant at the 5% level, and a group comprising those for whom we do not reject. We then consider whether each of the five characteristics differ between the two groups. And then we repeat for Score 2, etc. The last row in each panel of Table 4 presents the p-value of the Welch Two Sample t-test. Based on the Welch Two-Sample t-test, there is no significant difference in the mean of $\hat{\alpha}_j$ and $\tilde{\sigma}_j$. However, we find strong evidence that the mean of sign($\hat{\alpha}_j$) is higher in the "significant group" for Score 2, Score 3, and for the group for which we reject for all the Scores together. This indicates a positive association between believing that higher economic slack reduces inflation pressure *and* paying heed to the forecasts of others (at least as measured by Scores 2 and 3). Put differently, a stronger belief in the Phillips curve (measured by a negative unemployment rate coefficient in the Phillips curve model) is correlated with being influenced by the forecasts of others. It is not obvious why "theory-consistency" and "paying attention to others" should go together, and this merits further investigation.

We find little evidence that $\hat{\beta}_j$ varies across the significant and insignificant groups for the Scores, except for the Score 2 (with a 1.4% p-value for the Two-Sample t-test).

4 Robustness checks

We have made a number of modeling choices in deriving the results in Section 3. In this section we consider the robustness of those findings to various alternative choices and extensions, with a key emphasis on whether the application of FDA provides additional insight into forecaster behavior relative to the use of non-FDA methods. To this end, the first set of checks include the "consensus forecast", i.e., the mean of the cross-sectional distribution forecasts, in the individual Phillips curve models (Section 4.1), and we also include the higher cross-sectional moments (Section 4.2).

Next, we allow for the possibility that a simple linear relationship may not hold over the whole sample period, as suggested by Fendel et al. (2011) and others (albeit that Fendel et al. (2011) consider consensus forecasts as opposed to forecasts at the level of the individual). In Section 4.3 we allow for the possibility that the relationship differs between "normal" and "abnormal" times, where these descriptors are made operational as NBER-dated expansions and contractions. Finally, some authors suggest modelling inflation in "gaps" (see, e.g., Faust and Wright (2013)), and we check the robustness of our findings to adopting such a formulation for respondents' inflation forecasts in Section 4.4.

In an Online Supplement, we provide robustness checks additional to those included here: namely, we replace the expectations-augmented Phillips curve by a hybrid-Phillips curve model; decompose Score 2; consider the effects of allowing horizon fixed effects; of including VIX in the Phillips curve models; of specifying the Phillips curve models with

Table 4: Difference between respondents' characteristics

			$\hat{\alpha}_j$	sign($\hat{\alpha}_j$)	$\hat{\beta}_j$	$Period_j$	$\tilde{\sigma}_j$
Score 1	Significant Group	mean	-0.08	-0.80	0.86	0.21	1.03
		sd	*(0.07)*	*(0.61)*	*(0.32)*	*(0.28)*	*(0.30)*
		no. of respondent	30	30	30	30	30
	Insignificant Group	mean	-0.06	-0.71	0.80	0.33	1.13
		sd	*(0.06)*	*(0.72)*	*(0.28)*	*(0.39)*	*(0.32)*
		no. of respondent	34	34	34	34	34
	Two-Sample t-test	p-value	31.8%	57.3%	42.8%	13.9%	21.1%
Score 2	Significant Group	mean	-0.05	-1.00	0.64	0.23	1.16
		sd	*(0.05)*	*(0.00)*	*(0.25)*	*(0.27)*	*(0.29)*
		no. of respondent	12	12	12	12	12
	Insignificant Group	mean	-0.07	-0.69	0.87	0.28	1.07
		sd	*(0.07)*	*(0.73)*	*(0.30)*	*(0.36)*	*(0.32)*
		no. of respondent	52	52	52	52	52
	Two-Sample t-test	p-value	31.8%	0.4%	1.4%	61.2%	33.2%
Score 3	Significant Group	mean	-0.09	-1.00	0.80	0.37	1.19
		sd	*(0.10)*	*(0.00)*	*(0.29)*	*(0.39)*	*(0.43)*
		no. of respondent	5	5	5	5	5
	Insignificant Group	mean	-0.07	-0.73	0.83	0.27	1.07
		sd	*(0.07)*	*(0.69)*	*(0.30)*	*(0.34)*	*(0.30)*
		no. of respondent	59	59	59	59	59
	Two-Sample t-test	p-value	70.9%	0.4%	82.3%	60.4%	57.0%
F-test on Scores 1-3	Significant Group	mean	-0.08	-0.93	0.79	0.25	1.08
		sd	*(0.06)*	*(0.37)*	*(0.28)*	*(0.29)*	*(0.29)*
		no. of respondent	29	29	29	29	29
	Insignificant Group	mean	-0.06	-0.60	0.85	0.29	1.09
		sd	*(0.08)*	*(0.81)*	*(0.32)*	*(0.39)*	*(0.34)*
		no. of respondent	35	35	35	35	35
	Two-Sample t-test	p-value	33.2%	3.6%	44.9%	65.7%	89.6%

Note: This table shows the mean, the standard deviation, and the number of respondents of the five respondents' characteristics in the significant group and the insignificant group, according to Score 1, Score 2, and Score 3, individually and jointly. The last row in each panel of this table shows the p-value of the Welch Two-Sample t-test.

the unemployment rate gap; and the treatment of outliers.

4.1 Including the cross-sectional mean of the inflation forecasts

A potential issue with our analysis is that we have not included the "consensus forecast" in the models for each individual's inflation forecasts. The consensus is often taken to be the (cross-sectional) mean. Our FPCA scores are not highly correlated with the cross-sectional mean of the inflation forecasts (see Table 2), and hence would not appear to be simply capturing the effect of the mean.[16] Nevertheless, to determine whether its exclusion accounts for the significance of the scores (e.g., if the disagreement and the level of inflation were highly correlated), we run:

$$E_{j,t}\pi_{t+h} = \beta_j E_{j,t}\pi_{t+h+1} + \theta_j M_{t+h+1|t-1} + \sum_{k=1}^{3} \hat{\xi}_{k,t}\gamma_{j,k} + \alpha_j E_{j,t}u_{t+h} + e_{j,t}, \tag{9}$$

where

$$M_{t+h+1|t-1} = \frac{1}{N_{t-1}} \sum_{j=1}^{N_{t-1}} E_{j,t-1}\pi_{t+h+1},$$

is the cross-sectional mean, and N_{t-1} is the number of forecasts made at $t-1$ for π_{t+h+1}.

Table 5 presents the summary statistics (across j) for Model (9) estimated for each respondent. The percentage of F-test rejections is a little lower, at 38% compared to 45%, but clearly the exclusion of the mean does not account for the relevance of the scores for most respondents.

4.2 Including higher cross-sectional moments of the inflation forecasts

Our second robustness check acknowledges that the FPCA scores are highly correlated (whether linearly or not) with the higher moments (SD, skewness, and kurtosis) of the cross-sectional distribution of the inflation forecasts in Section 3.1. Hence we examine the extent to which the scores contain additional information, after controlling for the higher moments. To this end, we run the regression:

$$E_{j,t}\pi_{t+h} = \beta_j E_{j,t}\pi_{t+h+1} + \psi_j mom_{t+h+1|t-1} + \sum_{k=1}^{3} \hat{\xi}_{k,t}\gamma_{j,k} + \alpha_j E_{j,t}u_{t+h} + e_{j,t}, \tag{10}$$

where $mom_{t+h+1|t-1}$ corresponds to one of the higher cross-sectional moments (SD, skewness, and kurtosis) of the forecasts made at $t-1$ for π_{t+h+1}.

[16] There are a number of reasons why one might in principle expect the mean to be a relevant variable: see, for example, Clements (2018).

Table 5: Summary of parameter estimates of Model (9)

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	-0.200	1.037	-0.746	-0.032	0.358	18.8%	32.8%	40.6%
CPI	0.706	0.288	0.541	0.731	0.921	92.2%	93.8%	95.3%
Mean	0.499	0.433	0.183	0.400	0.756	54.7%	73.4%	79.7%
Score 1	0.099	0.454	-0.158	0.178	0.327	26.6%	34.4%	48.4%
Score 2	0.049	0.458	-0.185	0.039	0.243	4.7%	9.4%	15.6%
Score 3	-0.004	0.444	-0.164	0.016	0.223	3.1%	7.8%	9.4%
UNEMP	-0.055	0.082	-0.099	-0.053	-0.004	23.4%	42.2%	51.6%
F-test on Scores 1-3						18.8%	35.9%	40.6%
Adj. R^2	49.3%	24.3%	29.1%	50.8%	66.6%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), for each respondent. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

Table 6 provides the summary statistics for Model (10) for each respondent. Firstly, we observe that SD is statistically significant (at the 5% level) for 41% of respondents, which is perhaps as expected given that SD is a standard linear measure of forecaster disagreement. Despite Score 1 being highly correlated with SD, it is still statistically significant (at the 5% level) for 25% of the respondents, indicating it contains useful additional information even after controlling for SD. Secondly, the percentage for whom Score 2 is statistically significant is a little reduced (at 13% from 19%), after controlling for the cross-sectional skewness. Thirdly, the cross-sectional kurtosis is generally not significant (the percentages of rejections are less than significance levels). Lastly, the F-test rejection rate (of the three scores taken together) is reduced to 25% after controlling for SD, marginally reduced to 37.5% after controlling for skewness, and remains the same at 45.3% after controlling for kurtosis. Overall, we conclude that our FPCA scores not only partly reflect the cross-sectional moments, but also include valuable additional information from the distribution of individuals' forecasts.

4.3 Normal versus Abnormal times

We define the Abnormal times as the quarter before the onset of a recession, the recessionary period itself, and the quarter following the end of a recession, where recessions are dated from NBER US Business Cycle Expansions and Contractions.[17] Other periods

[17] `https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions`, accessed on 18 Nov 2023. We are grateful to an anonymous referee for this suggestion.

Table 6: Summary of parameter estimates of Model (10)

Controlling SD	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.977	1.233	0.270	0.813	1.657	37.5%	48.4%	53.1%
CPI	0.809	0.289	0.677	0.849	1.013	95.3%	96.9%	96.9%
SD	-0.279	1.435	-1.210	-0.131	0.611	18.8%	40.6%	46.9%
Score 1	0.325	0.881	-0.239	0.268	0.971	9.4%	25.0%	35.9%
Score 2	0.386	0.473	0.033	0.297	0.654	10.9%	26.6%	32.8%
Score 3	0.092	0.556	-0.148	0.003	0.405	4.7%	15.6%	21.9%
UNEMP	-0.066	0.068	-0.099	-0.051	-0.011	32.8%	39.1%	45.3%
F-test on Scores 1-3						14.1%	25.0%	42.2%
Adj. R^2	47.6%	24.7%	29.8%	47.6%	64.1%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			
Controlling Skewness	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.757	0.759	0.302	0.820	1.079	45.3%	62.5%	75.0%
CPI	0.824	0.303	0.680	0.868	1.030	95.3%	96.9%	98.4%
Skewness	-0.036	0.148	-0.108	-0.036	0.002	1.6%	7.8%	12.5%
Score 1	0.062	0.459	-0.187	0.131	0.388	18.8%	39.1%	45.3%
Score 2	0.516	0.767	0.158	0.385	0.986	0.0%	12.5%	23.4%
Score 3	-0.025	0.467	-0.220	0.018	0.236	4.7%	9.4%	10.9%
UNEMP	-0.068	0.070	-0.100	-0.055	-0.015	32.8%	39.1%	45.3%
F-test on Scores 1-3						18.8%	37.5%	48.4%
Adj. R^2	46.8%	24.8%	28.3%	46.7%	64.0%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			
Controlling Kurtosis	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.740	0.742	0.291	0.772	1.052	40.6%	57.8%	65.6%
CPI	0.824	0.300	0.679	0.878	1.025	95.3%	96.9%	96.9%
Kurtosis	0.008	0.026	-0.003	0.006	0.017	1.6%	3.1%	9.4%
Score 1	0.078	0.455	-0.215	0.170	0.347	21.9%	37.5%	53.1%
Score 2	0.365	0.452	0.027	0.278	0.592	7.8%	20.3%	29.7%
Score 3	-0.075	0.649	-0.383	-0.074	0.188	3.1%	12.5%	17.2%
UNEMP	-0.071	0.070	-0.107	-0.056	-0.018	32.8%	42.2%	48.4%
F-test on Scores 1-3						28.1%	45.3%	51.6%
Adj. R^2	46.7%	25.0%	28.3%	46.6%	64.1%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), for each respondent. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

are Normal times. We could just as well define Abnormal times as periods of high uncertainty, and Normal times as low uncertainty, given that uncertainty is typically found to be higher in recessions.[18] We run our baseline regression of Model (7) separately for the two subperiods, and the results are presented in Table 7. Our findings suggest the responsiveness of inflation to unemployment is greater in high-uncertainty (recessionary) periods: the mean coefficient on the unemployment rate is -0.188, compared to -0.043 in low-uncertainty (expansionary) periods. However, the FDA scores have a greater impact in low uncertainty Normal times. The rejection of the F-test (for scores 1-3) is 50.0% for the low economic uncertainty subperiod, and 12.5% for the subperiod of high economic uncertainty. In Normal times forecasters are more attentive to others and put less weight on their own unemployment rate forecasts. Given that Normal time observations outnumber Abnormal times approximately 4:1, the former periods hold greater sway and drive the baseline findings for the whole period. These findings are suggestive of differences in forecaster behaviour in the two different environments, and merit further research. However, the average number of observations is much smaller in the high-uncertainty recessionary periods, which will at least partially explain differences in statistical significance (e.g., of the F-test of scores 1-3.

4.4 Regression based on the inflation gap

Our inflation-gap formulation of the Phillips curve model is:

$$E_{j,t}\pi_{t+h} - E_{j,t}\tau_t = \delta_j + \beta_j(E_{j,t}\pi_{t+h+1} - E_{j,t}\tau_t) + \sum_{k=1}^{3} \hat{\xi}_{k,t}\gamma_{j,k} + \alpha_j E_{j,t}u_{t+h} + e_{j,t}, \qquad (11)$$

where $E_{j,t}\tau_t$ is respondent j's forecast of the trend rate of inflation at time t. This is the long-horizon forecast CPI inflation rate, which does not depend on h.

To empirically construct the variable $E_{j,t}\tau_t$, we use the combination of SPF ten-year CPI inflation forecasts and Blue Chip Economic Indicators. For the period 1991:Q4 to 2022:Q2 inclusive, we use the individual forecasts of average ten-year CPI inflation (CPI10). Prior to 1991:Q4 this variable was not collected, and for the surveys from 1981:Q3 to 1991:Q3, we use the twice yearly long-term inflation forecasts from Blue Chip Economic Indicators, made available on the SPF web site (as Additional-CPIE10.xslx). We lag the data, so that the 1981:Q3 survey uses the 1981:Q2 Blue Chip figure, etc. Thus the trend inflation rate varies across respondents from 1991:Q4 onwards, but prior to that date is the same for all respondents.

[18] In the Supplement we also check whether our findings are robust to an uncertainty measure relating to broad economic conditions, which might proxy judgmental adjustment - see Croushore and Stark (2019). We are grateful to an anonymous referee for this suggestion.

Table 7: Summary of parameter estimates of Model (7) across respondents during high/low economic uncertainty

High economic uncertainty								
	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	1.607	3.346	0.000	0.924	2.170	17.2%	29.7%	35.9%
CPI	0.814	0.598	0.531	0.892	1.158	65.6%	73.4%	81.3%
Score 1	-0.222	1.311	-0.587	-0.037	0.511	3.1%	20.3%	32.8%
Score 2	-0.021	1.109	-0.592	-0.010	0.684	0.0%	6.3%	10.9%
Score 3	0.838	1.595	-0.047	0.909	1.862	1.6%	7.8%	14.1%
UNEMP	-0.188	0.441	-0.242	-0.093	-0.013	21.9%	34.4%	45.3%
F-test on scores 1-3						3.1%	12.5%	14.1%
Adj. R2	47.7%	26.6%	28.0%	46.2%	69.2%			
No. of Obs.	46.0	18.5	35.0	48.0	60.0			

Low economic uncertainty								
	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.811	0.841	0.270	0.739	1.179	50.0%	54.7%	62.5%
CPI	0.741	0.360	0.602	0.805	0.992	85.9%	87.5%	89.1%
Score 1	0.146	0.408	-0.076	0.172	0.423	28.1%	39.1%	46.9%
Score 2	0.438	0.478	0.088	0.305	0.655	15.6%	31.3%	35.9%
Score 3	-0.015	0.516	-0.282	0.007	0.240	3.1%	14.1%	14.1%
UNEMP	-0.043	0.060	-0.090	-0.040	-0.005	18.8%	34.4%	42.2%
F-test on scores 1-3						32.8%	50.0%	59.4%
Adj. R2	46.0%	27.5%	24.8%	47.7%	65.9%			
No. of Obs.	197.5	74.2	132.0	180.0	253.0			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$) during subperiods of high/low economic uncertainty, for each respondent j. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

Table 8 shows the summary of parameter estimates of Model (11) across respondents. Compared to our baseline results, there is a marginal reduction in the percentage of rejections of the t-test of Score 1, and of the F-test of all scores together. However, the percentage of rejections for the t-test of Score 3 is almost doubled, from 7.8% to 15.5%. We conclude that adoption of an inflation gaps formulation does not qualitatively alter our findings.

One might argue that the unemployment rate term should enter as a 'gap', that is, inflation rate forecasts depend on the difference between the unemployment rate and the natural rate. However, forecasts of the natural rate were only collected by the U.S. SPF from 1996:Q3 onwards, and then only for the third-quarters of the year. Clements (2023) uses the estimate of the natural rate available at the time of the survey as the forecast, and finds the results are little affected. In the Supplement we replace the forecast of the unemployment rate with the first difference of the unemployment rate forecast, with little change to the results.

Table 8: Summary of parameter estimates of Model (11) across respondents

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.349	0.479	0.072	0.322	0.632	34.5%	43.1%	50.0%
CPI Gap	0.752	0.282	0.576	0.785	0.949	93.1%	94.8%	98.3%
Score 1	0.035	0.494	-0.192	0.110	0.317	27.6%	37.9%	48.3%
Score 2	0.344	0.439	0.097	0.258	0.563	3.4%	17.2%	27.6%
Score 3	0.039	0.431	-0.127	0.064	0.274	3.4%	15.5%	19.0%
FD UNEMP	-0.074	0.071	-0.119	-0.069	-0.025	36.2%	44.8%	56.9%
F-test on scores 1-3						27.6%	39.7%	50.0%
Adj. R2	38.7%	24.0%	20.0%	35.1%	48.2%			
No. of Obs.	232.9	84.0	161.0	212.0	285.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), for each respondent j. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

5 Conclusions

We have set out a way of applying FDA to a specific application using survey expectations, but as we expand on below, expect FDA to be useful more widely when analyzing survey expectations data. We estimate Phillips curve models at an individual level for a sample of respondents to the U.S. SPF over the last 40 years. While there is evidence that some respondents' inflation and unemployment rate forecasts conform to the belief that higher unemployment has a moderating influence on inflation, it is also plausible to suppose that an individual forecaster may be influenced by the forecasts of others. We use FDA to allow the data to determine the way in which the forecasts of others impact the individual's forecasts. We find that for nearly a half of the professional forecasters in the sample there is a role for the earlier inflation forecasts of the other respondents.

We assess whether the FDA "scores" are simply capturing the first and second-moments of the cross-sectional distribution of prior forecasts. One might suppose that forecasters adjust their forecasts towards (i.e., "herd on") the consensus view. Or indeed deliberately move away form the consensus position ("anti-herd"). However, the FDA scores transpire to be statistically significant for some respondents even when the consensus forecast is explicitly included. In fact, they remain significant for some respondents even when the cross-sectional standard deviation and higher moments of the forecast distribution are included. The FDA scores appear to capture useful additional information, even though they are correlated with interpretable characteristics of the cross-sectional distribution, such as the standard deviation, or "disagreement".

Alongside the empirical application, we provide a discussion of some aspects of FDA in the hope that this will encourage its wider use. We know of only one other study that makes use of FDA with survey expectations data (Meeks and Monti (2019)), and its application in economics more generally has been more limited (but see the papers referred to in the Introduction) than in some other disciplines.

Although we have provided one application of the use of FDA - to determine whether Phillips curve models of respondents' inflation and unemployment rate forecasts also allow a role for others' forecasts - we suspect FDA could be usefully applied to a number of areas in the analysis of survey expectations. These would be relatively simple to implement, given the approach to applying FDA to survey expectations described in this paper. They might include: testing fixed-event forecast rationality, following Nordhaus (1987) and Clements (1995), *inter alia*; testing expectations for over-reaction to new information; and an explicit testing of herding behavior. The first might be accomplished by regressing a forecast revision, say, the change in j's forecasts of π_{t+h} between forecast origins $t-1$ and t, namely, $E_{j,t}\pi_{t+h} - E_{j,t-1}\pi_{t+h}$, on the forecasts of others made at time $t-2$, expressed in general terms as $\int \gamma_j(x)\, dF_{t+h|t-2}(x)$ (c.f., (2)). The flexible parameterization of others' expectations accorded by FDA might also be useful in the recent literature on the over-reaction to new information (see, e.g., Bordalo et al. (2020), Kohlhas and Walther (2021) and Broer and Kohlhas (2021)) as a source of new information, and also in the literature on herding (see e.g., Clements (2018), which tends to rely on the consensus or mean forecast).

FDA is found to provide a sound statistical framework within which to address the issues of interest. In some instances, as here, it may prove a useful complement to alternative approaches, such as choosing a selection of moments from the cross-sectional distribution of responses. Whether it ends up complementing or simply replacing such approaches, in any instance, it has the potential to improve practice.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control 19*(6), 716–723.

Backus, D. K. and G. W. Smith (1993, November). Consumption and real exchange rates in dynamic economies with non-traded goods. *Journal of International Economics 35*(3-4), 297–316.

Bardsley, P., L. Horváth, P. Kokoszka, and G. Young (2017). Change point tests in functional factor models with application to yield curves. *The Econometrics Journal 20*(1), 86–117.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020, September). Overreaction in Macroeconomic Expectations. *American Economic Review 110*(9), 2748–2782.

Bouri, E., C. K. M. Lau, T. Saeed, S. Wang, and Y. Zhao (2021). On the intraday return curves of bitcoin: Predictability and trading opportunities. *International Review of Financial Analysis 76*, 101784.

Broer, T. and A. Kohlhas (2021, April). Forecaster (Mis-)Behavior. CEPR Discussion Papers 12898/2018, C.E.P.R. Discussion Papers.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics 12*, 383–398.

Cao, R., L. Horváth, Z. Liu, and Y. Zhao (2020). A study of data-driven momentum and disposition effects in the chinese stock market by functional data analysis. *Review of Quantitative Finance and Accounting 54*(1), 335–358.

Casey, E. (2020). Do macroeconomic forecasters use macroeconomics to forecast? *International Journal of Forecasting 36*(4), 1439 – 1453.

Chen, Z., Y. Bao, H. Li, and B. F. Spencer Jr (2019). Lqd-rkhs-based distribution-to-distribution regression methodology for restoring the probability distributions of missing shm data. *Mechanical Systems and Signal Processing 121*, 655–674.

Clements, M. P. (1995). Rationality and the role of judgement in macroeconomic forecasting. *Economic Journal 105*, 410–420.

Clements, M. P. (2018). Do macroforecasters herd? *Journal of Money, Credit and Banking 50*(2-3), 265–292.

Clements, M. P. (2023). Do professional forecasters believe in the Phillips curve? *International Journal of Forecasting*. Forthcoming. Available at SSRN: https://ssrn.com/abstract=4476230.

Clements, M. P., R. Rich, and J. Tracy (2022). Surveys of Professionals, Chapter 3. In Ruediger Bachmann, Giorgio Topa, Wilbert Klaauw (Ed.), *Handbook of Economic Expectations*, pp. 71–106. Academic Press, Elsevier.

Coibion, O., Y. Gorodnichenko, and R. Kamdar (2018, December). The Formation of Expectations, Inflation, and the Phillips Curve. *Journal of Economic Literature 56*(4), 1447–1491.

Croushore, D. (1993). Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review, November*, 3–15.

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

Croushore, D. and T. Stark (2019, October). Fifty Years of the Survey of Professional Forecasters. *Economic Insights 4*(4), 1–11.

Dräger, L. and G. Nghiem (2021, July). Are Consumers' Spending Decisions in Line with A Euler Equation? *The Review of Economics and Statistics 103*(3), 580–596.

Faust, J. and J. H. Wright (2013). Forecasting Inflation, chapter 1. In G. Elliott and

A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2A*. North Holland.

Fendel, R., E. M. Lis, and J.-C. Rülke (2011). Do professional forecasters believe in the Phillips curve? evidence from the G7 countries. *Journal of Forecasting 30*(2), 268–287.

Gali, J. and M. Gertler (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics 44*, 195–222.

Ganics, G., B. Rossi, and T. Sekhposyan (2023). From fixed-event to fixed-horizon density forecasts: Obtaining measures of multihorizon uncertainty from survey density forecasts. *Journal of Money, Credit and Banking*.

Haddow, A., C. Hare, J. Hooley, and T. Shakir (2013). Macroeconomic uncertainty: what is it, how can we measure it and why does it matter? *Bank of England Quarterly Bulletin*, 100–109. 2013 Q2.

Hall, P. and C. Vial (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(4), 689–705.

Hooper, P., F. S. Mishkin, and A. Sufi (2019, May). Prospects for inflation in a high pressure economy: Is the phillips curve dead or is it just hibernating? Working Paper 25792, National Bureau of Economic Research.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*, Volume 200. Springer Science & Business Media.

Horváth, L., P. Kokoszka, J. VanderDoes, and S. Wang (2022). Inference in functional factor models with applications to yield curves. *Journal of Time Series Analysis*.

Horváth, L., Z. Liu, G. Rice, and S. Wang (2020). A functional time series analysis of forward curves derived from commodity futures. *International Journal of Forecasting 36*(2), 646–665.

Jain, M. (2019). Perceived inflation persistence. *Journal of Business & Economic Statistics 37*(1), 110–120.

Jones, M. C. (1992). Estimating densities, quantiles, quantile densities and density quantiles. *Annals of the Institute of Statistical Mathematics 44*(4), 721–727.

Karhunen, K. (1947). *Über lineare Methoden in der Wahrscheinlichkeitsrechnung: akademische Abhandlung*. Sana.

Kearney, F. and H. L. Shang (2020). Uncovering predictability in the evolution of the wti oil futures curve. *European Financial Management 26*(1), 238–257.

Kohlhas, A. N. and A. Walther (2021, September). Asymmetric Attention. *American Economic Review 111*(9), 2879–2925.

Kokoszka, P., H. Miao, A. Petersen, and H. L. Shang (2019). Forecasting of density functions with an application to cross-sectional and intraday returns. *International Journal of Forecasting 35*(4), 1304–1317.

Kokoszka, P., H. Miao, and X. Zhang (2015). Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics 13*(2), 456–477.

Kokoszka, P. and M. Reimherr (2017). *Introduction to functional data analysis*. Chapman and Hall/CRC.

Loève, M. (1960). *Probability theory*. D. Van Nostrand Company.

McCallum, B. T. (1976, January). Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates. *Econometrica 44*(1), 43–52.

Meeks, R. and F. Monti (2019). Heterogeneous beliefs and the Phillips curve. *Bank of England Working Paper*.

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

Oomen, R. (2019). Price signatures. *Quantitative Finance 19*(5), 733–761.

Petersen, A., C.-J. Chen, and H.-G. Müller (2019). Quantifying and visualizing intraregional connectivity in resting-state functional magnetic resonance imaging with correlation densities. *Brain connectivity 9*(1), 37–47.

Petersen, A. and H.-G. Müller (2016). Functional data analysis for density functions by transformation to a hilbert space. *The Annals of Statistics 44*(1), 183–218.

Phillips, A. W. H. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica 25*, 283–299.

Ramsay, J. and B. W. Silverman (2005). *Functional Data Analysis*. Springer.

Reiss, P. T., J. Goldsmith, H. L. Shang, and R. T. Ogden (2017). Methods for scalar-on-function regression. *International Statistical Review 85*(2), 228–249.

Riesz, F. et al. (1990). *Functional Analysis*. Dover.

Shang, H. L. and F. Kearney (2022). Dynamic functional time-series forecasts of foreign exchange implied volatility surfaces. *International Journal of Forecasting 38*(3), 1025–1049.

Zarnowitz, V. and L. A. Lambros (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy 95(3)*, 591–621.

SUPPLEMENT

This is the supplementary material for the article "Do Professional Forecasters' Phillips Curves Incorporate the Beliefs of Others?". In Section A we present the results for a hybrid-Phillips curve model. In Section B, we decompose Score 2 into positive and negative parts. Section C checks for horizon fixed effects, and Section D whether our findings are robust to allowing for a measure of the uncertainty concerning the economic environment. In Section E we replace the unemployment rate with the difference of the unemployment rate to proxy the unemployment rate gap, and finally, in Section F consider the robustness of the results to the treatment of outliers.

A The specification of the Phillips curve model

From a theoretical perspective, a Phillips curve model with a forward-looking inflation term is better grounded than one with a backward-looking term, although one could consider a hybrid-Phillips curve model which includes both. For example, Clements (2023) used a hybrid-Phillips curve model that includes both a forward looking term ($E_{j,t}\pi_{t+h+1}$) and a backward looking term ($E_{j,t}\pi_{t+h-1}$). Following this approach, we modify our model (7) to a hybrid-Phillips curve version as:

$$E_{j,t}\pi_{t+h} = \delta_j + \beta_j E_{j,t}\pi_{t+h+1} + \psi_j E_{j,t}\pi_{t+h-1} + \sum_{k=1}^{3} \hat{\xi}_{k,t}\gamma_{j,k} + \alpha_j E_{j,t}u_{t+h} + e_{j,t}. \tag{12}$$

Note that we can only consider $h = 1, 2, 3$ (without $h = 0$) because we do not use $E_{j,t}\pi_{t-1}$ in this study. As a corresponding adjustment, we select the respondents who made 90 or more forecasts (for all $h = 1, 2, 3$) which is still at least 30 forecasts per h on average. By this adjustment, the same 64 respondents are selected as in Section 3.2.

Table 9 presents the summary statistics of the parameter estimates for Model (12) for each respondent. Comparing to the baseline model, we find that the percentage of respondents for whom we find the scores are individually statistically significant is reduced, as is the percentage of rejections for the three scores jointly. Nevertheless, we still find we reject the joint insignificance of the scores for over a quarter of respondents in this more highly parameterized model.

B Decompose Score 2 into positive and negative values

Given the recent interest in tail-end growth risks in the recent literature (see, e.g., Adrian et al. (2019) and Adams et al. (2021)), we consider the possibility that forecaster behaviour

Table 9: Summary of parameter estimates of Model (12)

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.511	0.767	0.073	0.273	0.734	25.0%	42.2%	54.7%
CPI_forward	0.585	0.250	0.471	0.637	0.726	92.2%	96.9%	96.9%
CPI_backward	0.206	0.122	0.146	0.220	0.272	90.6%	96.9%	96.9%
Score 1	-0.088	0.360	-0.160	-0.021	0.087	14.1%	21.9%	31.3%
Score 2	0.259	0.401	0.014	0.167	0.403	14.1%	18.8%	29.7%
Score 3	-0.042	0.407	-0.161	-0.017	0.097	1.6%	15.6%	17.2%
UNEMP	-0.014	0.046	-0.035	-0.006	0.007	4.7%	23.4%	25.0%
F-test on Scores 1-3						18.8%	26.6%	31.3%
Adj. R^2	64.9%	25.2%	50.8%	66.7%	85.6%			
No. of Obs.	182.3	67.2	122.3	166.5	234.5			

Note: The estimates are based on individual systems of 3 equations (i.e., one equation for $h = 1, 2, 3$), for each respondent. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

might be affected differently by upside and downside risk. In our setting, this means allowing that forecasters might respond differently to positive versus negative skew in (last period's) cross-section of forecasts. To investigate this possibility, we decompose Score 2 as:

$$\hat{\xi}_{2,t}^{(+)} = \max\left\{0,\ \hat{\xi}_{2,t}\right\},$$
$$\hat{\xi}_{2,t}^{(-)} = \min\left\{0,\ \hat{\xi}_{2,t}\right\},$$

where $\xi_{2,t}^{(+)}$ captures the positive skew and $\xi_{2,t}^{(-)}$ measures the negative skew. Then we run the following regression:

$$E_{j,t}\pi_{t+h} = \delta_j + \phi_j E_{j,t}\pi_{t+h+1} + \hat{\xi}_{1,t}\gamma_{1,k} + \hat{\xi}_{2,t}^{(+)}\gamma_{2,k}^{(+)} + \hat{\xi}_{2,t}^{(-)}\gamma_{2,k}^{(-)} + \hat{\xi}_{3,t}\gamma_{3,k} + \alpha_j E_{j,t}u_{t+h} + e_{j,t}. \quad (13)$$

Table 10 provides the summary statistics of the parameter estimates for Model (13) across respondents. We find fewer rejections for either $\gamma_{2,k}^{(+)}$ or $\gamma_{2,k}^{(-)}$, compared to the baseline model, suggesting there is no evidence of a different response to positive and negative values.

Table 10: Summary of parameter estimates of Model (7)

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.751	0.744	0.284	0.748	1.135	43.8%	57.8%	68.8%
CPI	0.825	0.300	0.679	0.879	1.020	95.3%	96.9%	96.9%
Score 1	0.096	0.464	-0.225	0.164	0.364	28.1%	46.9%	54.7%
Score 2 - positive	0.427	0.771	-0.099	0.252	0.730	3.1%	6.3%	14.1%
Score 2 - negative	0.316	0.694	-0.017	0.238	0.616	0.0%	6.3%	9.4%
Score 3	-0.001	0.467	-0.212	0.022	0.211	4.7%	7.8%	10.9%
UNEMP	-0.069	0.069	-0.099	-0.056	-0.015	93.8%	98.4%	100.0%
F-test on Scores 1-3						25.0%	37.5%	48.4%
Adj. R^2	46.7%	25.0%	28.4%	46.6%	64.0%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), for each respondent. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

C Fixed effects of h

A simple way to check the assumption of the constancy of parameters over the forecast horizon h, in the main text, is to check for the significance of horizon fixed effects, via:

$$E_{j,t}\pi_{t+h} = \delta_j + \beta_j E_{j,t}\pi_{t+h+1} + \sum_{k=1}^{3} \hat{\xi}_{k,t}\gamma_{j,k} + \alpha_j E_{j,t}u_{t+h} + \eta_{j,h} + e_{j,t}. \tag{14}$$

where $\eta_{j,h}$ contains the fixed effects h for respondent j.

Table 11 shows the summary of parameter estimates of Model (14) across respondents. The results allowing for fixed effects are very similar to the results for the baseline model without fixed effects. Moreover, the estimated coefficients on the horizon dummies are significant for only 12-17% of the 64 respondents.

D Inclusion of an uncertainty measure capturing economic conditions - VIX

We add the VIX as the additional aggressor in our baseline model and check whether our results are still robust. We use the quarterly (end of quarter) VIX data downloaded from Federal Reserve Economic Data (FRED). Note that VIX starts from 1990:Q1, and thus our results with VIX is based on the period 1990:Q1 - 2022:Q2. Table 12 presents the

Table 11: Summary of parameter estimates of Model (14) across respondents

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.756	0.787	0.285	0.727	1.237	40.6%	64.1%	68.8%
CPI	0.826	0.303	0.683	0.888	1.024	95.3%	96.9%	96.9%
Score 1	0.081	0.467	-0.213	0.153	0.363	25.0%	46.9%	53.1%
Score 2	0.379	0.458	0.028	0.279	0.606	7.8%	21.9%	28.1%
Score 3	0.010	0.450	-0.182	0.000	0.286	3.1%	9.4%	14.1%
UNEMP	-0.066	0.069	-0.098	-0.053	-0.017	34.4%	40.6%	40.6%
FE: $h = 1$	-0.017	0.216	-0.106	0.011	0.114	4.7%	14.1%	21.9%
FE: $h = 2$	-0.016	0.225	-0.093	0.004	0.133	4.7%	17.2%	23.4%
FE: $h = 3$	-0.009	0.228	-0.109	0.002	0.096	6.3%	12.5%	21.9%
F-test on scores 1-3						25.0%	43.8%	56.3%
Adj. R2	46.9%	24.8%	28.1%	46.6%	64.0%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$ of (14)), for each respondent j. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

summary of parameter estimates across respondents of Model (7) with VIX included as the additional regressor. The results with VIX is very similar to the baseline result, suggesting the robustness of our results to the addition of VIX.

E Use the first difference of the unemployment rate forecasts

We replace the level of the unemployment rate forecast with the first difference of the unemployment rate forecast, as shown below

$$E_{j,t}\pi_{t+h} = \delta_j + \beta_j E_{j,t}\pi_{t+h+1} + \sum_{k=1}^{3} \hat{\xi}_{k,t}\gamma_{j,k} + \alpha_j \Delta E_{j,t}u_{t+h} + e_{j,t} \tag{15}$$

where $\Delta E_{j,t}u_{t+h} = E_{j,t}u_{t+h} - E_{j,t-1}u_{t+h-1}$.

The result is shown in Table 13. There is a marginal drop in the percentage of rejection in the t-test of Score 1 and the F-test.

Table 12: Summary of parameter estimates of Model (7) with VIX across respondents

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.817	0.839	0.304	0.864	1.319	45.8%	61.0%	67.8%
CPI	0.814	0.314	0.647	0.849	1.034	94.9%	96.6%	96.6%
VIX	-0.004	0.012	-0.011	-0.003	0.002	11.9%	22.0%	28.8%
Score 1	0.096	0.476	-0.201	0.165	0.396	27.1%	45.8%	52.5%
Score 2	0.349	0.481	0.009	0.238	0.573	5.1%	16.9%	28.8%
Score 3	-0.037	0.508	-0.240	-0.013	0.177	3.4%	6.8%	10.2%
UNEMP	-0.062	0.069	-0.098	-0.046	-0.009	32.2%	39.0%	47.5%
F-test on scores 1-3						22.0%	40.7%	59.3%
Adj. R2	44.8%	24.1%	28.5%	44.3%	59.9%			
No. of Obs.	241.6	88.8	172.0	223.0	300.5			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), for each respondent j. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

Table 13: Summary of parameter estimates of Model (15) across respondents

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.334	0.605	-0.089	0.331	0.703	45.3%	60.9%	68.8%
CPI	0.833	0.298	0.695	0.884	1.016	95.3%	96.9%	98.4%
Score 1	-0.060	0.410	-0.206	0.024	0.170	23.4%	34.4%	39.1%
Score 2	0.390	0.465	0.044	0.314	0.550	12.5%	18.8%	29.7%
Score 3	0.011	0.470	-0.165	-0.011	0.278	3.1%	9.4%	12.5%
FD UNEMP	-0.082	0.143	-0.169	-0.062	0.000	29.7%	45.3%	45.3%
F-test on scores 1-3						20.3%	35.9%	45.3%
Adj. R2	46.8%	24.7%	27.5%	46.8%	60.9%			
No. of Obs.	241.6	89.6	161.5	220.5	304.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), for each respondent j. FD UNEMP is the first difference of the unemployment rate forecast. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels.

F Remove outliers at 5% level

When working with the LQD, it is necessary to remove outliers to prevent values of infinity occurring (as when any $s \in (0,1)$ gives $f_t(Q_t(s)) = 0$). However, the choice of the level at which to remove outliers is somewhat arbitrary. In this section we check that the results are not overly sensitive to the choice made. In the baseline analysis, outliers at the 1% level in each quarter t for a given h were removed. We repeat the analysis with outliers removed at the 5% level. The summary of parameter estimates of the baseline model is presented in Table 14. We find that the mean of the parameter estimates is similar to the base case. The F-test rejection percentage (at 5% significance level) of 47% is very close to the base case. The rejection percentage for the individual scores is slightly lower for Score 1 but a little higher for Scores 2 and 3. We conclude that the findings are not overly sensitive to the treatment of outliers.

Table 14: Summary of parameter estimates of Model (7) with outliers removed at 5% level

	mean	s.d.	l.q.	median	u.q.	rej. 1%	rej. 5%	rej. 10%
Intercept	0.735	0.750	0.325	0.785	1.066	48.4%	60.9%	75.0%
CPI	0.826	0.298	0.675	0.867	1.016	95.3%	98.4%	98.4%
Score 1	0.087	0.463	-0.228	0.164	0.358	28.1%	39.1%	50.0%
Score 2	0.472	0.532	0.083	0.348	0.679	17.2%	35.9%	39.1%
Score 3	0.078	0.487	-0.149	0.012	0.381	1.6%	10.9%	14.1%
UNEMP	-0.066	0.066	-0.095	-0.052	-0.017	31.3%	39.1%	43.8%
F-test on Scores 1-3						35.9%	46.9%	57.8%
Adj. R^2	47.0%	24.7%	28.9%	46.5%	63.6%			
No. of Obs.	243.5	89.6	162.8	223.0	312.8			

Note: The estimates are based on individual systems of 4 equations (i.e., one equation for $h = 0, 1, 2, 3$), or each respondent. For each parameter, we present the summary statistics of the cross-sectional distribution over j, including mean, standard deviation (s.d.), lower quartiles (l.q.), median, and upper quartiles (u.q.). In the last three columns, we report the proportion of the 64 regressions for which we reject the null hypothesis of the parameter equal to zero at 1%, 5%, and 10% levels. Outliers at the 5% level in each quarter t for a given h are removed.

Reference

Adams, P. A., T. Adrian, N. Boyarchenko, and D. Giannone (2021). Forecasting macroeconomic risks. *International Journal of Forecasting 37*(3), 1173–1191.

Adrian, T., N. Boyarchenko, and D. Giannone (2019, April). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Clements, M. P. (2023). Do professional forecasters believe in the Phillips curve? *International Journal of Forecasting*. Forthcoming. Available at SSRN: https://ssrn.com/abstract=4476230.