Survey respondents' inflation forecasts and the COVID period

Received: 19 September 2023 | Revised: 27 March 2024 | Accepted: 19 May 2024

DOI: 10.1002/for.3169

RESEARCH ARTICLE

WILEY

Survey respondents' inflation forecasts and the COVID period

Michael P. Clements

ICMA Centre, Henley Business School, University of Reading, Reading, UK

Correspondence
Michael P. Clements, ICMA Centre, Henley Business School, University of Reading, Reading RG6 6BA, UK.
Email: m.p.clements@reading.ac.uk

Abstract

How do professionals forecast in uncertain times, when the relationships between variables that held in the past may no longer be useful for forecasting the future? For inflation forecasting, we answer this question by measuring survey respondents' adherence to their pre-COVID-19 Phillips curve models during the pandemic. We also ask whether professionals *ought* to have put their trust in their Phillips curve models over the COVID-19 period. We address these questions allowing for heterogeneity in respondents' forecasts and in their perceptions of the Phillips curve relationship.

KEYWORDS

COVID-19 pandemic, inflation forecasting, model heterogeneity, Phillips curve

1 | INTRODUCTION

How do professionals forecast in uncertain times, when the relationships between variables that held in the past may no longer be useful for forecasting the future? Castle et al. (2016) review "a body of research that seeks to provide viable strategies for economic forecasting when past relationships can no longer be relied upon." They remind us that as long ago as the first half of the 20th century authors were struggling with forecasting in the presence of structural beaks: for example, Shoup et al. (1941) ("The times are so different now [i.e., October 1941] from 1935 to 1939 that relations existing then may not exist at all today."—even prior to the United States entering World War II) and Klein (1947) ("Would the econometrician merely substitute into his equations of peacetime behavior patterns in order to forecast employment in a period during which there will be a war?").

The review by Castle et al. (2016) suggests a number of partial remedies for forecast failure caused by structural breaks, which are aimed at avoiding making systematic forecast errors after a break has occurred and (to a lesser extent) using information on the break as it unfolds. Our interest is in the extent to which professionals adhere to their existing models or adopt more robust forecasting models, and if so, whether they are justified in doing so.

We consider the COVID-19 pandemic which began in the first quarter of 2020. Unlike the onset of business cycle downturns, which are often difficult to call in real time, COVID-19 was heralded by lockdowns and other restrictions and measures aimed at restricting the transmission of the virus. Hence, the respondents in the US Survey of Professional Forecasters (SPF) would have been left in little doubt that something was afoot. We compare the forecasts made prior to 2020:1 with those made during the period 2020:1 to 2022:4.

Hence, our paper differs from a growing body of literature that looks at ways of obtaining and utilizing more timely information on activity and other key indicators (than that afforded by official statistics), such as high-frequency "alternative" data or "Big Data."[1] Our interest is in how professional forecasters actually responded, as measured by surveys of their expectations, and allows for the possibility that new types of data and forecasting models were adopted.

A number of complicating factors arise when we try to determine whether forecasters changed their behavior confronted with the disruption to the economy and financial system caused by COVID-19. The first is determining how forecasters forecast in normal times. That is, the extent to which their forecasts are based on well-founded theory or models of how the economy operates. There is little direct (self-reported) evidence on how survey respondents form their expectations.[2] Consequently, we infer the respondents' models from the forecasts (and realizations) alone.

Second, there appears to be considerable heterogeneity in individual respondents' "models," as described by Clements (2024), in his analysis of respondents' Phillips curve models prior to COVID-19. This apparent heterogeneity is perhaps exacerbated by the short data samples available for many respondents and by respondents being active survey participants at different times. In addition, many economic relationships are beset by nonlinearities or have parameters that change over time. The Phillips curve is a case in point: There is evidence that the curve may have "flattened" over time, as well as that the slope may depend on the tightness of the labor market (see, e.g., Hooper et al., 2019). In principle, it may be possible to estimate models with these features using individual-level forecasts, but in practice, small samples with missing observations due to non-participation suggest it would be challenging to accurately capture such complex interactions. A recent study by Cerrato and Gitti (2022) considers whether the *actual* Phillips curve changed at the time of COVID. Their study suggests that the slope of the Phillips curve became flat during the COVID period, before steepening dramatically from March 2021 onwards. Although forecasters' inflation models would be expected to mirror changes in the economy which persist over time, it is debatable whether their forecasts would reflect the dramatic changes found by Cerrato and Gitti (2022) as they unfolded in real time.

Following Clements (2024), we consider the Phillips curve but focus on the extent to which survey respondents altered their reliance on the Phillips curve from 2020:Q1 onwards. We use the approach discussed in Clements (2024) to estimate individual PC models for the respondents. We acknowledge the importance of the heterogeneity in individuals' PC models *before* COVID-19. We consider the extent to which these individuals changed their behavior over the COVID-19 period. Individuals who were not operative over the COVID-19 period are discarded, as are individuals who did not participate during the COVID period. This avoids the possible distortionary effects of considering aggregates of forecasters pre-COVID-19 and during COVID-19, whereby compositional effects may give a misleading picture.[3] If forecasters were relatively homogeneous, we might simply consider the behavior of the consensus forecaster across the two periods. However, the diversity in actual behavior cautions against this approach.[4]

There are a number of papers on the theory-consistency of survey expectations, some of which are reviewed by Clements (2024). Of those, a number consider whether survey expectations are consistent with a Phillips curve relationship. These include Fendel et al. ((2011), p. 286), Dräger et al. (2016), and Casey (2020), in addition to Clements (2024).[5] Our paper differs from this literature by explicitly focusing on the possible changes in expectations behavior resulting from an upheaval: in our case the upheaval caused by COVID-19.

The paper is also related to a literature on *perceived* persistence in expectations formation. As examples, Aguiar and Gopinath (2007), Bluedorn and Leigh (2018), Krane (2011), and Clements (2020) consider the permanency of shocks to output, and Jain (2019) examines the heterogeneity of inflation persistence perceptions. We consider perceptions of inflation persistence indirectly through the lens of the Phillips curve, following Clements (2024).[6]

The plan of the rest of the paper is as follows. In Section 2, we describe how we estimate individual PC models for the respondents, following Clements (2024). Section 3 explains how we measure the change in respondents' inflation forecasting behavior in response to the onset of COVID-19. Section 4 reports our empirical findings. Section 5 takes a more aggregated approach and tracks the forecasts through the COVID-19 period. Section 6 considers the robustness of the findings to a different specification of the Phillips curve model and to a different (more precise but potentially biased) way of estimating the unknown model parameters. Section 7 offers some concluding remarks. There are Appendices detailing the calculation of Spearman's Rank Correlation coefficient, and a description of the forecast data and actual data used in the paper.

2 | ESTIMATING PHILLIPS CURVES FOR THE INDIVIDUAL RESPONDENTS.

We follow Clements (2024), which contains the details. A "hybrid" PC can be written as

$$\pi_t = \beta_b \pi_{t-1} + \beta_f E_t \pi_{t+1} + \gamma u_t + \varepsilon_t, \tag{1}$$

where π_t is quarterly inflation and $E_t\pi_{t+1}$ is the expectation given the information available at t. The unemployment rate u_t is the measure of slack in the economy. When estimated on "actual" data, a proxy for, or way of estimating, the expectations term has to be found. Coibion et al. (2018) review studies using survey expectations, whereas McCallum (1976) is an early example of instrumenting this term. This issue does not arise in our approach, because all the terms in (1) are forecasts.

We estimate (1) for each respondent who made more than a minimum number of forecasts of inflation and the unemployment rate. As remarked by Clements (2024), model heterogeneity will arise for a number of reasons. If a respondent's forecasts do not reflect a linkage between inflation and the unemployment rate, we would expect γ to be zero for that respondent—his/her "model" is not PC theory consistent.

For respondent j, we estimate (1) using as data the *forecasts* of that respondent:

$$E_{j,t}\pi_{t+h} = \beta_{b,j}E_{j,t}\pi_{t+h-1} + \beta_{f,j}E_{j,t}\pi_{t+h+1} + \gamma_j E_{j,t}u_{t+h} + e_{j,t,h}, \tag{2}$$

where $E_{j,t}\pi_{t+h}$ are j's h-step-ahead forecasts of π_{t+h} made at time t, and $E_{j,t}u_{t+h}$ is j's h-step-ahead forecasts of u_{t+h} made at time t. The model relates three different step-ahead inflation forecasts to a single horizon unemployment rate forecast. Our source of forecast data, the US SPF (described in Appendix B) allows us to estimate (2) for left-hand-side forecast horizons of $h = 0,1,2$, and 3,[7] which we can estimate as a system:

$$\begin{aligned}
E_{j,t}\pi_t \;\; &= \;\zeta_j + \beta_{b,j}E_{j,t}\pi_{t-1} + \beta_{f,j}E_{j,t}\pi_{t+1} + \gamma_j E_{j,t}u_t + e_{j,t,0}\\
E_{j,t}\pi_{t+1} &= \;\zeta_j + \beta_{b,j}E_{j,t}\pi_t + \beta_{f,j}E_{j,t}\pi_{t+2} + \gamma_j E_{j,t}u_{t+1} + e_{j,t,1}\\
E_{j,t}\pi_{t+2} &= \;\zeta_j + \beta_{b,j}E_{j,t}\pi_{t+1} + \beta_{f,j}E_{j,t}\pi_{t+3} + \gamma_j E_{j,t}u_{t+2} + e_{j,t,2}\\
E_{j,t}\pi_{t+3} &= \;\zeta_j + \beta_{b,j}E_{j,t}\pi_{t+2} + \beta_{f,j}E_{j,t}\pi_{t+4} + \gamma_j E_{j,t}u_{t+3} + e_{j,t,3},
\end{aligned} \tag{3}$$

requiring that the parameters are the same across the equations. Jain (2019) used such a system approach to improve the precision of her estimates of inflation persistence, although this will come at a cost if the parameters differ across equations. We compare results using both the individual equation estimates (from 2, for each h separately) and using the system (for all h together, from 3). When we use a system approach, we estimate the equations by GLS assuming the following error structure: $E(e_{j,t,i}e_{j,s,k}) = \sigma_{ik}$ when $t = s$, but is zero otherwise. That is, the errors in the equations for the forecasts made by individual j in response to the same surveys are allowed to be correlated, reflecting the impact of unmodeled factors at time t on forecasts at all horizons. See Clements (2024) for details.

3 | MEASURING THE RESPONSE TO COVID-19

We consider the set of respondents who responded to at least 20 surveys during the period 1981:Q3 to 2019;Q4 and then went on to respond to at least four of the surveys during 2020:Q1 to 2022:Q4, inclusive.

For each respondent, we calculated a hybrid PC for their $h = 0$, $h = 1$, $h = 2$, and $h = 3$ forecasts, either singly, or as a system, for the pre-COVID-19 period (up to and including 2019:Q4), and recorded estimates of the slope parameter and the R^2. The slope measures the responsiveness of the inflation forecasts to the unemployment rate forecasts, and the R^2 the proportion of the variation in the respondent's inflation forecasts which is attributable to his/her PC model. Taken together, these two statistics measure the extent of the respondent's reliance on a PC relationship in the generation of their expectations.

Given the estimates of the unknown parameters in (3), for a given respondent j, we calculate j's PC model forecasts as follows. The 2020:Q1 current-quarter model-based forecast is the value predicted by the first equation of (3) using the estimated values of $\zeta_j, \beta_{b,j}, \beta_{f,j}, \gamma_j$ and the reported forecasts as the right-hand-side variables: the 2020:Q1 forecast of inflation in 2019:Q4, the 2020:Q1 forecast of inflation in 2020:Q2, and the 2020:Q1 forecast of the unemployment rate in 2020:Q1. The 2020:Q2 current-quarter model-based forecast uses the same parameter estimates, but the right-hand-side variables are the 2020:Q2 forecast of inflation in 2020:Q1, the 2020:Q2 forecast of inflation in 2020:Q3, and the 2020:Q2 forecast of the unemployment rate in 2020:Q2. We repeat up to 2022:Q4.

The same approach is used for the $h = 3$ forecasts, but now, we use the last equation of (3), again estimated on the surveys up to 2019:Q4. The $h = 3$ forecast made at time 2020:Q1 is of target quarter 2020:Q4 and uses the 2020:Q1 forecast of inflation in 2020:Q3, the 2020:Q1 forecast of inflation in 2021:Q1, the 2020:Q1 forecast of the unemployment rate in 2020:Q4, and so on.

There are two main focuses of our investigation.

First, we measure respondents' adherence to their pre-COVID-19 PC models. We do so by measuring how well the PC models explain the COVID-19 period reported forecasts. We calculate the (square root) of the average squared error between the reported and model-implied forecasts, during 2020:Q1–2022:Q4 and compare this to the models' estimated standard errors, which measure how well the models explain the reported forecasts during the pre-COVID-19 in-sample period. Comparing a model's squared forecast errors with the model's error variance (multiplied by a function of the explanatory variables) underpins the Chow (1960) test statistic, which was used by Clements and Hendry (2002) as a test of "forecast failure" (FF) Clements and Hendry (2002) define FF "as significant mis-forecasting relative to the previous record (in-sample, or earlier forecasts), whereas poor forecasting is judged relative to some standard, either absolute (perhaps because of a policy requirement for accuracy), or relative to a rival model." The notion of FF suits our purpose here, interpreted as whether or not the PC model "fails" to explain forecast behavior over the COVID-19 period. We might expect an individual's PC model forecasts of her reported forecasts over COVID-19 to be "worse" than anticipated, based on the in-sample fit, if the reported forecasts reflect a perceived breakdown in the pre-COVID-19 relationship between inflation and unemployment, and, or if, there are abnormally large idiosyncratic "errors" resulting from one-off factors that affect the relationship.

We also consider whether in-sample fit is correlated with the out-of-sample RMSFE between the reported forecasts and model forecasts *across* respondents. If respondents do not change their forecasting behavior much we would expect a positive correlation: Those with small (large) in-sample model standard errors will tend to have small (large) differences between their model and reported forecasts.[8] For each respondent, we calculate a measure of FF: the ratio of the square root of the averaged PC model errors (in terms of predicting the reported forecasts), out-of-sample, to the in-sample fit of their PC models. We also consider whether FF depends on in-sample characteristics of the PC models. That is, whether FF is related to the R^2 and the slope parameter. Of interest is whether those whose forecasts bear a strong PC model imprint are more or less likely to exhibit FF.

Second, we consider whether professionals *ought* to have put their trust in the PC over the COVID-19 period. Comparing the forecast accuracy of the reported forecasts to that of implied PC model forecasts is unlikely to provide a fair assessment of the value of the PC relationship. This is because the reported forecasts may draw on knowledge of the PC relationship and in addition will likely also incorporate information on the current state of the economy and idiosyncratic factors, giving them an edge over the model forecasts.[9] One might expect a close adherence to a PC model to count against forecast accuracy if it comes with a failure to take on board COVID-19-related developments.

We compare the PC model forecasts during the COVID-19 period to the simple "no-change" predictor proposed for US inflation by Atkeson and Ohanian (2001) (henceforth referred to as AO forecasts). Atkeson and Ohanian (2001) suggested using a simple average of the four quarterly inflation rates up to the forecast origin, a device found by the authors to be competitive with standard Phillips curve forecasting models.[10] AO and PC forecasts are on a level-playing field in that neither incorporate information about current and unfolding events.[11] We consider whether the relative accuracy of the PC model and AO forecasts depends on the characteristics of the PC model (such as the activity slope parameter).[12]

The COVID-19 period - taken here to comprise forecasts reported to surveys in 2020:1 to 2022:4 inclusive - is necessarily short, and heterogeneous. This suggests caution in interpreting and generalizing the results.

As an important check on whether there is a "COVID effect" on forecaster behavior we compare our results with those obtained for a non-COVID period. After all, FF as defined above might simply reflect worse out-of-sample performance than expected given the model's in-sample performance, and this may occur at any time and not be the result of a response to COVID. For this reason, we repeat the empirical analysis for the period 2017:1 to 2019:4, having estimated the models on the forecasts from the surveys up to 2016:4. This 3-year out-of-sample period is adjacent to our COVID forecasting period, because this choice maximizes the size of the in-sample period, and hence, the availability of individual forecasts to estimate the models.[13]

4 | EMPIRICAL FINDINGS

4.1 | Do respondents adhere to their pre-COVID-19 PC models?

Figure 1 shows the relationship between in-sample PC model fit, and the extent to which "out-of-sample" COVID-period forecasts deviate from PC model forecasts, for each forecaster, for $h=0$ and $h=3$.[14]

The figure shows that $RMSFE_{SPF,PC}$ exceeds $\hat{\sigma}$ for most forecasters for $h=0$, and there is some suggestion that the two are positively correlated across respondents (formally tested below). For $h=3$, the two are more broadly in line, than for $h=0$, indicating a clear difference between the short and medium horizons in this regard. At the shortest horizon, the reported inflation forecasts during COVID-19 differ more from the PC model than during the in-sample period. For $h=3$, the post-2020 forecasts are more in line with what is expected based on pre-2020 behavior.

As a supplement to Figure 1, Table 1 reports the averages of the individual FF measures across respondents, by forecast horizon, as well as the quartiles of the FF

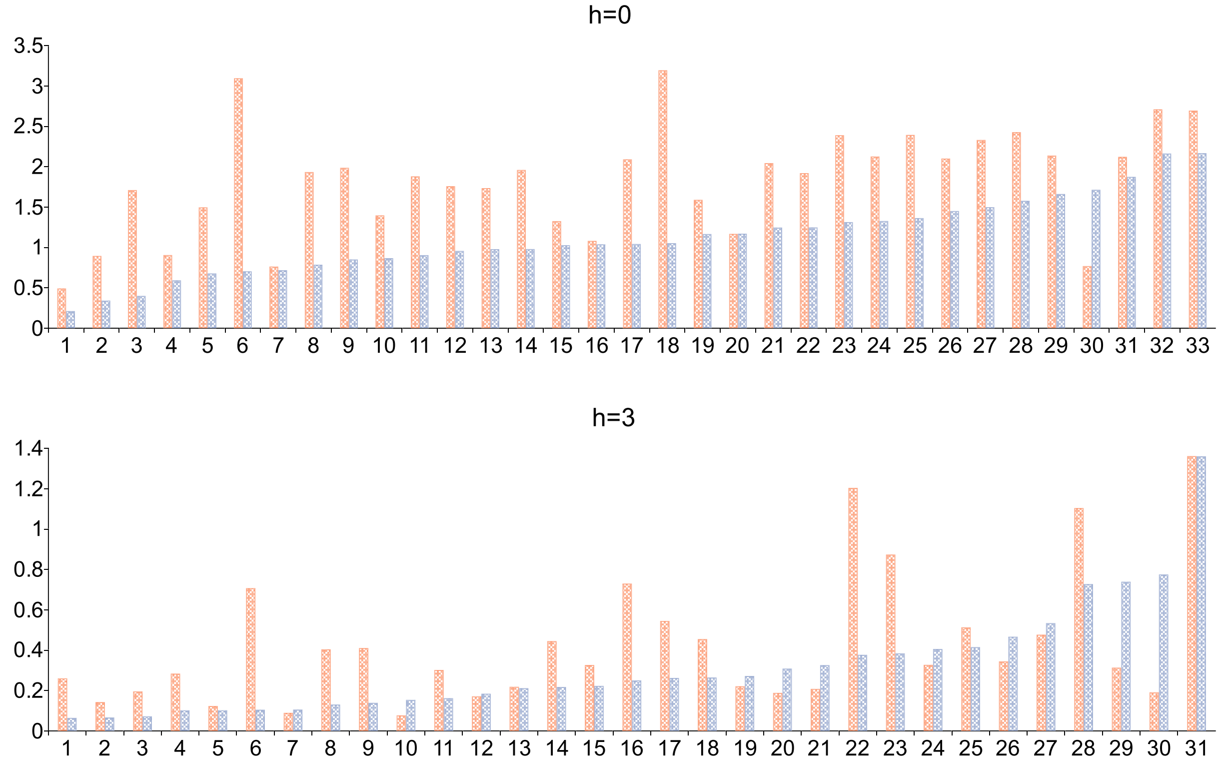

FIGURE 1 In-sample and out-of-sample performance of individuals' Hybrid PC models. The first bar is $RMSFE_{SPF,PC}$, the "distance" between the reported and model forecasts out-of-sample, and the second is the in-sample PC model estimated standard error. Individuals are ordered by the second variable. We graph the findings for $h=0$ and $h=3$, the shortest and longest horizons. Here and in the remaining figures, the PC model is estimated separately for each forecast horizon.

TABLE 1 Forecast failure (FF): Hybrid PC models (single-equation estimation).

	COVID 2020–2022				2017–2019			
h	0	1	2	3	0	1	2	3
LQ	1.288	1.152	0.947	0.812	0.506	0.567	0.502	0.627
Median	1.606	1.614	1.417	1.350	0.675	0.770	0.676	0.885
UQ	2.060	2.542	2.014	2.381	0.857	1.025	1.128	1.106

Note: The entries are the quartiles of the distribution of individual FF values for the horizon h given in the column.

values, as an indicator of the heterogeneity across respondents.[15] We report these summary statistics for the COVID forecast period, 2020–2022, as well as for the "normal" period, 2017–2019. The table clearly shows much higher FF values during COVID, as indicated by a comparison of the median values across the two periods. Notwithstanding the relatively wide dispersion of FF values across respondents, for $h=0,1$, the COVID-period lower quartiles exceed the normal period upper quartiles.

For a more formal assessment of the COVID period, Table 2 reports the rank correlations[16] (across respondents) between $\hat{\sigma}$, and $RMSFE_{SPF,PC}$.[17] We find a positive rank correlation at all horizons, suggesting that individuals' whose forecasts were well described by their estimated PC models (small estimated standard error) tend to have more closely matching out-of-sample (COVID-19 period) PC model and reported forecasts. This suggests a tendency to persevere with PC models during the COVID-19 period. These results are based on estimating the PC separately for each h (Equation 2). In Section 6, we investigate whether this finding is robust to estimating the PC models by pooling over h.

Table 2 also suggests FF—the ratio of $RMSFE_{SPF,PC}$ to $\hat{\sigma}$—is related to the in-sample slope coefficient $\hat{\gamma}$ and the in-sample R^2 (where in-sample is again the pre-COVID-19 period 1981:3 to 2019:4) at the longer forecast

TABLE 2 Pre-pandemic belief in the Phillips curve and forecasting during the pandemic: Hybrid Phillips curve, with horizon-specific estimates.

h	$\hat{\sigma}$ and $RMSFE_{SPF,PC}$		FF and $\hat{\gamma}$		FF and R^2	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
0	0.574	0.000	0.092	0.312	0.190	0.153
1	0.550	0.001	−0.240	0.900	0.090	0.318
2	0.496	0.002	0.339	0.032	0.466	0.004
3	0.466	0.005	0.392	0.017	0.382	0.019

Note: The statistics are the Spearman rank correlation coefficients.

horizons. The second panel of the table shows that FF is systematically positively related to the slope at $h=3$, such that a larger negative γ is correlated with less FF (smaller ratio of $RMSFE_{SPF,PC}$ to $\hat{\sigma}$). The third panel suggests the higher the R^2, the higher FF, at $h=2,3$: Those whose forecasts were well approximated by a PC model in-sample are more likely to experience FF.

For the normal period, we also find a statistically significant positive rank correlation between $\hat{\sigma}$ and $RMSFE_{SPF,PC}$ across respondents at all horizons, but in contrast to the COVID period, FF is no longer related to the slope coefficient (results not shown).

4.2 | Ought respondents put their trust in their PC models?

For the COVID period, at the shortest horizon, Figure 2 shows that reported forecasts are more accurate than the model forecasts and AO forecasts for most respondents, as expected given the additional information that the reported forecasts would be expected to draw on. There is little to choose between the model forecasts and AO forecasts. However, there is again a divergence between the short- and medium-term horizons. At $h=3$, the reported model forecasts and AO forecasts are more closely in tune.[18]

Table 3 complements Figure 2 by providing the medians and quartiles of the cross-sectional distributions of the individual ratios of the RMSFEs of reported to the PC model forecasts, and of the PC model to AO forecasts. Table 3 panel A shows that for the COVID period, the median ratio of the RMSFE of the reported forecasts to the model forecasts is 0.754. At the two longer horizons, there is little to choose between the reported and model forecasts. The same is broadly true of the 2017–2019 normal forecast period: The reported forecasts are more accurate than the model forecasts at $h=0$, but similar at $h=2,3$. One difference is that the distribution of the ratios of the RMSFEs of the reported to model forecasts is markedly smaller for the COVID period (e.g., compare the inter-quartile ranges for $h=3$).

If we view these ratios as reflecting the role of judgment and/or recent developments, then judgment has a similar effect in the two forecast periods. This interpretation supposes respondents act "as if" they have models, which they use to generate forecasts, and then adjust. Recall that we have inferred their models by estimating PC relationships on an in-sample period, so we do not know if this is valid as a behavioral description.[19]

What of the AO forecasts? Table 3 panel B indicates the AO forecasts are markedly more accurate on average than the PC model forecasts for $h=1,2$ in the COVID period compared with normal times (compare the median ratios in the two periods), supporting the use of the "robust" forecasting device in uncertain times. In fact, as indicated by panel C, the AO forecasts are more accurate than the reported forecast except at $h=0$, by around 10% on RMSE at $h=1,2$.

Next, regardless of which model or method is more accurate, we consider whether the relative accuracy of the model forecasts to the AO forecasts varies across the cross-section depending upon the closeness of the reported and model forecasts and characteristics of the (in-sample) PC models, such as the R^2 and slope parameter. See Table 4. We find that $RMSFE_{PC}/RMSFE_{AO}$ is positively correlated with $RMSFE_{SPF,PC}$ at the shortest horizon.[20] Model forecasts that are closer to the reported forecasts are more accurate than the AO forecasts out-of-sample.

A more negative $\hat{\gamma}$ is associated with less accurate model forecasts, that is, a higher $RMSFE_{PC}/RMSFE_{AO}$, at $h=3$, counting against the value of the Phillips curve. However, a higher R^2 is significantly associated with a lower value $RMSFE_{PC}/RMSFE_{AO}$ at $h=0$. Hence, a greater reliance on the PC relationship in-sample gives rise to more accurate model forecast (relative to AO, out-of-sample) when the strength of the PC model is defined by the R^2 but not when it is defined by the degree of responsiveness of inflation to the measure of slack. We consider this issue in more depth subsequently by

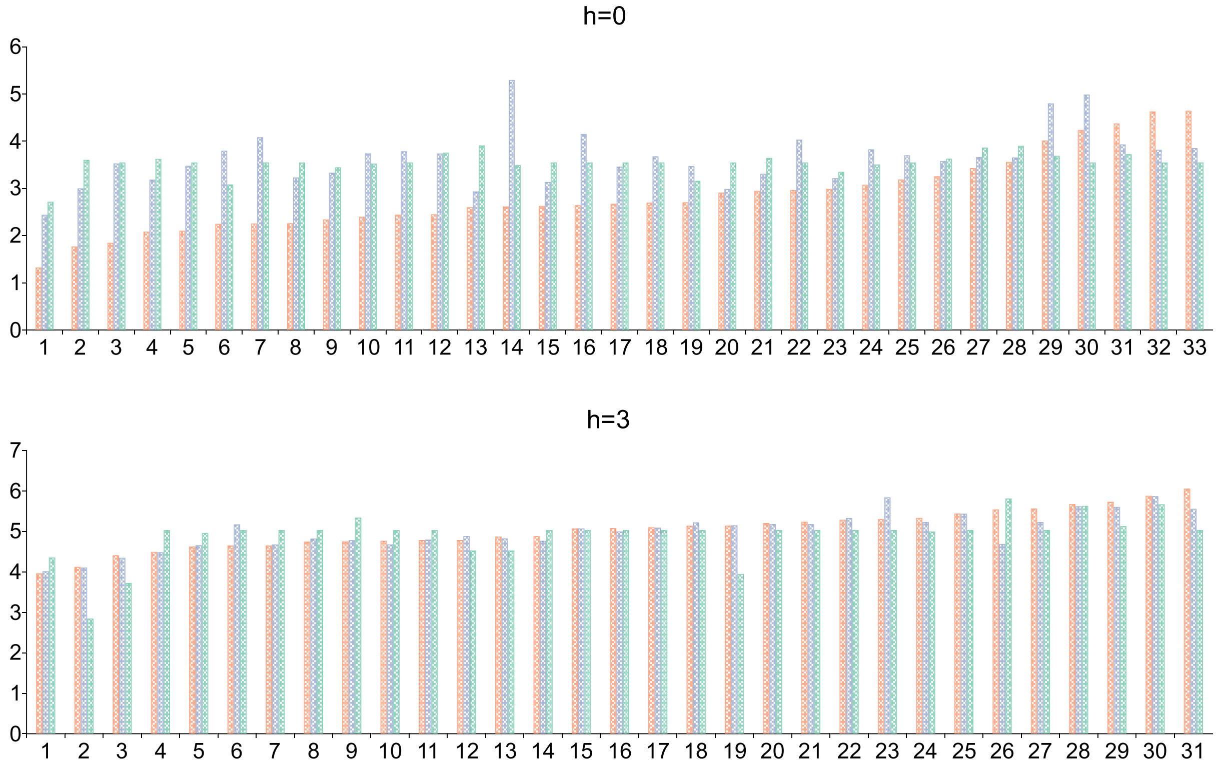

FIGURE 2 Out-of-sample RMSFEs. The first bar is the RMSFE of the actual SPF forecasts, the second is the RMSFE of the hybrid PC model forecasts, and the third is the RMSFE of the AO forecasts. Individuals are ordered by the first variable. We graph the findings for $h = 0$ and $h = 3$, the shortest and longest horizons.

TABLE 3 Relative forecast accuracy: Hybrid PC models (single-equation estimation).

	COVID 2020–2022				**2017–2019**			
h	**0**	**1**	**2**	**3**	**0**	**1**	**2**	**3**
	Panel A. Reported forecast RMSFE to PC model RMSFE							
LQ	0.638	0.899	0.989	0.993	0.545	0.915	0.945	0.949
Median	0.754	0.955	1.004	1.002	0.715	0.984	1.021	1.014
UQ	0.889	0.980	1.022	1.015	0.832	1.059	1.098	1.103
	Panel B. PC model RMSFE to AO forecasts RMSFE							
LQ	0.917	1.089	1.005	0.945	0.861	0.982	0.915	0.967
Median	0.995	1.154	1.074	1.029	0.921	1.053	0.963	1.044
UQ	1.091	1.210	1.158	1.069	1.010	1.148	1.101	1.131
	Panel C. Reported forecast RMSFE to AO forecasts RMSFE							
LQ	0.656	1.010	1.009	0.946	0.536	0.901	0.835	0.978
Median	0.751	1.098	1.091	1.017	0.643	1.024	0.949	1.062
UQ	0.891	1.170	1.153	1.069	0.832	1.135	1.105	1.248

Note: The entries are the quartiles of the distribution of the individual ratios, specified in the given panels, for the horizon *h* given in the column.

replacing the hybrid PC with a "backward-looking" PC, as has been used in some studies of survey expectations.

In summary, our key findings are that respondents' forecasts are much less aligned with their (imputed) PC model forecasts during COVID than during the earlier "normal" period. However, across forecasters, we find that those whose forecasts are more closely aligned with a PC model in-sample are more likely to produce out-

TABLE 4 Phillips curve inflation forecasts versus the AO benchmarks during the pandemic: hybrid Phillips curve, with horizon-specific estimates. Rank correlation between the ratio of the PC model MSFE to AO MSFE and the closeness of the reported and model forecasts, and characteristics of the (in-sample) PC models.

	$RMSFE_{SPF,PC}$		$\hat{\gamma}$		R^2	
h	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
0	0.644	0.000	0.015	0.467	−0.484	0.998
1	0.196	0.149	−0.066	0.635	−0.188	0.840
2	0.074	0.349	−0.384	0.983	−0.327	0.962
3	0.126	0.258	−0.483	0.997	−0.185	0.832

Note: The statistics are the Spearman rank correlation coefficients.

of-sample forecasts which more closely align with that model, during both the normal and COVID periods.

For the majority of individuals, we found that the accuracy of the reported forecasts was close to that of the PC model forecasts and AO forecasts at the longest horizon (three quarters ahead) but not at the shortest. There is evidence that using a no-change predictor (AO) improves on the reported forecasts, and the PC model forecasts, beyond the shortest horizon forecasts, and to a greater extent during COVID than during normal times. Hence, although the reported forecasts are superior at the shortest horizon, this superiority does not extent beyond "nowcasting" of the current-quarter value. The gain from the exercising of judgment (if we interpret the reported forecasts as model forecasts *plus* judgment) is short lived.

5 | A CLOSER LOOK IN THE TIME DIMENSION

Figures 3 and 4 show the current-quarter and three-step-ahead median forecasts, respectively, for the COVID-19 period. The dates on the horizontal axis refer to the dates of the surveys. These correspond to the target periods for the current-quarter forecasts but not for the longer horizon forecasts. For example, in Figure 4, the 2020:Q2 forecasts are of 2021:Q1, and the actual value is the 2021:Q1 inflation rate (which is the actual corresponding to the 2021:Q1 survey in Figure 3). Figure 3 shows the actual rate of inflation fell sharply in 2020:Q2, before increasing by around 5% in the third quarter, and increasing further and remaining high until the second half of 2022. The reported median SPF forecasts anticipated the fall in the second quarter, and some of the rebound in the third quarter, and were markedly more accurate than the median PC forecasts, or the AO forecasts, for these two quarters. Thereafter, the reported forecasts are not much better than the others. In 2021:Q2, for instance, all forecasts were far too low,[21] and all the forecasts underpredicted inflation throughout 2021. (Recall that Figure 3 depicts current-quarter forecasts.) Clear divergences between reported and PC model forecasts occurred in 2020:Q2 and Q3, but not subsequently, although the model forecasts were generally too low.

In contrast, the longer horizon median reported and PC model forecasts (Figure 4) generally move in lockstep and fail to anticipate the high rates of inflation that occurred. The median reported and PC model $h=3$ forecasts deviate little from 2% or so throughout the period. The AO forecasts pick up as they start to feed off the higher realized inflation rates we move through 2021.

These results suggest the main differences between the short-horizon reported and PC model forecasts *on average* resulted from the forecasts made in 2020:Q2 and Q3. For the longer horizon, the differences mainly stem from 2020:Q3. Of course, the averages hide differences between individuals, and these are studied in Section 4.

6 | ROBUSTNESS OF THE FINDINGS

6.1 | The specification of the Phillips curve

As discussed in Section 1, some studies simply relate forecasts of inflation to forecasts of the unemployment rate and do not include a forward-looking inflation term. To see whether our results using the forward-looking, hybrid specification carry over to the simpler specification, we repeat the analysis with the PC model:

$$\pi_t = \beta_{bw}\pi_{t-1} + \gamma_{bw}u_t + \varepsilon_{bw,t}, \tag{4}$$

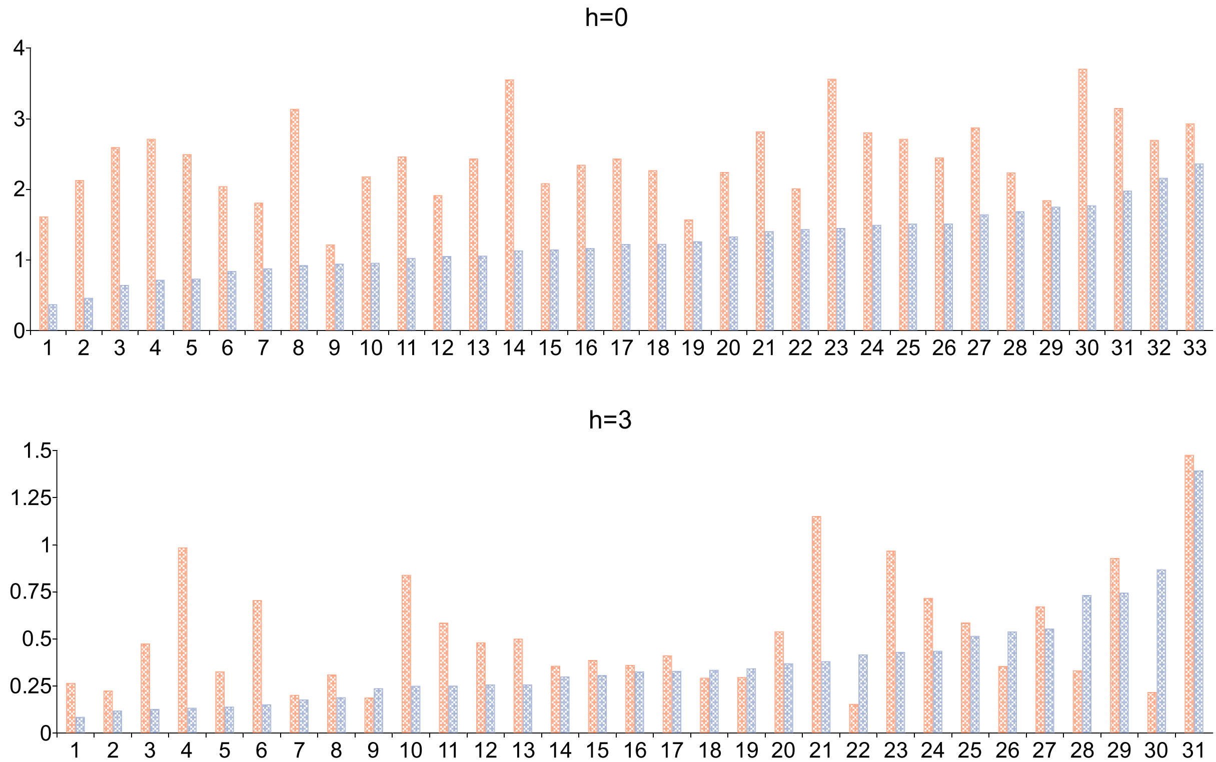

FIGURE 3 Current-quarter median forecasts for surveys 2020:Q1 to 2022:4.

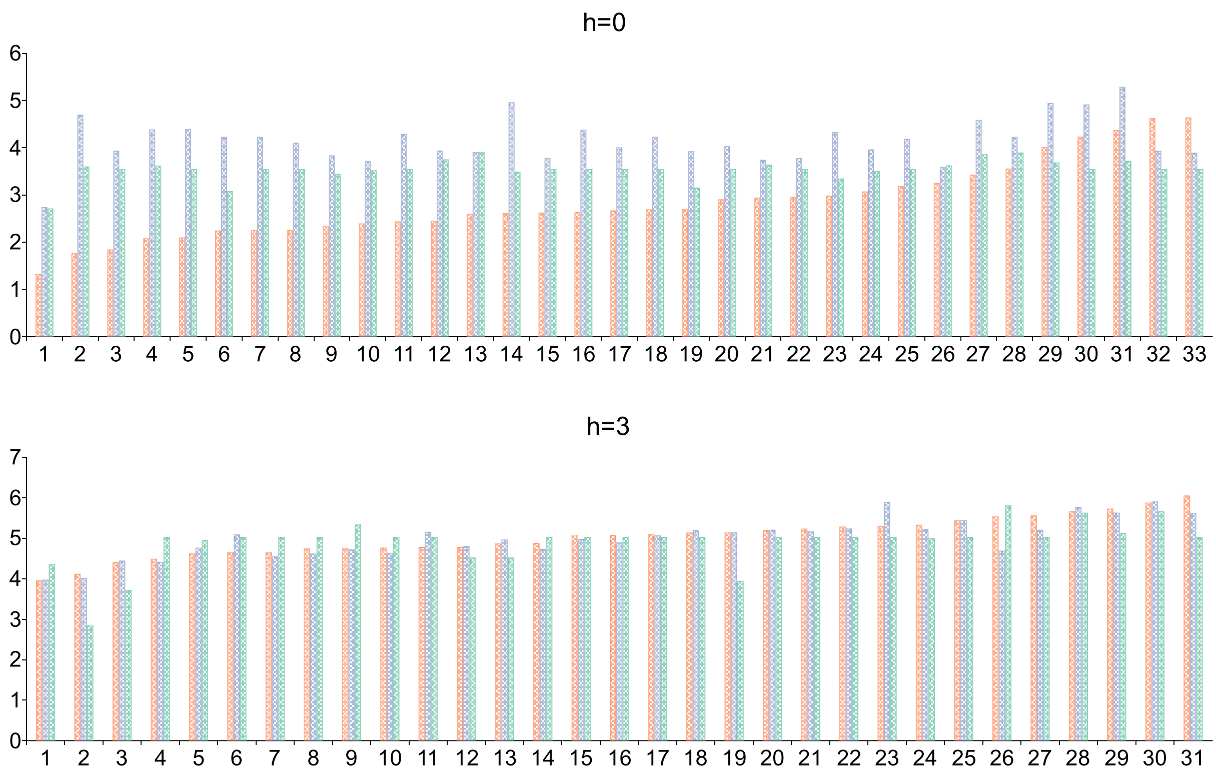

FIGURE 4 $h = 3$ median forecasts for surveys 2020:Q1 to 2022:4.

in place of (1). Without the forward-looking term, the SPF data would allow a five-equation system to be estimated, with an equation for the $h = 4$ inflation rate (compare to the four-equation system 3 for the hybrid model). However, for ease of comparability to the results for the hybrid model, we do not use this additional equation.

The results for the backward model are recorded in Figures 5 and 6, and Tables 5 to 8. Using such models, we replicate the finding for the hybrid model that the FF values are much higher in the COVID period than the normal period (compare Table 5 to Table 1).

Table 6 suggests FF is not related to the PC model slope for the backward model (compare to Table 2), although as before, FF does appear to depend on the in-sample R^2.

The findings for the backward model are also similar to the hybrid when we compare the accuracy of the model forecasts to the reported forecasts (compare Table 7 panel B to Table 3 panel B). The median ratio is 0.71 (COVID period, and 0.67, normal times) at the shortest horizon, and beyond that is much closer to 1. As before, the model forecasts are worse than the AO forecasts during the COVID period. (For example, the median ratio of the model to AO RMSFE is some 15% higher at $h = 0,1$ during the COVID period.) Table 8 is broadly similar to Table 4, suggesting that it is also the case that in the backward model, more negative slopes (in-sample) are associated with less accurate PC mode forecasts relative to the AO and that higher R^2 values are associated with more accurate PC model forecasts.

We conclude that the main findings are not sensitive to the choice of a hybrid versus a backwards Phillips curve model.

6.2 | Single-equation versus systems estimation

If an individual's PC is the same for each h, instead of estimating the model separately for each h, we would obtain more precise estimates by pooling and estimating the system (Equation 3), rather than estimating each equation separately, as hitherto. In this section we compare results for the4 hybrid model. Table 9 replicates table 2, but using pooled estimates. (The comparison between reported and AO forecasts would of course be identical to panel C of Table 2, and so is omitted from Table 9). The results for pooled estimation are qualitatively similar to the single-equation results. The reported forecasts are more accurate at $h = 0$, and the AO forecasts generally out-perform the model forecasts by a greater margin during COVID (approximately 5% and 14% more accurate) than in normal times.

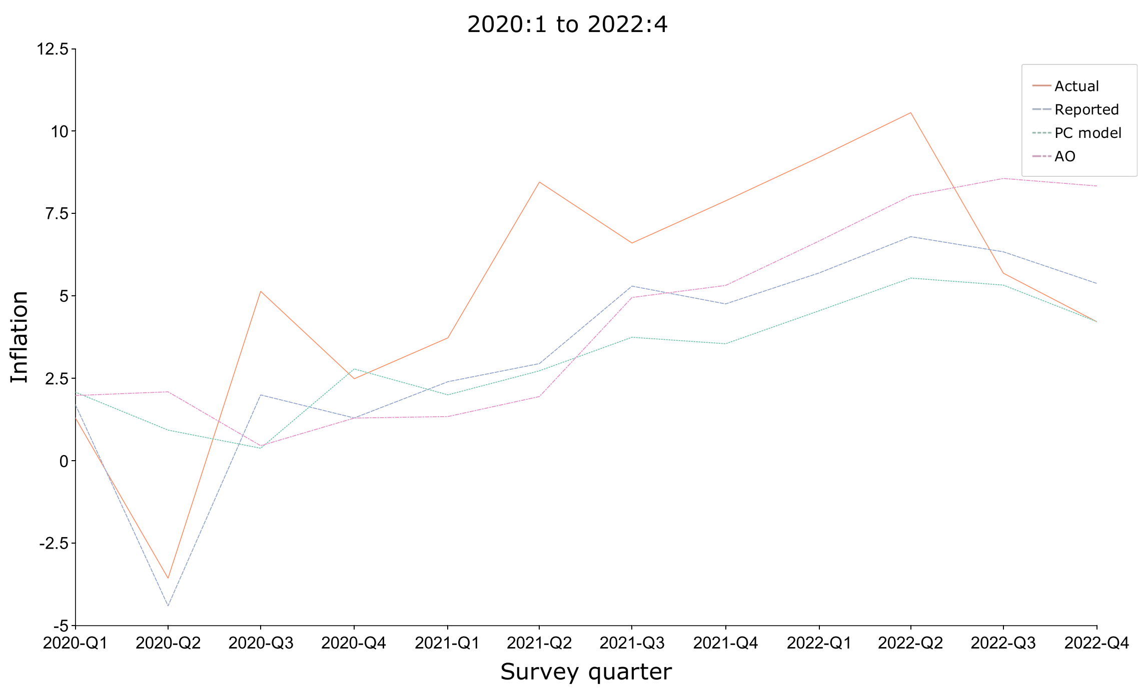

FIGURE 5 In-sample and out-of-sample performance of individuals' backward-looking PC models. The first bar is $RMSFE_{SPF,PC}$, the "distance" between the reported and model forecasts out-of-sample, and the second is the in-sample PC model estimated standard error. Individuals are ordered by the second variable. We graph the findings for $h = 0$ and $h = 3$, the shortest and longest horizons.

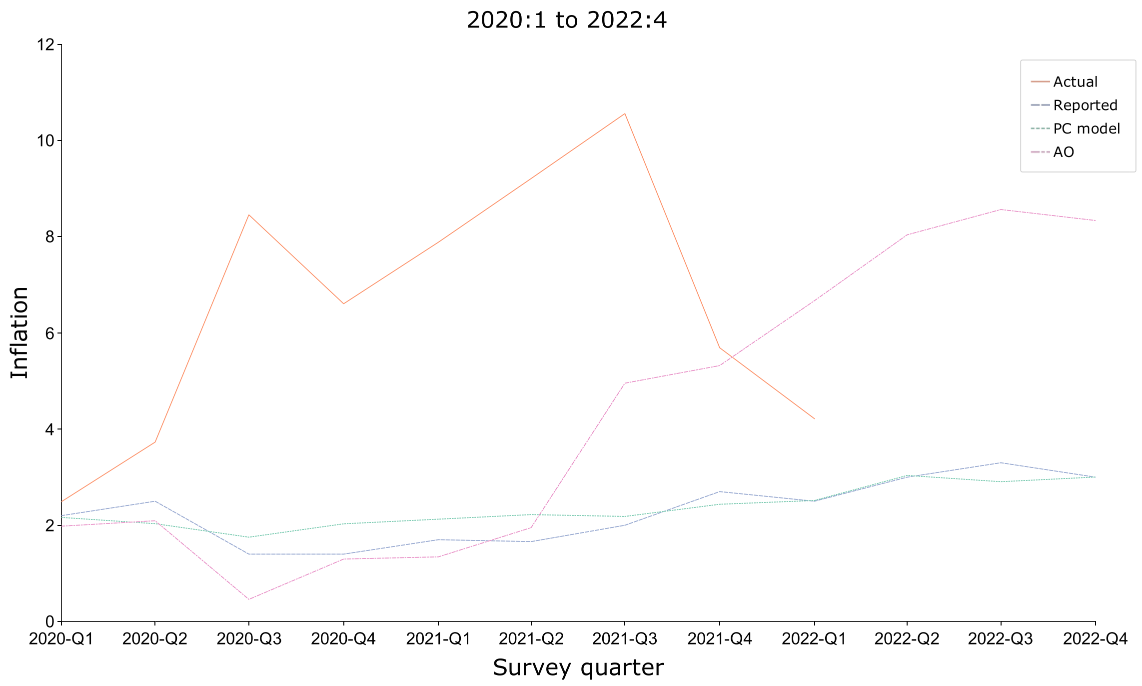

FIGURE 6 Out-of-sample RMSFEs. The first bar is the RMSFE of the actual SPF forecasts, the second is the RMSFE of the backward-looking PC model forecasts, and the third is the RMSFE of the AO forecasts. Individuals are ordered by the first variable. We graph the findings for $h=0$ and $h=3$, the shortest and longest horizons.

TABLE 5 Forecast failure: Backward PC models (single-equation estimation).

	COVID 2020–2022				2017–2019			
h	0	1	2	3	0	1	2	3
LQ	1.598	1.167	0.780	1.013	0.490	0.520	0.593	0.576
Median	1.932	1.757	1.304	1.256	0.609	0.773	0.746	0.769
UQ	2.419	3.069	1.936	2.270	0.761	0.965	1.182	1.135

Note: The entries are the quartiles of the distribution of individual FF values for the horizon *h* given in the column.

TABLE 6 Pre-pandemic belief in the Phillips curve and forecasting during the pandemic: Backward Phillips curve, with horizon-specific estimates.

h	$\hat{\sigma}$ and $RMSFE_{SPF,PC}$		FF and $\hat{\gamma}$		FF and R^2	
	Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
0	0.382	0.016	0.123	0.255	0.245	0.091
1	0.144	0.223	−0.234	0.893	0.346	0.029
2	0.234	0.106	0.277	0.069	0.408	0.012
3	0.257	0.088	0.157	0.208	0.473	0.004

Note: The statistics are the Spearman rank correlation coefficients.

TABLE 7 Relative forecast accuracy: Backward PC models (single-equation estimation).

	COVID 2020–2022				2017–2019			
h	**0**	**1**	**2**	**3**	**0**	**1**	**2**	**3**
Reported forecast RMSFE to PC model RMSFE								
LQ	0.538	0.872	1.002	0.991	0.586	0.914	0.918	0.937
Median	0.666	0.954	1.019	1.006	0.706	0.975	1.000	1.033
UQ	0.782	1.000	1.039	1.023	0.963	1.070	1.080	1.117
PC model RMSFE to AO forecasts RMSFE								
LQ	1.071	1.117	1.016	0.935	0.832	0.996	0.933	0.989
Median	1.147	1.164	1.066	1.027	0.880	1.047	0.966	1.049
UQ	1.238	1.229	1.137	1.067	0.910	1.122	1.126	1.164

Note: The entries are the quartiles of the distribution of the individual ratios, specified in the given panels, for the horizon h given in the column.

TABLE 8 Phillips curve inflation forecasts versus the AO benchmarks during the pandemic: Backward Phillips curve, with horizon-specific estimates. Rank correlation between the ratio of the PC model MSFE to AO MSFE and the closeness of the reported and model forecasts, and characteristics of the (in-sample) PC models.

	$RMSFE_{SPF,PC}$		$\hat{\gamma}$		R^2	
h	**Statistic**	***p*-value**	**Statistic**	***p*-value**	**Statistic**	***p*-value**
0	0.484	0.002	−0.051	0.607	−0.612	1.000
1	0.433	0.008	−0.463	0.996	−0.179	0.828
2	0.440	0.007	−0.380	0.982	−0.365	0.977
3	−0.127	0.745	−0.457	0.994	−0.193	0.843

Note: The statistics are the Spearman rank correlation coefficients.

TABLE 9 Relative forecast accuracy: Hybrid PC models (pooled-over-horizon estimates).

	COVID 2020-2022				2017–2019			
h	**0**	**1**	**2**	**3**	**0**	**1**	**2**	**3**
Panel A. Reported forecast RMSFE to PC model RMSFE								
LQ	0.626	0.956	0.987	0.984	0.567	0.907	0.963	0.942
Median	0.787	0.981	1.005	1.003	0.705	0.940	1.018	1.008
UQ	0.877	1.007	1.014	1.014	0.857	1.023	1.078	1.088
Panel B. PC model RMSFE to AO forecasts RMSFE								
LQ	0.923	1.048	1.021	0.948	0.842	0.979	0.869	0.960
Median	0.973	1.109	1.080	1.020	0.894	1.057	0.943	1.056
UQ	1.077	1.164	1.154	1.070	1.003	1.133	1.097	1.217
Panel C. Reported forecast RMSFE to AO forecasts RMSFE								
LQ	0.656	1.010	1.009	0.946	0.536	0.901	0.835	0.978
Median	0.751	1.098	1.091	1.017	0.643	1.024	0.949	1.062
UQ	0.891	1.170	1.153	1.069	0.832	1.135	1.105	1.248

Note: The entries are the quartiles of the distribution of the individual ratios, specified in the given panels, for the horizon h given in the column.

7 | CONCLUSIONS

We consider whether professional forecasters react differently in the COVID-19 period, in terms of forecasting CPI inflation. Do their forecasts of inflation and the unemployment rate conform to Phillips curve relationships to the same degree during the pandemic, as they did during "normal times"? Naturally, forecasters are expected to have made larger forecast errors than usual due to the effects of unforeseen events. Faced by this additional uncertainty, do they persevere with their PC models, supplemented by judgment regarding the likely effects of unfolding events, or look to alternative ways to forecast inflation?

We find that by and large professionals do not radically alter their behavior, in the sense that respondents whose forecasts conformed closely to a Phillips curve model pre-COVID-19 are more likely to produce forecasts which are consistent with that model during COVID-19. Naturally, the reported forecasts deviate from the PC model forecasts because they reflect additional information on how the evolving pandemic is expected to impact on inflation and do so to a greater extent than during "normal times." In this respect, our results are consistent with professional forecasters drawing on additional sources of information during COVID-19, such as those discussed by Ferrara and Sheng (2022) (and the papers therein), and using that information to chart the effects of the pandemic, and the attempts of governments at mitigation, on the macroeconomy. Our results do not suggest a wholesale rejection of forecasting with Phillips curves.

Comparing reported forecasts, PC model forecasts and "no-change" AO forecasts during COVID-19, we find that for most respondents, the short-term ($h=0$) reported forecasts are markedly more accurate, and the PC model and AO forecasts similar. This is consistent with the reported forecasts being able to draw on relevant information, explaining their superior performance at $h=0$. By $h=3$, this superiority has vanished when we consider the forecasters en masse, in the sense that the median ratio of the RMSFE of the reported and AO forecasts favors the AO forecasts. We find that the relative gains to AO over either the reported or model forecasts is greater during COVID-19 than in the normal comparator period (2017–2019).

Hence, there is evidence that the respondents would have done better, in terms of accuracy, to switch to a "robust" forecasting device (the AO forecasts). The comparison of the AO forecasts to the PC model forecasts is fairer than comparisons involving reported forecasts, in that neither draw on current-survey-quarter information, although that information only provides an edge at $h=0$. The AO forecast is just one example of a robust device—see Castle et al. (2015) for details—and alternatives may do better, but we do not pursue them here.[22]

Our findings are broadly the same if we replace the hybrid PC model with a backward PC model or pool across horizons when we estimate the PC models' unknown parameters.

ACKNOWLEDGMENTS

Helpful comments from conference participants at the International Symposium of Forecasting in Oxford, 2022, are gratefully acknowledged, as are the comments from a referee. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST STATEMENT

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Survey of Professional Forecasters at https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters and Real-time Data Research at https://www.philadelphiafed.org/surveys-and-data/real-time-data-research.

ORCID

Michael P. Clements https://orcid.org/0000-0001-6329-1341

ENDNOTES

[1] Ferrara and Sheng (2022) present an edited sample of such papers. "Big Data" often includes data from online information-gathering services: See, for example, Diebold (2003).

[2] Questions on how expectations are formed are not routinely included but are occasionally asked in special one-off surveys. For the US SPF respondents, Stark (2013) report: "We found that almost all respondents use a combination approach to forecasting: Twenty of 25 respondents said they use a mathematical/computer model plus subjective adjustments to that model in reporting their projections. (One respondent reported using pure model-generated forecasts, and four respondents said they use only their experience and intuition.) One interpretation of these results is that SPF panelists, like many macroeconomists in general, think models are useful but should not be fully trusted to deliver reasonable results in every circumstance."

[3] For example, if the proportions of PC model and non-PC model forecasters differed between the two periods. See Engelberg et al. (2011) and Manski (2011) on compositional effects.

[4] Clements (2022) highlights the importance of forecaster heterogeneity.

[5] These papers consider different formulations for the relationship between inflation and the unemployment rate. For example, Dräger et al. (2016) consider the number of times inflation and

unemployment forecasts move in opposite directions (at each point in time, across the cross-section of forecasters). Here, we consider forecasters adherence to Phillips curve relationships.

[6] Binder et al. (2021) show that perceived persistence also affects the term structure of uncertainty.

[7] $h=0$ denotes a forecast of the survey–quarter value. This is forecast because the "actual" value will not be revealed until the following quarter. $h=-1$ is a "forecast" of the previous quarter. These are reported by the SPF, and the respondents usually write in the advance actual values.

[8] We naturally expect that out-of-sample fit (to the forecasts) will generally worsen across the board.

[9] On the value of the timeliness of the information embedded in survey forecasts, see Faust and Wright (2009) and Clements (2015), inter alia.

[10] An update by Stock and Watson (2009) reports more nuanced findings, in particular that Phillip curve forecasting models display episodic good performance.

[11] The "no-change" predictor was one of the robust forecasting methods proposed by Clements and Hendry (1999) and Castle et al. (2015, 2016), inter alia.

[12] Note that when we calculate RMSFEs to compare the accuracy of the reported forecasts and the AO forecasts, for example, we only include AO forecasts for the periods for which the respondent in question made a forecast. Thus, it is a fair comparison in the sense that we are comparing forecasts for the same forecast origins (surveys).

[13] Earlier periods, such as the financial crisis period, would further restrict the individual-level forecast data required to estimate the PC models.

[14] The in-sample PC model fit, $\hat{\sigma}$, is the estimated standard error of the PC model, on the reported forecasts for the period 1981:Q3 to 2019:Q4: although there will typically be many missing values. The deviation between the model forecasts "out-of-sample" and the reported forecasts, denoted $RMSFE_{SPF,PC}$, is the square root of the average squared difference between the two sets of forecasts over the period 2020:Q1–2022:Q4.

[15] Recall that FF is $RMSFE_{SPF,PC}$ divided by $\hat{\sigma}$.

[16] The rank correlation coefficient is explained in Appendix A.

[17] There is a maximum of 12 current-quarter forecasts ($h=0$) but only nine $h=3$ forecasts. This is because the latest actual value is for 2022:4, so that the last $h=3$ forecast is made at time 2022:1. For some individuals there will we fewer than these maximum numbers, because of non-participation.

[18] Equality of two sets of forecasts on (R)MSFE does not of course imply that the two sets of forecasts are the same and leaves open the possibility that each set may continue useful information not included in the other. This can be made operational by testing whether or not a linear combination of the two forecasts is more accurate than either one alone (the notion of forecast encompassing of Chong and Hendry; 1986) but is not explored here.

[19] The role of judgment in macroforecasting has been studied by Turner (1990) and Clements (1995), inter alia.

[20] $RMSFE_{PC}$ and $RMSFE_{AO}$ denote the PC model and AO forecast RMSFEs. For respondent i, the calculation of the AO RMSFE includes only AO forecasts for surveys for which i made a forecast. $RMSFE_{PC}$ depends on h and i, whereas for a given origin, the AO forecasts do not depend on h. (As implemented, the AO forecasts are simply the average of the inflation rates of the previous four quarters, for all future periods.) Hence, $RMSFE_{AO}$ will be different for respondents who do not make forecasts for the same surveys.

[21] As shown in Figure 3, the actual inflation rate was a little over 8%, and the median reported forecast a little over 3%.

[22] AO forecasts will adapt more slowly than devices which average over shorter periods, for example.

REFERENCES

Aguiar, M., & Gopinath, G. (2007). Emerging market business cycles: The cycle is the trend. *Journal of Political Economy*, *115*, 69–102.

Atkeson, A., & Ohanian, L. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, *25*(1), 2–11.

Binder, C., McElroy, T., & Sheng, X. (2021). The term structure of uncertainty: New evidence from survey expectations. *Journal of Money, Credit and Banking*, *54*(1), 39–71.

Bluedorn, J. C., & Leigh, D. (2018). Is the cycle the trend? Evidence from the views of international forecasters. (*Working Paper 18/163*): International Monetary Fund. https://ideas.repec.org/p/imf/imfwpa/18-163.html

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

Castle, J. L., Clements, M. P., & Hendry, D. F. (2015). Robust approaches to forecasting. *International Journal of Forecasting*, *31*, 99–112.

Castle, J. L., Clements, M. P., & Hendry, D. F. (2016). An overview of forecasting facing breaks. *Journal of Business Cycle Research*, *12*(1), 3–23. https://doi.org/10.1007/s41549-016-0005-2

Cerrato, A., & Gitti, G. (2022). Inflation since covid: Demand or supply. https://doi.org/10.2139/ssrn.4193594

Chong, Y. Y., & Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, *53*, 671–690. Reprinted in Granger, C. W. J. (ed.) (1990), Modelling Economic Series. Oxford: Clarendon Press.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*, 591–605.

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

Clements, M. P. (2015). Are professional macroeconomic forecasters able to do better than forecasting trends? *Journal of Money, Credit and Banking*, *47*(2–3), 349–381.

Clements, M. P. (2020). Individual forecaster perceptions of the persistence of shocks to GDP. *Journal of Applied Econometrics*, *37*, 640–656.

Clements, M. P. (2022). Forecaster efficiency, accuracy and disagreement: Evidence using individual-level survey data. *Journal of Money, Credit and Banking*, *54*(2–3), 537–567.

Clements, M. P. (2024). Do professional forecasters believe in the Phillips curve? *International Journal of Forecasting*, *40*(3), 1238–1254.

Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*, The Zeuthen Lectures on Economic Forecasting. MIT Press.

Clements, M. P., & Hendry, D. F. (2002). Modelling methodology and forecast failure. *The Econometrics Journal*, *5*, 319–344.

Clements, M. P., Rich, R., & Tracy, J. (2023). Surveys of professionals, Chapter 3. In Ruediger Bachmann, G. T. (Ed.), *Handbook of economic expectations* (pp. 71–106). Academic Press, Elsevier.

Coibion, O., Gorodnichenko, Y., & Kamdar, R. (2018). 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*, *1993*, 3–15.

Curran, P. A. (2015). Monte Carlo error analyses of Spearman's rank test. International Centre for Radio Astronomy Research.

Diebold, F. X. (2003). Big data dynamic factor models for macroeconomic measurement and forecasting. In Dewatripont, M., Hansen, L. P., & Turnovsky, S. (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, pp. 115–122.

Dräger, L., Lamla, M. J., & Pfajfar, D. (2016). Are survey expectations theory-consistent? The role of central bank communication and news. *European Economic Review*, *85*(C), 84–111.

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

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

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

Ferrara, L., & Sheng, X. S. (2022). Guest editorial: Economic forecasting in times of COVID-19. *International Journal of Forecasting*, *38*(2), 527–528.

Hooper, P., Mishkin, F. S., & Sufi, A. (2019). 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.

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

Klein, L. R. (1947). The use of econometric models as a guide to economic policy. *Econometrica*, *15*, 111–151.

Krane, S. D. (2011). Professional forecasters' view of permanent and transitory shocks to GDP. *American Economic Journal: Macroeconomics*, *3*(1), 184–211.

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

McCallum, B. T. (1976). Rational expectations and the natural rate hypothesis: Some consistent estimates. *Econometrica*, *44*(1), 43–52.

Shoup, C., Friedman, M., & Mack, R. P. (1941). Amount of taxes needed in June 1942 to avert inflation: A preliminary report submitted to a joint committee of the Carnegie Corporation and the Institute of Public Administration. Institute of Public Administration.

Stark, T. (2013). SPF Panelists' Forecasting Methods: A note of the aggregate results of a November 2009 Special Survey. Research Department, Federal Reserve Bank of Philadelphia. Available at: http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/

Stock, J., & Watson, M. W. (2009). Phillips curve inflation forecasts. In Fuhrer, J., Kodrzycki, Y., Little, J., & Olivei, G. (Eds.), *Understanding inflation and the implications for monetary policy* (pp. 99–202). MIT Press.

Turner, D. S. (1990). The role of judgement in macroeconomic forecasting. *Journal of Forecasting*, *9*, 315–345.

AUTHOR BIOGRAPHY

Michael P. Clements is a professor of Econometrics at the ICMA Centre, Henley Business School, University of Reading. He holds a DPhil in Economics from Nuffield College Oxford. He is a distinguished author of the Journal of Applied Econometrics, an honorary fellow of the International Institute of Forecasters, a fellow of the International Association for Applied Econometrics, and a series editor of the Palgrave Texts in Econometrics and Palgrave Advanced Texts in Econometrics. His research interests include forecast evaluation, non-linear modeling and business cycle analysis, real-time modeling and forecasting, and the analysis of survey expectations.

How to cite this article: Clements, M. P. (2024). Survey respondents' inflation forecasts and the COVID period. *Journal of Forecasting*, 1–16. https://doi.org/10.1002/for.3169

APPENDIX A: SPEARMAN'S RANK CORRELATION COEFFICIENT

Spearman's rank correlation coefficient has been used to determine whether there are cross-respondent relationships in various quantities of interest: for example, between the in-sample fit and the distance between COVID-19 reported and model forecasts. The rank correlation does not require that the relationship between the variables is linear, because it works off the ranks of the variables. It will detect an association provided only that the relationship is monotonic.

The Spearman rank correlation r lies between -1 and 1, where 0 indicates no relationship. The rank correlation given by

$$r = 1 - \frac{6R}{N(N^2-1)},$$

where R is the sum of squared differences between the ranks (of the forecasters by sample size and by the value of $\hat{\gamma}_j$). It is common to calculate the Fisher transformation,

$$F(r) = \frac{1}{2}\ln\frac{1+r}{1-r},$$

such that $z = F(r).\sqrt{\frac{N-3}{1.06}} \sim N(0,1)$ under the null of statistical independence. As well as reporting r, we report the probability of the test statistic z being at least as large as we obtained if the null hypothesis (of a zero correlation) is true. Probabilities less than 0.025 or greater than 0.975 indicate rejections of the null in a two-sided test at the 5% level. (High probabilities suggest a negative relationship and low probabilities a positive relationship.)

The test statistics we report are based on estimates of the quantities of interest, rather than the population values. For example, when we consider the relationship between PC model unemployment rate parameter γ, and the expected squared error of the reported forecasts during COVID-19, we use estimates of these quantities - $\hat{\gamma}_i$ and the MSFE of the forecasts which individual i made during this period.

We could account for the uncertainty in the estimate $\hat{\gamma}_i$, because these quantities are realizations of random variables with known "measurement" uncertainties: See, for example, the perturbation method of Curran (2015). Or alternatively we could perform a (block) bootstrap on individual i's reported pre-COVID-19 forecasts of inflation and the unemployment rate, to determine the bootstrap distribution of the unemployment rate parameter (or of the in-sample standard error, σ_i). The problem arises in that we do not know the measurement uncertainties in the COVID-19 period quantities. Simply re-sampling the COVID-19 period forecasts and actual values to estimate the distribution of an individual's expected squared loss is unlikely to be successful, given the small number of forecasts (from 4 to a maximum of 12), and will not capture the impact of, say, an individual filing a forecast in response to the 2020:Q2 survey but not the 2020:Q1 survey, or vice versa. In short, the COVID-19 period in our dataset (2020:Q1 to 2022:Q4) is a singular period. For this reason, we do not attempt to posit a distribution for the COVID-19 period data that underlies the RMSFE and related measures and which would be necessary to calculate the uncertainty in the estimates.

APPENDIX B: DATA APPENDIX

B.1 | Forecast data

The forecast data are from the US Survey of Professional Forecasters (SPF). The SPF is a quarterly survey of macroeconomic forecasters of the US economy that began in 1968, administered by the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER). Since June 1990, it has been run by the Philadelphia Fed, renamed as the SPF: See Croushore (1993). It is one of the principal sources for academic research of macroeconomic expectations. See, for example, Clements et al. (2023). The forecast data are freely available from the Philadelphia Fed website http://www.phil.frb.org/econ/spf/.

We use the individual expectations from 1981:Q3 to 2022:Q4 for:

Civilian Unemployment Rate (UNEMP). Percentage points. Seasonally adjusted. Quarterly average.

CPI Inflation Rate (CPI) Headline. Annualized quarter on quarter percentage change. Seasonally adjusted. Based on quarterly average index level.

B.2 | Actual data

The matching real-time data were downloaded from http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/.

The Real-Time Data Set for Macroeconomists (RTDSM) provides quarterly vintages of monthly data. From the quarterly vintages of monthly data, we construct quarterly vintages of quarterly data, by averaging the months. We then transform the CPI data to match the forecasts. Although the revisions to CPI and UNEMP are small, we use the advance estimates. (For the CPI, vintages are not available prior to 1994:Q3, so the data up to 1994:Q2 are taken from the 1994:Q3 vintage.)