The role of survey-based expectations in real-time forecasting of US inflation

Andriantomanga, Zo

University of Wisconsin-Milwaukee

November 2023

The Role of Survey-Based Expectations in Real-Time Forecasting of US Inflation *

Zo Andriantomanga †
University of Wisconsin-Milwaukee

November, 2023

Abstract

This paper performs a real-time forecasting exercise for US inflation from 1992Q1 to 2022Q2. We reinvestigate the literature on autoregressive (AR) inflation gap models - the deviation of inflation from long-run inflation expectations. The findings corroborate that, while simple models remain hard to beat, the multivariate extensions to the AR gap models can improve forecasting performance at short horizons. The results show that (i) forecast combination improves forecast accuracy over simpler models, (ii) aggregating survey measures, using dynamic principal components, improves forecast accuracy, (iii) and the additional information obtained from the error correction process between inflation and long run inflation expectations can improve forecasting performance. In spite of our models providing more accurate one-step ahead forecasts on average, fluctuation tests reveal that over unstable time periods - mainly during the GFC and the Covid-19 pandemic - the AR(1) benchmark performed better.

JEL Classification : C53, E31, E37

Keywords: Inflation, survey forecasts, forecast combination, inflation expectations, error correction, real-time data

*We thank Kundan Kishor, and the participants of the 2023 Midwest Economics Association Conference for their comments.

†University of Wisconsin - Milwaukee, Bolton Hall 8th floor, 3210 N Maryland Ave, Milwaukee, WI 53211, USA. E-Mail Address: andrian2@uwm.edu

1 Introduction

After an extended period of low inflation and interest rates, the economic recovery over the Covid-19 pandemic has led costs of living to soar. The U.S. Consumer Price Index (CPI) rose by 9.1% year-over-year in June 2022, the biggest jump since 1981, and core CPI (which excludes food and energy) climbed by 5.9%. The Personal Consumption Expenditures (PCE) Index rose 7%, and core PCE rose 5% from a year ago. While remaining high, inflation started to subdue in the subsequent months partly thanks to successive aggressive rate hikes by the Federal Reserve. Nonetheless, markets are worried about inflation as these numbers are well above the Federal Open Market Committee (FOMC)'s long-run inflation rate objective of 2%.

The economic uncertainty surrounding high levels of inflation bring up concerns about inflation expectations becoming unanchored (Rudd, 2022). Indeed, inflation expectations play a key role in the dynamics of inflation (Döpke et al., 2008; Faust & Wright, 2013; Chan et al., 2018; Feldkircher & Siklos, 2019). They refer to the general belief about the future prices in the economy. If people expect prices to rise in the future, they are likely to start spending more money now, in order to take advantage of the lower prices before they actually go up. This increased spending can lead to higher demand for goods and services, which can in turn drive up prices. Policymakers worry that such phenomenon would trigger a self-fulfilling inflation spiral.

To keep prices stable, optimal monetary policy depends on reliable inflation forecasts which is an essential task for inflation targeting central banks (Svensson, 1997; Svensson & Woodford, 2004). The literature on inflation forecasting is vast as researchers have developed different models and data in order to predict inflation (Stock & Watson, 1999; Atkeson et al., 2001; Ang et al., 2007; Groen et al., 2009; Bernanke et al., 2007; Faust & Wright, 2013; Groen et al., 2013; Hauzenberger et al., 2022). Nonetheless, inflation remains hard to predict and there is no one consensual method to forecast inflation (Stock & Watson, 2007).

In this paper, we forecast US inflation using Personal Consumption Expenditure (PCE), core PCE and Consumer Price Index (CPI) real-time data from 1992Q1 to 2022Q2. We explore the role of long run inflation expectations survey in improving forecast accuracy of inflation. To conduct the out-of-sample forecast evaluation, we estimate all models -univariate and multivariate - based on a recursively expanding sample. The estimation starts in 1992Q1, with the first estimation sample ending in 1999Q4. Our work relates to three recent strands of the inflation forecasting literature.

First, we revisit the autoregressive inflation gap (AR gap henceforth) literature. We build on the comprehensive work of Faust and Wright (2013), which argues that good forecasts of inflation must include a local mean of inflation. In their original paper, they approximated trend inflation using the 7-11 years Blue Chip survey of long run inflation expectations. We extend their work by considering alternative long run surveys of inflation expectations such as the Michigan Survey of Consumers (MSC), the Survey of Professional Forecasters (SPF) and the Livingston Survey.

Additionally, we extend the inflation gap models to a multivariate setting. We use a vector autoregressive (VAR) framework to model the endogenous relationship between inflation and inflation expectations. Then, we augment the VAR models with an error-correction term. The vector error-correction (VECM) framework elicits the long-run relationship between inflation and inflation expectations. The additional information brought by the error correction term helps improving the forecast accuracy of inflation.

A second strand of the literature seeks to provide a measure of trend inflation. The literature has widely used surveys of long run inflation expectations as a proxy for trend inflation(e.g. Faust & Wright, 2013; Fulton & Hubrich, 2021; Kishor et al., 2022). Given the multitude of survey measures of inflation expectations, it is increasingly difficult to choose which survey to use. We propose a uniform measure of long run inflation expectations by combining information from all four surveys of inflation expectations considered in this paper using the generalized dynamic principal component approach. While we are not the first to adopt this approach, our newly created measure of trend inflation stands out for two reasons: (i) it integrates information on inflation expectations from consumers and professionals, (ii) and covers a longer time period unlike any other constructed measure of trend inflation.

Third, the literature investigates the role of forecast combination in improving inflation forecasts. This approach is in line with the theory of portfolio diversification. Forecast combination of a set of different forecasts have been proven to improve the forecast of inflation (Wright, 2009; Faust & Wright, 2013; Hubrich & Skudelny, 2017; Fulton & Hubrich, 2021; Bravo & El Mekkaoui, 2022). In this paper, we provide combined forecasts for the univariate and multivariate models. We follow two forecast combination techniques: (i) the simple average method which gives equal weight to the set of forecasts considered and (ii) the Bates and Granger (1969) method which allocates optimal weights to the forecasts.

The findings corroborate that while simple models remain hard to beat, the Vector Autoregressive (VAR) and Vector Error Correction Model (VECM) can improve forecast performance by 2 to 7% at short horizons. The results show that (i) forecast combination improves forecast accuracy over simpler models by 8 to 10%, (ii) aggregating survey measures, using dynamic principal components, improves forecast accuracy by 4 to 6%, (iii) and the additional information obtained from the error correction process between inflation and long run inflation expectations can improve forecasting performance by an average of 4%. We also compare the accuracy of model-based forecasts to subjective forecasts.

We conduct a series of robustness check to our forecasting methods. First, we explore the use of subjective forecasts as an alternative to model-based forecasts. We utilize the long run surveys of inflation expectations as direct forecasts for inflation. Then, we introduce rolling forecast as an alternative to the recursive forecasting mechanism used for the baseline forecasts. Under these two exercises, the results show that the forecast accuracy is more or less the same as in the baseline and therefore support the main findings above.

Then, we extend our analysis to the local relative out-of-sample forecasting performance of the AR gap, VAR and VECM models. By implementing the fluctuation test outlined in Giacomini and Rossi (2010), we found out that despite our models performing better on average over the 200Q1 to 2022Q2 period, there are times when the AR(1) benchmark produces better forecasts: mostly during unstable time periods like the Great Financial Crisis and the Covid-19 pandemic.

The remainder of this paper is organized as follows: Section 2 describes data and methods. Section 3 describes the forecasting methodology. Section 4 presents the results of the real-time out of sample forecasting exercise. Section 5 discusses the out-of-sample forecasting performance in unstable environments. Section 6 concludes.

2 Data

2.1 Inflation Measures

We consider three different measures of inflation: the personal consumption expenditure (PCE), the core personal consumption expenditure (CPCE), and the consumer price index (CPI). Figure 1 shows the evolution of these measures. The sample period is 1992Q1 - 2022Q2. We measure inflation as the annualized quarter-over-quarter percent change in the price index.

$$\pi_t = 400 * \ln\left(\frac{P_t}{P_{t-1}}\right)$$

The first inflation measure is the personal consumption expenditures (PCE) chain-type price index produced by the Bureau of Economic Analysis (BEA). In January 2012, the Federal reserve adopted the personal consumption expenditures price index for their monetary policy target. Panel A of Table 1 shows the summary statistics for all inflation measures and inflation expectations survey data over the sample period of 1992Q1 to 2022Q2.

The core PCE strips out the volatile components of headline PCE. It is obtained using all items in the PCE basket less food and energy. Core inflation is widely used as a measure of the underlying price trends. It also plays a central role in the policy makers' decision for the future path of monetary policy. While the Federal Reserves targets headline inflation (PCE), they also monitor changes in the underlying price inflation measured by the core PCE.

Nonetheless, other measures of inflation remain important, both as economic indicators and for our exercise here. The Consumer Price Index (CPI) provides an

Table 1: Summary statistics and correlation

Panel A: Summary Statistics							
	Obs.	Mean	Std. Dev.	Median	Min	Max	Range
PCE	122	2.02	1.66	2.03	-6.44	6.84	13.28
Core PCE	122	1.92	0.92	1.83	-0.97	5.88	6.84
CPI	122	2.46	2.23	2.53	-9.27	10.01	19.28
MSC	122	2.89	0.28	2.90	2.30	3.80	1.50
SPF	122	2.56	0.42	2.50	2.03	3.90	1.87
Livingston	122	2.60	0.41	2.50	2.00	4.00	2.00
Blue Chip	122	2.59	0.42	2.50	2.20	3.90	1.70

Panel B: Correlation							
	PCE	Core PCE	CPI	MSC	SPF	Livingston	Blue Chip
PCE	1.00	0.81	0.98	0.21	0.12	0.10	0.00
Core PCE		1.00	0.75	0.28	0.24	0.23	0.11
CPI			1.00	0.22	0.14	0.12	0.02
MSC				1.00	0.85	0.86	0.79
SPF					1.00	0.97	0.95
Livingston						1.00	0.96
Blue Chip							1.00

Note: Panel A presents the summary statistics for different variables from 1992Q1 to 2022Q2. Panel B presents the pairwise Pearson correlation coefficients. Inflation rates for PCE, core PCE and CPI are annualized quarter-over-quarter log changes.

alternative measure of inflation. It is a price index of a basket of goods and services paid by urban consumers. The consumer Price Index is monthly published by the Bureau of Labor and Statistics (BLS). We use the Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL) from the FRED database.

Panel A of Figure 1 illustrates the historical series of our three measures of inflation: headline PCE, headline CPI, and core PCE. We can observe that the headline measures of inflation follow a similar pattern. Panel B of Table 1 shows that they have a raw correlation of 0.98. However, it appears that headline CPI is more volatile compared to headline PCE. Panel of Table 1 corroborates this information as the standard deviation for headline CPI and PCE are, respectively, 2.46 and 2.02. As expected, core inflation is less volatile compared to the measures of headline inflation.

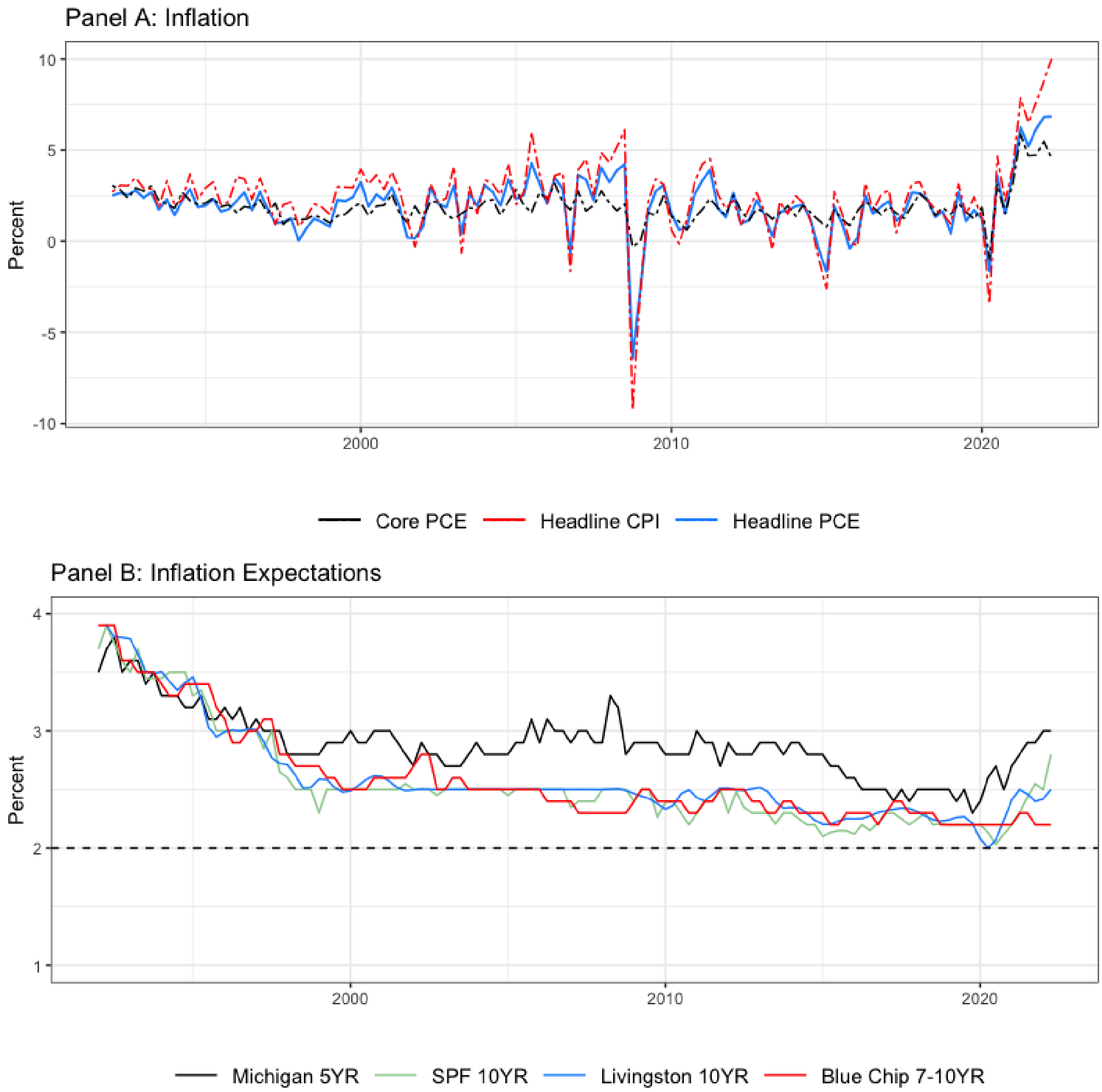
2.2 Inflation expectations data

Panel B of Figure 1 shows four long-run inflation expectations measures: the 5 year-ahead expectation from the Michigan Survey of Consumers (MSC), the 10 year-ahead CPI forecast from the Survey of Professional Forecasters, the 10 year-ahead CPI forecast from the Livingston Survey and the 7-11 years Blue Chip inflation forecast. These surveys ask similar questions on the expectations of average inflation in the long run but differ in their respondents, coverage and frequency. [1]

Our sample coverage is based on the simultaneous availability of the four surveys of long run inflation expectations. The Michigan Survey of Consumers (MSC) is a monthly survey that asks households on their expected inflation in 5 years since 1978. The Livingston survey (LVG) asks professionals from various sectors including academia, non-commercial banks, consulting, investment, non-financial sectors, etc. This semi-annual survey began asking for long run inflation expectations since June 1991. The Survey of Professional Forecasters (SPF) is a quarterly survey of economists that began asking questions on long run inflation expectations since 1991 Q4. The long term measure for the Blue Chip (BC) survey is released on a semi-annual basis. The long run inflation expectations data from SPF only starts on 1992Q1 while all other surveys have data prior that date. Therefore, we start our sample in 1992Q1

[1]The literature also identifies market-based measures of inflation expectations as an alternative to survey data. This paper focuses on survey data of long-run inflation expectations.

Figure 1: Inflation and Long Run Inflation Expectations 1992Q1 - 2022Q2



Note: Panel A presents the annualized quarter-over-quarter inflation rates for three different measures: CPI, PCE and core PCE. Panel B shows the survey data of long run inflation expectations: 5-year-ahead Michigan Survey of Consumer, 10-year-ahead Survey of Professional Forecasters (SPF), 10-year-ahead Livingston Survey, and 7 to 10-year-ahead Blue Chip Survey. The dashed line represents the long run inflation target of the Federal Reserve of 2%. The sample is 1992Q1 - 2022Q2.

for which data is available for all four surveys.

As noted above, the survey data measures come with different frequencies. For our exercise, we require the series to be transformed at a quarterly frequency. Both the Blue Chip and the Livingston Survey only come out twice a year. We implement an interpolation algorithm to transform the data at a quarterly frequency. We aggregate the monthly Michigan Survey of Consumers (MSC) at a quarterly level with simple average.

Figure 1 shows that the professional surveys - SPF, Livingston and Blue Chip - present a similar behavior. Panel B of Table 1 shows that the SPF has a raw correlation of 0.97 with Livingston, and 0.95 with Blue Chip. The household survey behaves differently compared to the professional ones. The Michigan survey has a lower correlation with the surveys of professional forecasters: 0.85 with SPF, 0.86 with Livingston and 0.79 with Blue Chip. Moreover, Panel B of Figure 1 shows that households consistently expect a higher level of inflation in the long run compared to professionals. Nonetheless, the Michigan survey has a lower volatility despite having a higher mean. The Michigan survey has a mean of 2.9% over our sample, compared to 2.6% for the three professional surveys. The standard deviation is 0.3 for MSC against 0.4 for the professional surveys. These results suggest the information set at our disposal is rich to make forecasts of inflation.

3 Forecasting Methodology

The real-time exercise focuses on three different measures of quarterly inflation: Personal Consumption Expenditure (PCE), core PCE and Consumer Price Index (CPI).

Since these series, especially the PCE, are subject to potentially large revisions, a real-time forecasting exercise is necessary. The data were drawn from the Real-Time Data Research Center database hosted by the Federal Reserve Bank of Philadelphia.[2]

To conduct the out-of-sample forecast evaluation, we estimate all models based on a recursively expanding sample. The estimation starts in 1992Q1, with the first estimation sample ending in 1999Q4. The final set of forecast is made in 2022Q2. Because PCE price data are heavily revised, there is no single source of true data against which to compare the model forecasts. Following Tulip (2009), Faust and Wright (2013) and Fulton and Hubrich (2021), we use PCE price inflation as measured in the release two quarters after the reference quarter as the true value from which forecast errors are constructed.

We use the root mean squared prediction error (RMSPE) to assess the forecasting performance of our models. The results are usually shown relative to the benchmark model AR(1) model. A relative RMSPE number less than one indicates improvement compared to the benchmark. We consider forecasts for each period $t + h$ up to one year ahead ($h = 1,...,4$). We assess whether a model forecast is significantly different compared to the benchmark AR(1). We conduct the non-nested models comparison test following Diebold and Mariano (2002). The null hypothesis is that the RMSPE of the two models are equal to each other. The alternative of the one-sided test is that the model forecasts are more accurate than the benchmark. Cases in which the root mean square prediction error is significantly lower than the benchmark model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively.

[2] https://www.philadelphiafed.org/surveys-and-data/real-time-data-research

3.1 Model Specification

We construct forecasts for inflation π_{t+h} using the following models.

Autoregressive Model - AR(1). The Autoregressive Model of order 1 serves as the benchmark model. Simple models have always remained hard to beat (Meese & Rogoff, 1983).

$$\pi_t = c + \pi_{t-1} + u_t$$

Inflation Gap Model - AR GAP. A common approach in the literature is to decompose inflation into two components: trend inflation and inflation gap. Researchers have assumed that trend inflation could be approximated with long run inflation expectations. This idea is consistent with inflation expectations being anchored, meaning that they do not respond to short term news. They define $g_t = \pi_{t-1} - \pi^e_{t-1}$ and fit an AR gap as follows.

$$g_t = \rho_0 + \sum_{j=1}^{p} \rho_j g_{t-j} + \epsilon_{t+1} \tag{1}$$

To make forecasts of inflation, they add to the trend the AR fitted gap. We then proceed as in Faust and Wright (2013) by taking the predictions of the gap – the forecasts g_{t+1} – and adding back the final observation of the trend to get the implied prediction.

This approach has been widely used in the literature (e.g. Faust & Wright, 2013; Fulton & Hubrich, 2021; Kishor et al., 2022). Faust and Wright (2013) argue that good forecasts of inflation need account for a local mean of inflation. Survey measures have been used as a proxy for such inflation trend. In this exercise, we will explore five measures of trend inflation.[3] While the literature acknowledged this approach to be successful at outperforming the benchmark , we believe that there are gains to extend it to a multivariate settings.

Vector Autoregressive Model - VAR. We consider an extension to the AR Gap model to a multivariate setting. In contrast to the univariate inflation gap model, the VAR explicitly specifies the endogeneity between inflation and inflation expectations. That is, future inflation π_{t+h} depends not only on its lags but also lags on inflation expectations.

$$\begin{cases} \Delta\pi_t & = \mu_1 + \phi_{11}^1 \Delta\pi_{t-1} + \phi_{12}^1 \Delta\pi_{t-1}^e + \cdots + u_{1t} \\ \Delta\pi_t^e & = \mu_2 + \phi_{21}^1 \Delta\pi_{t-1} + \phi_{22}^1 \Delta\pi_{t-1}^e + \cdots + u_{2t} \end{cases} \tag{2}$$

Vector Error Correction Model - VECM. We now consider another multivariate model. We augment the VAR model with an auto-correction term. The VECM framework elicits the long-run relationship between inflation and inflation expectations.

$$\begin{cases} \Delta\pi_t & = \mu_1 + \phi_{11}^1 \Delta\pi_{t-1} + \phi_{12}^1 \Delta\pi_{t-1}^e + \cdots + \alpha_1 \hat{\varepsilon}_{t-1} + u_{1t} \\ \Delta\pi_t^e & = \mu_2 + \phi_{21}^1 \Delta\pi_{t-1} + \phi_{22}^1 \Delta\pi_{t-1}^e + \cdots + \alpha_2 \hat{\varepsilon}_{t-1} + u_{2t} \end{cases} \tag{3}$$

[3]Survey of Professional Forecasters (SPF), Michigan Survey of Consumer (MSC), Livingston, Blue Chip, and a constructed measure of trend inflation

The multivariate models provide forecasts for the change in inflation, denoted $\Delta\hat{\pi_{t+1}}$. To construct a forecast for inflation $\hat{\pi_{t+1}}$, we add the predicted change in inflation to the current period inflation π_t.

3.2 Survey Data Aggregation

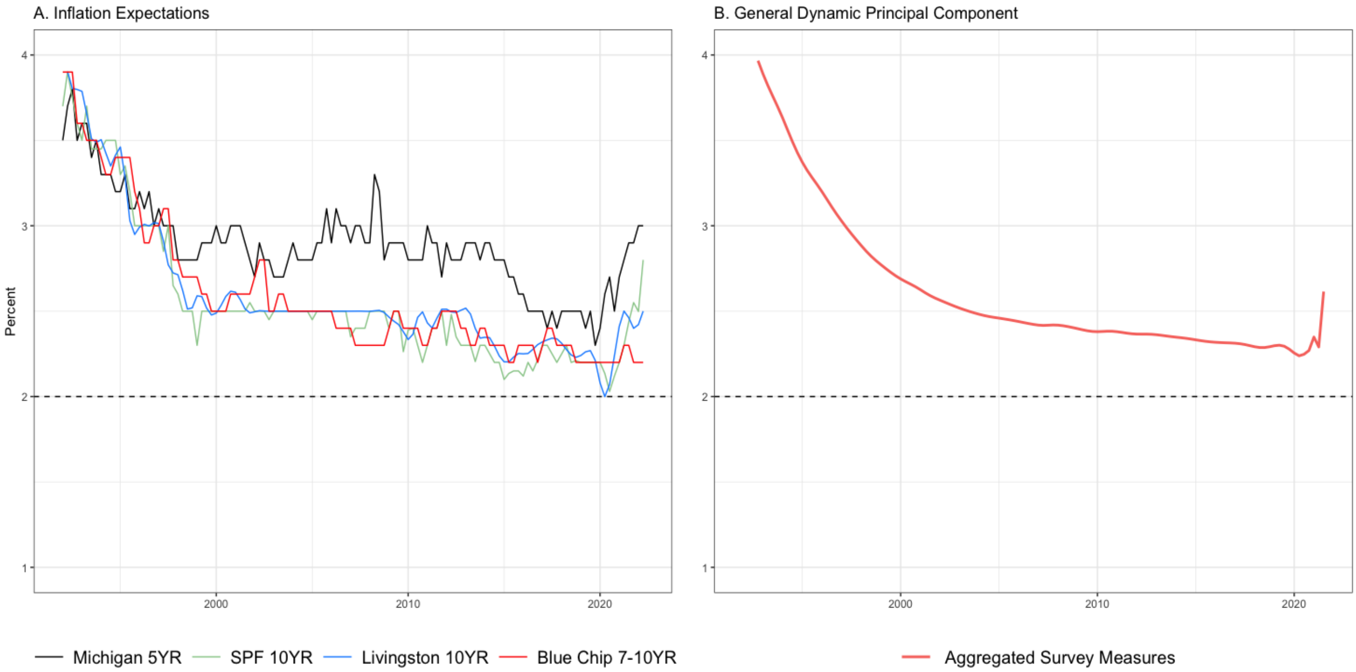
The literature has traditionally used long-run surveys of inflation expectations as a proxy for trend inflation. In their prominent paper, Faust and Wright (2013) use the five-to-ten-year-ahead inflation forecast from Blue Chip to capture the trend of inflation. The authors did not motivate this choice further while there are several other inflation expectations surveys. A large strand of the literature has attempted to come up with a measure of trend inflation (e.g Ahn, Fulton, et al., 2020; Kishor et al., 2022).

We address this question by combining our four survey measures to construct a single measure of trend inflation. We chose these survey measures based on the availability of the data for the sample we consider in this paper. The measures of inflation expectations are: Survey of Professional Forecasters (SPF), Michigan Survey of Consumer (MSC), Livingston, Blue Chip. We used the General Dynamic Principal Component (GDPC) method to extract the common component of trend inflation. We call the newly obtained measure as "Aggregate".

The General Dynamic Principal Component is a dimensionality reduction technique widely used in multivariate time series analysis (see Brillinger (2001); Hörmann et al. (2015); Peña et al. (2019)). We follow the method outlined by Peña and Yohai (2016).[4] We are able to extract a non-linear combination of long-run inflation ex-

[4]We use the GDPC package in R developed by Peña et al. (2020).

Figure 2: Survey Data Aggregation



Note: Panel A shows the survey data of long run inflation expectations: 5-year-ahead Michigan Survey of Consumer, 10-year-ahead Survey of Professional Forecasters (SPF), 10-year-ahead Livingston Survey, and 7 to 10-year-ahead Blue Chip Survey. Panel B shows the newly created measures of inflation expectations using general dynamic principal components. The dashed line represents the long run inflation target of the Federal Reserve of 2%. The sample is 1992Q1 - 2022Q2.

pectations. The number of lags is chosen following the Akaike Information Criteria (AIC). The unique factor we extracted explains 90% of the variation of the original data.

3.3 Forecast Combination

Forecast combination has been proven to improve the forecast of inflation (Wright, 2009; Faust & Wright, 2013; Hubrich & Skudelny, 2017; Fulton & Hubrich, 2021; Kishor, 2021; Bravo & El Mekkaoui, 2022). We provide combined forecasts in one hand for the five univariate models, and on the other hand, for the 10 multivariate models. We consider two methods of forecast combination. First, we generate a ”simple average” forecast with equal weights assigned to the considered models. Despite its straightforward approach to forecast combination, the simple average method has proven to be successful at beating the forecasting performance of benchmark models (Ang et al., 2007; Faust & Wright, 2013; Fulton & Hubrich, 2021).

Second, we follow the forecast combination method developed by Bates and Granger (1969). Similar to the simple average approach, the idea behind the Bates and Granger (1969) method is to combine a set of n forecasts to construct a better forecast. Intuitively, their method allocates optimal weights to the set of n model forecasts with more weights assigned to the better performing ones. This method uses the relative variances and covariances to construct a weighted average of the forecasts that minimized the mean square error of the combined forecast.

Three steps are involved in the Bates-Granger Forecast Combination method:

- First, get the out of sample mean squared error (MSE) for each model
- Then, calculate the inverse of MSE. $\bar{\sigma}_i^2 = \frac{1}{\sigma_i^2}$

- And calculate the optimal weight associated to each forecast as follows:

$$\omega_i^{BG} = \frac{\sigma_i^{-2}}{\sum_{i=1}^{N} \sigma_i^{-2}}$$

where σ_i^{-2} is the estimated mean squared prediction error of the i-th model. This forecast combination method has been successful at improving forecast accuracy(Ang et al., 2007; Faust & Wright, 2013; Fulton & Hubrich, 2021).

4 Results: Real-Time Forecasting of US Inflation

4.1 Baseline Results - PCE

Table 2 presents the out of sample forecasting results for PCE price inflation. The forecasting period considered is 2000Q1 to 2021Q4. The competing forecasts are evaluated using the root mean squared error ratio. A value less than one indicates that the considered model outperforms the AR (1) benchmark.

The main message from Table 2 is that the benchmark AR(1) is generally difficult to improve upon except for horizon h = 1 and 2. Indeed, the autoregressive inflation gap models using several measures of inflation expectations do better than AR(1) at horizon h = 1 and 2. The multivariate VAR and VECM models only outperform the benchmark at horizon h=1. Finally, the forecasting methods we employ, in particular survey forecast aggregation and forecast combination, show substantial improvements over the benchmark.

First, Panel A of Table 2 indicates that the inflation gap models are modestly doing better than the AR(1) benchmark at 1 and 2 steps ahead. Inflation gap models using Michigan and Livingston surveys improve the forecast by 4% at one step ahead, while models using the Survey of Professional Forecasters and Blue Chip outperform the benchmark by 3%. These models perform badly at horizon h=3 and 4.

The forecast combination approach is the most successful at beating the benchmark. The simple average and the Bates-Granger yield similar results. Both forecast combination methods improve upon the benchmark by as much as 7% at horizon h=1 and 3% at h=2. Additionally, the yearly average forecast, i.e at h=1-4, outperforms the benchmark. The inflation gap with aggregated survey data - which combines the four surveys via generalized dynamic principal component - improves on the benchmark at horizon h=1. While this model outperforms the benchmark, it does not do better than forecasts from individual surveys. The aggregated forecast does not improve on the benchmark at horizon h=2.

Second, Panel B of Table 2 shows that the multivariate models only do well at the short term horizon h=1. The forecasts from the VAR models outperforms the benchmark forecast by about 3 to 4% depending on which survey data is used. Next, we discuss the results using the Vector Error Correction (VECM) framework. The VECM models outperform the benchmark forecast by about 3 to 4% on average. While the forecasting performance using MSC and SPF are similar (around 3%), there is a slight improvement up to 4% using the Livingston or the Blue Chip survey. Nonetheless, the additional information brought by the error term does not seem to improve forecasting performance much more. The multivariate models do not perform well at horizon 3 and 4.

The fact that the VECM forecasts do not outperform the AR(1) benchmark at 3 and 4-step-ahead is not very surprising. The most plausible explanation is that the error correction, which helps improve the forecast, only happens at short term horizons. Beyond two quarters, the error correction process vanishes and the models behave like regular VAR models. Additionally, prior studies using real-time data, like the prominent paper by Faust and Wright (2013), fail to improve the forecast accuracy of inflation beyond horizon h=2.

The forecast provided by the multivariate VAR and VECM models using the aggregated survey data both improve upon the benchmark at 1-step-ahead. The VECM *Aggregate* model in particular fares better than the VECM models using the individual surveys. Next, combining the forecasts from all multivariate models - VAR and VECM - improve on the benchmark. Forecasts from the Simple Average and the Bates-Granger are similar - both yielding a 4% improvement. Forecast combination is doing better than the benchmark at only horizon h=1.

4.2 Alternative inflation measures

While our main focus is to forecast the Personal Consumption Expenditure (PCE) price inflation, other measures of inflation remain important, both as economic indicators and for our exercise here.

4.2.1 Consumer Price Index - CPI

Table 3 reports the results for CPI price inflation. While most entries in Table 3 are greater than one, the inflation gap models seem to do well in the short term. The univariate inflation gap models improve upon the benchmark at horizon h=1 and 2 but fail do to so at horizon 3 and 4. Nonetheless, the yearly average forecast - at horizon h= 1-4 - very modestly outperform the benchmark AR(1). The multivariate models perform badly at every horizon considered in our exercise.

The inflation gap models improve upon the benchmark in the short term at horizon h=1 and 2. The models yield similar performance. The Michigan Survey of Consumer improves the benchmark by 7% while the professional surveys outperform the AR(1) by 8% at horizon 1. It is worth noting that the inflation gap models do a better job at tracking CPI at horizon 1 than PCE where the average improvement was about 3 to 4%. Beyond horizon 2, the models fail to improve on the benchmark. Nonetheless, the yearly average forecast - at horizon h= 1-4 - very modestly outperform the benchmark AR(1).

The inflation gap model with aggregated survey data - which combines the four surveys via generalized dynamic principal component improves on the benchmark at horizon h=1 and 2. While this model outperforms the benchmark, it yields similar forecasts to the models using survey data individually. The forecast combination approach is also successful at beating the benchmark. The simple average and the Bates-Granger provide similar results. Both forecast combination methods improve upon the benchmark by as much as 8% at horizon h=1 and 2.

The multivariate models perform poorly at every horizon considered in our exercise. Both the VAR and VECM models do not do better than the benchmark. The forecast combination approach does not help in improving forecast accuracy since individual models do not do well.

4.2.2 Core PCE

This section presents the results for the Core PCE price inflation. Core PCE measures the underlying inflation trend by stripping out food and energy which are the volatile part of headline PCE. While the Federal Reserve's objectives target headline PCE, the Fed also monitors core inflation. The forecasting results are presented in Table 4.

Most entries in Table 4 are greater than one. Neither the univariate inflation gap models nor their multivariate extension do better than the benchmark AR(1). The models using the aggregated survey data, which proved to be successful for PCE and CPI, also perform badly. Moreover, forecast combination methods do not improve upon the benchmark.

The additional information brought by long-run inflation expectations does not help improve forecast accuracy for core inflation. One possible reason for such failure is the nature of the question asked in the survey itself. Broadly, the question would look like the following: What do you expect the level of inflation to be ten years from now? Here, "inflation" refers to headline inflation including food and energy. The measure provided whether by consumers or professional forecasters consider the future possible fluctuations in food and energy.

4.3 Different subsamples: Pre vs post GFC Analysis

The Great Financial Crisis (GFC) disrupted the US economy at several levels. Not only it brought a slowdown of economic activities but also changes in the dynamics of inflation. A potential structural break might have happened. There is a need to investigate whether the forecasting methods we used to improve forecast accuracy deliver similarly accurate forecast before and after the financial crisis.

In the previous sections, our full sample forecasting spanned from 2000Q1 - 2022Q2. Here, we consider two forecasting subsamples. The first subsample covers the period 2000Q1 to 2009Q4. It covers the era of the Great Financial Crisis. The second Period covers the era post great financial crisis: 2010Q1 - 2019Q4. We omit the period after 2019Q4 in our second subsample. As inflation started to pick up in 2021, we suspect the presence of a structural break in the dynamics of inflation as seen in Figure 1.

Table 5 presents the forecasting results for the Personal Consumption Expenditure (PCE) price inflation when comparing the prior and post crisis period. The main takeaway from Table 5 is that our primary results using the full sample agree with those of the pre-crisis period. Unlike the results using the full sample, many of the models do not outperform the benchmark in the post crisis period at horizon 1 and 2.

In the post-crisis sample period, the univariate models perform as good as the in the full sample for horizon h=1. Where we see the most improvement is on horizon h=2, 3, and 4 as the forecasts beat the benchmark by 4 to 8%. The yearly average forecast for horizon h= 1-4, improves by 12% upon the benchmark and the Diebold - Mariano test is statistically significant at 5% level.

Furthermore, the forecast combination methods of the univariate models outperform the benchmark . Many of VAR and VECM models begin to break down in this subsample despite doing well in the full sample. These results indicate that forecast accuracy of PCE inflation can be improved even more in the post crisis period.

Next, we discuss the results for CPI price inflation. Table A.1 show that in the pre-crisis and post-crisis periods, the models slightly improve overall forecasting accuracy compared to the full sample. The univariate models do better in the pre-crisis period, while both the univariate and multivariate models perform better in the post-crisis period.

In the pre-crisis period, the univariate models slightly provide better forecasts compared to the full sample. In particular, the inflation gap models do better at horizon 1-4, significantly improving on the yearly average of CPI inflation forecast. The forecast combination of the univariate models also do better than the benchmark.

In the post-crisis period, both the univariate and the multivariate models perform better. In particular, the inflation gap models provide striking results. They improve upon the benchmark and the full sample results at horizon 1, 2, 3 and 4. For example, at horizon h=1, the univariate AR gap models beat the benchmark by as much as 13% and the Diebold-Mariano test is significant at 1% level. Additionally, the forecast for the yearly average inflation sees an improvement of as much as 11 percent with the Bates-Granger forecast combination method yielding the best results. For the multivariate models, the VAR and VECM models using SPF, Blue Chip and the aggregated survey data perform well compared to the full sample.

Finally, we discuss the results for Core PCE inflation. Table A.2 show that the forecasting results for the pre-crisis period agree with the full sample - the models cannot improve on the benchmark. For the post-crisis period, the multivariate models and the forecast combination methods improve on the benchmark at horizon 1. The VAR and VECM models using the SPF, Blue Chip and the aggregated survey data, are seen to improve forecast accuracy. Additionally, combining the multivariate forecasts also improve on the benchmark.

To summarize, by splitting our full sample into two subperiods, we find mixed results. For PCE inflation, the models deliver about the same forecasting performance as the full sample in the pre-crisis period and downgraded performance in the post-crisis period. Dividing our forecasting sample into two subsamples improved the forecasting accuracy for CPI inflation. The univariate models do better in the pre-crisis period, while both the univariate and multivariate models perform better in the post-crisis period. Finally, for Core PCE inflation, the forecasting results for the pre-crisis period agree with the full sample - the models cannot improve on the benchmark. For the post-crisis period, the multivariate models and the forecast combination methods improve on the benchmark at horizon 1.

4.4 The Covid-19 period

In this section, we divide our sample to focus on the Covid-19 period. The estimation sample is from 2020Q1 to 2022Q2. During this time period, the dynamics of inflation have changed as the costs of living started to soar in early 2021. Rising food and energy prices, supply chain disruptions, expansionary fiscal and monetary policies to support the recovery from the pandemic all contributed to the inflation spikes. We investigate the ability of our models in forecasting inflation during the Covid-19 period.

We begin the discussion with the out-of-sample forecasts for PCE price inflation which are presented in Table 6. The most striking result from panel A Table 6 is that combining the forecasts of the univariate models allows to outperform the benchmark at all horizons h=1,..,4 and 1-4 - unlike the results obtained using the full sample. The simple average and Bates-Granger combination methods yield similar performance and improve upon the benchmark for as much as 6 to 13%. At horizons h=2,3 and 4, the Diebold Mariano forecast comparison test is significant. Nonetheless, it must be noted that individually the univariate AR gap models do not improve upon the benchmark over the Covid-19 subsample.

The multivariate models yield more or less similar performance as when using the full samples. Panel B of Table 6 shows that the models only improve upon the benchmark at short term horizons h=1 and 2. In particular, forecasts from the models using the aggregated survey data perform the best, yielding 6 to 10% improvement on the benchmark. Additionally, the forecast combination methods do better than the benchmark at horizon 2.

Next, we discuss the results for CPI price inflation. Table 7 indicates that it is difficult to forecast CPI over the Covid-19 period. There are two main takeaways from Table 7. First, the univariate models perform badly for the considered sample. Second, the multivariate models very slightly improve upon the benchmark at horizon h=2. The forecasting techniques we employ, which includes survey data aggregation and forecast combination do improve upon the benchmark by around 6% at horizon h=2. However, the Diebold Mariano comparison test shows that these forecasts are not statistically different from the benchmark.

Finally, we discuss the results for Core PCE inflation. The key takeaway from Table 7 is that forecast combination applied to the univariate models outperforms the benchmark at all horizons considered. The improvements range from 5 to 19%. For example, the simple average method improves upon the benchmark by 19% and the Diebold Mariano test is significant at 99% confidence level. It must be noted that the Diebold Mariano test is not significant for the forecasts using the Bates-Granger method.

To sum up our analysis over the Covid-19 period, we saw that the forecast of PCE inflation is very good at all horizons using either the simple average or the

Bates-Granger combination techniques. These results are in line with Hubrich and Skudelny (2017) who argue that forecast combination hedges against bad forecasts in times of crisis. Both the univariate and multivariate models fail to improve the forecast of CPI over the pandemic. For core PCE inflation, the forecast combination methods improve upon the benchmark at all horizons.

5 Robustness Check

5.1 Survey-based Forecasts

In this section, we introduce an alternative methodology to model-based forecasts. We explore the use of long run surveys as direct forecasts for inflation. These subjective forecasts are not based on an explicit forecasting model. For this exercise, we consider the Michigan Survey of Consumers (MSC), the Survey of Professional Forecasters (SPF), the Livingstone Survey (LVG), Blue Chip (BC), and our newly created Aggregate survey measure (see subsection 3.2).

The inflation literature has considered the use of subjective survey-based forecasts (e.g. Ang et al., 2007; Faust & Wright, 2013; Fulton & Hubrich, 2021). Ang et al. (2007) investigated whether macroeconomic variables, asset markets, or surveys forecast inflation better. They found that surveys provide more accurate forecasts of inflation. In their analysis, they consider survey forecasts from the Survey of Professional Forecasters (SPF), Livingston (LVG) and the Michigan Survey of Consumer (MSC). We extend this paper by using additional survey data measures and considering the Covid-19 period. As part of this robustness check, we consider three sample periods: the full sample 2000Q1 - 2022Q2, before and after the Covid-19.

We begin by discussing our results for the full sample period 2000Q1 - 2022Q2. Table 9 presents the findings for PCE, CPI and core CPI. The key takeaways are: (i) the survey forecasts can beat the AR(1) benchmark in the short term at horizon h=1 and 2 for PCE and CPI, (ii) surveys from professional forecasters and our newly constructed aggregate measure of inflation expectations perform slightly better than the MSC.

Next, we restrict the sample to the period prior to the Covid-19 period. The forecasting sample considered is now 2000Q1 - 2019Q4. The results reported in Table B.3 suggest a better forecast accuracy the survey measures at short horizon h=1 and 2 for PCE and CPI. The AR(1) benchmark remains hard to beat at longer horizon. It must also be noted that the survey forecasts do no predict core CPI well.

Finally, we analyze the forecast accuracy of the survey measures during the Covid-19 period. Table B.3 suggest that the survey forecasts cannot improve on the forecast accuracy of inflation at all horizons. This finding corroborates what we found in the previous section as we highlighted that in times of crisis it is hard to beat the forecasts from simple models.

To sum up, we conducted a series of robustness checks across different time periods to see whether using survey data can improve the forecast accuracy of inflation. This allows us to compare the forecasting performance of the model-based forecasts to subjective forecasts. The findings are twofold. First, at short horizon h=1 and 2, the survey forecasts do slightly better than the AR - Gap models discussed in section 4. Second, over the Covid-19 period, and similar to model-based, forecasts, the survey forecasts fail to beat the AR(1) benchmark.

5.2 Rolling Forecasts

In this section, we introduce a new forecasting mechanism: Rolling forecasts. Our goal is to explore how this alternative forecasting scheme may change the baseline model forecasts. The rolling window is set for 32 observations and initial estimation sample is 1992 Q1 to 1999 Q4.

Table 10 present the results for PCE inflation. Overall, the models do not outperform the benchmark using the rolling forecast mechanism. Nonetheless, the univariate forecast combination provide more accurate forecast at all horizons considered. The Bates - Granger forecasts perform slightly better than the simple average. For example, the one step-ahead forecast improves upon the benchmark by as much as 7%. These improvements become modest as the forecast horizon is extended to h=2,3 and 4. However, it must be noted that the average yearly forecast is better than the AR(1) by 6%.

Despite all the values in Table 10 being greater than the unity, the univariate model forecasts beat the benchmark by an average of 4% at horizon h=1. This finding corroborates the results of the baseline scenario in section 4.1 such that (i) simple models provide better forecasts and (2) that our models cannot beast the benchmark at long-term horizon.

The rolling forecasts results for CPI inflation are presented in Table 10. The key insight from this table is that the univariate models provide better forecasts at short term horizons h= 1 and 2. The univariate forecasts from professionals do better compared to the household based univariate forecast. Moreover, aggregating survey data slightly improves forecasting performance. Finally, the forecast combination techniques, i.e. the simple average and the Bates-Granger, help in improving upon the benchmark forecast.

To sum up, in this section, we explored how introducing the alternative rolling forecasting method may affect the performance of our model forecasts. We found that model forecasts were closely similar to the ones obtained using the recursive forecast technique. To be specific, the findings in this section corroborate that (i) simple models produce more accurate forecasts, (ii) it is hard to beat the AR(1) benchmark at long term horizon and (iii) the forecasting techniques - i.e. forecast combination and survey data aggregation - we employ improve forecast accuracy.

Table 2: PCE Forecasts: Full Sample 2000Q1 - 2022Q2

Horizon	1	2	3	4	1-4
Panel A: Univariate Models					
AR GAP MSC	0.956*	0.987	1.024	1.024	1.015
AR GAP SPF	0.967	0.996	1.028	1.026	1.023
AR GAP LVG	0.961*	0.992	1.031	1.032	1.022
AR GAP BC	0.968	0.993	1.029	1.026	1.020
AR GAP Aggregate	0.979	1.001	1.033	1.033	1.033
Combination					
Simple Average	0.929	0.963	1.011	0.998	0.972
Bates-Granger	0.933	0.968	1.013	0.998	0.982
Panel B: Multivariate Models					
VAR MSC	0.965	1.031	1.079	1.069	1.079
VAR SPF	0.963	1.041	1.074	1.060	1.077
VAR LVG	0.961	1.028	1.090	1.067	1.081
VAR BC	0.961	1.041	1.071	1.058	1.077
VAR Aggregate	0.965	1.009	1.093	1.068	1.055
VECM MSC	0.966	1.032	1.092	1.070	1.082
VECM SPF	0.972	1.048	1.074	1.064	1.085
VECM LVG	0.963	1.032	1.092	1.068	1.085
VECM BC	0.957	1.046	1.075	1.056	1.081
VECM Aggregate	0.960	1.012	1.093	1.032	1.039
Combination					
Simple Average	0.954	1.021	1.076	1.057	1.072
Bates-Granger	0.953	1.021	1.076	1.057	1.072

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly different than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 3: CPI Forecasts: Full Sample 2000Q1 - 2022Q2

Horizon	1	2	3	4	1-4
Panel A: Univariate Models					
AR GAP MSC	0.926**	0.975	1.015	1.010	0.989
AR GAP SPF	0.921**	0.975	1.015	1.008	0.984
AR GAP LVG	0.917**	0.972	1.016	1.012	0.984
AR GAP BC	0.918**	0.972	1.016	1.008	0.980
AR GAP Aggregate	0.922*	0.972	1.014	1.009	0.982
Combination					
Simple Average	0.924	0.965	1.017	1.012	0.980
Bates-Granger	0.918	0.960	1.014	1.009	0.972
Panel B: Multivariate Models					
VAR MSC	1.117	1.049	1.184	1.058	1.097
VAR SPF	1.049	1.099	1.092	1.051	1.092
VAR LVG	1.226	1.143	1.178	1.092	1.128
VAR BC	1.091	1.107	1.093	1.058	1.093
VAR Aggregate	1.212	1.105	1.167	1.160	1.083
VECM MSC	1.136	1.043	1.203	1.072	1.097
VECM SPF	1.066	1.105	1.100	1.052	1.100
VECM LVG	1.229	1.142	1.179	1.091	1.133
VECM BC	1.081	1.106	1.094	1.053	1.096
VECM Aggregate	1.155	1.120	1.183	1.077	1.037
Combination					
Simple Average	1.093	1.043	1.094	1.045	1.088
Bates-Granger	1.086	1.038	1.088	1.038	1.086

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Lag length are selected using the AIC selection criteria. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 4: Core PCE Forecasts: Full Sample 2000Q1 - 2022Q2

Horizon	1	2	3	4	1-4
Panel A: Univariate Models					
AR GAP MSC	1.037	1.033	1.046	1.068	1.075
AR GAP SPF	1.070	1.073	1.086	1.087	1.125
AR GAP LVG	1.052	1.060	1.082	1.085	1.112
AR GAP BC	1.070	1.067	1.083	1.084	1.119
AR GAP Aggregate	1.017	1.074	1.102	1.132	1.128
Combination					
Simple Average	1.144	1.076	1.094	1.102	1.155
Bates-Granger	1.436	1.350	1.318	1.316	1.515
Panel B: Multivariate Models					
VAR MSC	1.127	1.002	0.994	1.054	1.016
VAR SPF	1.152	0.996	1.021	1.033	1.016
VAR LVG	1.161	0.968	1.064	1.036	1.024
VAR BC	1.143	1.012	1.034	1.031	1.013
VAR Aggregate	1.189	1.080	0.998	1.176	1.006
VECM MSC	1.129	0.998	1.000	1.062	1.016
VECM SPF	1.131	1.021	1.025	1.039	1.021
VECM LVG	1.175	0.977	1.036	1.056	1.018
VECM BC	1.126	1.018	1.037	1.032	1.014
VECM Aggregate	1.172	1.090	1.009	1.197	0.998
Combination					
Simple Average	1.120	0.971**	0.999	1.053	1.009
Bates-Granger	1.110	0.963*	0.995	1.047	1.002

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 5: PCE Forecasts: Subsamples

Panel A: Sample 2000 Q1 -2009 Q4					
Horizon	1	2	3	4	1-4
AR GAP MSC	0.952	0.968	1.016	1.003	0.988
AR GAP SPF	0.959	0.984	1.021	1.009	0.999
AR GAP LVG	0.957	0.981	1.027	1.017	1.004
AR GAP BC	0.961	0.982	1.021	1.010	0.994
AR GAP Aggregate	0.977	0.998	1.038	1.031	1.039
Combination					
Simple Average	0.884	0.930*	1.010	0.997	0.899*
Bates-Granger	0.884	0.933*	1.008	0.992	0.911*
VAR MSC	0.930	0.961	1.078	1.030	1.040
VAR SPF	0.939	1.026	1.059	1.029	1.042
VAR LVG	0.934	0.981	1.082	1.050	1.035
VAR BC	0.944	1.027	1.056	1.029	1.044
VAR Aggregate	0.971	0.989	1.077	1.017	1.040
VECM MSC	0.942	0.977	1.086	1.033	1.048
VECM SPF	0.952	1.037	1.061	1.037	1.062
VECM LVG	0.940	0.988	1.085	1.054	1.049
VECM BC	0.945	1.033	1.062	1.028	1.056
VECM Aggregate	0.963	0.992	1.073	1.011	1.042
Combination					
Simple Average	0.942	0.995	1.064	1.027	1.044
Bates-Granger	0.941	0.993	1.063	1.026	1.043
Panel B: Sample 2010 Q1 - 2019 Q4					
AR GAP MSC	0.948**	1.011	0.994	0.961	0.970
AR GAP SPF	0.946*	0.968	0.940	0.914	0.879**
AR GAP LVG	0.927**	0.963	0.946	0.923	0.873**
AR GAP BC	0.941*	0.965	0.958	0.919	0.887**
AR GAP Aggregate	0.947*	0.969	0.944	0.914	0.881*
Combination					
Simple Average	1.099	1.164	1.106	1.096	1.282
Bates-Granger	1.108	1.167	1.106	1.087	1.287
VAR MSC	0.998	1.219	1.178	1.105	1.298
VAR SPF	0.982	1.213	1.178	1.106	1.297
VAR LVG	1.059	1.223	1.133	1.088	1.322
VAR BC	0.976	1.214	1.178	1.106	1.297
VAR Aggregate	0.982	1.213	1.177	1.123	1.273
VECM MSC	1.002	1.240	1.182	1.121	1.324
VECM SPF	0.982	1.212	1.181	1.113	1.302
VECM LVG	1.053	1.218	1.129	1.083	1.312
VECM BC	0.970	1.222	1.189	1.106	1.302
VECM Aggregate	0.976	1.210	1.175	1.079	1.249
Combination					
Simple Average	0.986	1.209	1.165	1.101	1.295
Bates-Granger	0.979	1.206	1.160	1.100	1.288

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, five forecast horizons, and two subsamples. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 6: PCE Forecasts: COVID 19 Sample 2020Q1 - 2022Q2

Horizon	1	2	3	4	1-4
Panel A: Univariate Models					
AR GAP MSC	1.042	1.026	1.062	1.066	1.055
AR GAP SPF	1.073	1.074	1.095	1.091	1.092
AR GAP LVG	1.058	1.064	1.096	1.094	1.088
AR GAP BC	1.070	1.067	1.091	1.088	1.087
AR GAP Aggregate	1.059	1.059	1.083	1.081	1.079
Combination					
Simple Average	0.921	0.870**	0.928**	0.942**	0.897
Bates-Granger	0.919	0.866**	0.928**	0.942**	0.903
Panel B: Multivariate Models					
VAR MSC	1.066	1.034	0.995	1.061	1.029
VAR SPF	1.027	0.984	1.041	1.054	1.024
VAR LVG	0.993	0.945	1.068	1.061	1.027
VAR BC	1.028	0.984	1.040	1.054	1.023
VAR Aggregate	0.961	0.899	1.082	1.056	0.981
VECM MSC	0.997	0.998	1.032	1.052	1.017
VECM SPF	1.032	0.987	1.040	1.052	1.023
VECM LVG	0.995	0.946	1.040	1.052	1.017
VECM BC	1.023	0.980	1.037	1.050	1.019
VECM Aggregate	0.942	0.905	1.085	0.976	0.952
Combination					
Simple Average	1.002	0.963	1.048	1.047	1.012
Bates-Granger	0.996	0.957	1.046	1.045	1.008

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 7: CPI Forecasts: COVID 19 Sample 2020Q1 - 2022Q2

Horizon	1	2	3	4	1-4
Panel A: Univariate Models					
AR GAP MSC	1.046	1.001	1.043	1.047	1.038
AR GAP SPF	1.071	1.038	1.070	1.066	1.067
AR GAP LVG	1.061	1.030	1.070	1.066	1.063
AR GAP BC	1.068	1.033	1.067	1.064	1.063
AR GAP Aggregate	1.058	1.025	1.060	1.058	1.055
Combination					
Simple Average	1.130	1.031	1.088	1.084	1.076
Bates-Granger	1.129	1.029	1.088	1.084	1.076
Panel B: Multivariate Models					
VAR MSC	1.333	0.993	1.061	1.065	1.047
VAR SPF	1.263	0.998	1.058	1.060	1.053
VAR LVG	1.186	0.994	1.151	1.076	1.070
VAR BC	1.287	0.995	1.061	1.060	1.052
VAR Aggregate	1.150	0.932	1.122	1.171	1.002
VECM MSC	1.335	0.945	1.102	1.052	1.039
VECM SPF	1.279	1.000	1.058	1.056	1.051
VECM LVG	1.187	0.997	1.146	1.071	1.065
VECM BC	1.286	0.992	1.058	1.058	1.050
VECM Aggregate	1.104	0.953	1.150	1.021	0.947
Combination					
Simple Average	1.219	0.957*	1.089	1.063	1.035
Bates-Granger	1.192	0.935*	1.068	1.056	1.017

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 8: Core PCE Forecasts: COVID 19 Sample 2020Q1 - 2022Q2

Horizon	1	2	3	4	1-4
Panel A: Univariate Models					
AR GAP MSC	1.093	1.048	1.080	1.097	1.088
AR GAP SPF	1.131	1.117	1.132	1.134	1.144
AR GAP LVG	1.106	1.097	1.128	1.135	1.133
AR GAP BC	1.117	1.102	1.120	1.123	1.130
AR GAP Aggregate	0.915	0.998	1.069	1.119	1.040
Combination					
Simple Average	0.879	0.813***	0.886	0.908*	0.842*
Bates-Granger	0.951	0.837	0.892	0.915	0.833**
Panel B: Multivariate Models					
VAR MSC	1.186	1.022	0.978	1.080	1.021
VAR SPF	1.236	0.987	1.029	1.045	1.018
VAR LVG	1.198	0.917	1.080	1.057	1.024
VAR BC	1.239	0.984	1.041	1.039	1.017
VAR Aggregate	1.227	1.105	0.922	1.275	0.997
VECM MSC	1.174	1.000	0.989	1.083	1.015
VECM SPF	1.222	1.013	1.036	1.044	1.021
VECM LVG	1.247	0.917	1.027	1.082	1.022
VECM BC	1.220	0.989	1.035	1.036	1.011
VECM Aggregate	1.197	1.102	0.928	1.318	0.984
Combination					
Simple Average	1.179	0.939	0.978	1.084	1.010
Bates-Granger	1.115	0.890	0.964	1.069	0.972

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 9: Survey Data Direct Forecasts Full Sample 2000Q1 - 2022Q2

Panel A: PCE

Horizon	1	2	3	4	1-4
MSC	0.935*	0.983	1.025	1.024	1.011
SPF	0.926*	0.987	1.031	1.026	1.013
LVG	0.922*	0.984	1.034	1.032	1.014
BC	0.923	0.984	1.034	1.027	1.011
Aggregate	0.932	0.990	1.038	1.033	1.025

Panel B: CPI

Horizon	1	2	3	4	1-4
MSC	0.914**	0.970*	1.014	1.010	0.982
SPF	0.902**	0.969	1.013	1.008	0.973*
LVG	0.900**	0.966	1.015	1.012	0.974*
BC	0.899**	0.967	1.015	1.008	0.971*
Aggregate	0.903**	0.966	1.014	1.009	0.973

Panel C: Core PCE

Horizon	1	2	3	4	1-4
MSC	1.033	1.025	1.046	1.068	1.072
SPF	1.054	1.055	1.083	1.086	1.113
LVG	1.039	1.044	1.080	1.085	1.103
BC	1.054	1.046	1.080	1.084	1.108
Aggregate	1.040	1.067	1.112	1.140	1.149

Note: This table presents the forecast accuracy of five surveys at five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Three inflation measures are considered: PCE, CPI and core PCE. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 10: PCE Rolling Forecasts: Full Sample 2000Q1 - 2022Q2

Horizon	1	2	3	4	1-4
AR(1)	1.000	1.000	1.000	1.000	1.000
AR GAP MSC	0.968	1.012	1.056	1.061	1.071
AR GAP SPF	0.959	1.006	1.052	1.054	1.058
AR GAP LVG	0.954	1.004	1.058	1.063	1.062
AR GAP BC	0.956	1.004	1.057	1.058	1.059
AR GAP Aggregate	0.958	1.006	1.057	1.060	1.062
Combination					
Average	0.926	0.954	0.989	0.971	0.938
Bates-Granger	0.927	0.955	0.990	0.969	0.942
VAR MSC	0.956	1.020	1.1038	1.055	1.076
VAR SPF	0.943	1.056	1.069	1.064	1.073
VAR LVG	1.211	1.397	1.393	1.494	1.150
VAR BC	0.972	1.040	1.073	1.057	1.081
VAR Aggregate	1.032	1.022	1.141	1.105	1.079
VECM MSC	1.388	1.420	1.433	1.427	1.776
VECM SPF	1.371	1.403	1.413	1.369	1.733
VECM LVG	1.900	2.524	2.016	3.753	2.342
VECM BC	1.363	1.396	1.404	1.360	1.717
VECM Aggregate	1.554	1.607	1.556	1.617	1.952
Combination					
Average	1.075	1.174	1.198	1.238	1.286
Bates-Granger	0.994	1.074	1.130	1.108	1.166

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 11: CPI Rolling Forecasts: Full Sample 2000Q1 - 2022Q2

Horizon	1	2	3	4	1-4
AR(1)	1.000	1.000	1.000	1.000	1.000
AR GAP MSC	0.955	1.000	1.045	1.045	1.050
AR GAP SPF	0.948	0.996	1.042	1.039	1.041
AR GAP LVG	0.944	0.993	1.045	1.045	1.042
AR GAP BC	0.941	0.993	1.045	1.043	1.041
AR GAP Aggregate	0.944	0.993	1.043	1.042	1.039
Combination					
Average	0.945	0.963	1.009	0.996	0.968
Bates-Granger	0.940	0.960	1.006	0.994	0.963
VAR MSC	1.152	1.105	1.222	1.030	1.101
VAR SPF	1.066	1.131	1.077	1.056	1.091
VAR LVG	2.412	2.277	2.713	2.963	1.543
VAR BC	1.147	1.110	1.110	1.054	1.106
VAR Aggregate	1.343	1.310	1.364	1.378	1.136
VECM MSC	1.325	1.359	1.400	1.336	1.703
VECM SPF	1.315	1.354	1.358	1.302	1.691
VECM LVG	1.639	2.083	2.018	2.333	2.326
VECM BC	1.306	1.344	1.355	1.298	1.673
VECM Aggregate	1.530	1.568	1.494	1.591	1.912
Combination					
Average	1.088	1.163	1.221	1.215	1.270
Bates-Granger	1.043	1.068	1.120	1.065	1.166

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, and five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

6 Out-of-Sample Forecast Stability

In the previous sections, our forecast comparisons were based on measures of global performance. For example, we were able to say that forecasts from one model were better than those from the AR(1) benchmark model on average. Rather, we want to compare the out-of-sample forecasting performance of two competing models in the presence of possible instabilities. Under unstable environments, the forecasting performance of the two models may vary over time. We implement the fluctuation test outlined in Giacomini and Rossi (2010).

We apply the fluctuation test to analyze the time variation in the out-of-sample forecasting performance of our AR gap, VAR and VECM models relative to the AR(1) benchmark. The advantage of using this approach is the ability to focus on the relative performance of the competing models over time. Therefore, we may recover useful information that is lost when looking for the model that provides the best forecast on average. We consider three measures of inflation: PCE, CPI and core PCE. We use the one-step ahead forecasts from our models using our newly created aggregate survey data outlined in section 3.2.

The fluctuation test measures the local relative forecasting performance of the two models. The null hypothesis is that the mean squared prediction error difference equals zero at each point in time. This means that the forecasting performance of the two models is the same at each point in time. The alternative states that the forecasting performance differs at least at one point in time. We apply the two sided version of the test and set the rolling window to be equal to 8 quarters.

In Figure 3, the solid blue line indicates a sequence of differences between the MSPE of the AR(1) benchmark and the MSPE of our models. Following Giacomini and Rossi (2010), each MSPE difference is rescaled by its standard deviation, to abstract from unit of measurement issues. Positive (negative) values of MSPE differences indicate that the AR(1) benchmark produced better (worse) forecasts than our models. The dashed red lines represent the critical values which, if crossed, signals that the null of equal forecasting performance at each point in time is rejected. Therefore, the alternative that one forecast is significantly better than the other one holds.

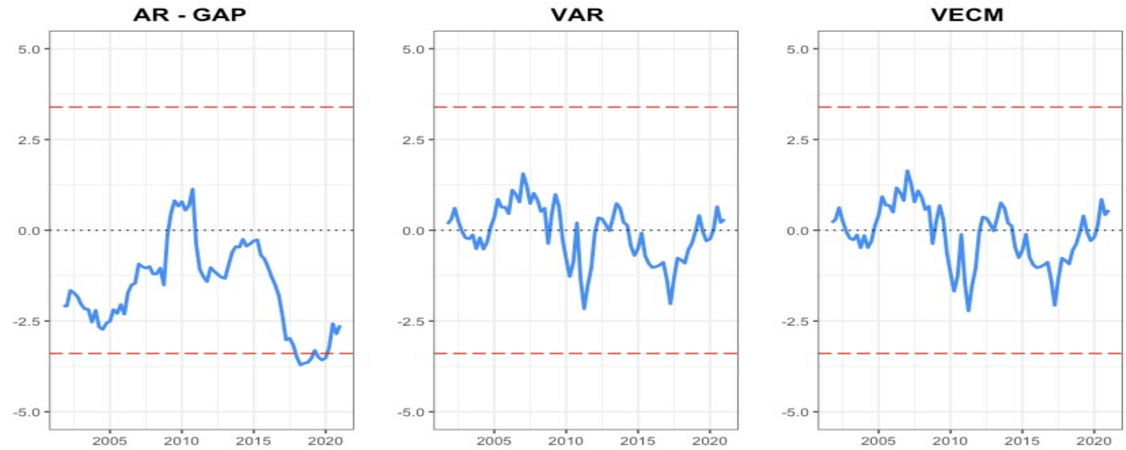
Let us now present the main finding of the fluctuation test depicted in Figure 3 with our three measures of inflation PCE, CPI and Core PCE. Despite that our models provide a better average forecast over the 2000Q1 to 2022Q2 period, there are times when the AR(1) benchmark produces better forecasts: mostly during unstable time periods like the Great Financial Crisis and the Covid-19 pandemic.

First, we discuss the results for PCE price inflation. Panel (a) of Figure 3 shows that the AR gap model produces significantly better forecast than the benchmark AR(1) model. Since the local relative MSPE exceeds the critical value around 2018-2019, we reject the null hypothesis of equal forecast performance. It must be noted that most of local MSPEs are located below the zero line, which confirms the findings of Table 2 that on average the AR gap model outperforms the AR(1) benchmark. However, we must note that there were times were the AR(1) benchmark produced better forecast: mainly during 2007-2011 period.

The multivariate VAR and VECM models performed better than the benchmark between 2015 and 2019. Again, we see that during unstable time periods like the Great Financial Crisis, and the Covid-19 period, the AR(1) benchmark provided better forecasts. Since the critical values are not crossed by the sequence of rolling

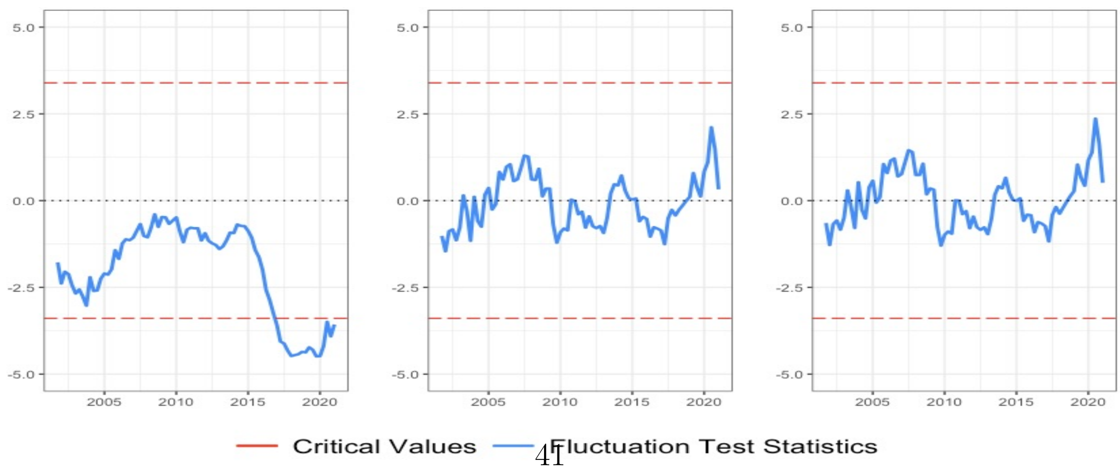
Figure 3: Fluctuation Test

(a) PCE



(b) CPI

(c) Core PCE



Note: The solid blue line depicts a sequence of differences between the MSPE of the considered model and the MSPE of the AR(1) benchmark model computed over rolling windows of 8 observations. The dashed red lines indicate the critical values of the test. Positive (negative) values of MSPE differences indicate that the AR(1) benchmark produced better (worse) forecasts than our models.

MSPE there is no evidence that these multivariate models produced significantly better forecasts than the AR(1).

Next, we discuss the results of the fluctuation test for CPI price inflation. The AR gap model provides better forecasts than the AR(1) on average except for the Great Financial Crisis time period. During such unstable time, the AR(1) provided better forecasts. Moreover, Table 3 shows that on average the multivariate VAR and VECM models do not perform better than the AR(1) benchmark. Panel (b) of Figure 3 shows that there are times were the mutivariate models do better than the AR(1) benchmark: mainly during the 2014 to 2018 period.

Finally, we present the results for core PCE price inflation. The AR gap model produces significantly better forecasts than the benchmark since the test statistics are clearly outside of the critical values. We therefore reject the null of equal forecasts at each point in time. The fluctuation test also shows that the multivariate VAR and VECM models are outperformed by the AR(1) benchmark during the Great financial Crisis and Covid-19 pandemic time periods.

7 Conclusion

This paper explored the role of inflation expectations survey data in real-time forecasting of US inflation. We first revisit the autoregressive inflation gap literature using different surveys of long-run inflation expectations. Then, we extend the autoregressive gap models to a multivariate setting. The VAR framework embeds the endogenous relationship between inflation and inflation expectations while the VECM model integrates additional information on their long run relationship to improve forecasting performance.

The findings corroborate that while simple models remain hard to beat, the Vector Autoregressive (VAR) and Vector Error Correction Model (VECM) can improve forecast performance at short horizons. Additionally, our work provides insights on the following three topics. First, forecast combination of competing models improve forecast accuracy over simpler models. This is in line with portfolio diversification theory and provides an hedge against bad forecasts. Second, aggregating survey data of inflation expectations from consumers and professional forecasters, using dynamic principal components, provides a new trend inflation measure that improves the forecast accuracy of our models. Third, the additional information obtained from the error correction process between inflation and long run inflation expectations can improve forecasting performance in a multivariate framework. These findings were robust to alternative forecasting methodologies including the use of surveys as subjective forecasts and a rolling forecasting scheme.

We also examined the local relative out-of-sample forecasting performance of the AR gap, VAR and VECM models to the AR(1) benchmark. By implementing the fluctuation test outlined in Giacomini and Rossi (2010), we found that despite our models performing better on average over the 200Q1 to 2022Q2 period, there are times when the AR(1) benchmark produces better forecasts: mostly during unstable time periods like the Great Financial Crisis an the Covid-19 pandemic.

Appendix

A Alternative inflation measures: Pre vs Post GFC

Table A.1: CPI Forecasts: Subsamples

Panel A: Sample 2000 Q1 -2009 Q4					
Horizon	1	2	3	4	1-4
AR GAP MSC	0.923	0.958*	1.011	0.993	0.947*
AR GAP SPF	0.912	0.964	1.009	0.990	0.940
AR GAP LVG	0.909	0.959	1.012	0.995	0.939
AR GAP BC	0.908*	0.960	1.010	0.989	0.930
AR GAP Aggregate	0.916	0.965	1.016	0.999	0.954
Combination					
Simple Average	0.929	0.952	1.012	0.989	0.925
Bates-Granger	0.919	0.944	1.005	0.982	0.905
VAR MSC	1.147	0.942	1.099	1.030	1.102
VAR SPF	1.068	1.095	1.080	1.021	1.080
VAR LVG	1.236	1.009	1.129	1.085	1.104
VAR BC	1.103	1.127	1.102	1.042	1.086
VAR Aggregate	1.277	1.059	1.127	1.078	1.117
VECM MSC	1.174	0.941	1.155	1.074	1.101
VECM SPF	1.092	1.103	1.094	1.027	1.101
VECM LVG	1.242	1.008	1.129	1.089	1.124
VECM BC	1.090	1.123	1.101	1.033	1.094
VECM Aggregate	1.186	1.063	1.122	0.982	1.073
Combination					
Simple Average	1.141	1.009	1.043	1.003	1.090
Bates-Granger	1.129	0.999	1.037	0.994	1.088
Panel B: Sample 2010 Q1 - 2019 Q4					
AR GAP MSC	0.892***	1.000	0.987	0.965	0.949*
AR GAP SPF	0.877***	0.954	0.939*	0.925**	0.857***
AR GAP LVG	0.871***	0.958	0.947	0.937*	0.868***
AR GAP BC	0.874***	0.959	0.954	0.932*	0.870***
AR GAP Aggregate	0.876***	0.955	0.940*	0.926	0.857**
Combination					
Simple Average	0.836**	0.967	0.927	0.897*	0.838**
Bates-Granger	0.831**	0.957	0.922	0.893*	0.827**
VAR MSC	0.975	1.232	1.190	1.062	1.219
VAR SPF	0.941	1.195	1.148	1.062	1.220
VAR LVG	1.253	1.192	1.094	1.042	1.294
VAR BC	0.938	1.188	1.148	1.061	1.223
VAR Aggregate	1.002	1.213	1.195	1.119	1.185
VECM MSC	0.959	1.239	1.172	1.064	1.233
VECM SPF	0.939	1.193	1.147	1.064	1.225
VECM LVG	1.250	1.185	1.089	1.039	1.285
VECM BC	0.937	1.195	1.155	1.060	1.228
VECM Aggregate	0.991	1.214	1.197	1.029	1.137
Combination					
Simple Average	0.954	1.155	1.123	1.051	1.219
Bates-Granger	0.910	1.132	1.095	1.046	1.168

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, five forecast horizons, and two subsamples. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table A.2: Core PCE Forecasts: Subsamples

Panel A: Sample 2000Q1 -2009Q4					
Horizon	1	2	3	4	1-4
AR GAP MSC	1.036	1.005	1.012	1.000	1.051
AR GAP SPF	1.048	1.062	1.063	1.027	1.122
AR GAP LVG	1.051	1.059	1.075	1.045	1.134
AR GAP BC	1.068	1.080	1.082	1.051	1.157
AR GAP AGGREGATE	1.166	1.219	1.229	1.198	1.409
Combination					
Simple Average	1.269	1.192	1.292	1.344	1.546
Bates-Granger	1.765	1.677	1.751	1.793	2.351
VAR MSC	1.135	0.977	1.023	1.023	1.018
VAR SPF	1.133	1.007	1.029	1.024	1.028
VAR LVG	1.243	1.020	1.088	1.061	1.054
VAR BC	1.108	1.064	1.051	1.026	1.017
VAR AGGREGATE	1.276	1.096	1.150	1.066	1.055
VECM MSC	1.156	1.003	1.018	1.039	1.029
VECM SPF	1.085	1.038	1.026	1.029	1.029
VECM LVG	1.192	1.040	1.094	1.049	1.036
VECM BC	1.102	1.071	1.063	1.031	1.046
VECM AGGREGATE	1.247	1.124	1.166	1.052	1.044
Combination					
Simple Average	1.142	1.010	1.050	1.027	1.027
Bates-Granger	1.127	1.003	1.042	1.025	1.023
Panel B: Sample 2010Q1 - 2019Q4					
AR GAP MSC	0.964	1.053	1.047	1.093	1.150
AR GAP SPF	1.011	1.028	1.048	1.063	1.132
AR GAP LVG	0.990	1.018	1.051	1.069	1.122
AR GAP BC	1.014	1.011	1.052	1.067	1.128
AR GAP Aggregate	1.074	1.125	1.204	1.213	1.463
Combination					
Simple Average	1.315	1.382	1.301	1.152	1.727
Bates-Granger	1.531	1.811	1.750	1.360	2.439
VAR MSC	1.035	1.052	1.093	1.081	1.180
VAR SPF	0.999	1.072	1.094	1.081	1.175
VAR LVG	1.010	1.076	1.083	1.031	1.158
VAR BC	0.986	1.079	1.095	1.085	1.173
VAR Aggregate	0.971	1.092	1.083	1.025	1.081
VECM MSC	1.049	1.047	1.111	1.112	1.247
VECM SPF	0.992	1.083	1.093	1.099	1.192
VECM LVG	1.025	1.077	1.110	1.074	1.174
VECM BC	0.973	1.081	1.107	1.085	1.172
VECM Aggregate	0.991	1.101	1.091	1.000	1.081
Combination					
Simple Average	0.980	1.059	1.085	1.064	1.153
Bates-Granger	0.975	1.053	1.079	1.061	1.147

Note: This table presents the forecast accuracy for seven univariate models, twelve multivariate models, five forecast horizons, and two subsamples. The RMSPE ratios against the AR(1) benchmark are reported. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

B Survey-based Forecasts

Table B.3: Survey Data Direct Forecasts - 2000Q1 to 2019Q4

Panel A: PCE

Horizon	1	2	3	4	1 to 4
MSC	0.921*	0.977	1.014	0.994	0.981
SPF	0.900*	0.970	1.008	0.989	0.954
LVG	0.899*	0.968	1.012	0.996	0.959
BC	0.898*	0.969	1.013	0.990	0.956
Aggregate	0.912*	0.980	1.023	1.006	0.991

Panel B: CPI

Horizon	1	2	3	4	1 to 4
MSC	0.894**	0.967	1.006	0.987	0.944*
SPF	0.874**	0.957	0.994	0.976	0.900**
LVG	0.873**	0.955	0.998	0.982	0.905**
BC	0.871**	0.956	0.999	0.977	0.901**
Aggregate	0.876**	0.957	0.999	0.982	0.912*

Panel C: Core PCE

Horizon	1	2	3	4	1 to 4
MSC	1.051	1.020	1.020	0.980	1.057
SPF	1.038	1.023	1.038	0.989	1.067
LVG	1.028	1.012	1.032	0.984	1.049
BC	1.052	1.029	1.054	0.998	1.090
Aggregate	1.193	1.165	1.180	1.123	1.356

Note: This table presents the forecast accuracy of five surveys at five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Three inflation measures are considered: PCE, CPI and core PCE. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table B.4: Survey Data Direct Forecasts - 2020Q1 to 2022Q2

Panel A: PCE

Horizon	1	2	3	4	1 to 4
MSC	1.040	1.011	1.060	1.065	1.048
SPF	1.072	1.059	1.093	1.090	1.084
LVG	1.060	1.050	1.094	1.093	1.082
BC	1.068	1.050	1.088	1.087	1.079
GDPC	1.055	1.040	1.080	1.079	1.069

Panel B: CPI

Horizon	1	2	3	4	1 to 4
MSC	1.052	0.990	1.042	1.047	1.032
SPF	1.075	1.027	1.069	1.065	1.060
LVG	1.066	1.019	1.068	1.066	1.057
BC	1.073	1.020	1.065	1.063	1.056
GDPC	1.062	1.012	1.058	1.057	1.048

Panel C: Core PCE

Horizon	1	2	3	4	1 to 4
MSC	1.083	1.029	1.078	1.097	1.078
SPF	1.126	1.097	1.129	1.132	1.134
LVG	1.106	1.079	1.125	1.134	1.125
BC	1.114	1.076	1.115	1.121	1.117
GDPC	0.964	0.987	1.081	1.125	1.057

Note: This table presents the forecast accuracy of five surveys at five forecast horizons. The RMSPE ratios against the AR(1) benchmark are reported. Three inflation measures are considered: PCE, CPI and core PCE. Cases in which the relative root mean square prediction error is significantly lower than the baseline model at the 10, 5 and 1 percent significance levels are denoted with one, two or three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Declarations

Disclosure of Potential Conflicts of Interest.

The author declares that there is no conflict of interest.

Funding.

No funding was received to assist with the preparation of this manuscript.

Data availability.

The author confirms that the data supporting the findings of this study are available within its supplementary materials.

References

Ahn, H. J., Fulton, C., et al. (2020). *Index of common inflation expectations* (Tech. Rep.). Board of Governors of the Federal Reserve System (US).

Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of monetary Economics*, *54*(4), 1163–1212.

Atkeson, A., Ohanian, L. E., et al. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, *25*(1), 2–11.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the operational research society*, *20*(4), 451–468.

Bernanke, B. S., et al. (2007). Inflation expectations and inflation forecasting. In *Speech at the monetary economics workshop of the national bureau of economic research summer institute, cambridge, massachusetts* (Vol. 10).

Bravo, J. M., & El Mekkaoui, N. (2022). Short-term cpi inflation forecasting: Probing with model combinations. In *World conference on information systems and technologies* (pp. 564–578).

Brillinger, D. R. (2001). *Time series: data analysis and theory*. SIAM.

Chan, J. C., Clark, T. E., & Koop, G. (2018). A new model of inflation, trend inflation, and long-run inflation expectations. *Journal of Money, Credit and Banking*, *50*(1), 5–53.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134–144.

Döpke, J., Dovern, J., Fritsche, U., & Slacalek, J. (2008). The dynamics of european inflation expectations. *The BE Journal of Macroeconomics*, *8*(1).

Faust, J., & Wright, J. H. (2013). Forecasting inflation. In *Handbook of economic* *forecasting* (Vol. 2, pp. 2–56). Elsevier.

Feldkircher, M., & Siklos, P. L. (2019). Global inflation dynamics and inflation expectations. *International Review of Economics & Finance*, *64*, 217–241.

Fulton, C., & Hubrich, K. (2021). Forecasting us inflation in real time. *Econometrics*, *9*(4), 36.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*(4), 595–620.

Groen, J. J., Kapetanios, G., & Price, S. (2009). A real time evaluation of bank of england forecasts of inflation and growth. *International Journal of Forecasting*, *25*(1), 74–80.

Groen, J. J., Paap, R., & Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, *31*(1), 29–44.

Hauzenberger, N., Huber, F., & Klieber, K. (2022). Real-time inflation forecasting using non-linear dimension reduction techniques. *International Journal of Forecasting*.

Hörmann, S., Kidziński, Ł., & Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *77*(2), 319–348.

Hubrich, K., & Skudelny, F. (2017). Forecast combination for euro area inflation: a cure in times of crisis? *Journal of Forecasting*, *36*(5), 515–540.

Kishor, N. K. (2021). Forecasting real-time economic activity using house prices and credit conditions. *Journal of Forecasting*, *40*(2), 213–227.

Kishor, N. K., Koenig, E. F., et al. (2022). Finding a role for slack in real-time inflation forecasting. *International Journal of Central Banking*, *18*(2), 245–282.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics*, *14* (1-2), 3–24.

Peña, D., Smucler, E., & Yohai, V. J. (2019). Forecasting multiple time series with one-sided dynamic principal components. *Journal of the American Statistical Association*.

Peña, D., Smucler, E., & Yohai, V. J. (2020). gdpc: an r package for generalized dynamic principal components. *Journal of Statistical Software*, *92*, 1–23.

Peña, D., & Yohai, V. J. (2016). Generalized dynamic principal components. *Journal of the American Statistical Association*, *111* (515), 1121–1131.

Rudd, J. B. (2022). Why do we think that inflation expectations matter for inflation?(and should we?). *Review of Keynesian Economics*, *10* (1), 25–45.

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44* (2), 293–335.

Stock, J. H., & Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, *39*, 3–33.

Svensson, L. E. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European economic review*, *41* (6), 1111–1146.

Svensson, L. E., & Woodford, M. (2004). Implementing optimal policy through inflation-forecast targeting. In *The inflation-targeting debate* (pp. 19–92). University of Chicago Press.

Tulip, P. (2009). Has the economy become more predictable? changes in greenbook forecast accuracy. *Journal of Money, Credit and Banking*, *41* (6), 1217–1231.

Wright, J. H. (2009). Forecasting us inflation by bayesian model averaging. *Journal of Forecasting*, *28* (2), 131–144.