**WORKING PAPER NO. 17-26
DO PHILLIPS CURVES CONDITIONALLY HELP TO
FORECAST INFLATION?**

Michael Dotsey
Research Department
Federal Reserve Bank of Philadelphia

Shigeru Fujita
Research Department
Federal Reserve Bank of Philadelphia

Tom Stark
Research Department
Federal Reserve Bank of Philadelphia

This version: August 2017

Do Phillips Curves Conditionally Help to Forecast Inflation?

Michael Dotsey, Shigeru Fujita, and Tom Stark*

This version: August 2017

Abstract

This paper reexamines the forecasting ability of Phillips curves from both an unconditional and conditional perspective by applying the method developed by Giacomini and White (2006). We find that forecasts from our Phillips curve models tend to be unconditionally inferior to those from our univariate forecasting models. Significantly, we also find conditional inferiority, with some exceptions. When we do find improvement, it is asymmetric – Phillips curve forecasts tend to be more accurate when the economy is weak and less accurate when the economy is strong. Any improvement we find, however, vanished over the post-1984 period.

1 Introduction

The Phillips curve has long been used as an important guide for monetary policy. Its use was recently well articulated by Chair Janet Yellen at the Philip Gamble Memorial Lecture Series at the University of Massachusetts Amherst where she stated "Economic theory suggests, and empirical analysis confirms, that such deviations of inflation from trend depend partly on the intensity of resource utilization in the economy − as approximated, for example, by the gap between the actual unemployment rate and its so-called natural rate..." One need only read the transcripts or the minutes of FOMC meetings to realize the central role that the Phillips curve occupies in monetary policy discussions. However, given the recent, numerous studies indicating that, over the past 20 years or so, inflation forecasts based on the Phillips curve generally do not predict inflation any better than a naive forecast or a forecast based on either an unobserved stochastic volatility model or an IMA(1,1) model raises the question of whether the Phillips curve should continue to occupy such an important place in policy discussions. One of the first papers to cast doubt on the usefulness of Phillips curve forecasts was that of Atkeson and Ohanian (2001) who found that naive forecasts of inflation generally outperform those based on a Phillips curve model.[1] Since then, the question of relative forecasting performance has been explored in a variety of papers, most notably by Stock and Watson (2007, 2008). Thus, a reasonable impression regarding the usefulness of Phillips curve models for forecasting inflation is fairly bleak. Of note is that theoretical work by Benigno and Ricci (2011) provides persuasive reasons why this outcome may be expected.

Stock and Watson, however, pose an interesting hypothetical question: Despite the rich evidence against the usefulness of Phillips curve forecasts, would you change your forecast of inflation if you were told that the economy was going to enter a recession in the next quarter with the unemployment rate jumping by 2 percentage points? There is strong evidence that many forecasters and monetary policymakers would, in fact, change their forecasts. For example, the June 4 2010 issue of Goldman Sachs' *US Economics Analyst* posits, "Under any reasonable economic scenario, this gap − estimated at 6.5% of GDP as of year-end 2009 by the Congressional Budget Office − will require years of above-trend growth to eliminate. Accordingly, we expect the core consumer inflation measures ⋯ to trend further, falling close to 0% by late 2011." These sentiments were echoed in the April 27−28, 2010 minutes of the Federal Open Market Committee: "In light of stable longer-term inflation expectations and the likely continuation of substantial resource slack, policymakers anticipated that both

---

[1]To be more precise, there are some papers in the literature prior to Atkeson and Ohanian (2001) that point out the early "warning signs" about deterioration of the Phillips curve forecasts in the late 1990s (e.g., Brayton et al. (1999) and Stock and Watson (1999)). See Section 3.1 in Stock and Watson (2008) for a discussion on this topic.

overall and core inflation would remain subdued through 2012."

Although most studies that examine the comparative forecasting performance of Phillips curve models place emphasis on the performance over entire sample periods and specific subsamples, there has been little work that sheds light on the question posed by Stock and Watson. Dotsey and Stark (2005) examine whether large decreases in capacity utilization add any forecasting power to inflation forecasts and find that they do not. However, Stock and Watson (2008) provide some rough evidence that large deviations of the unemployment gap are associated with periods when Phillips curve-based forecasts are relatively good. Fuhrer and Olivei (2010) also examine the Stock and Watson evidence and find that a threshold model of the Phillips curve outperforms a naive model. This paper statistically investigates the strength of the Stock and Watson observation along a number of dimensions and in great depth.

We do so in a variety of ways using both real-time and final data and by formally comparing forecast accuracy of our Phillips curve-based forecasts with those of various univariate models using the methodology developed by Giacomini and White (2006). We use their procedure because (i) it can be used when comparing the forecasts from misspecified models, (ii) it allows for both unconditional and conditional tests, and (iii) it is relevant for testing both nested and nonnested models. To explore whether it is primarily large deviations of the unemployment gap that are informative for inflation forecasting, we look at a threshold model as well as use the conditional forecast comparison procedures developed by Giacomini and White (2006).

Our basic results indicate that forecasts from our baseline Phillips curve model or the model augmented with a threshold unemployment gap are unconditionally inferior to those of our naive forecasting models, and the difference is sometimes statistically significant, especially over a post-1984 sample period. We generally also do not find that conditioning on various measures of the state of the economy improves the performance of the Phillips curve model relative to the IMA(1,1) model in a statistically significant way with an exception of the SPF (Survey of Professional Forecasters) recession downturn probabilities. With respect to a random walk forecast, conditioning on various states of the economy does improve the relative forecasting power of the Phillips curve model with more regularity, but the relative improvement is far from a universal outcome of the test. Of interest is that improvement is more likely to occur over the entire sample period 1969Q1−2014Q2 than occurs over the later sample 1984Q1-2014Q2. Further, we find little or no evidence that supports the conjecture in Stock and Watson (2008) that the size of the unemployment gap improves forecasts. Importantly, over the later sample, there are no conditioning variables that significantly help

to improve the forecast of the Phillips curve model relative to the IMA(1,1) model, indicating that the answer to the Stock and Watson question is no. Thus, our results indicate that monetary policymakers should at best be very cautious in their reliance on the Phillips curve when gauging inflationary pressures.

Following a brief literature review, we lay out the various forecasting models. We then discuss the procedures used for comparing forecasts. We follow this with the body of our statistical analysis and then provide a brief summary and conclusion.

2   Literature Review

Our literature review is fairly focused, concentrating on those papers that help inform our particular approach. An excellent and in-depth literature review on inflation forecasts can be found in Stock and Watson (2008).[2] A departure point for our inquiry is the work of Atkeson and Ohanian (2001). In that paper, the authors compare the root-mean-square errors (RMSEs) of out-of-sample forecasts of 12-month-ahead inflation generated by a Phillips curve model using either the unemployment rate or a monthly activity index developed at the Federal Reserve Bank of Chicago with those of a naive model, which predicts that 12-month-ahead inflation will be the same as current 12-month inflation. They examine the relative RMSEs for forecasts over the period between Jan. 1984 and Nov. 1999 and find that the forecasts generated by the Phillips curve models do not outperform those of the naive model. Therefore, they conclude that the Phillips curve approach is not useful for forecasting inflation. Stock and Watson (1999) look at two subsamples when comparing the relative forecasting power of Phillips curve specifications with a naive forecast and one based on an autoregressive specification of the inflation rate. Over the first subsample, 1970−1983, the Phillips curve-based forecasts are superior, whereas over the second subsample 1984−1996, the Phillips curve-based forecasts outperform the naive forecast but are no better than forecasts based on lagged inflation only.

This is in stark contrast to Atkeson and Ohanian (2001), and as reported in Stock and Watson (2008), it is due to the different sample period. In particular, Phillips curve forecasts did not do well in the latter half of the 1990s. Further, over the 1984−1999 sample period, the naive forecast outperforms forecasts based on simple autoregressive specifications, which prompts Stock and Watson to adopt an unobserved components stochastic volatility (UCSV) model as their benchmark for comparison. They find that there is not much dif-

---

[2]On the methodological side, the literature on forecast comparisons was initiated by Diebold and Mariano (1995). See, for example, Clark and McCracken (2013) for a review of this literature.

ference between the naive forecasts over the 1984−1999 subsample, but that subsequently the forecasts generated by the two methods diverge, at which point the UCSV forecasts are superior. Fisher et al. (2002) use rolling regressions with a 15-year window rather than recursive procedures. They also document that Phillips curve-based forecasts outperform naive forecasts over the period 1977−1984 and that, for a PCE-based inflation measure, the Phillips curve forecasts improved on naive forecasts over the period of 1993−2000. They also indicate that the 1985−1992 and 1993−2000 periods may represent different forecasting environments. Another intriguing result from Fisher et al. (2002) is that Phillips curve forecasts do better at two-year horizons, which is in stark contrast to the findings in Stock and Watson (2007), who find that Phillips curve forecasts tend to do better at horizons of less than one year. Ang et al. (2007), however, tend to confirm the Atkeson-Ohanian results that Phillips curve models offer no improvement over naive forecasts for the periods 1985−2002 and 1995−2002, a result that is consistent with those found in Stock and Watson (2008) when the latter use UCSV as the atheoretical benchmark.

Clark and McCracken (2006) reach a more cautious conclusion, pointing out that the out-of-sample confidence bands for ratios of RMSEs are fairly wide and that rejecting Phillips curve models based on ratios should be approached with care. However, some of the ratios found in studies such as Atkeson and Ohanian (2001) (AO), Stock and Watson (2007), and Ang et al. (2007) are so large that they probably imply failure to reject the null of no forecast improvement. However, many of the ratios reported in Fisher et al. (2002) are only slightly greater than one and most likely do not imply a rejection of the null hypothesis. From a practical point of view, one can interpret much of the evidence in these papers as indicating that activity gaps are not reliable predictors of inflation and that inflation forecasts are not overly sensitive to whether or not a Phillips curve is relied upon.

Like ours, some studies use real-time data. Orphanides and van Norden (2005) find that Phillips curve-based forecasts using an output gap measure of real activity outperform an autoregressive benchmark prior to 1983 but offer no improvement over the 1984-2002 period. In addition, a number of studies have found that the Phillips curve specification has been unstable over time. Stock and Watson (1999, 2007) find that the instability is largely confined to the coefficients on lagged inflation, whereas Clark and McCracken (2006) find instability in the coefficients on the output gap. Dotsey and Stark (2005) also find instability in coefficients on capacity utilization, with those coefficients becoming smaller and insignificant as they rolled their sample forward. Giacomini and Rossi (2009) find evidence of forecast failure in real-time Phillips curve projections, caused by changes in inflation volatility as well as changes in the monetary policy regime.

Finally, Stock and Watson (2008) present an interesting finding, which indicates that, although inflation forecasts based on the Phillips curve do not outperform forecasts based on inflation alone, there are episodes when that is not the case. In particular, they notice that the RMSEs from Phillips curve forecasts tend to be lower than those from an unconditional stochastic volatility model when the unemployment gaps are larger than 1.5 in absolute value. This finding motivates our interest in conditional forecasting tests.

3   Forecasting models

To investigate what appears to be a particular type of nonlinearity associated with forecasting performance, we use standard Phillips curve models together with the conditional forecast comparison methods of Giacomini and White (2006) to indicate whether Phillips curve models provide better forecasts of inflation when conditional on the state of the economy. Because Stock and Watson (2008) indicate that the measure of real activity is of secondary importance when evaluating forecast performance, we concentrate on unemployment rates and unemployment gaps. We also use real-time data on unemployment as our benchmark data set but investigate whether the use of real-time data as opposed to final data affects our results. We also concentrate our forecasting exercise on headline PCE (Personal Consumption Expenditures) inflation and do so for two reasons. One is that PCE inflation is often considered to be the most relevant measure of inflation for policy purposes. It is also less affected by commodity price shocks than the CPI. Using the headline as opposed to the core allows us to extend our sample period further back in time, and we can, therefore, include data from the 1969 and 1973 recessions.

3.1   The Benchmark Models

Our two benchmark models will be the naive forecasting model of Atkeson and Ohanian (2001) and the rolling IMA(1,1) model of Stock and Watson (2007).[3] Following Stock and Watson (2008), the naive forecast is based on the following specification:

$$E_t(\pi_{t+h}^h - \pi_{t-1}^4) = 0, \tag{1}$$

---

[3]One might wonder that another natural candidate would be a finite-order autoregressive model. The literature has shown, however, that univariate autoregressive models are consistently beaten by the two univariate models considered here and thus not part of our univariate reference models. See Table 3 in Stock and Watson (2008) that compares the forecasting performance of various models for overall PCE inflation including 10 different autoregressive models over six different sample periods.

where $\pi_t^h = (400/h)[\log(p_t) - \log(p_{t-h})]$ and p_t is the PCE price index and $h = 2, 4, 6,$ and 8. The IMA(1,1) specification for quarter-over-quarter inflation is given by

$$\Delta \pi_t = \epsilon_t + \theta \epsilon_{t-1}. \tag{2}$$

In estimating the model, we use only the real-time observations that would have been available at the date when the forecast was made.[4]

3.2 Phillips Curve Models

To investigate the benefits of a Phillips curve model for forecasting inflation, we examine a simple autoregressive Phillips curve model given by:

$$\pi_{t+h}^h - \pi_t = a^h(L)\Delta \pi_t + b^h(L)\widetilde{u}_t + v_{t+h}^h, \tag{3}$$

where π_{t+h}^h is the h-quarter-ahead forecast of an h-quarter-annualized average of inflation and $\widetilde{u}_t$ is the unemployment gap. We will use time-varying estimates of NAIRU based on real-time measures that are constructed using an HP (Hodrick and Prescott) filter where we pad future observations with forecasts from an AR(4) model for unemployment (see below). In addition we shall append the model with a threshold term. The threshold model is, therefore, an extension of the Phillips curve with a threshold effect on the unemployment gap. The threshold variable is an absolute value of the unemployment gap:

$$\pi_{t+h}^h - \pi_t = \alpha^h(L)\Delta \pi_t + 1(|\tilde{u}_t| > u)\gamma(L)\tilde{u}_t + 1(|\tilde{u}_t| \leq u)\delta(L)\tilde{u}_t + \nu_{t+h}, \tag{4}$$

where u is a threshold value and $1(|\tilde{u}_t| > u)$ takes the value of unity when $|\tilde{u}_t| > u$ and zero otherwise. Initially, we intended to use the TAR (Threshold Autoregressive) model of Hansen (1997). However, there was insufficient variation in the data to identify the threshold over any of our rolling windows. Therefore, we imposed a value of 1.2, which implied that the absolute value of the unemployment gap exceeds the threshold one-third of the time, by using the one standard deviation value of the real-time gap estimated using the latest vintage unemployment over the period 1954Q1−2014Q2. Doing so provided us with enough threshold measures to conduct our conditioning tests.[5]

---

[4]Stock and Watson (2008) indicate that the IMA(1,1) model performs about as well as a more sophisticated UCSV model.

[5]Although we cannot meaningfully estimate the threshold values for each rolling sample, as argued above, we did estimate one threshold value for the entire sample and for each h. The estimated threshold values for the whole sample period are indeed around 1.2 for all forecasting horizons, ranging between 1.07 ($h = 8$)

3.3 Forecast Comparison

Statistical forecast comparisons are made using the methods developed by Giacomini and White (2006), whose procedure can be used for nested and nonnested models as well as for constructing both unconditional and conditional tests of forecast accuracy. Using their procedure requires limited memory estimators such as fixed windows. This allows them to formulate test statistics that come from a chi-square distribution. Given the apparent instability in the Phillips curve, the rolling window methodology appears superior to a recursive forecasting procedure. For unconditional tests, the null hypothesis is for equal predictability of forecasting methods, which can be formerly stated as $E(\delta_{t+h}) = 0$, where δ_{t+h} is the difference in the squared h-step-ahead forecast errors between any two forecasting methods.[6] The relevant test statistic is as follows:

$$n\Big(n^{-1}\sum_t \delta_{t+h}\Big)\widehat{V}_h^{-1}\Big(n^{-1}\sum_t \delta_{t+h}\Big) \xrightarrow{d} \chi_1^2, \tag{5}$$

where h denotes the forecast horizon, n is the size of the forecast sample, and $\widehat{V}_h$ is the HAC variance of $n^{-1}\sum_t \delta_{t+h}$. Note that the HAC correction is necessary, since we are looking at multiple-period-ahead forecast errors. We apply a uniform lag window with truncation parameter set to $h-1$.[7] For conditional tests, we examine the test statistic:

$$n\Big(n^{-1}\sum_t x_t\delta_{t+h}\Big)'\widehat{V}_h^{-1}\Big(n^{-1}\sum_t x_t\delta_{t+h}\Big) \xrightarrow{d} \chi_k^2, \tag{6}$$

where x is a $k \times 1$ vector of conditioning variables and $\widehat{V}_h$ is the HAC-corrected estimator of the variance of $n^{-1}\sum_t x_t\delta_{t+h}$.

The unconditional test statistic tells us only if the forecasts are statistically different from one another on average over the sample. To ascertain which of any two models is giving the better forecast, we examine the sign of the coefficient in the regression:

$$\delta_{t+h} = \beta_0 + e_{t+h}. \tag{7}$$

A negative coefficient indicates that model one, which we denote the reference model, produces the better forecast on average. We shall refer to model two as the alternative model.

---

and 1.33 ($h = 2$).

[6] To be precise, note that forecast error differences between the two *methods* could arise due to estimation uncertainty as well as model differences, as discussed in Giacomini and White (2006).

[7] When the uniform lag window produces a nonpositive definite variance, we use the Bartlett lag window and increase the truncation lag to $2(h-1)$.

When comparing the forecasts of our two Phillips curve models with the two benchmark models, we also examine when there are statistically significant differences conditional on (i) whether the economy is in recession, (ii) the probability of recession from the SPF data set, (iii) our real-time estimate of the unemployment gap, (iv) the four-quarter change in the unemployment gap, (v) the absolute value of the real-time gap, and (vi) whether the gap is bigger than a specified threshold. It is important to note that the conditional GW test is a marginal test. It tells us whether conditioning on a certain value significantly improves one forecast relative to another, not whether the forecast is actually better. For example, if the IMA(1,1) model gave an unconditionally better forecast and we find that conditioning on a recession significantly improves the Phillips curve forecast relative to the IMA(1,1) forecast, our results do not indicate that the Phillips curve is conditionally providing a better forecast, only that conditioning significantly improves its forecast relative to that of the IMA(1,1) model. To infer which forecast is better, we need to look at the size and sign of the regression coefficient, β_1, on the conditioning variable in the regression:

$$\delta_{t+h} = \beta_0 + \beta_1 x_{i,t} + e_{t+h}, \tag{8}$$

where x_i is one of our conditioning variables. For the first four conditioning variables, when the slope coefficient is statistically significant, we calculate the cut off value that implies that the alternative model's forecast is better. It is important to note that, because we are generally conditioning on variables that were known at the time of the forecasts, the fact that relative forecast accuracy depends on this information implies that none of our models are true data-generating mechanisms and that each is to some degree misspecified. Constructing the true model is likely to be an extremely difficult exercise, and the conditioning tests are a simple, straightforward alternative for analyzing whether the state of the economy affects the relative usefulness of Phillips curve forecasting models. One could argue that conditioning on whether the economy is in recession or not is conditioning on information that forecasters are unlikely to possess in real time. That is true, strictly speaking, but the SPF recession probabilities indicate that forecasters are generally cognizant in real time as to whether the economy is or is not in recession. The evidence in a recent paper by Kotchoni and Stevanovic (2016) also supports this argument. These authors compute recession probabilities based on Probit models that are estimated only on real-time data and show that their real-time recession probabilities also exhibit significant increase during the recessions. Note also that, as mentioned above, the SPF recession probability is also used as one of the conditioning variables and that exercise is from the issues surrounding the NBER recession dummy.[8]

---

[8]As a real-time measure of the recession probability, we prefer the SPF measure over the one based on

Figure 1: Unemployment and Unemployment Gap Series: The unemployment gap is based on the HP filter with smoothing parameter 10^5. The final estimate of the gap series uses the 2014Q3 vintage of data. Shading indicates periods of the NBER recession.
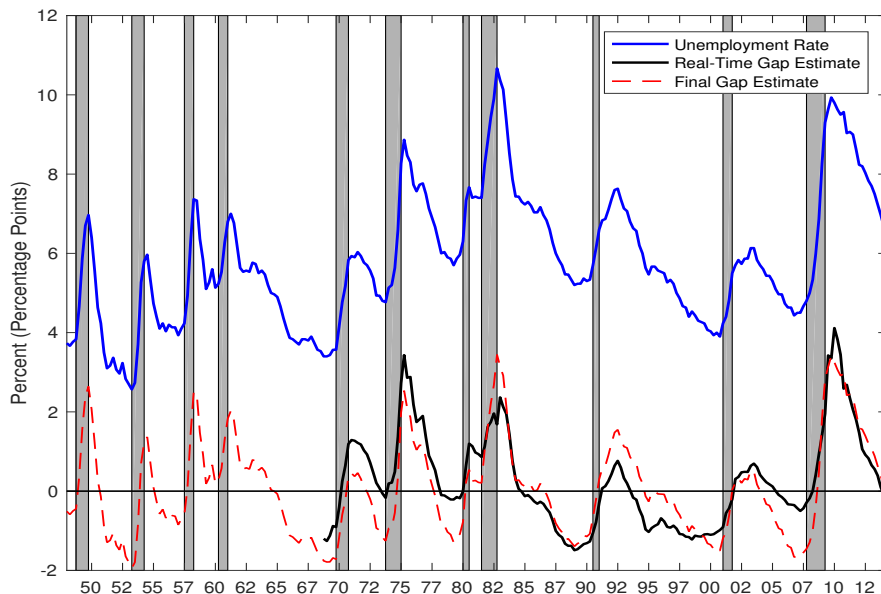
4 Data Definitions and Transformations

Our analysis uses real-time data on unemployment and PCE inflation constructed from vintage data available to the public in the middle of the quarter.[9] Thus, a regression run at date t uses observations on unemployment and inflation, as they were known as of that date. As regressions are rolled forward, updated data are used from the vintage that were available as of the new date. The quarter-over-quarter inflation rate is defined as $\pi_t = 400 \log(P_t/P_{t-1})$, and the h-quarter annual average inflation rate at time t is given by $\pi_t^h = (400/h) \log(P_t/P_{t-h})$.

---

a particular statistical model as in Kotchoni and Stevanovic (2016) because the SPF series captures the consensus of many forecasters rather than a prediction from one particular model, which is more prone to some idiosyncratic errors (such as model misspecification). Moreover, even though one can construct a "real-time" recession probability using only the data available in real time, the model itself can be chosen ex post with the benefit of hindsight, and thus, strictly speaking, the model-based measure may not constitute a true real-time measure.

[9]The real-time data used in this paper are available at the Philadelphia Fed's website "Real-Time Data Set for Macroeconomists" (www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/). Note also that unlike the large revisions that occur for the real-time unemployment gap, which we document later, the revisions to the unemployment series are fairly small. They are due only to revisions in seasonals and rarely result in a revision of more than 0.1 percentage point.

A key variable in our analysis is the unemployment gap, $\tilde{u}_t$, defined as the difference between the unemployment rate and the HP estimate of trend unemployment. Specifically, we use the smoothing parameter of 10^5 to identify the trend component.[10] In constructing trend unemployment, we use an HP filter with 20 quarters of forecast values beyond the sample endpoint. The forecasts are from an AR model of unemployment where the maximum lag length is four and the fixed window for the regression is 84 quarters. The lag length is selected separately each period using the SIC criteria. The unemployment gap is given by:

$$\tilde{u}_t = u_t - u_t^{HP}, \tag{9}$$

where u_t^{HP} is the HP trend, which we associate with a time-varying NAIRU. Orphanides and van Norden (2005) and Orphanides and Williams (2005) indicate that there are significant differences between real-time and final estimates of the unemployment gap, and we find similar results for our construct over our sample period. The final time estimates are constructed by HP-filtering the unemployment rate over the entire sample.

As can be seen in Figure 1, revisions to the unemployment gap are significant. The solid black line depicts the real-time estimates of the unemployment gap, and the dotted red line shows the unemployment gap using the 2014Q3 vintage of data. The largest revisions do not seem to follow any particular pattern. For example, in both the latter half of the 1970s and the latter half of the 2000s, the unemployment gap is a good deal higher than the final estimate, and these are periods of falling unemployment. The opposite is true of the 1990s, however, when the real-time gap is lower than the final estimate and again unemployment is falling.

The dependent variable in the main body of our analysis is various averages of real-time headline PCE inflation, and these are depicted in Figure 2. Comparing Figure 1 with Figure 2, it is evident that the unemployment gap is a much more heavily revised series than is inflation.

---

[10]Stock and Watson (2007, 2008) use a high-pass filter that filters out frequencies of less than 60 quarters. The value of the smoothing parameter (10^5) is often used in the recent labor search literature (see Shimer (2005)). There is variation in the literature regarding what frequency should be used, and we recognize that the properties of the unemployment gap are sensitive to the choice of the smoothing parameter. In general, most studies use an unemployment gap that is constructed by including frequencies significantly lower than those associated with the traditional business cycle frequencies as in this paper.

| (a) Two-Quarter Average | (b) Four-Quarter Average |
| (c) Six-Quarter Average | (d) Eight-Quarter Average |
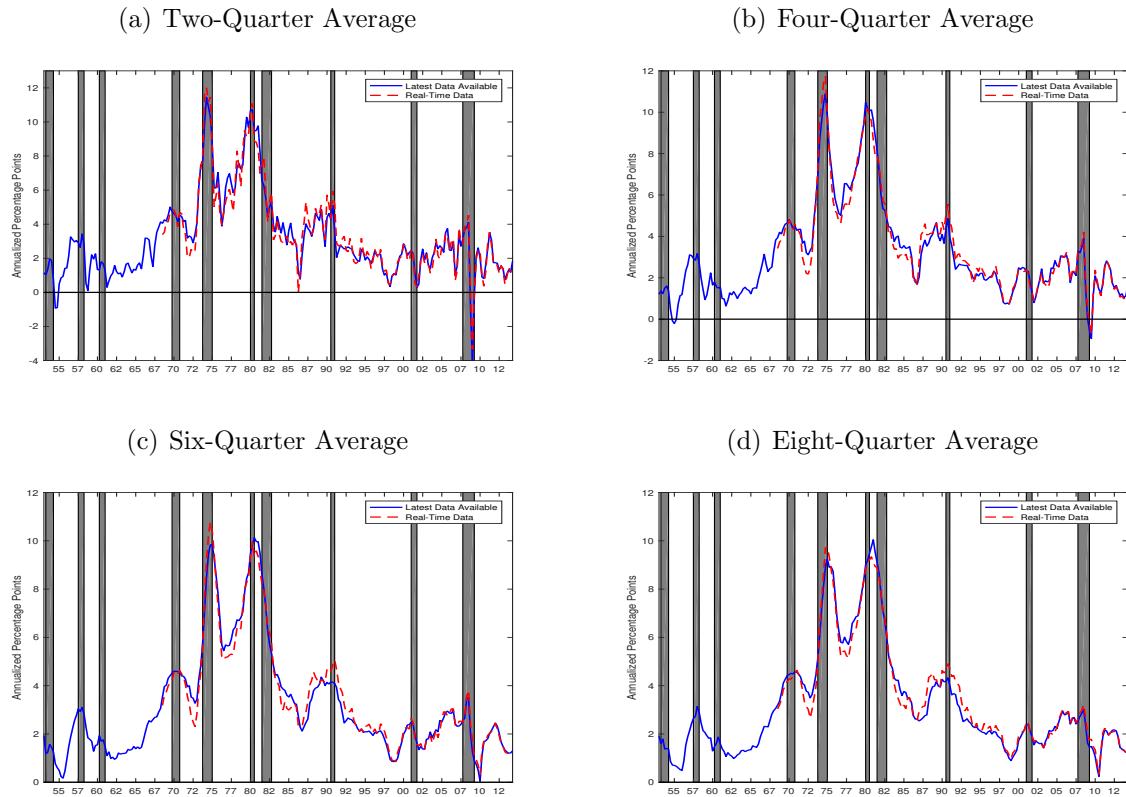
Figure 2: Headline PCE Inflation Realizations: Shading indicates periods of NBER-dated recessions.

5   The Usefulness of Phillips Curve Forecasts

In this section, we analyze how useful Phillips curve models are for forecasting inflation in real time. Our motivation for emphasizing the use of real-time data is twofold. First, these data are relevant for policy purposes. Second, the work of Orphanides and van Norden (2005) on the output gap and our own analysis of real-time unemployment gaps make the strong case for incorporating the measurement error associated with the real-time gap. Our investigation focuses on whether unemployment gaps provide useful information in extreme circumstances. The exploration of whether Phillips curve models estimated on final data generally help predict inflation has already been exhaustively explored in the literature.[11] In a subsequent section, we will analyze the role that using real-time data plays by comparing our results with those using final data.

Here we compare the Phillips curve forecasts from (3) and (4) with our two benchmark models (1) and (2) where we use unemployment gaps based on the current real-time vintage

---

[11]For an excellent summary as well as an exhaustive set of experiments, see Stock and Watson (2008).
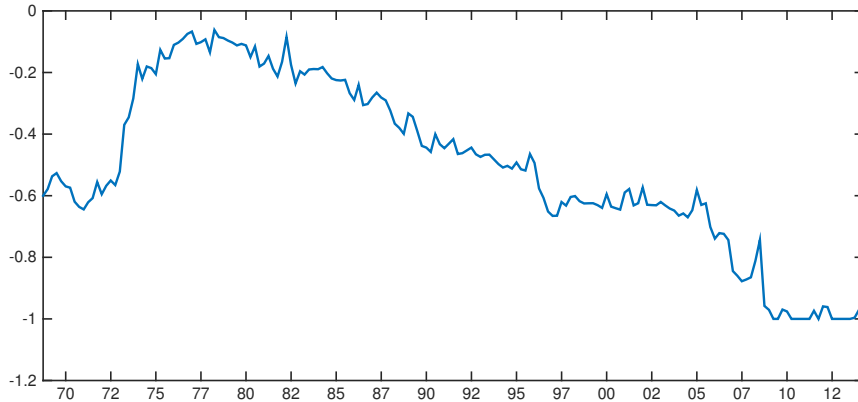
Figure 3: IMA(1,1) Coefficient Estimates: Estimated on a fixed window of 60 quarters. Coefficient estimates are aligned at sample endpoints.

as of period t. The lag length is reestimated each period using the SIC lag selection method, and lag lengths are allowed to vary across the variables. In statistically comparing forecasts, we use both the unconditional and conditional forecast tests developed in Giacomini and White (2006). We do this for four forecast horizons, namely two-, four-, six-, and eight-quarter-ahead average forecasts of inflation. We also compare the forecasts over two sample periods: the entire sample period from 1969Q1 to 2014Q2 and a later sample period that includes forecasts from 1984Q1 through 2014Q2. The entire sample begins in 1969Q1 for the two-step horizons, as it is the earliest date that we can make a forecast based on a 60-quarter window. We break the sample at 1984 because that latter sample is associated with the Great Moderation and consistently low and less variable inflation.

5.1   An Analysis of Our Regression Results

Before turning to the forecast comparison tests, it is useful to examine some of the properties of our forecasting models. First, we note that the estimates of the moving average coefficient, θ, in the real-time fixed window IMA(1,1) model vary over time (Figure 3). Early in the sample (the mid to late 1970s), a 1 percentage point inflation shock is associated with a long-run multiplier $(1 + \hat{\theta})$ on the level of inflation of roughly .90. The multiplier then declines fairly consistently. At present, the long-run multiplier is about zero, implying that the persistence of the inflation process has declined significantly over our sample period. Over recent 60-quarter windows, inflation shocks have had only a negligible long-run effect on the level of inflation. Thus, over our sample, the behavior of inflation changes from

13

(a) Two-Quarter Average Inflation

(b) Four-Quarter Average Inflation

(c) Six-Quarter Average Inflation

(d) Eight-Quarter Average Inflation
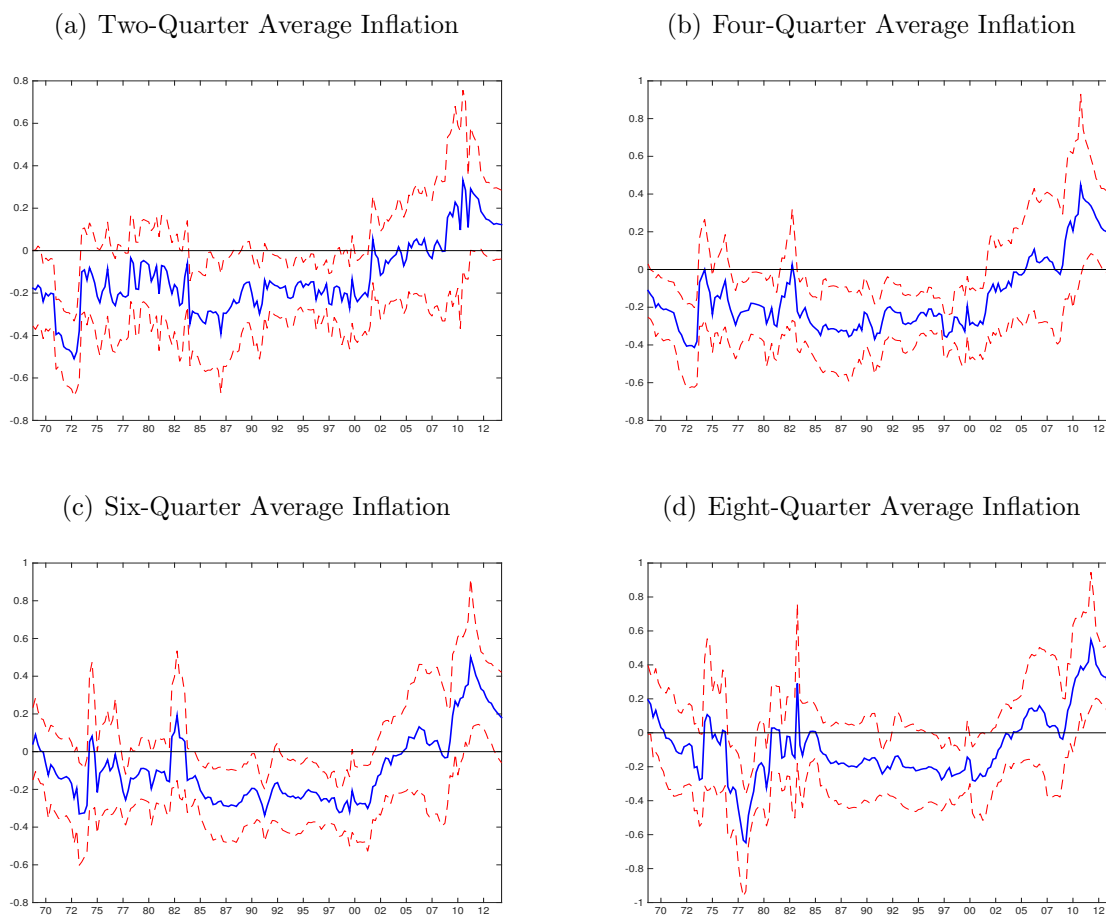
Figure 4: Coefficients on the Unemployment Gap in the Phillips Curve: Dashed lines indicate the 90 percent confidence interval based on HAC standard errors.

something close to a random walk to a process that more closely resembles white noise.[12]

Importantly, we also find evidence of instability in the coefficient estimates on the gap in the Phillips curve (Figure 4). In particular, the in-sample effect of the unemployment gap on inflation varies over time and across forecast horizons. The Phillips curve literature suggests that a larger gap precedes lower inflation. The estimate of the sum-of-coefficients is typically negative for inflation equations at all horizons and is significantly so for the four-quarter- and six-quarter-ahead forecasts, but it becomes less negative and statistically insignificantly different from zero around 2000 as we roll the regressions forward. Surprisingly, the coefficient actually takes on the incorrect sign during the Great Recession. The falling significance may in part be due to the more transitory nature of changes in inflation that we documented

---

[12]Our result is consistent with evidence in Stock and Watson (2007) and occurs because the volatility of the permanent component of inflation has been decreasing over time.

previously along with the observation that the movements in the gap remain highly persistent over the entire sample.

As mentioned, in recent 60-quarter windows, the sign of the sum of the coefficients turns positive. In these latter samples, a larger gap is associated with higher, not lower, inflation. This is significantly so for all but the two-quarter-ahead forecasts. Graphically, we display the coefficient instability at our four horizons. The instability we find in the coefficients in both the univariate model and the Phillips curve is consistent with evidence presented in Ang et al. (2007) and serves as justification for using a rolling windows methodology as is also done in Fisher et al. (2002). Using more formal statistical techniques, Giacomini and Rossi (2009) indicate that the Phillips curve relationship suffers from forecast breakdowns due to instabilities in the data-generating process.

5.2 Forecast Comparisons

In this section, we compare both the unconditional and conditional forecasting performances of our four models. We first take a general look at the forecasts and document the contribution of unemployment gaps to these forecasts. Subsequently, we perform the statistical forecast comparison exercise developed by Giacomini and White (2006).

5.2.1 An Initial Look at the Forecasts

An initial examination of the relative forecasting ability of the various models is shown in Table 1. We see that the IMA(1,1) forecasts are preferred to those of the AO model and both Phillips curve specifications over both the full sample and the more recent sample. The findings regarding the relative forecasting ability of our two benchmark models generally agree with the analysis of Stock and Watson (2007).

In Figure 5, we show the forecasts for each horizon, along with actual inflation. The largest disparities between the IMA(1,1) and the Phillips curve forecasts at all horizons occur in the early 1980s and the entire Great Recession period and subsequent recovery when the IMA(1,1) model indicates that inflation itself is white noise. During the most recent period, the Phillips curve forecast is overpredicting inflation.

We next examine the unemployment gap's contribution to the forecasts, which is depicted in Figure 6. Specifically, the contribution of the unemployment gap is given by $\sum_{j=1}^{n(h)} \hat{b}_j^h \tilde{u}_{t-(j-1)}$, where the summation goes from one to the SIC minimizing lag length $n(h)$, calculated at each forecast horizon h, using the appropriate vintage of data. As shown in Figure 6, the contribution of the gap (blue line) is similar across all forecast horizons, but especially so for the four-, six-, and eight-quarter horizons. During the 1970s and early 1980s,

(a) Two-Quarter Average Inflation
(b) Four-Quarter Average Inflation

(c) Six-Quarter Average Inflation
(d) Eight-Quarter Average Inflation
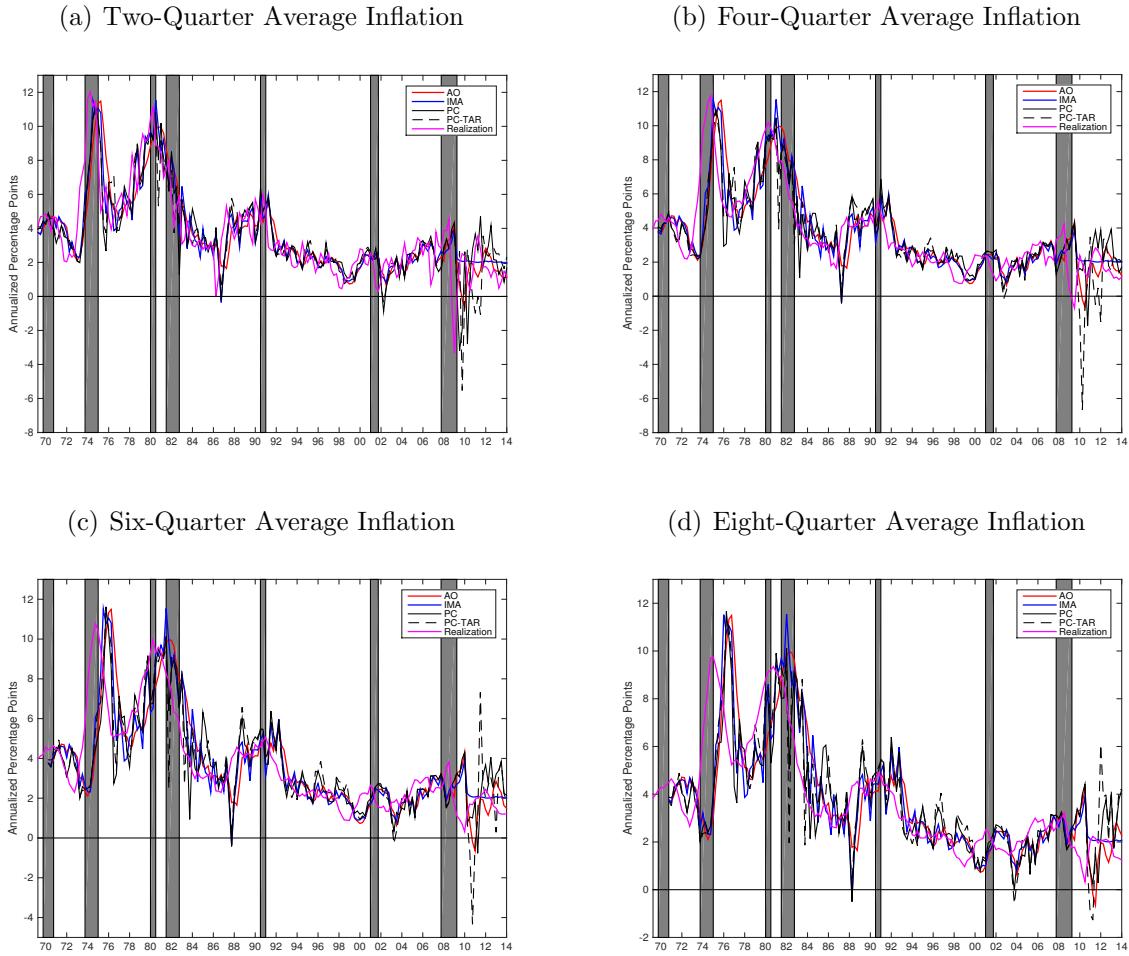
Figure 5: Inflation Projections: Shading indicates periods of NBER-dated recessions.

the unemployment gap makes a pronounced contribution to the Phillips curve projections at all horizons. These periods are characterized by large unemployment gaps that pull down the forecast of inflation. Also, following the 1991 recession, the gap is again high, and it contributes negatively to forecasted inflation. This is true in the early 2000s as well. Recently, the gap (red line) is also high, but it is contributing to higher than expected inflation due to the perverse sign of the estimated coefficient, which, as shown in Figure 4, is now insignificantly positive. Further, Figure 6 points to the reason that the gap is becoming less of a factor in forecasting inflation. Inflation has become much less volatile and less persistent, while the gap has continued to fluctuate and these fluctuations are persistent. The relative stability of inflation makes it less likely that other economic variables will have significant explanatory power with respect to its behavior.

The results in Table 1 and Figures 4 and 5 are suggestive regarding the unconditional test

(a) Two-Quarter Average Inflation   (b) Four-Quarter Average Inflation
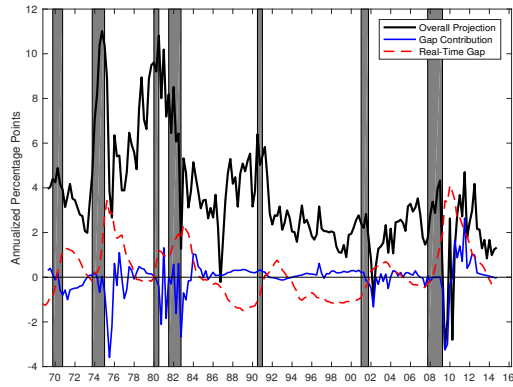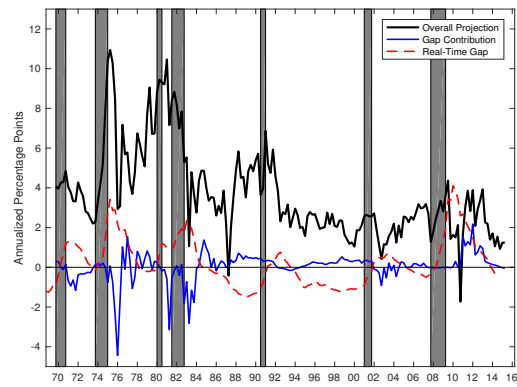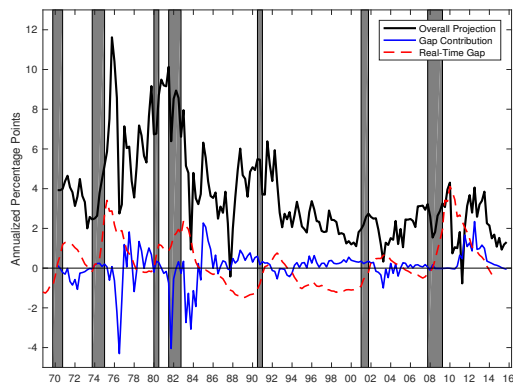
(c) Six-Quarter Average Inflation   (d) Eight-Quarter Average Inflation
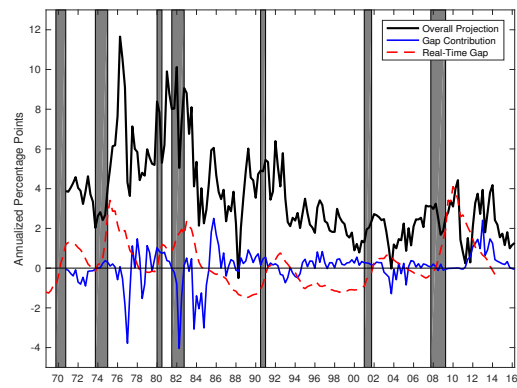
Figure 6: Effect of Unemployment Gap on Phillips Curve Forecasts: The real-time unemployment gap is aligned at the date when the forecast was made. The contribution term is plotted at the date forecasted. Shading indicates periods of NBER-dated recessions.

proposed by Giacomini and White (2006). The explanatory power of the gap seems not to be that significant and appears to be becoming less so, and the forecasting differences between the benchmark models and the Phillips curve models do not appear especially large. These observations, however, are not overly informative about the conditional tests. We do see a few periods where the gap is large, and its contribution to the inflation forecast is helpful relative to benchmark models (at least at the four- and six-quarter horizons), namely after the 1970, 1973, 1980, and 1982 recessions. It remains to be seen if that help is statistically significant.

5.2.2 Theoretical Underpinnings

One important finding of our initial look is the instability in the slope coefficient of the Phillips curve. Since 2000, the slope has steadily declined in absolute value, eventually becoming statistically insignificant and of the wrong sign. And as we show in the next section, as a reduced form unconditional forecasting device the Phillips curve does not outperform naive forecasting models, and there may be good theoretical reasons for this unreliability. Work by Benigno and Ricci (2011) develops a model that gives theoretical foundations, at least qualitatively, to our findings. They incorporate downward nominal wage rigidities (DNWR) into a fairly simple New Keynesian setting in which DNWR is the only nominal friction. In order to avoid encountering the constraint that future nominal wages cannot decline, wage setters less aggressively raise wages relative to what they would do in the absence of this friction. The degree to which the constraint may bind is primarily governed by the growth rate of the economy, average nominal wage inflation, the variance of shocks, and the elasticity of labor supply.

Most relevant to our empirical findings is that the higher average nominal wage inflation is and the lower volatility is, the less likely the constraint will bind and the closer wages will be set to their flexible wage counterpart. Thus, the long-run Phillips curve becomes vertical at high rates of inflation and approaches verticality as the variance of shocks goes to zero. Another feature of the Benigno-Ricci (BR) model is that for high elasticities of labor supply, which require less changes in wages to adjust labor supply, the constraint is less costly and wages are closer to the flexible counterpart as well. Importantly, as volatility rises the Phillips Curve shifts out, because for any inflation rate the constraint is more likely to bind, requiring more inflation to be associated with any given level of unemployment. Firms set current wages lower, which ends up producing more future wage growth at any unemployment rate. This outward shift has the additional implication that absolute value of the slope coefficient increases with volatility, so all things equal, a great moderation should induce a decline in the coefficient's absolute value. We do see a gradual decline in the volatility of nominal expenditure beginning in 1990, so the timing of the decline in the slope coefficient is pretty much in line with what the theory predicts (recall we use a 60-quarter rolling window).

Regarding short-run Phillips curves, it is possible for employment in the BR benchmark model to exceed what would occur with flexible prices. This outcome can occur if wages are initially fairly close to the flexible wage counterpart and the economy was hit by persistently low productivity. With DNWR, the wage would be higher and unemployment lower than if there were no constraint. Hence a portion of the short-run Phillips curve lies to the left of

18

the natural rate of unemployment. Because at very high rates of wage inflation the short-run Phillips curve is vertical at the natural rate of unemployment (for reasons analogous with the LRPC), the short-run curve must bend back and its slope is positive. Thus, it is possible in this theory to produce "wrongly" signed coefficients. However, for inflation and volatility levels consistent with our data, these would be counterfactually high. That is not to say that a more elaborate theory that builds on the BR mechanism would not be compatible with our findings.

An important message of the BR theory though is that the slope coefficient in a standard Phillips curve specification is likely to be unstable because it is a complicated function of deeper structural parameters that are likely to vary over time. We have witnessed a great moderation over our sample period, and medium-term trend inflation has varied as well. Demographic changes have most likely also affected the elasticity of labor supply. Thus, the theoretical model advanced by BR provides underpinnings for some of our results and should provide additional caution for using simple Phillips Curve models as the basis for making unconditional forecasts of inflation.

6   Statistical Comparisons

We now examine the relative forecasting performance of the various models in a precise statistical sense. To do this, we use the unconditional and conditional tests for comparing forecast methods developed by Giacomini and White (2006).

6.1   Unconditional Comparison

First, we investigate whether the results concerning forecast accuracy presented in Table 1 are statistically significant. The unconditional forecasting performance is shown in Table 2, where the left portion of the table refers to our entire sample and the right portion of the table refers to results over the more recent sample. Each row of the table corresponds to a particular benchmark model. For example, in the second row of each panel the IMA(1,1) model is the benchmark. The columns indicate the alternative model, so the second column indicates that the basic Phillips curve model is the alternative. Thus, the (2,2) element of the left half of panel (a) compares the IMA(1,1) model's forecast with that of the Phillips curve. In comparing forecasts we use both a 5% and 10% significance level. Over the entire sample, there is only one statistically significant difference in forecast ability between the naive models and the two Phillips curve specifications, and that occurs for a significantly better forecast performance of the IMA(1,1) model at the two-quarter-ahead horizon. However,

the constant in the second rows of Table 2 is generally negative. With regard to the more recent sample, both the AO and IMA(1,1) specifications are preferred to the Phillips curve specification, while AO is preferred to the Phillips curve threshold specification as well. Thus, from the unconditional tests, there is little to suggest the use of a Phillips curve specification for forecasting headline PCE inflation.

6.2 Conditional Forecasting Tests

In light of the Stock and Watson (2008) findings, we first tried conditioning on the absolute value of the unemployment gap. This is a symmetric test because it analyzes whether conditioning on both large and small values of the gap affects the relative forecasting properties of the two models. Similarly in spirit, we also condition on a threshold dummy that equals one when the absolute value of the gap is greater than 1.20. Alternatively, it may be that the unemployment gap affects the conditional forecasting properties asymmetrically. For example, the forecasts of the Phillips curve model may improve conditional on the output gap being large and positive. To test this type of hypothesis, we conditioned on two measures of the unemployment gap: its level and its four-quarter change. Along these lines, we also condition on recession dates and the estimated recession probabilities from the SPF. The behavior of these conditioning variables is depicted in Figure 7.

The results of our conditional forecast comparison tests are given in Tables 3 through 8. The tables are laid out as follows: The rows refer to the reference models and the columns refer to the alternative model. We report the p-values of the GW χ^2 test statistic, and we report the adjusted R^2 and the estimates of the constant and slope coefficient on the conditioning variable in equation (8). To help highlight the salient features of the exercise, we use three different shadings. The darkest shading indicates that the slope coefficient on the conditioning variable is positive and significant and that the GW χ^2 statistic is significant, indicating that the two forecast methods are significantly different. The middle shading includes cases in which the slope coefficient is positive and statistically significant but the GW χ^2 is not. The lightest shading is where the slope coefficient has a positive sign but is not significant and where the GW χ^2 statistic went from being significant unconditionally to insignificant conditionally. When the conditioning variable is the lagged recession dummy, a positive coefficient implies that the alternative model's accuracy increases in recessions. With respect to the probability of a recession, when the regression coefficient is positive, it means that the higher the probability of recession, the better the Phillips curve forecasts are. In terms of the conditioning variables using the unemployment gap, a positive coefficient implies that high unemployment gaps improve the Phillips curve forecasts but that negative

(a) Absolute Value of Real-Time Gap   (b) Threshold Dummy

(c) Recession Dummy   (d) SPF Recession Probability

(e) Real-Time Unemployment Gap   (f) Four-Quarter Change in Real-Time Gap
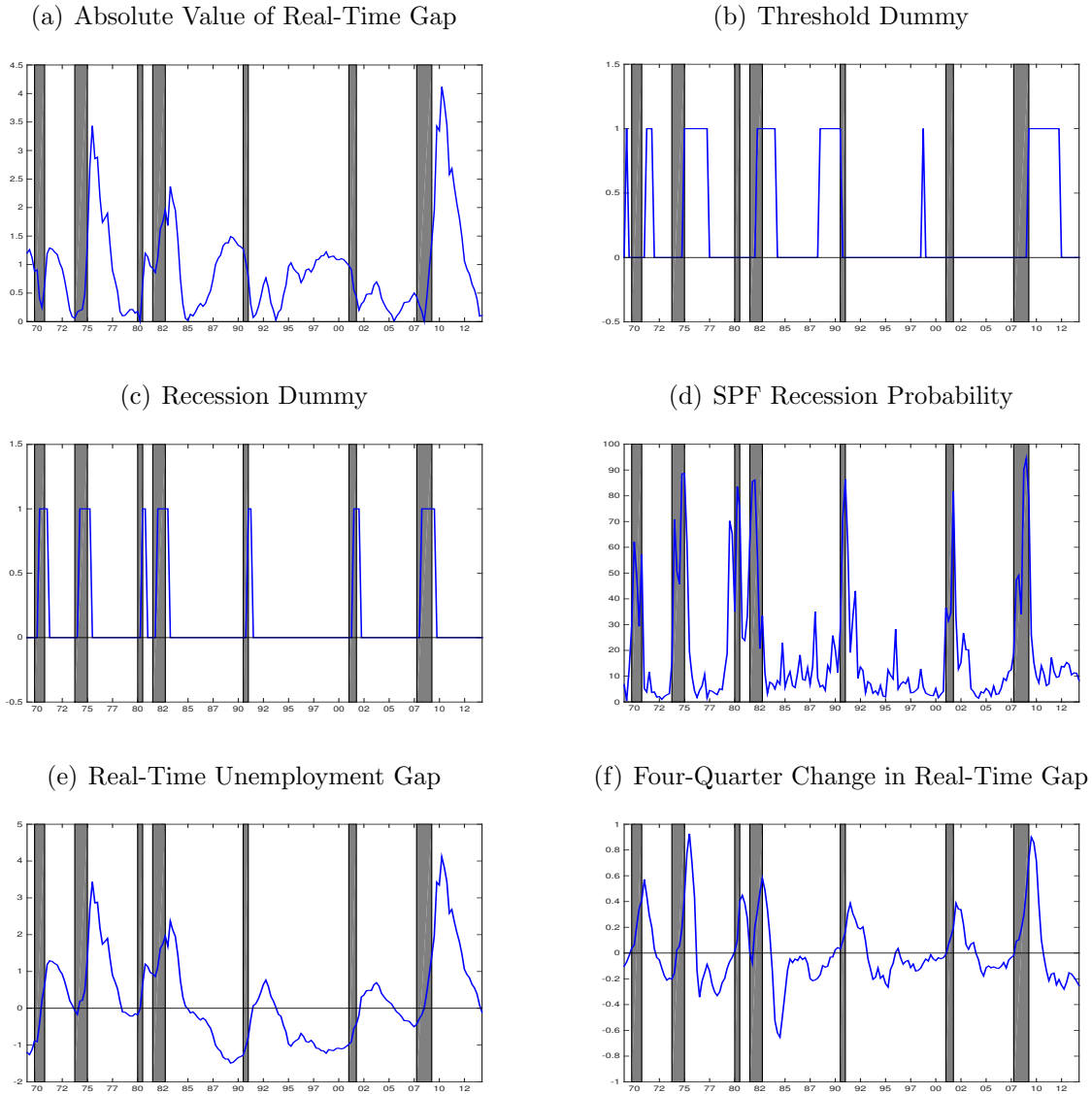
Figure 7: Conditioning Information for Giacomini-White Tests: Each plot shows our GW conditioning information. The data are aligned (using the timing conventions discussed in the paper) at the forecast date (not the date forecasted). Shading shows periods of NBER-dated recessions.

unemployment gaps worsen the Phillips curve forecasts. When assessing the conditional performance of the absolute value of the gap, a positive coefficient means that both high and low gaps tend to improve the Phillips curve forecasts. For the three continuous conditioning variables, we compute the cutoff value of the variable that implies that the alternative forecast outperforms the reference model.

6.2.1 Basic Results

The first basic result is that conditioning on gap-type measures in a symmetric way does not generally improve the forecast performance of Phillips curve models. Table 3 presents the results when the absolute value of the unemployment gap is used as a conditioning variable. Over both the full sample and the more recent sample, we found no cases in which conditioning on this variable improved Phillips curve forecasts relative to those of our two benchmark models. Similarly, conditioning on the threshold dummy does not improve the forecast performance of the Phillips curve models relative to the benchmark models (Table 4). In fact, the slope coefficient is generally negative, and sometimes significantly so. This is especially true with regard to the IMA(1,1) benchmark and the Phillips curve model.

The second basic result is that when the conditioning tends to be asymmetric, we find that, in recessions, over the full sample there is a tendency for improvement in inflation forecasts from the Phillips curve models, especially with regard to the SPF downturn probabilities. There is, however, no evidence that these variables conditionally help Phillips-curve-type forecasts over the more recent sample period.

Using the real-time gap series $\tilde{u}_{t-1}$, there is no evidence that it significantly improves Phillips-curve based forecasts. The slope coefficient is generally of the wrong sign and significantly so in the more recent sample. Table 8 presents the results when the four-quarter change in the real-time gap is used as a conditioning variable. These results indicate that, with respect to the AO forecasts, the change in the gap generally significantly improves the relative conditional forecasting accuracy of the Phillips curve model, especially over the more recent sample period. There is, however, little indication that it does so when the reference model is the IMA(1,1) model.

However, it is important to point out that these findings reflect the average forecast behavior over the sample periods of the GW regressions. As we discussed with respect to Figure 4, coefficients on the unemployment gap in the Phillips curve model are close to zero and not statistically significant in recent years, which implies little statistical difference in recent years between the Phillips curve forecasts and AO or IMA(1,1) forecasts. This suggests that the presence of a large unemployment gap in recent years does not contribute to the superior forecast performance of the Phillips curve models.

6.2.2 When Should One Rely on the Phillips Curve?

It is also important to go beyond a classification of statistical inference and examine when the use of a Phillips curve model is preferred. For example, we saw over the entire sample the slope coefficient on the recession dummy is significant for the Phillips curve model at the four-

step-ahead, six-step-ahead, and eight-step-ahead forecast horizons when compared with AO. Table 9 selects the cases from Table 5 in which both the constant and slopes are statistically significant and calculates the squared error difference conditional on the recession dummy being zero or one. The implication is that in these cases the reference model is preferred when the dummy is turned off and the alternative model is preferred when the dummy is turned on. Thus, with the exception of two-step-ahead horizon forecasts, the AO model is preferred during expansions, while during recessions one is better off using the Phillips curve models for forecasting inflation. However, it is again worth pointing out that these results occur only for the full sample and that there is little persuasive evidence that this has remained the case over the more recent sample period − the lone exception being the four-step-ahead horizon.

With regard to the SPF downturn probability, we find that the slope coefficient is significant at the four-, six-, and eight-step-ahead horizon when comparing the forecasts from both the AO and IMA(1,1) models. Further, when a continuous conditioning variable is used in the regression, we can calculate the cutoff value for each conditioning variable that turns the squared error difference from negative to positive. Table 10 presents the cutoff values for the SPF downturn probabilities above which the Phillips curve models are producing lower forecast errors relative to both the AO and IMA(1,1) models. With respect to the threshold model, the calculation is only relevant for the two-step-ahead horizon and indicates that recession probabilities greater than 24.4% improve the forecast of the Phillips curve threshold model relative to the AO model. The analogous probability for the IMA(1,1) model is 38.4%. Regarding the basic Phillips curve model, downturn probabilities that exceed 29.1% imply that Phillips curve predictions of inflation should be carefully considered. Again, these numbers are relevant for the full sample results, and there is no indication that Phillips curve predictions are useful when basing the analysis on the most recent sample period.

Results from conditioning on the change in the real-time gap (Table 11) lend support to using Phillips curve forecasts as opposed to forecasts from the AO model in both the full and recent sample periods. Concentrating on the situation that is indicated by the darkest shading (i.e., the cases in which both the GW test statistic is significant and the slope coefficients are also significantly positive), we see that for the change in the real-time gap it pays to look at the Phillips curve forecasts over the full and later sample period at four-, six-, and eight-quarter-ahead horizons when compared with the AO model, even when the change in the unemployment gap is only slightly negative over the full sample. Thus, although Phillips curve forecasts do not generally outperform the benchmark forecasts, there are a few situations when they prove useful. Unfortunately for the advocates of the Phillips

(a) AO Minus Phillips Curve

(b) IMA Minus Phillips Curve

(c) AO Minus Threshold Phillips Curve

(d) IMA Minus Threshold Phillips Curve
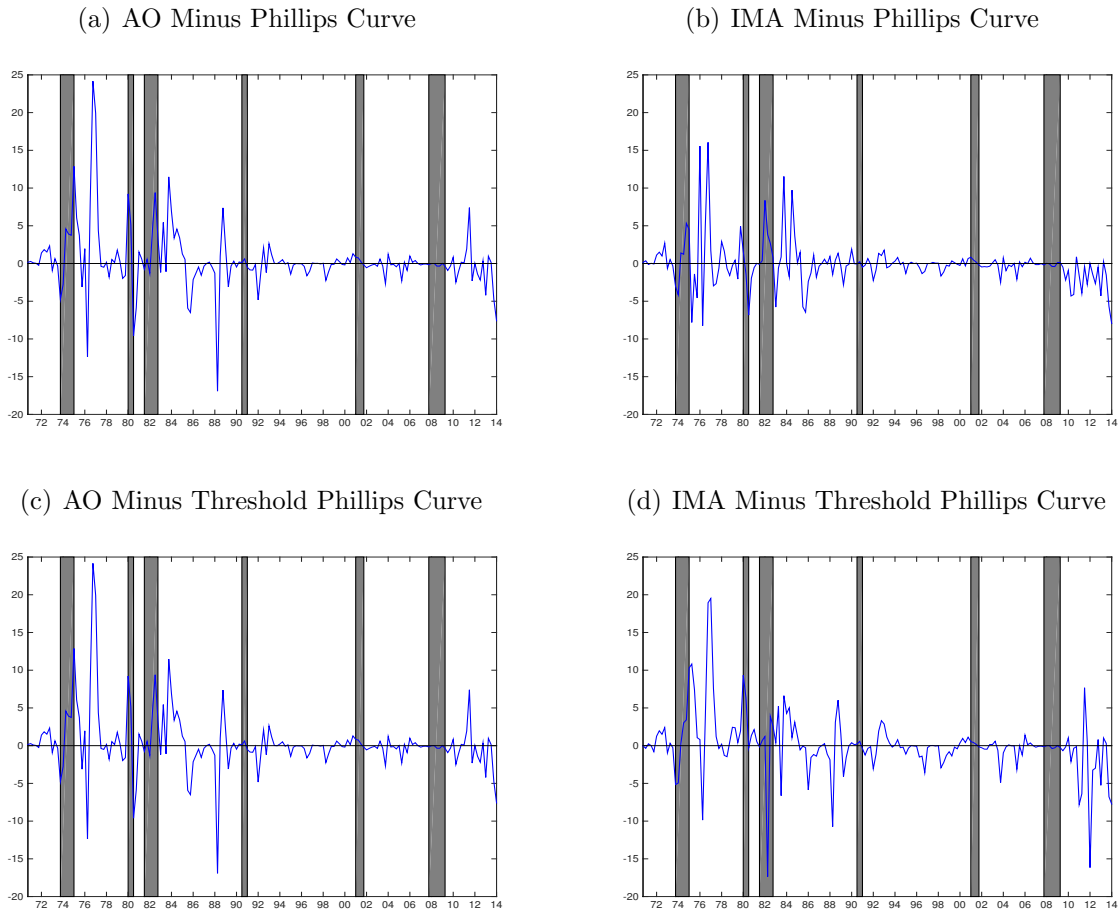
Figure 8: Squared Error Difference: Eight-quarter-ahead forecasts. Shading shows periods of NBER-dated recessions. The squared error difference is aligned at the date forecasted.

curve, these situations are much less prevalent over the most recent sample period.

6.2.3 Inspecting the Mechanism

We inspect the mechanism for these results in Figure 8, where we graph the squared forecast error differences for (i) the AO and the two Phillips curve models (the two left panels) and for (ii) the IMA(1,1) model and the two Phillips curve models (the two right panels) at the eight-quarter-ahead forecast horizon. Thus, positive values indicate that the Phillips curve forecast was more accurate. The squared error differences are associated with the date forecasted. Thus, the large positive spike that appears in 1977 in all the figures reflects that the Phillips curve forecasts made in 1975 were more accurate than the forecasts made with either of the pure time series models. The greater forecasting accuracy of the Phillips curves

24

models is also associated with the end of the 1982 recession. It also appears to help with the Great Recession, but recall that the sign on the unemployment gap is perverse during this episode. There is no evidence that using a Phillips curve model helped in prediction for most of the post-1984 sample period. Further, as seen in Figure 6, since the Phillips curve term generally makes a significant contribution to the inflation forecasts around recessions in the early part of the sample period, it is not surprising that both SPF downturn probabilities and a large increase in the output gap conditionally improve forecast accuracy in the early part of the sample, but that this is not the case in the more recent sample period when the unemployment gap generally contributes little to the forecast of inflation.

6.3   The Phillips Curve With Short-Term Unemployment

When estimating Phillips curve models, we use the real-time unemployment gap measure that is based on the overall unemployment rate. The past literature, however, has raised a theoretical possibility that short-term and long-term unemployment exert different inflationary pressures. The idea is that those who are jobless for a long period might have lost their skills or desire to work, thus contributing less to wage and price pressures than the short-term unemployed do.[13] In light of this theoretical consideration, several papers examine the hypothesis by using short-term unemployment as a measure of the inflationary pressure in estimating Phillips curve models (e.g., Llaudes (2005) and Kiley (2015)). We check the robustness of our results with respect to this variation of the Phillips curve specification, by replacing the gap measure in our baseline Phillips curve model with the gap measure constructed from the "short-term unemployment rate." We define the short-term unemployment rate as the number of unemployed less than or equal to 26 weeks normalized by the total labor force. We construct the real-time gap measure, following exactly the same procedure as in our baseline model.[14]

We provide a brief summary of the results here (all results are available upon request). Overall, the results from this exercise are very similar to our main results with some minor

---

[13]See, for example, Blanchard and Summers (1986) and Blanchard and Diamond (1994) for underlying theories.

[14]More precisely speaking, there is one minor difference in the procedure. Remember that the official unemployment rate is revised retrospectively once a year due to the revision of the seasonal factors. In our baseline model, we do take into account this revision so that our gap measure is truly a real-time measure. For the current exercise with short-term unemployment, however, the real-time vintages are not readily available and thus we use the most recent vintage as of January 2017. Importantly, however, the revision of the data due to updated seasonal factors is typically very small, and the "real-time" nature of the unemployment gap largely comes from the uncertainty about its trend. We do incorporate this uncertainty in the current exercise as well.

differences. Specifically, when the short-term unemployment gap is used in the Phillips curve model, the overall forecasting performance measured by MAEs and RMSEs (as in Table 1) indeed improves over the baseline Phillips curve. However, the improvement is at most less than 5 basis points, and thus ranking relationships among the forecasting models remain exactly the same. Regarding the GW tests, there are now several more cases (than in our main results) that are favorable to this modified Phillips curve model. Moreover, for the cases in which the original Phillips curve model is preferred over the other models, the same was true for this modified Phillips curve model. These improvements are quantitatively very small, and thus none of these results change our overall conclusion regarding the usefulness of the Phillips curve model.[15]

7   Results Using Latest Vintage Data

In this section, we look at whether and to what extent the use of the latest vintage data for the unemployment gap influences our conclusions. To do this, we reestimate the Phillips curve using final estimates of the unemployment gap, compute new forecasts, and rerun our forecast evaluation tests using the revised unemployment gaps to construct our conditioning variables. We first characterize the relative unconditional forecasting ability of the two benchmark specifications and the Phillips curve model. As shown in Table 12, using final revised data does not change our perception regarding the accuracy of Phillips curve inflation forecasts, and the changes are not large enough to overturn the relative ranking of the forecasting models that were examined earlier in Table 1 using the real-time data. Note, however, that the forecast accuracy of the Phillips curve model does improve when using final estimates of the unemployment gap.

We now examine GW tests comparing the forecasting performance of the AO, IMA(1,1), and Phillips curve models using the latest vintage of unemployment gaps. The overall message is the same as in the real-time results, but there are a few notable differences. The results of the GW tests are given in Table 13. With regard to the unconditional forecast evaluation presented in the first two columns of that table, there is only one qualitative change in results: Namely, the Phillips curve specification is preferred at the eight-step-ahead forecast horizon for the entire sample but not significantly so. The IMA(1,1) specification is still preferred over the later sample period at all forecast horizons.

When we examine the conditional forecast results with respect to the recession dummy,

---

[15]Our results are consistent with Kiley's (2015) conclusion that long-term unemployment and short-term unemployment have exerted similar inflationary pressure.

there is now evidence that this variable conditionally improves Phillips curve forecasts with respect to the IMA(1,1) model as well as continues to improve on the AO model as it did when using real-time data. This change in results should imply that results based on the latest vintage of unemployment gaps do not always reflect what is implied by real-time analysis. On the other hand, there is qualitatively little change in forecast evaluation when we condition on the SPF downturn probability. Lastly, with respect to the four-quarter change in the gap, that variable no longer helps improve the conditional forecast of the Phillips curve at the four-step-ahead horizon but continues to be of use at the six-step-ahead and eight-step-ahead horizons. Thus, with the exception of the recession dummy, replacing real-time data with the latest vintage data does not substantially alter any of the conclusions drawn from our earlier analysis.

8   Summary and Conclusion

In this paper, we explored in a formal statistical way the inflation forecasting properties of Phillips curve models relative to the naive model of Atkeson and Ohanian (2001) and the IMA(1,1) model. Our results comparing the forecasts support the preponderance of evidence indicating that, if anything, Phillips curve models are not relatively good at forecasting inflation on average. For the 1969Q1−2014Q2 sample, we find, as did Stock and Watson (2007, 2008), that over the entire sample an IMA(1,1) model outperforms Phillips curve models but not in a statistically significant way. For the 1984−2014 sample, the IMA(1,1) model remains the preferred forecast model and significantly so at all forecast horizons with respect to the basic Phillips curve specification. Using the latest revised output gaps as opposed to final time output gaps does not appreciably change the general thrust of our results.

Of note, however, is that, conditional on variables that capture the state of the economy, the Phillips curve model can prove useful for forecasting, but that conclusion is tied to analysis regarding the entire sample. It is only with regard to the SPF downturn probabilities that we find any conditional improvement in Phillips curve forecasts. Importantly, we find that its usefulness is asymmetric in the sense that it helps improve the accuracy of inflation forecasts when the economy is weak while it hurts the accuracy during expansionary periods. The statistically significant improvement tends to be concentrated over the entire sample period, which agrees with the general perception one obtains from the existing literature. Thus our evidence may indicate that using the Phillips curve may add value to the monetary policy process during downturns, but the evidence is far from conclusive given

that conditional forecasting improvement is not evident in the more recent sample period. We find no evidence for relying on the Phillips curve during normal times, such as those currently facing the U.S. economy.

Finally, we focused our analysis strictly on headline PCE inflation because it has a longer history than the core and allows us to look at real-time measures. We also confined our Phillips curve analysis to unemployment gaps, and it would be interesting to see if our results carry over to other gap measures. Our reading of the literature, in which many inflation and gap measures have been explored, leads us to believe our results will turn out to be general but that conjecture awaits confirmation.

References

Andrew Ang, Geert Bekaert, and Min Wei. Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, 54(4):1163–1212, 2007.

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

Pierpaolo Benigno and Luca Antonio Ricci. The inflation-output trade-off with downward wage rigidities. *American Economic Review*, 101(4):1436–1466, 2011.

Olivier Blanchard and Peter Diamond. Ranking, unemployment duration, and wages. *Review of Economic Studies*, 61(3):417–34, 1994.

Olivier Blanchard and Lawrence Summers. Hysteresis and the European unemployment problem. In Stan Fischer, editor, *NBER Macroeconomics Annual*. MIT Press, 1986.

Flint Brayton, John Roberts, and John Williams. What's happened to the Phillips curve? FEDS Working Paper No. 99-49, 1999.

Todd Clark and Michael McCracken. The predictive content of the output gap for inflation: Resolving in-sample and out-of-sample evidence. *Journal of Money, Credit and Banking*, 38(5):1127–1148, 2006.

Todd Clark and Michael McCracken. Advances in forecast evaluation. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, chapter 20, pages 1107–1201. Elsevier, 2013.

Francis Diebold and Roberto Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–63, 1995.

Michael Dotsey and Tom Stark. The relationship between capacity utilization and inflation. *Business Review*, Q2:8–17, Federal Reserve Bank of Philadelphia, 2005.

Jonas Fisher, Chin Te Liu, and Ruilin Zhou. When can we forecast inflation? *Economic Perspectives*, 1Q:30–42, Federal Reserve Bank of Chicago, 2002.

Jeffrey Fuhrer and Giovanni Olivei. The role of expectations and output in the inflation process: An empirical assessment. *Public Policy Briefs*, 10(12), Federal Reserve Bank of Boston, 2010.

Raffaella Giacomini and Barbara Rossi. Detecting and predicting forecast breakdowns. *Review of Economic Studies*, 76(2):669–705, 2009.

Raffaella Giacomini and Halbert White. Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578, 2006.

Bruce Hansen. Inference in TAR models. *Studies in Nonlinear Dynamics & Econometrics*, 2(1):1–14, 1997.

Michael Kiley. An evaluation of the inflationary pressure associated with short- and long-term unemployment. *Economics Letters*, 137:5–9, 2015.

Rachidi Kotchoni and Dalibor Stevanovic. Forecasting U.S. recessions and economic activity. Unpublished Manuscript, 2016.

Ricardo Llaudes. The Phillips curve and long-term unemployment. European Central Bank Working Paper Series No. 44, 2005.

Athanasios Orphanides and Simon van Norden. The reliability of inflation forecasts based on output gap estimates in real time. *Journal of Money, Credit and Banking*, 37(3):583–600, 2005.

Athanasios Orphanides and John Williams. The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations. *Journal of Economic Dynamics and Control*, 29:1927–1950, 2005.

Robert Shimer. The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, 95(1):25–49, 2005.

James Stock and Mark Watson. Forecasting inflation. *Journal of Monetary Economics*, 44 (2):293–335, 1999.

James Stock and Mark Watson. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(1):3–34, 2007.

James Stock and Mark Watson. Phillips curve inflation forecast. In Jeff Fuhrer, Yolanda Kodrzycki, Jane Sneddon Little, and Giovanni Olivei, editors, *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, pages 101–187. MIT Press, 2008.

Table 1: Real-Time Forecast Error Comparisons for the Inflation Rate

Forecast	1969Q1−2014Q2				1984Q1−2014Q2			
horizon	AO	IMA	PC	PC-TAR	AO	IMA	PC	PC-TAR
(a) Mean Absolute Errors (MAEs)								
2	1.193	1.136*	1.260	1.264	0.966	0.941*	1.111	1.152
4	1.212	1.089*	1.182	1.284	0.844	0.819*	0.972	1.123
6	1.286	1.150*	1.271	1.318	0.842	0.797*	0.982	1.086
8	1.323	1.183*	1.295	1.341	0.837	0.744*	0.972	1.048
(b) Root-Mean-Square Errors (RMSEs)								
2	1.690	1.592*	1.748	1.757	1.378	1.342*	1.617	1.687
4	1.733	1.554*	1.648	1.795	1.158	1.119*	1.285	1.587
6	1.841	1.689*	1.771	1.844	1.099	1.073*	1.268	1.441
8	1.915	1.782*	1.811	1.883	1.141	1.066*	1.270	1.380

Notes: MAEs and RMSEs are calculated by estimating each model with a fixed window size of 60 quarters. The model that gives the smallest MAE or RMSE is indicated by the asterisk.

Table 2: GW Unconditional Test

		1969Q1−2014Q2			1984Q1−2014Q2		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
		(a) 2-Step-Ahead Forecast					
AO	P-Value	0.245	0.607	0.633	0.630	0.021**	0.053*
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.322	−0.197	−0.232	0.099	−0.714**	−0.948*
IMA	P-Value		0.091*	0.239		0.031**	0.097*
	R^2		0.000	0.000		0.000	0.000
	Const.		−0.519*	−0.554		−0.813**	−1.047*
PC	P-Value			0.907			0.547
	R^2			0.000			0.000
	Const.			−0.034			−0.234
		(b) 4-Step-Ahead Forecast					
AO	P-Value	0.075*	0.389	0.726	0.614	0.022**	0.081*
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.587*	0.288	−0.220	0.088	−0.310**	−1.178*
IMA	P-Value		0.266	0.172		0.001**	0.109
	R^2		0.000	0.000		0.000	0.000
	Const.		−0.299	−0.807		−0.399**	−1.266
PC	P-Value			0.315			0.218
	R^2			0.000			0.000
	Const.			−0.508			−0.867
		(c) 6-Step-Ahead Forecast					
AO	P-Value	0.038**	0.395	0.984	0.756	0.013**	0.064*
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.537**	0.255	−0.011	0.056	−0.401**	−0.870*
IMA	P-Value		0.178	0.257		0.021**	0.135
	R^2		0.000	0.000		0.000	0.000
	Const.		−0.282	−0.547		−0.457**	−0.926
PC	P-Value			0.532			0.358
	R^2			0.000			0.000
	Const.			−0.265			−0.469
		(d) 8-Step-Ahead Forecast					
AO	P-Value	0.075*	0.357	0.799	0.490	0.105	0.081*
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.491*	0.387	0.120	0.165	−0.313	−0.604*
IMA	P-Value		0.681	0.370		0.045**	0.116
	R^2		0.000	0.000		0.000	0.000
	Const.		−0.103	−0.370		−0.478**	−0.769
PC	P-Value			0.333			0.275
	R^2			0.000			0.000
	Const.			−0.267			−0.291

Notes: Entries in each block present the p-value for the GW χ^2 test statistic and, for the GW regressions, the adjusted R^2 and the coefficient estimate from the regression specified in (7). The dependent variable is the time-t squared forecast error differential between the model listed in the row and the model listed in the column. * (**) indicate statistical significance at the 10% (5%) level. P-values and test statistics use HAC standard errors.

Table 3: GW Conditional Test: Absolute Value of Initial Unemployment Gap

		1969Q1−2014Q2			1984Q1−2014Q2		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
(a) 2-Step-Ahead Forecast							
AO	P-Value	0.421	0.459	0.312	0.539	0.055*	0.154
	R^2	0.007	−0.001	0.016	0.026	0.002	0.044
	Const.	−0.049	0.185	0.764	−0.311	−0.306	0.497
	Slope	0.397	−0.409	−1.065	0.465	−0.463	−1.638**
IMA	P-Value		0.182	0.235		0.094*	0.252
	R^2		0.024	0.038		0.036	0.063
	Const.		0.234	0.813**		0.006	0.809
	Slope		−0.806**	−1.462*		−0.928**	−2.104*
PC	P-Value			0.203			0.285
	R^2			0.006			0.021
	Const.			0.579			0.803
	Slope			−0.656			−1.175
(b) 4-Step-Ahead Forecast							
AO	P-Value	0.197	0.690	0.631	0.510	0.027**	0.175
	R^2	0.009	0.001	0.004	0.003	−0.006	0.038
	Const.	0.152	−0.031	0.493	−0.125	−0.193	0.425
	Slope	0.464	0.340	−0.759	0.234	−0.129	−1.759**
IMA	P-Value		0.153	0.390		0.003**	0.270
	R^2		−0.005	0.019		0.046	0.045
	Const.		−0.183	0.342		−0.068	0.550*
	Slope		−0.124	−1.222*		−0.363**	−1.992**
PC	P-Value			0.406			0.391
	R^2			0.015			0.027
	Const.			0.525*			0.618**
	Slope			−1.098*			−1.629**
(c) 6-Step-Ahead Forecast							
AO	P-Value	0.111	0.610	0.778	0.695	0.016**	0.000**
	R^2	0.021	0.000	0.002	0.005	−0.006	0.066
	Const.	−0.079	−0.085	0.516	−0.179	−0.275	0.410
	Slope	0.654	0.361	−0.559	0.253	−0.136	−1.378**
IMA	P-Value		0.338	0.523		0.039**	0.157
	R^2		−0.001	0.041		0.020	0.088
	Const.		−0.006	0.595		−0.096	0.589
	Slope		−0.293	−1.213		−0.388*	−1.631**
PC	P-Value			0.651			0.655
	R^2			0.021			0.046
	Const.			0.601			0.685**
	Slope			−0.920			−1.242
(d) 8-Step-Ahead Forecast							
AO	P-Value	0.178	0.563	0.844	0.319	0.114	0.080*
	R^2	0.030	0.004	−0.004	0.024	−0.005	0.023
	Const.	−0.232	−0.084	0.295	−0.270	−0.469**	−0.038
	Slope	0.768**	0.501	−0.185	0.463**	0.166	−0.602
IMA	P-Value		0.724	0.534		0.037**	0.102
	R^2		−0.001	0.044		0.009	0.094
	Const.		0.148	0.527		−0.199	0.233
	Slope		−0.267	−0.952*		−0.297	−1.065**
PC	P-Value			0.453			0.548
	R^2			0.030			0.071
	Const.			0.379			0.432*
	Slope			−0.686*			−0.768**

Notes: Entries in each block present the p-value for the GW χ^2 test statistic and, for the GW regressions, the adjusted R^2 and the coefficient estimates from the regression specified in (8). The dependent variable is the time-t squared forecast error differential between the model listed in the row and the model listed in the column. *(**) indicate statistical significance at the 10% (5%) level. P-values and test statistics use HAC standard errors. See Subsection 6.2 for explanations of the shading.

Table 4: GW Conditional Test: Threshold Dummy

		1969Q1−2014Q2			1984Q1−2014Q2		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
		(a) 2-Step-Ahead Forecast					
AO	P-Value	0.473	0.683	0.599	0.634	0.063*	0.124
	R^2	0.004	0.005	0.008	0.013	0.061	0.060
	Const.	0.167	0.081	0.156	−0.050	−0.239	−0.196
	Slope	0.620	−1.115	−1.551	0.719	−2.303*	−3.638*
IMA	P-Value		0.231	0.477		0.044	0.184
	R^2		0.034	0.023		0.114	0.072
	Const.		−0.086	−0.011		−0.189	−0.147
	Slope		−1.735	−2.171		−3.022*	−4.357*
PC	P-Value			0.803			0.742
	R^2			−0.004			0.002
	Const.			0.075			0.042
	Slope			−0.436			−1.335
		(b) 4-Step-Ahead Forecast					
AO	P-Value	0.201	0.596	0.611	0.664	0.069*	0.000**
	R^2	0.006	−0.001	0.004	0.001	0.003	0.055
	Const.	0.397	0.163	0.139	−0.001	−0.189*	−0.288**
	Slope	0.752	0.495	−1.421	0.399	−0.545	−3.985
IMA	P-Value		0.358	0.295		0.003	0.034
	R^2		−0.004	0.017		0.089	0.061
	Const.		−0.235	−0.258		−0.188**	−0.288**
	Slope		−0.257	−2.173		−0.944**	−4.384
PC	P-Value			0.599			0.218
	R^2			0.013			0.034
	Const.			−0.023			−0.100
	Slope			−1.916			−3.440
		(c) 6-Step-Ahead Forecast					
AO	P-Value	0.115	0.513	0.922	0.936	0.018**	0.001**
	R^2	0.009	−0.004	−0.004	−0.006	0.008	0.042
	Const.	0.303	0.158	0.102	0.010	−0.228**	−0.348**
	Slope	0.915	0.378	−0.440	0.190	−0.724	−2.178
IMA	P-Value		0.361	0.479		0.045**	0.018**
	R^2		−0.001	0.011		0.034	0.047
	Const.		−0.144	−0.201		−0.238*	−0.359**
	Slope		−0.538	−1.356		−0.913*	−2.368
PC	P-Value			0.813			0.104
	R^2			0.000			0.012
	Const.			−0.056			−0.121*
	Slope			−0.818			−1.454
		(d) 8-Step-Ahead Forecast					
AO	P-Value	0.201	0.650	0.948	0.292	0.191	0.106
	R^2	0.015	−0.002	−0.006	0.020	−0.008	0.008
	Const.	0.210	0.236	0.095	−0.041	−0.296*	−0.398*
	Slope	1.087*	0.583	0.098	0.806*	−0.065	−0.804
IMA	P-Value		0.772	0.666		0.076*	0.155
	R^2		0.000	0.010		0.033	0.058
	Const.		0.027	−0.114		−0.255*	−0.357
	Slope		−0.503	−0.989		−0.871	−1.610
PC	P-Value			0.605			0.439
	R^2			−0.001			0.012
	Const.			−0.141			−0.102
	Slope			−0.486			−0.739

Notes: See notes to Table 3. The threshold dummy takes 1 when the absolute value of the real-time gap is larger than 1.2; otherwise it takes 0.

Table 5: GW Conditional Test: Recession Dummy

		1969Q1−2014Q2			1984Q1−2014Q2		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
		(a) 2-Step-Ahead Forecast					
AO	P-Value	0.407	0.707	0.685	0.468	0.034**	0.073*
	R^2	0.021	−0.006	−0.005	0.017	0.083	0.047
	Const.	0.129	−0.202	−0.268	−0.002	−0.376**	−0.529*
	Slope	1.286	0.031	0.240	1.107	−3.725	−4.604
IMA	P-Value		0.067*	0.299		0.015**	0.178
	R^2		0.009	−0.001		0.150	0.061
	Const.		−0.331	−0.397		−0.374**	−0.528
	Slope		−1.254	−1.045		−4.832	−5.711
PC	P-Value			0.954			0.813
	R^2			−0.005			−0.006
	Const.			−0.066			−0.154
	Slope			0.209			−0.879
		(b) 4-Step-Ahead Forecast					
AO	P-Value	0.169	0.075*	0.925	0.542	0.015**	0.068*
	R^2	0.016	0.085	−0.003	0.033	0.006	0.086
	Const.	0.397	−0.133	−0.090	−0.023	−0.389**	−0.538**
	Slope	1.251	2.775*	−0.857	1.225	0.862*	−7.034
IMA	P-Value		0.138	0.202		0.004**	0.162
	R^2		0.029	0.009		−0.002	0.108
	Const.		−0.531	−0.487		−0.366**	−0.515
	Slope		1.524	−2.108		−0.363	−8.259
PC	P-Value			0.526			0.455
	R^2			0.039			0.097
	Const.			0.043			−0.150
	Slope			−3.632			−7.896
		(c) 6-Step-Ahead Forecast					
AO	P-Value	0.057*	0.187	0.982	0.952	0.045**	0.000**
	R^2	0.005	0.071	−0.005	−0.007	0.010	0.020
	Const.	0.391	−0.200	−0.046	0.031	−0.512**	−0.627**
	Slope	0.947	2.960**	0.227	0.236	1.037	−2.261
IMA	P-Value		0.115	0.269		0.048**	0.045**
	R^2		0.042	−0.002		0.009	0.024
	Const.		−0.591	−0.437		−0.543**	−0.658**
	Slope		2.014	−0.720		0.800	−2.498
PC	P-Value			0.071*			0.253
	R^2			0.042			0.047
	Const.			0.154			−0.115
	Slope			−2.734**			−3.298
		(d) 8-Step-Ahead Forecast					
AO	P-Value	0.141	0.112	0.602	0.331	0.075*	0.050**
	R^2	0.009	0.094	0.007	0.028	0.037	−0.003
	Const.	0.322	−0.160	−0.078	0.015	−0.514**	−0.680**
	Slope	1.087*	3.525**	1.278	1.212**	1.627	0.614
IMA	P-Value		0.011**	0.506		0.071*	0.162
	R^2		0.080	−0.005		−0.003	−0.003
	Const.		−0.482	−0.400		−0.530**	−0.695*
	Slope		2.439	0.192		0.415	−0.598
PC	P-Value			0.258			0.514
	R^2			0.071			0.014
	Const.			0.082			−0.165
	Slope			−2.247*			−1.013*

Notes: See notes to Table 3.

Table 6: GW Conditional Test: SPF Downturn Probability

		1969Q1−2014Q2			1984Q1−2014Q2		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
		(a) 2-Step-Ahead Forecast					
AO	P-Value	0.440	0.069*	0.064*	0.850	0.010**	0.049**
	R^2	0.018	0.011	0.034	−0.007	0.029	−0.008
	Const.	−0.043	−0.700*	−1.194**	0.033	−0.139	−1.050**
	Slope	0.019	0.026	0.049*	0.004	−0.036	0.006
IMA	P-Value		0.002**	0.045**		0.000**	0.078*
	R^2		−0.004	0.010		0.039	−0.008
	Const.		−0.657**	−1.151**		−0.172	−1.083**
	Slope		0.007	0.030*		−0.040	0.002
PC	P-Value			0.319			0.340
	R^2			0.007			0.014
	Const.			−0.494			−0.911
	Slope			0.023*			0.043
		(b) 4-Step-Ahead Forecast					
AO	P-Value	0.203	0.106	0.217	0.800	0.024**	0.105
	R^2	0.014	0.114	0.007	0.001	−0.006	−0.002
	Const.	0.222	−0.679**	−0.822*	−0.050	−0.395**	−0.748
	Slope	0.019	0.049*	0.031	0.009	0.005	−0.027
IMA	P-Value		0.074*	0.072*		0.004**	0.161
	R^2		0.054	−0.004		−0.006	0.001
	Const.		−0.901**	−1.044**		−0.345**	−0.698
	Slope		0.031*	0.012		−0.003	−0.036
PC	P-Value			0.480			0.448
	R^2			−0.001			−0.001
	Const.			−0.143			−0.353
	Slope			−0.019			−0.033
		(c) 6-Step-Ahead Forecast					
AO	P-Value	0.112	0.335	0.770	0.952	0.045**	0.071*
	R^2	−0.001	0.060	0.001	−0.008	−0.006	−0.008
	Const.	0.349	−0.575	−0.360	0.016	−0.502**	−0.815**
	Slope	0.010	0.042*	0.018	0.003	0.006	−0.003
IMA	P-Value		0.033**	0.252		0.055*	0.140
	R^2		0.046	−0.004		−0.007	−0.008
	Const.		−0.923**	−0.709		−0.518**	−0.831*
	Slope		0.032**	0.008		0.004	−0.006
PC	P-Value			0.324			0.655
	R^2			0.010			−0.007
	Const.			0.215			−0.313
	Slope			−0.024**			−0.010
		(d) 8-Step-Ahead Forecast					
AO	P-Value	0.184	0.166	0.637	0.599	0.071*	0.054*
	R^2	−0.003	0.060	0.005	0.001	0.005	−0.003
	Const.	0.354	−0.486	−0.236	−0.011	−0.556**	−0.780**
	Slope	0.007	0.044**	0.018	0.011	0.015	0.011
IMA	P-Value		0.010**	0.314		0.079*	0.150
	R^2		0.078	0.000		−0.007	−0.008
	Const.		−0.840**	−0.590		−0.545**	−0.769*
	Slope		0.037**	0.011		0.004	0.000
PC	P-Value			0.406			0.506
	R^2			0.038			−0.007
	Const.			0.250			−0.223
	Slope			−0.026*			−0.004

Notes: See notes to Table 3.

Table 7: GW Conditional Test: Real-Time Gap

		1969Q1−2014Q2			1984Q1−2014Q2		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
		(a) 2-Step-Ahead Forecast					
AO	P-Value	0.294	0.871	0.521	0.275	0.066*	0.143
	R^2	0.025	−0.005	0.006	0.053	0.008	0.051
	Const.	0.215	−0.164	−0.098	0.098	−0.713**	−0.945**
	Slope	0.406	−0.126	−0.505	0.417*	−0.387*	−1.160**
IMA	P-Value		0.229	0.260		0.067*	0.252
	R^2		0.023	0.033		0.066	0.082
	Const.		−0.379	−0.313		−0.811**	−1.043*
	Slope		−0.532*	−0.911*		−0.804**	−1.577*
PC	P-Value			0.171			0.245
	R^2			0.003			0.020
	Const.			0.066			−0.232
	Slope			−0.379			−0.773
		(b) 4-Step-Ahead Forecast					
AO	P-Value	0.181	0.415	0.855	0.228	0.072*	0.070*
	R^2	0.034	0.019	−0.001	0.018	−0.008	0.050
	Const.	0.455	0.176	−0.128	0.081	−0.310**	−1.137**
	Slope	0.505	0.426	−0.351	0.233**	0.004	−1.301**
IMA	P-Value		0.436	0.392		0.001**	0.229
	R^2		−0.005	0.021		0.041	0.064
	Const.		−0.279	−0.583		−0.391**	−1.217**
	Slope		−0.079	−0.856*		−0.229**	−1.534**
PC	P-Value			0.318			0.305
	R^2			0.018			0.043
	Const.			−0.304			−0.826
	Slope			−0.777			−1.305**
		(c) 6-Step-Ahead Forecast					
AO	P-Value	0.097*	0.611	0.976	0.422	0.019**	0.000**
	R^2	0.042	0.010	−0.005	0.014	−0.005	0.058
	Const.	0.388**	0.152	0.013	0.045	−0.395**	−0.828**
	Slope	0.580**	0.399	−0.094	0.219	−0.119	−0.862*
IMA	P-Value		0.394	0.504		0.024**	0.082**
	R^2		−0.001	0.027		0.041	0.089
	Const.		−0.236	−0.375		−0.440**	−0.873*
	Slope		−0.181	−0.674		−0.338**	−1.081*
PC	P-Value			0.711			0.611
	R^2			0.012			0.036
	Const.			−0.139			−0.433
	Slope			−0.493			−0.743
		(d) 8-Step-Ahead Forecast					
AO	P-Value	0.121	0.330	0.966	0.144	0.061*	0.057*
	R^2	0.048	0.019	−0.005	0.041	−0.008	0.028
	Const.	0.336*	0.259	0.098	0.142	−0.316*	−0.578**
	Slope	0.623**	0.516	0.091	0.375**	0.046	−0.424
IMA	P-Value		0.902	0.620		0.025**	0.141
	R^2		−0.004	0.029		0.040	0.125
	Const.		−0.077	−0.238		−0.458**	−0.720**
	Slope		−0.107	−0.532		−0.329**	−0.799**
PC	P-Value			0.470			0.550
	R^2			0.025			0.060
	Const.			−0.161			−0.262
	Slope			−0.425*			−0.47*

Notes: See notes to Table 3.

Table 8: GW Conditional Test: Four-Quarter Change in Real-Time Gap

		1969Q1−2014Q2			1984Q1−2014Q2		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
		(a) 2-Step-Ahead Forecast					
AO	P-Value	0.352	0.871	0.891	0.459	0.041**	0.077*
	R^2	0.043	−0.001	−0.002	0.080	0.075	0.052
	Const.	0.275	−0.175	−0.207	0.148	−0.80**	−1.063**
	Slope	2.316	−1.132	−1.235	2.469*	−4.273*	−5.773
IMA	P-Value		0.187	0.466		0.015**	0.162
	R^2		0.054	0.023		0.206	0.092
	Const.		−0.450*	−0.482		−0.948**	−1.211*
	Slope		−3.448	−3.551		−6.742**	−8.242
PC	P-Value			0.993			0.831
	R^2			−0.006			−0.004
	Const.			−0.032			−0.263
	Slope			−0.102			−1.500
		(b) 4-Step-Ahead Forecast					
AO	P-Value	0.189	0.143	0.940	0.381	0.015**	0.052*
	R^2	0.047	0.082	0.001	0.044	0.012	0.106
	Const.	0.529*	0.207	−0.177	0.110	−0.294**	−1.298**
	Slope	2.614	3.649**	−1.926	1.638**	1.238*	−9.189*
IMA	P-Value		0.447	0.285		0.004*	0.142
	R^2		0.003	0.032		−0.003	0.133
	Const.		−0.322	−0.706		−0.404**	−1.408**
	Slope		1.035	−4.540		−0.400	−10.827*
PC	P-Value			0.578			0.459
	R^2			0.054			0.121
	Const.			−0.384			−1.004**
	Slope			−5.575			−10.427*
		(c) 6-Step-Ahead Forecast					
AO	P-Value	0.111	0.206	0.985	0.430	0.030**	0.000**
	R^2	0.049	0.090	−0.005	0.021	0.058	0.058
	Const.	0.467**	0.146	0.001	0.057	−0.398**	−0.875**
	Slope	2.822	4.417**	−0.497	1.234	2.468**	−4.259
IMA	P-Value		0.276	0.425		0.045**	0.003**
	R^2		0.011	0.033		0.018	0.095
	Const.		−0.321	−0.466		−0.456**	−0.932*
	Slope		1.595	−3.318		1.234	−5.493
PC	P-Value			0.226			0.379
	R^2			0.079			0.141
	Const.			−0.144			−0.477
	Slope			−4.913*			−6.727**
		(d) 8-Step-Ahead Forecast					
AO	P-Value	0.183	0.113	0.708	0.343	0.041**	0.065*
	R^2	0.062	0.101	0.012	0.066	0.114	0.000
	Const.	0.401*	0.249	0.064	0.143	−0.347*	−0.614*
	Slope	3.181**	4.920**	2.033	2.265**	3.498**	1.022
IMA	P-Value		0.292	0.669		0.069*	0.140
	R^2		0.018	0.002		0.020	0.005
	Const.		−0.152	−0.338		−0.490*	−0.757*
	Slope		1.739	−1.148		1.233	−1.242
PC	P-Value			0.207			0.485
	R^2			0.064			0.070
	Const.			−0.186			−0.267
	Slope			−2.886**			−2.475**

Notes: See notes to Table 3.

38

Table 9: Squared Forecast Error Differences Conditional on the Recession Dummy

Reference Model	Alternative Model	Forecast Horizon	Sample	Squared Error Diff. $D = 0$	$D = 1$
AO	PC	4	Full	-0.133	2.642
AO	PC	4	Post-84	-0.389	0.473
AO	PC	6	Full	-0.200	2.760
AO	PC	8	Full	-0.160	3.365

Notes: This table considers the cases with dark and middle shadings in Table 5 within the comparisons between each of the two Phillips curve models and either the AO or IMA models. The last two columns calculate the squared error difference when the recession dummy is zero and one, respectively.

Table 10: Cutoff Value of the SPF Downturn Probability

Reference Model	Alternative Model	Forecast Horizon	Sample	Cutoff Values (%)
AO	PC-TAR	2	Full	24.4
IMA	PC-TAR	2	Full	38.4
AO	PC	4	Full	13.9
IMA	PC	4	Full	29.1
AO	PC	6	Full	13.7
IMA	PC	6	Full	28.8
AO	PC	8	Full	11.0
IMA	PC	8	Full	22.7

Notes: The last column reports the cutoff value of the SPF recession probability above (below) which the alternative (reference) model gives the smaller forecast error. This table includes only the cases with the dark and middle shadings in Table 6 within the comparisons between each of the two Phillips curve models and either the AO or IMA models.

Table 11: Cutoff Values of the Four-Quarter Change in Real-Time Unemployment Gap

Reference Model	Alternative Model	Forecast Horizon	Sample	Cutoff Values
AO	PC	4	Full	−0.057
AO	PC	4	Post 84	0.237
AO	PC	6	Full	−0.033
AO	PC	6	Post 84	0.161
AO	PC	8	Full	−0.051
AO	PC	8	Post 84	0.099

Notes: The last column reports the cutoff value of the real-time unemployment gap above (below) which the alternative (reference) model gives the smaller forecast error. This table includes only the cases with dark and middle shadings in Table 8 within the comparisons between each of the two Phillips curve models and either the AO or IMA models. The cutoff values are in the units of the four-quarter change in the unemployment gap (expressed in percentage points).

Table 12: Phillips Curve Forecast Using Real-Time and Final Unemployment Gaps

Forecast horizon	1969Q1−2014Q2		1984Q1−2014Q2	
	Real-Time	Final	Real-Time	Final
(a) Mean Absolute Errors				
2	1.260	1.226	1.111	1.107
4	1.182	1.140	0.972	0.966
6	1.271	1.224	0.982	0.974
8	1.295	1.255	0.972	0.971
(b) Root-Mean-Square Errors				
2	1.748	1.712	1.617	1.628
4	1.648	1.588	1.285	1.282
6	1.771	1.707	1.268	1.259
8	1.811	1.740	1.270	1.257

Table 13: GW Test Results Using Final Revised Gap: Phillips Curve versus AO and IMA Models

Conditioning		None Full Sample	None Sub-sample	Recession Dummy Full Sample	Recession Dummy Sub-sample	SPF Downturn Prob. Full Sample	SPF Downturn Prob. Sub-sample	Final Revised Gap Full Sample	Final Revised Gap Sub-sample	4Q Change in Gap Full Sample	4Q Change in Gap Sub-sample
		(a) 2-Step-Ahead Forecast									
AO	P-Value	0.853	0.019**	0.883	0.012**	0.359	0.003**	0.373	0.014**	0.828	0.020**
	R^2	0.000	0.000	−0.005	0.070	0.007	0.049	0.008	0.008	0.003	0.075
	Const.	−0.072	−0.750**	−0.135	−0.427**	−0.506	−0.014	−0.034	−0.725**	−0.052	−0.824**
	Slope	—	—	0.419	−3.550	0.022	−0.046	−0.415	−0.374	−1.562	−4.304
IMA	P-Value	0.197	0.027**	0.177	0.006**	0.011**	0.003**	0.323	0.024**	0.431	0.013**
	R^2	0.000	0.000	0.001	0.128	−0.005	0.061	0.030	0.052	0.064	0.199
	Const.	−0.394	−0.849**	−0.264*	−0.425**	−0.463*	−0.047	−0.343	−0.801**	−0.347	−0.965**
	Slope	—	—	−0.867	−4.657	0.004	−0.050	−0.551	−0.719	−3.637	−6.717**
		(b) 4-Step-Ahead Forecast									
AO	P-Value	0.178	0.030**	0.263	0.026**	0.318	0.029**	0.400	0.092*	0.353	0.028**
	R^2	0.000	0.000	0.072	0.006	0.085	−0.006	−0.004	−0.008	0.038	0.009
	Const.	0.481	−0.304**	0.079	−0.387**	−0.393	−0.396**	0.472	−0.305**	0.440	−0.293*
	Slope	—	—	2.650*	0.910**	0.044*	0.006	0.098	0.005*	2.649	1.151*
IMA	P-Value	0.514	0.002**	0.080**	0.006**	0.064**	0.006**	0.777	0.004**	0.698	0.007**
	R^2	0.000	0.000	0.036	−0.003	0.053	−0.007	−0.005	0.005	−0.004	−0.003
	Const.	−0.107	−0.392**	−0.319**	−0.364**	−0.615**	−0.346**	−0.102	−0.380**	−0.112	−0.396**
	Slope	—	—	1.399*	−0.315	0.026*	−0.003	−0.051	−0.114	0.367	−0.376
		(c) 6-Step-Ahead Forecast									
AO	P-Value	0.189	0.008**	0.215	0.027**	0.349	0.028**	0.379	0.010**	0.270	0.022**
	R^2	0.000	0.000	0.075	0.017	0.050	−0.005	−0.003	−0.008	0.057	0.053
	Const.	0.476	−0.377**	−0.01	−0.515**	−0.326	−0.500**	0.463	−0.376**	0.412	−0.384**
	Slope	—	—	3.171**	1.290	0.041	0.008	0.175	−0.009	3.695**	2.386**
IMA	P-Value	0.730	0.023**	0.082*	0.031**	0.097**	0.034**	0.911	0.027**	0.589	0.036**
	R^2	0.000	0.000	0.06	0.019	0.048	−0.006	−0.004	−0.002	0.005	0.020
	Const.	−0.060	−0.433**	−0.402**	−0.546**	−0.675**	−0.516**	−0.053	−0.417**	−0.081	−0.437**
	Slope	—	—	2.224*	1.054	0.031**	0.005	−0.103	−0.113	1.180	1.284
		(d) 8-Step-Ahead Forecast									
AO	P-Value	0.249	0.129	0.196	0.078*	0.392	0.067*	0.496	0.028**	0.208	0.037**
	R^2	0.000	0.000	0.121	0.037	0.045	0.009	−0.004	−0.006	0.059	0.089
	Const.	0.639	−0.280	0.028	−0.488**	−0.119	−0.572**	0.632	−0.293**	0.563	−0.323**
	Slope	—	—	3.935**	1.677	0.038	0.018	0.116	0.091	3.745**	3.087**
IMA	P-Value	0.667	0.042**	0.134	0.049**	0.184	0.041**	0.449	0.038**	0.779	0.060*
	R^2	0.000	0.000	0.103	−0.001	0.05	−0.003	0.007	0.008	−0.001	0.012
	Const.	0.148	−0.445**	−0.294	−0.503**	−0.473*	−0.561**	0.164	−0.421**	0.132	−0.459**
	Slope	—	—	2.849*	0.464	0.031*	0.007	−0.276	−0.172	0.806	0.984

Notes: *(**) indicate statistical significance at the 10% (5%) level. P-values and test statistics use HAC standard errors. See Subsection 6.2 for explanations of the shading.