WORKING PAPERS

RESEARCH DEPARTMENT

WORKING PAPER NO. 11-40
DO PHILLIPS CURVES CONDITIONALLY HELP TO FORECAST INFLATION?

Michael Dotsey
Federal Reserve Bank of Philadelphia

Shigeru Fujita
Federal Reserve Bank of Philadelphia

Tom Stark
Federal Reserve Bank of Philadelphia

September 2011

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA

Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

Do Phillips Curves Conditionally Help to Forecast Inflation?*

Michael Dotsey, Shigeru Fujita, and Tom Stark†

September 2011

Abstract

The Phillips curve has long been used as a foundation for forecasting inflation. Yet numerous studies indicate that over the past 20 years or so, inflation forecasts based on the Phillips curve generally do not predict inflation any better than a univariate forecasting model. In this paper, we take a deeper look at the forecasting ability of Phillips curves from both an unconditional and conditional view. Namely, we use the test results developed by Giacomini and White (2006) to examine the forecasting ability of Phillips curve models. Our main results indicate that forecasts from our Phillips curve models are unconditionally inferior to those of our univariate forecasting models and sometimes the difference is statistically significant. However, we do find that conditioning on various measures of the state of the economy does at times improve the performance of the Phillips curve model in a statistically significant way. Of interest is that improvement is more likely to occur at longer forecasting horizons and over the sample period 1984Q1−2010Q3. Strikingly, the improvement is asymmetric − Phillips curve forecasts tend to be more accurate when the economy is weak and less accurate when the economy is strong. It, therefore, appears that forecasters should not fully discount the inflation forecasts of Phillips curve-based models when the economy is weak.

JEL codes: C53, E37
Keywords: Phillips curve, unemployment gap, conditional predictive ability

*We wish to thank Todd Clark, Jesus Fernandez Villeverde, Frank Schorfheide, Keith Sill, Simon van Norden, Mark Watson, and Jonathan Wright for numerous helpful discussions. The views expressed herein are the authors' and do not reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. This paper is available for free of charge at www.philadelphiafed.org/research-and-data/publications/working-papers/.

†Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall Philadelphia, PA 19106 (e-mail: michael.dotsey@phil.frb.org; shigeru.fujita@phil.frb.org; tom.stark@phil.frb.org).

1 Introduction

The Phillips curve has long been used as a foundation for forecasting inflation. Yet numerous studies indicate 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 a forecast from an unobserved stochastic volatility model or an IMA(1,1) model. This point was forcefully made by Atkeson and Ohanian (2001) in regard to naive forecasts and has subsequently been explored in great depth by Stock and Watson (2007, 2008). Thus, a reasonable impression regarding the usefulness of Phillips curve models for forecasting inflation is fairly bleak. Stock and Watson, however, pose an interesting hypothetical question: namely despite the rich evidence against the usefulness of Phillips curve forecasts, would you change your forecast of inflation if you were told that next quarter the economy was going to enter a recession 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 and Sachs' *US Economics Analyst* posits that "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 $\cdots$ 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 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 sub-samples, 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. Stock and Watson (2008), however, do 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 will statistically investigate 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 non-nested models. In order 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 employing 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 sometimes the difference is statistically significant. However, we do find that conditioning on various measures of the state of the economy does, at times, improve the performance of the Phillips curve model in a statistically significant way. Of interest is that improvement is more likely to occur over the sample period 1984Q1−2010Q3 and that the unemployment gap improves forecasts in an asymmetric way. Specifically, the greater the actual gap, the greater the improvement in the conditional forecasting accuracy of the Phillips curve model. Therefore, it appears that policymakers should not fully discount the inflation forecasts of Phillips curve-based models when the economy is weak.

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 will be fairly focused, concentrating on those papers that help inform our particular approach. An excellent and in-depth literature review can be found in Stock and Watson (2008). A departure point for our inquiry is the work of Atkeson and Ohanian (2001). In that paper, the authors compare the root-mean-square error (RMSE) 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 1984:1−1999:11 and find that the forecasts generated by the Phillips curve models do not outperform those of the naive model. They, therefore, 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 relative to both 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 model (UCSV) as their benchmark for comparison. They find that there is not much difference between the naive forecasts over the 1984−1999 subsample, but that subsequently the forecasts generated by the two methods diverge upon 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 1993−2000. They also indicate that the 1993−2000 and 1985−1992 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 over 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 like Atkeson and Ohanian (2001), 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 predictions of inflation are not overly sensitive to whether a Phillips curve is relied on or not.

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.

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 RMSE 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 conditional on the state of the economy Phillips curve models provide better forecasts of inflation. Because Stock and Watson (2008) indicate that the measure of real activity is of secondary importance when evaluating forecast performance, we will concentrate on unemployment rates and unemployment gaps. We will also use real-time data on unemployment as our benchmark data set but will investigate whether the use of real-time data as opposed to final data affects our results. We also concentrate our forecasting exercise on core-PCE inflation and do so for two reasons. One is that core-PCE is often considered to be the most relevant measure of inflation for policy purposes and is also less affected by commodity price shocks than headline measures of inflation. Concentrating on core-PCE means that we must use the latest vintage estimates of inflation because real-time vintage data on the core-measure has been reported only since 1996. Thus, there does not exist a long enough set of vintages to use real-time measures of inflation.

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). Following Stock and Watson (2008) the naive forecast is based on the following specification:

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

where $\pi^h_t = (400/h)[\log(p_t) - \log(p_{t-h})]$ and p_t is the price index for core personal consumption expenditures 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 observations that would have been available at the date when the forecast was made.[1]

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:

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

where π^h_{t+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 a HP filter where we pad future observations

[1] Stock and Watson (2008) indicate that the IMA(1,1) model performs about as well as a more sophisticated unobserved component model with stochastic volatility.

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 model of Hansen (1997). However, there was insufficient variation in the data to identify the threshold over any of our rolling windows. We, therefore, imposed a value of 1.19, which implied that the absolute value of the unemployment gap exceeds the threshold roughly one-third of the time, by using the one standard deviation value of the gap. Doing so provided us with enough threshold measures to conduct our conditioning tests.

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 non-nested 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, and the test statistic is:

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

where h denotes the forecast horizon, δ_{t+h} is the difference in the squared h-step ahead forecast errors between any two forecasting models, 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 a heteroskedastic autoregressive correction is necessary, since we are looking at multiple-period ahead forecast errors. Following Giacomini and White (2006), we apply a Newey-West estimator with truncation parameter set to $h-1$.[2] For conditional tests, we examine the test statistic:

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

where x is a $k \times 1$ vector of instruments and $\widehat{V}_h$ is 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. In order 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.

When comparing the forecasts of our two Phillips curve models with the two benchmarks, we will also examine when there are statistically significant differences conditional on (i) whether the economy is in recession, (ii) the probability of recession from the Survey of Professional Forecasters (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:

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

[2]It is common when using a Newey-West correction to employ a truncation parameter that is somewhat larger and that depends on the sample size. We follow Giacomini and White's methodology, because in footnote 5 of their paper they indicate that $h-1$ works well in practice.

where x 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 as the SPF recession probabilities indicate forecasters are generally cognizant in real time as to whether the economy is or is not in recession. Even if one were somewhat uncertain about whether the economy was in recession, a policymaker with an asymmetric loss function might want to condition on being in a recession if there was sufficient evidence indicating that the economy might be in a recession.

4 Data Definitions and Transformations

Our analysis uses real-time data on unemployment constructed from vintage data available to the public in the middle of the quarter and latest vintage data on inflation. Thus, a regression run at date t uses observations on unemployment as they were known as of that date and inflation as it was known in the last quarter of our data set. As regressions are rolled forward, updated data are used from the vintage that was 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 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.[3] In constructing

[3]Stock and Watson (2007, 2008) use a high-pass filter that filters out frequencies of less than 60 quarters. The value of 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

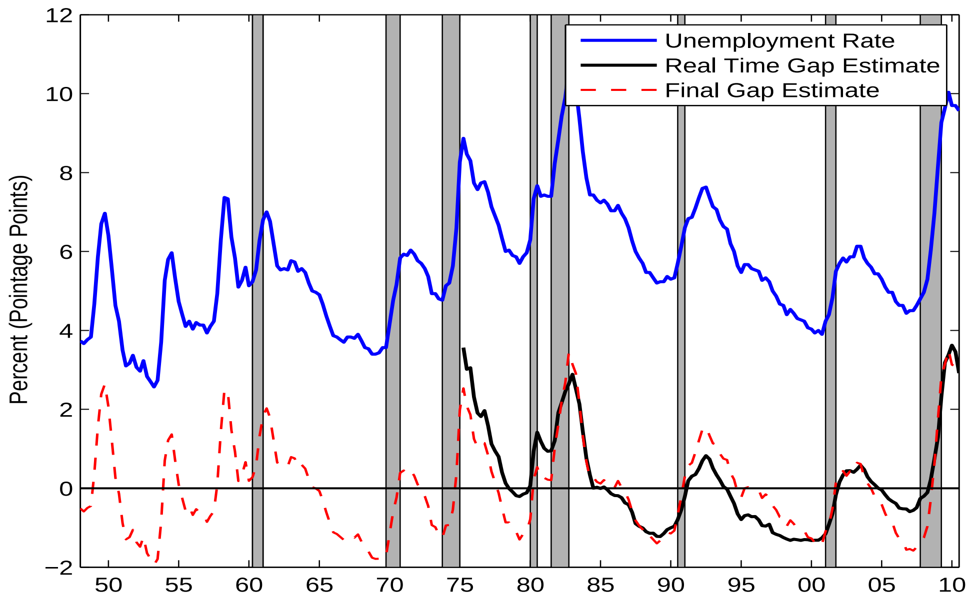

Figure 1: Unemployment and Unemployment Gap Series: Unemployment gap is based on the HP filter with smoothing parameter of 10^5. The final estimate of the gap series uses the 2010Q4 vintage of data. Shading indicates periods of the NBER recession.

trend unemployment, we use a 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 110 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 dotted-red

those associated with the traditional business cycle frequencies as in this paper. Importantly, we have also conducted the same analysis using the smoothing parameter of 1,600 and found that using 10^5 yields more cases in favor of the Phillips curve models relative to using $1,600$.

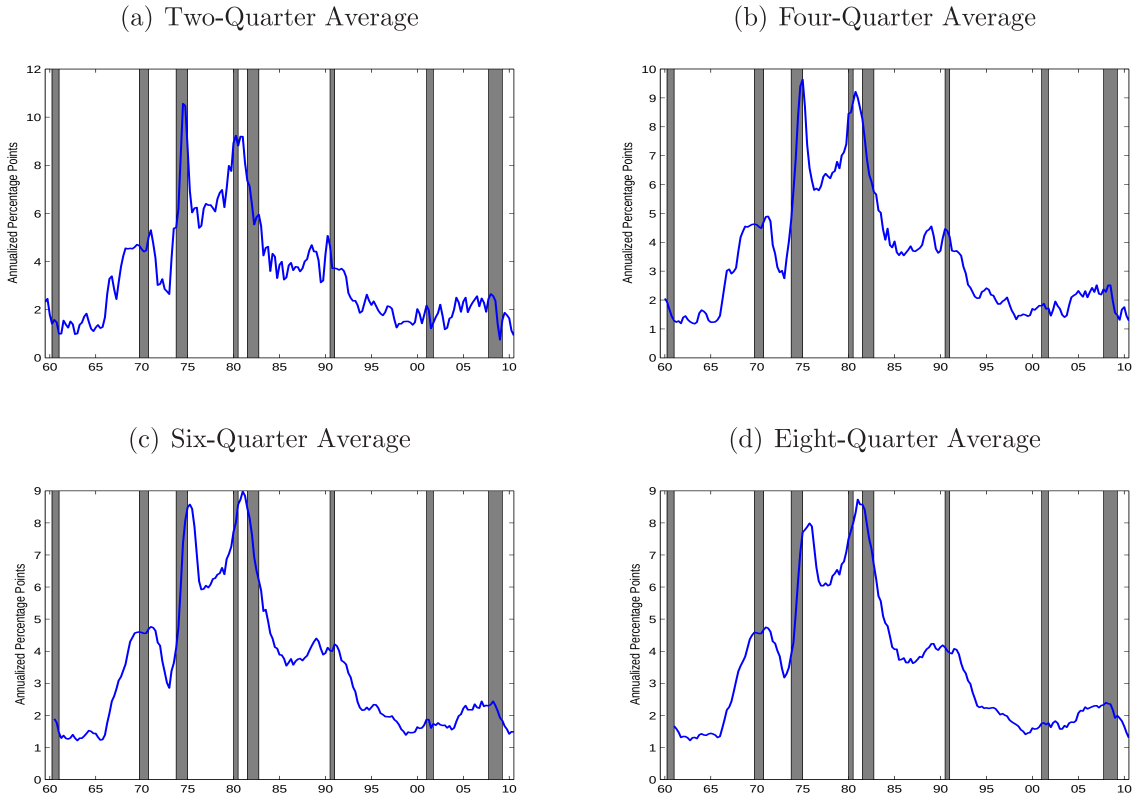

Figure 2: Core PCE Inflation Realizations: Shading indicates periods of NBER-dated recessions.

line shows the unemployment gap using the 2010Q4 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 latter half of the 2000s, the unemployment gap is a good deal higher than the final estimate and these are periods of falling unemployment, but the opposite is true of the 1990s where real-time gap is lower than the final estimate and again unemployment is falling.

The dependent variable in our analysis is various averages of final-revised core-PCE inflation, and these are depicted in Figure 2. We are forced to use final-revised data for this variable because vintage history begins in 1996, when U.S. Bureau of Economic Analysis first constructed the core-PCE price index. This is probably not a significant problem for our real-time focus because PCE inflation, for which we do have vintage data, does not suffer from the same sort of revision problems as the unemployment gap. Those revisions are due to not having knowledge of future unemployment rates and relying on a one-sided filter.

5 The Usefulness of Phillips Curve Forecasts

In this section, we analyze how useful Phillips Curve models are for forecasting inflation using the real-time unemployment gap. Our motivation for emphasizing the use of real-time data are twofold. The first is that these are the data that are relevant for policy purposes, and second the work of Orphanides and Van Norden (2005) on the output gap and our own analysis of real-time unemployment gaps makes the strong case for incorporating the measurement error associated with the real-time gap. Our investigation will focus 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.[4] 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 benchmarks (1) and (2) where we use unemployment gaps based on the current real-time vintage as of period t. Lag length is re-estimated 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 1975Q3 to 2010Q3 and a later sample period that includes forecasts from 1984Q1 through 2010Q3. The entire sample begins in 1975Q3 for the two-step horizon because 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.

With regard to the threshold Phillips curve model, we set the threshold of the real-time gap at a fixed value of 1.19 throughout our exercise. We initially intended to estimate the threshold value for each rolling window, applying the TAR model of Hansen (1997). However, the use of rolling regressions makes it difficult to tightly identify the threshold values that are reasonably stable over time. The value of 1.19 equals the standard deviation of the final revised unemployment gap series, and we have chosen this value to ensure that there is at least some variation in the threshold dummy for each estimation window.

[4] For an excellent summary as well as an exhaustive set of experiments, see Stock and Watson (2008).

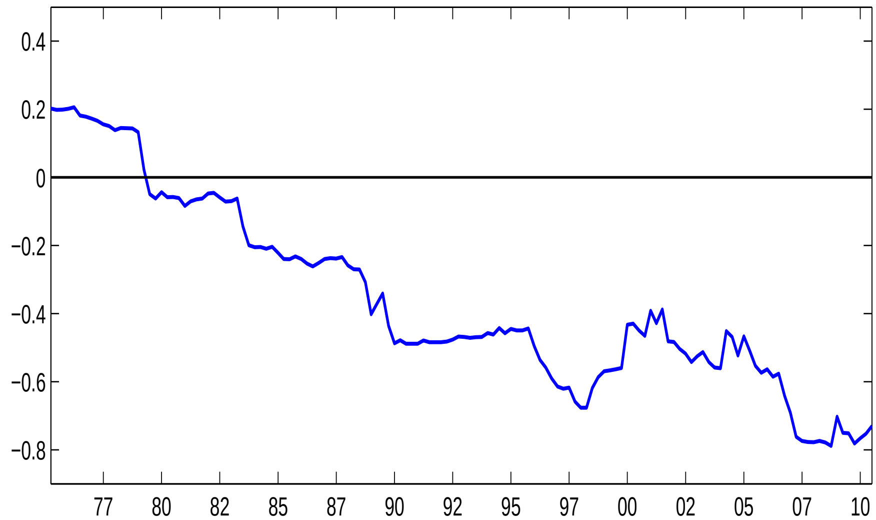

Figure 3: IMA (1,1) Coefficient Estimates: Estimated on a fixed window of 60 quarters. Coefficient estimates are aligned at sample endpoints.

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, a one-percentage point inflation shock is associated with a long-run multiplier $(1+\hat{\theta})$ on the level of inflation of 1.20. The multiplier then declines fairly consistently. At present, the long-run multiplier is about 0.2 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 small effect on the level of inflation. Thus, over our sample, the behavior of inflation changes from something close to a random walk to a process that more closely resembles white noise.[5]

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, but it becomes less negative as we roll the

[5] 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.

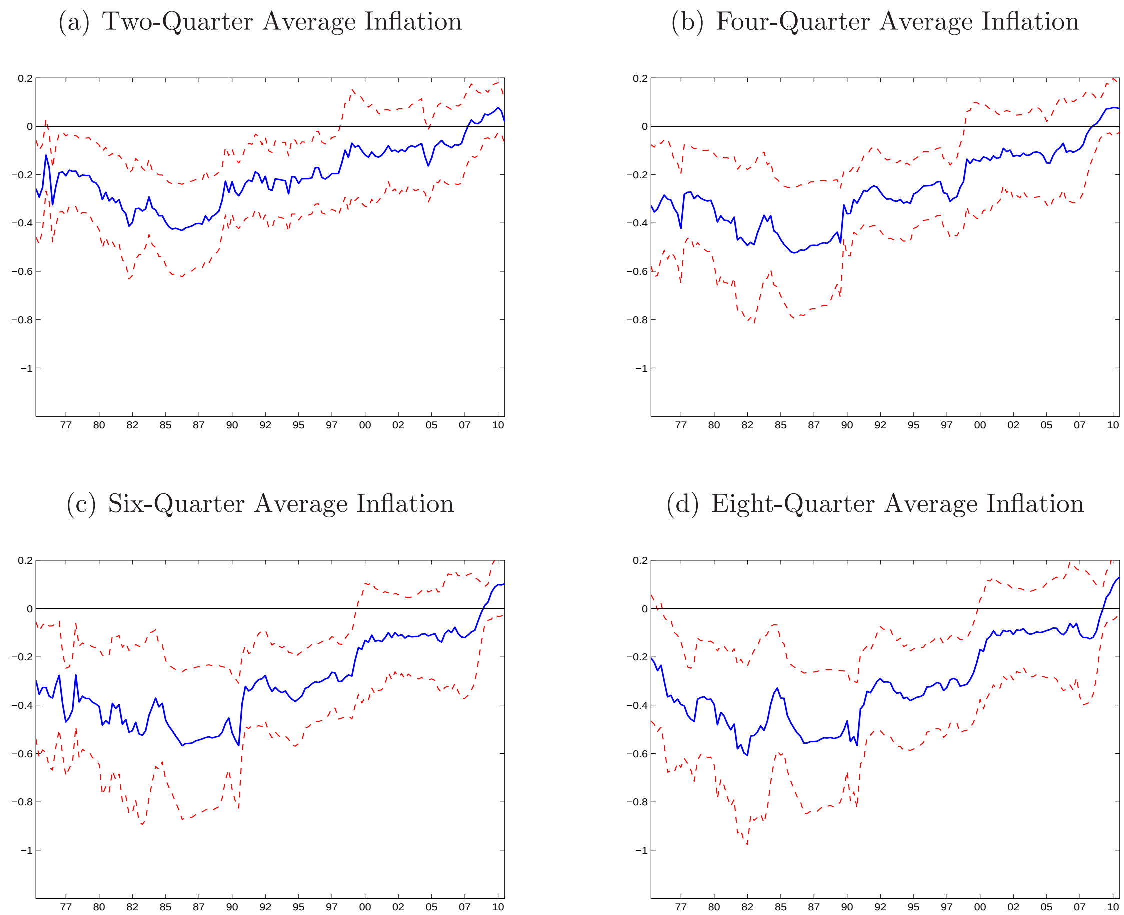

Figure 4: Coefficients on the Unemployment Gap in the Phillips Curve: Dashed lines indicate the 90 percent confidence interval based on HAC standard errors.

regressions forward and is statistically insignificantly different from zero beginning around 1999 at all horizons. The falling significance may in part be due to the more transitory nature of changes in inflation that we documented above along with the observation that the movements in the gap remain highly persistent over the entire sample.

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 although not significantly so. 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 are 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).

5.2 Forecast Comparisons

In this section, we compare both the unconditional and conditional forecasting performance of our four models. We first take a general look at the forecasts and document the unemployment gaps contribution 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 AO and both Phillips curve specifications over the full sample. With the exception of the eight-quarter forecast horizon, the AO specification is preferred over the more recent sample period. The findings regarding the relative forecasting ability of our two benchmarks generally agree with the analysis of Stock and Watson (2007). However, they run counter to the analysis of Fisher et al. (2002) who find that Phillips curve models help forecast inflation at two-year horizons for the core PCE.

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 late 1980s. There is also a large disparity between recent forecasts generated by the Phillips curve and the IMA(1,1) model. 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 is similar across all forecast horizons, but especially so for the four-, six-, and eight-quarter horizons. During the early 1980s, the unemployment gap makes a pronounced contribution to the Phillips curve projections at all horizons. This period is characterized by a large unemployment gap that pulls 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. However, recently, the gap is also high, but it is contributing to higher expected inflation due to the perverse sign of the estimated coefficient, which as shown in Figure 4 is now insignificantly positive. Further,

(a) Two-Quarter Average Inflation

(b) Four-Quarter Average Inflation

(c) Six-Quarter Average Inflation

(d) Eight-Quarter Average Inflation

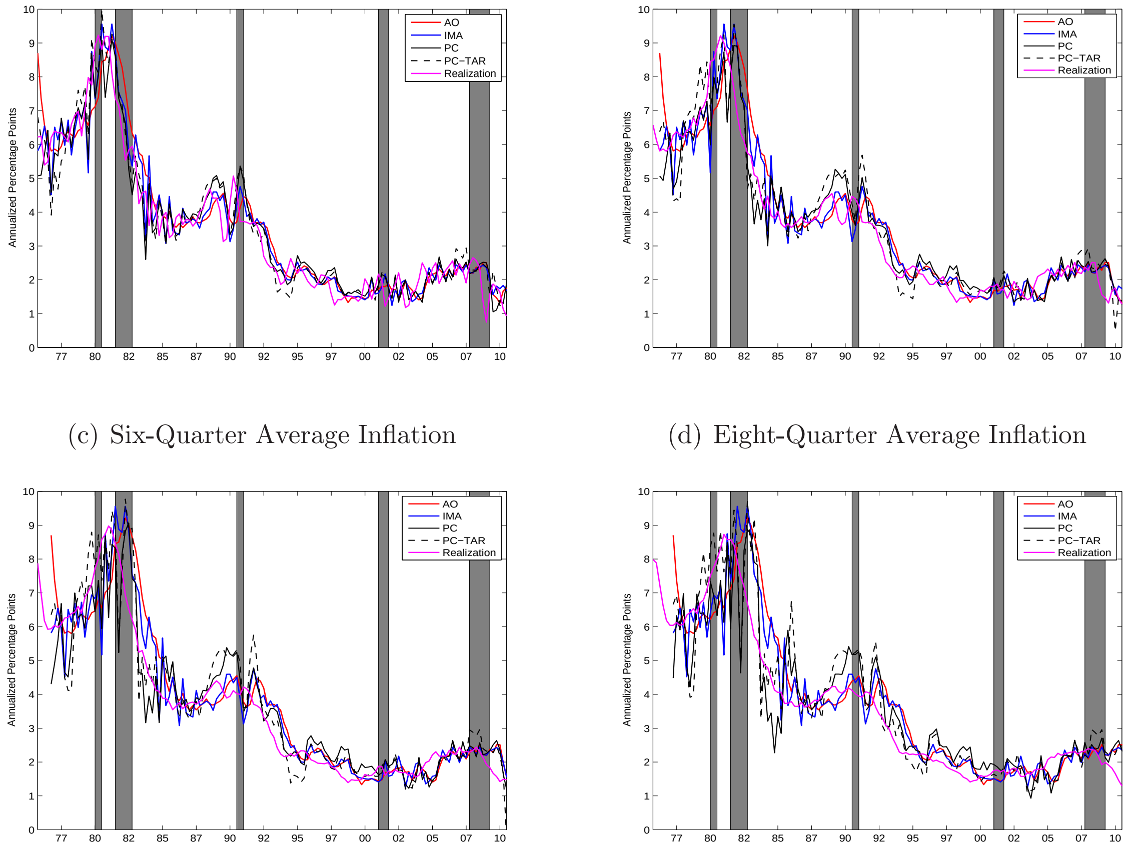

Figure 5: Inflation Projections: Shading indicates periods of NBER-dated recessions.

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 the 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 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 periods where the gap is large and its contribution to the inflation forecast is helpful relative to the benchmark forecast. It remains to be seen if that help is statistically significant.

(a) Two-Quarter Average Inflation

(b) Four-Quarter Average Inflation

(c) Six-Quarter Average Inflation

(d) Eight-Quarter Average Inflation

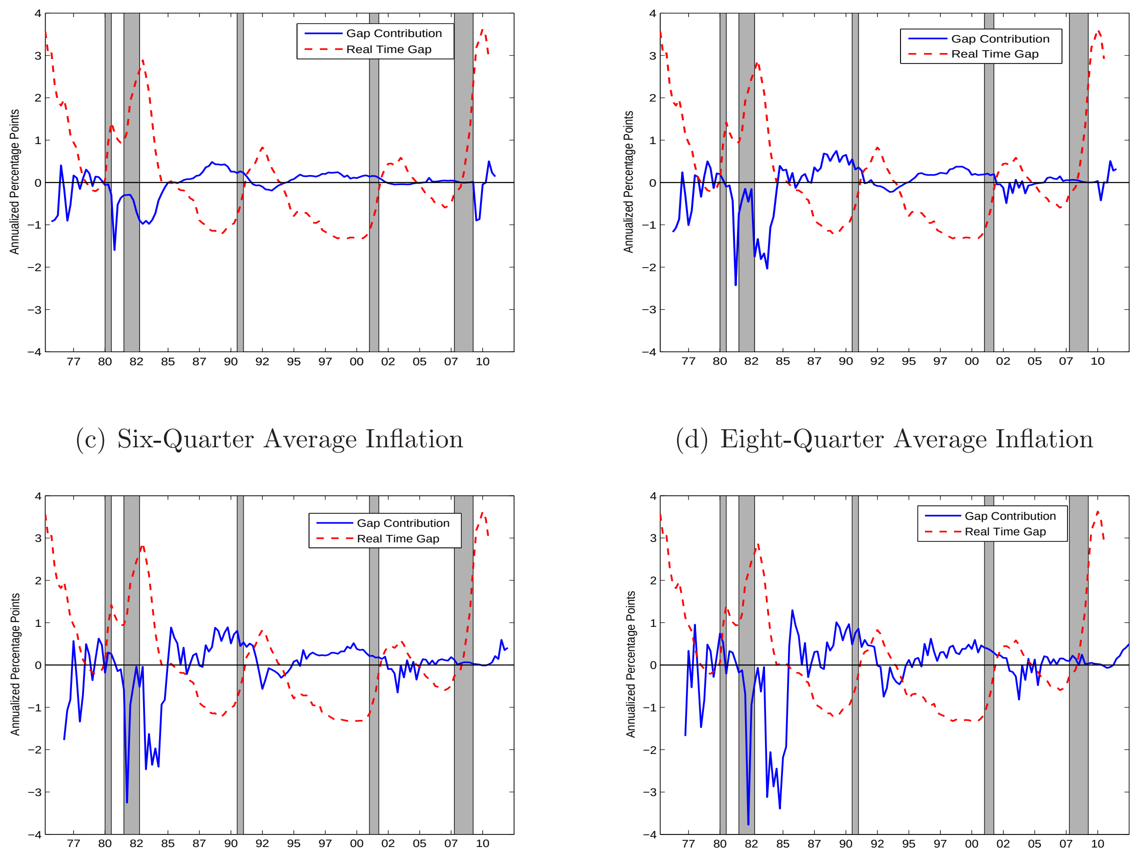

Figure 6: Effect of Unemployment Gap on Phillips Curve Forecasts: The real-time unemployment gap is aligned at the date when the forecast was made. The contribution term is plotted at the date forecasted. Shading indicates periods of NBER-dated recessions.

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 forecasts developed by Giacomini and White (2006).

6.1 Unconditional Comparison

First, we investigate whether the results concerning forecast accuracy 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 to that of the Phillips curve. In comparing forecasts we use both a 5% and 15% significance level. Over the entire sample, there is no statistically significant differences in forecast ability between the naive models and the two Phillips curve specifications, although the constant in (7) is generally negative. With regard to the more recent sample, the AO specification is preferred to both Phillips curve specifications at the two-quarter horizon, and the TAR specification at the four-quarter horizon, but not significantly preferred at longer horizons, while the IMA(1,1) forecast is significantly preferred to the Phillips curve model at the eight-quarter horizon. Thus, from the unconditional tests, there is little to suggest the use of a Phillips curve specification for forecasting core-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 affect the relative forecasting properties of 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 1.19. Alternatively, it may be that the unemployment gap may affect 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 row variable refers to the reference model and the columns refer to the alternative model. We report the p-values of the GW chi-square 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

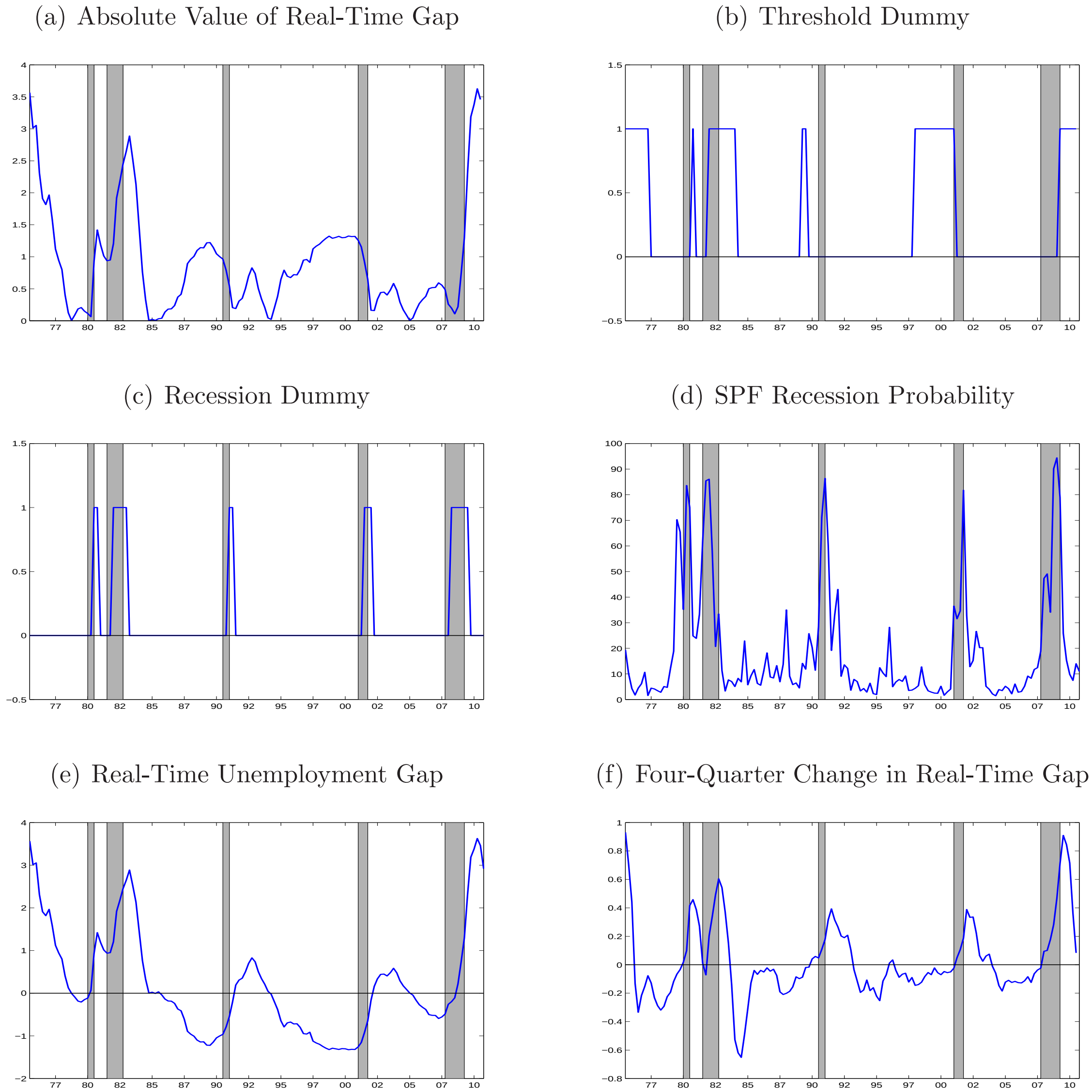

Figure 7: Conditioning Information for Giacomini-White Tests: Each plot shows our GW conditioning information. The data are aligned (using the timing conventions discussed in the paper) at the forecast date (not the date forecasted). Shading shows periods of NBER-dated recessions.

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 the GW χ^2 statistic is significant, indicating that the two forecasts 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, and 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 is now the better forecast. 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 are Phillips curve forecasts. In terms of the conditioning variables using the unemployment gap, a positive coefficient implies that high unemployment gaps improve the Phillips curve forecasts, but negative 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 the full sample, we found no cases in which conditioning on this variable improved Phillips curve forecasts relative to those of our two benchmarks. Over the more recent sample, conditioning on this variable did improve the PC-TAR specification relative to both the AO and IMA(1,1) models, but only at the two-quarter horizon. Similarly, conditioning on the threshold dummy does not improve the forecast performance of the Phillips curve models relative to the benchmarks (Table4).

The second basic result is that when the conditioning tends to be asymmetric, we find that in recessions there is a tendency for improvement in inflation forecasts from the Phillips curve models, especially over the more recent sample period. With regard to the recession dummy, there are some notable changes in the later forecast period (see Table 5). Namely both Phillips curve models provide better longer-term forecasts of inflation, and the PC-TAR model provides a better forecast relative to AO at the two-quarter horizon. For the full sample, there is no evidence of any statistically significant difference in the forecasts where previously, in the unconditional tests, the IMA(1,1) model was preferred at the two-quarter and eight-quarter forecast horizons. The results conditioning on the SPF probability of recession also indicate that this variable significantly improves both of our Phillips curve longer-term forecasts over both the entire and the later sample (see Table 6).

Using the real-time gap series, $\tilde{u}_{t-1}$ has even larger repercussions for both Phillips curve forecasts (Table 7). Over the later period, it significantly improves them relative to the two benchmark models at almost all forecast horizons, and for the entire sample it improves the basic Phillips curve forecast relative to AO at all but the shortest horizon, but not relative to the preferred IMA(1,1) benchmark. Table 8 presents the results when the four-quarter change in the real-time gap is used as a conditioning variable. These results further strengthen the preference for the Phillips curve forecasts. In particular, over the full sample, the significant improvements of the Phillips curve forecasts are identified at almost all forecast horizons when compared with the AO forecasts. For the later sample, the improvements of Phillips curve forecasts, relative to both AO and IMA(1,1) forecasts, continue to be observed, particularly at longer horizons.

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 in the later sample that the slope coefficient on the recession dummy is significant for the Phillips curve model at the six-step-ahead and eight-step-ahead forecast horizons when compared with AO and that this is also true at the eight-step-ahead horizon when comparing the forecasts from the IMA(1,1) model and the threshold model. Table 9 selects the cases from Table 5 in which both 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. Of particular interest is the case involving the eight-step-ahead forecast results comparing AO and both Phillips curves. In this case, during expansions the AO model is preferred, while during recessions one is better off using the Phillips curve models for forecasting.

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. Tables 10 through 12 present the cutoff values for the three continuous conditioning variables, focusing on the cases with the darkest shading and middle shading in the earlier tables.

The fourth and fifth rows of Table 10 indicate that the cutoff value on the SPF downturn probability, above which the Phillips curve models are producing lower forecast errors relative to the AO model, are 23.7% for the basic Phillips curve model and 22.6% for the threshold Phillips curve model at the eight-quarter forecast horizon over the post-1984 sample period. When SPF downturn probabilities exceed these numbers, one should carefully consider Phillips curve predictions of inflation.

Looking at results when conditioning on our two gap variables (Tables 11 and 12) lend support to using Phillips curve forecasts in even more circumstances. Concentrating on the situation that is indicated by the darkest shading, i.e., the cases in which both the GW test statistic is significant at the 15% level and the slope coefficients are also significantly positive, we see that for the real-time gap it pays to look at the Phillips curve forecasts over the later sample period at four-, six-, and eight-quarter-ahead horizons when compared with AO even when the unemployment gap is only slightly positive. Regarding the PC-TAR model over the later sample and two-, six-, and eight-quarter-ahead horizons, we draw a similar conclusion when AO is the benchmark.

These results are reinforced when conditioning on the four-quarter change in the real-time gap. When looking at the later sample and at four-quarter, six-quarter, and eight-quarter-ahead forecast horizons, we see that there exist cutoff values below which the respective benchmark is the preferred specification and above which the respective Phillips curve model is preferred. First consider the case when the AO model is the benchmark. At the four-quarter horizon, the Phillips curve is the preferred model when the change in the unemployment gap exceeds 0.155. At the six-quarter-ahead horizon, the Phillips curve becomes the preferred model when the change in the unemployment gap exceeds 0.014. For the eight-quarter-ahead horizon, the cutoff values are 0.036 and 0.012, respectively. Thus, when analyzing the later sample period, when the unemployment gap is rising, inflation forecasts at most horizons using the Phillips curve improve, and we are struck by the relatively small values of the gap that imply improvement. The conclusion when comparing IMA(1,1) forecasts and Phillips curve forecasts are roughly the same. Thus, although Phillips curve

(a) AO vs. Phillips Curve

(b) AO vs. Threshold Phillips Curve

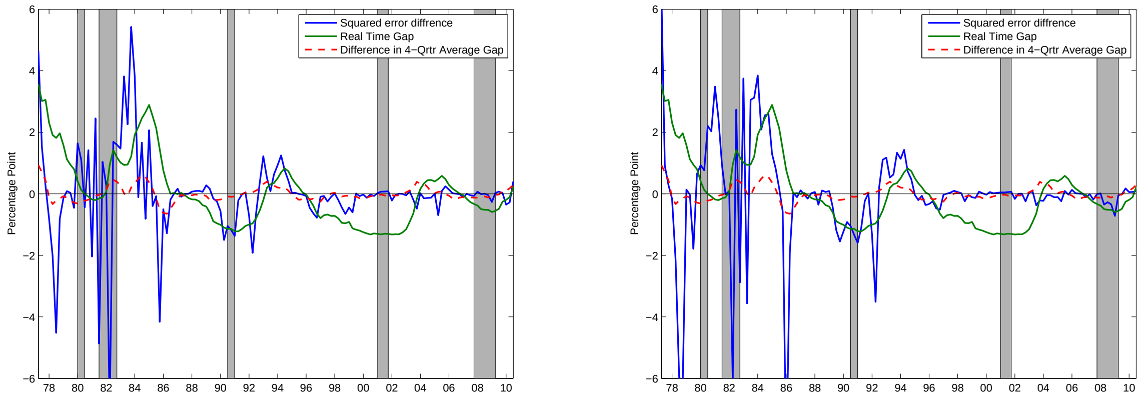

Figure 8: Squared Error Difference and Gap Measures: eight-quarter-ahead forecasts. Shading shows periods of NBER-dated recessions. The squared error difference is aligned at the date forecasted, and the observations for the gap measures at that date are the ones used in the GW regression.

forecasts do not generally outperform the benchmark forecasts, there do exist situations when they prove useful. These situations are much more 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: (i) the squared forecast error differences for the AO and Phillips curve model and the AO and threshold Phillips curve model at the eight-quarter-ahead forecast horizon, (ii) the real-time unemployment gap, and (iii) the four-quarter change in the gap. In the figure, the squared error difference is associated with the date forecasted, and the gap measures at that date are the ones used in the GW regressions. We are thus plotting the left- and right-hand sides of (8). There are several interesting observations in this figure. First, Phillips curve forecasts tend to be better than the AO benchmark around 1983−84 and 1993−95. Both are periods where rising unemployment helped forecast a decline in inflation.[6] Thus, in these periods, the gap is large and tends to improve the relative forecasting accuracy of the Phillips curve. Second, recall our earlier finding that the improvement of the conditional forecast ability of the Phillips curve models was largely concentrated in the post-1984 sample. Both panels of Fig-

[6]These dates are similar to the ones over which Stock and Watson (2008) also indicate that their Phillips curve models forecast relatively well.

ure 8 clearly illustrate that this result comes from the stronger positive correlations between the squared error differences and the gap series between 1984 through the mid-1990s. It is not surprising that including the observation prior to 1984 only weakens the result. Third, the 1989−91 period is one of the periods that contributes to the positive coefficient of the GW regression. Note, however, that during this period, the negative gap is associated with a worsening of the Phillips curve forecasts, implying that the usefulness of Phillips curve models is asymmetric.

7 Results Using Latest Vintage Data

In this section, we look at whether and to what extent the use of latest vintage data for the unemployment gap influences our conclusions. To do this, we re-estimate the Phillips curve using final estimates of the unemployment gap, compute new forecasts, and re-run 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 13, 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.

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 14. With regard to the unconditional forecast evaluation presented in the first two columns of that table, there are no qualitative changes in results. The AO specification is still preferred over the later sample period for the two-step-ahead forecast horizons. Also, the IMA(1,1) forecast remains statistically better for eight-quarter horizon forecasts over the later sample.

Examining the conditional forecast results with respect to the recession dummy, there is no longer any evidence that this variable conditionally improves Phillips curve forecasts as it did over the later sample at six- and eight-quarter horizons when using real-time data. On the other hand, there is qualitatively little change in forecast evaluation when we condition on the SPF downturn probability. Over the entire sample, there is now a statistical distinction between the quality of the forecasts between the AO benchmark model and the Phillips curve model, but only at the four-quarter forecast horizon. Over the later sample, the results of the forecast comparison are little changed. As in the real-time analysis, the recession probability improves the Phillips curve forecast relative to AO at the eight-quarter-ahead forecast horizon and now additionally at the six-quarter horizon. The comparisons with respect to IMA(1,1) are nearly identical. With respect to the gap variables, either final revised gaps or four-quarter changes in the gap, using the latest revised data does not appreciably change any of the conclusions. Thus, 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 have 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 an 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 1975−2010 sample, we find, as did Stock and Watson (2007, 2008), that an IMA(1,1) model outperforms Phillips curve models but not in a statistically significantly way. For the 1984−2010 sample, the AO model is the preferred forecast model and significantly so at the two-quarter-ahead forecast horizon. Using the latest revised output gaps as opposed to final time output gaps does not appreciably change the 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. Importantly, we find that its usefulness is asymmetric helping in times when the economy is weak and hurting the accuracy of inflation forecasts when the economy is growing. The variables that provide the biggest improvement pertain to unemployment gaps themselves both in their level and rate of change. The statistically significant improvement tends to be concentrated over the later sample period, which is in stark contrast to the general perception one obtains from the existing literature. It is important to note that this result refers largely to our conditional forecast exercises, so it is not directly comparable to results based on unconditional forecast comparisons.

Finally, we have focused our analysis strictly on core-PCE inflation because it is thought by many to be the most relevant inflation measure for monetary policy in the U.S. We have also confined our Phillips curve analysis to unemployment gaps, and it would be interesting to see if our results carry over to other inflation and 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

Ang, A., G. Bekaert and M. Wei, "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?," *Journal of Monetary Economics* 54 (2007), 1163–1212.

Atkeson, A. and L. Ohanian, "Are Phillips Curves Useful for Forecasting Inflation?," *Federal Reserve Bank of Minneapolis Quarterly Review* 25 (2001), 2–11.

Clark, T. and M. 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 (2006), 1127–1148.

Dotsey, M. and T. Stark, "The Relation Between Capacity Utilization and Inflation," *Federal Reserve Bank of Philadelphia Business Review* Q2 (2005), 8–17.

Fisher, J., C. Liu and R. Zhou, "When Can We Forecast Inflation?," *Federal Reserve Bank of Chicago Economic Perspectives* 1Q (2002), 30–42.

Fuhrer, J. and G. Olivei, "The Role of Expectations and Output in the Inflation Process: An Empirical Assessment," Federal Reserver Bank of Boston Public Policy Briefs No. 10-12, May 2010.

Giacomini, R. and H. White, "Tests of Conditional Predictive Ability," *Econometrica* 74 (2006), 1545–1578.

Hansen, B., "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics* 2 (1997), 1–14.

Orphanides, A. and S. Van Norden, "The Reliability of Inflation Forecast Based on Output Gap Estimates in Real Time," *Journal of Money, Credit and Banking* 37 (2005), 583–600.

Orphanides, A. and J. Williams, “The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning, and Expecations,” *Journal of Economic Dynamics and Control* 29 (2005).

Shimer, R., “The Cyclical Behavior of Equilibrium Unemployment and Vacancies,” *American Economic Review* 95 (2005), 25–49.

Stock, J. and M. Watson, “Forecasting Inflation,” *Journal of Monetary Economics* 44 (1999), 293–335.

———, “Why Has U.S. Inflation Become Harder to Forecast?,” *Journal of Money, Credit and Banking* 39 (2007), 3–34.

———, “Phillips Curve Inflation Forecast,” in J. Fuhrer, J. S. Little, Y. Kodrzycki and G. Olivei, eds., *Understanding Inflation and the Implications for Monetary Policy, a Phillips Curve Retrospective* (MIT Press, 2008), 101–187.

Table 1: Forecast Error Comparisons for the Inflation Rate

Forecast horizon	75Q3−10Q3 AO	75Q3−10Q3 IMA	75Q3−10Q3 PC	75Q3−10Q3 PC-TAR	84Q1−10Q3 AO	84Q1−10Q3 IMA	84Q1−10Q3 PC	84Q1−10Q3 PC-TAR
(*a*) Mean Absolute Errors								
2	0.524	0.515*	0.552	0.545	0.404*	0.429	0.466	0.456
4	0.531	0.493*	0.506	0.558	0.374*	0.413	0.411	0.459
6	0.567	0.532*	0.573	0.606	0.391*	0.422	0.434	0.486
8	0.632	0.587*	0.659	0.652	0.446*	0.447	0.516	0.513
(*b*) Root-Mean-Square Errors								
2	0.716	0.697*	0.727	0.709	0.540*	0.573	0.612	0.598
4	0.756	0.692	0.690*	0.744	0.492*	0.534	0.547	0.601
6	0.826	0.756*	0.806	0.845	0.528*	0.555	0.557	0.632
8	0.909	0.840*	0.917	0.928	0.616	0.591*	0.684	0.689

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

		75Q3−10Q3			84Q1−10Q3		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
		(a) 2-Step-Ahead Forecast					
AO	P-Value	0.756	0.853	0.922	0.266	0.026**	0.118*
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.027	−0.016	0.010	−0.037	−0.083**	−0.066*
IMA	P-Value		0.385	0.779		0.263	0.509
	R^2		0.000	0.000		0.000	0.000
	Const.		−0.043	−0.017		−0.046	−0.030
PC	P-Value			0.597			0.473
	R^2			0.000			0.000
	Const.			0.026			0.017
		(b) 4-Step-Ahead Forecast					
AO	P-Value	0.406	0.413	0.903	0.156	0.250	0.076*
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.093	0.096	0.018	−0.043	−0.057	−0.120*
IMA	P-Value		0.962	0.460		0.775	0.225
	R^2		0.000	0.000		0.000	0.000
	Const.		0.003	−0.075		−0.014	−0.077
PC	P-Value			0.248			0.158
	R^2			0.000			0.000
	Const.			−0.078			−0.063
		(c) 6-Step-Ahead Forecast					
AO	P-Value	0.326	0.812	0.849	0.254	0.680	0.253
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.110	0.033	−0.032	−0.029	−0.031	−0.121
IMA	P-Value		0.357	0.298		0.968	0.341
	R^2		0.000	0.000		0.000	0.000
	Const.		−0.077	−0.142		−0.003	−0.092
PC	P-Value			0.583			0.169
	R^2			0.000			0.000
	Const.			−0.065			−0.089
		(d) 8-Step-Ahead Forecast					
AO	P-Value	0.352	0.926	0.858	0.558	0.372	0.570
	R^2	0.000	0.000	0.000	0.000	0.000	0.000
	Const.	0.120	−0.014	−0.035	0.030	−0.088	−0.095
IMA	P-Value		0.098	0.366		0.094*	0.329
	R^2		0.000	0.000		0.000	0.000
	Const.		−0.134	−0.155		−0.118*	−0.125
PC	P-Value			0.872			0.937
	R^2			0.000			0.000
	Const.			−0.021			−0.007

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 model listed in the column. * (**) indicate statistical significance at the 15% (5%) level. P-values and test statistics use HAC standard errors.

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

		75Q3−10Q3			84Q1−10Q3		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
(a) 2-Step-Ahead Forecast							
AO	P-Value	0.645	0.583	0.569	0.537	0.071*	0.085*
	R^2	0.021	0.006	0.016	0.001	−0.009	0.006
	Const.	−0.147	−0.140	−0.166	0.002	−0.071*	−0.125**
	Slope	0.194	0.138	0.196	−0.052	−0.016	0.079*
IMA	P-Value		0.661	0.956		0.130*	0.102*
	R^2		−0.002	−0.007		−0.006	0.031
	Const.		0.008	−0.019		−0.073*	−0.127**
	Slope		−0.057	0.001		0.036	0.131*
PC	P-Value			0.761			0.288
	R^2			−0.001			0.063
	Const.			−0.026			−0.054
	Slope			0.058			0.095**
(b) 4-Step-Ahead Forecast							
AO	P-Value	0.597	0.610	0.758	0.362	0.344	0.195
	R^2	0.042	0.039	0.013	0.022	−0.009	−0.006
	Const.	−0.211	−0.179	−0.190	0.025	−0.062	−0.080
	Slope	0.354	0.319	0.241	−0.092	0.006	−0.055
IMA	P-Value		0.934	0.648		0.434	0.263
	R^2		−0.006	−0.002		0.009	−0.008
	Const.		0.032	0.021		−0.086	−0.104
	Slope		−0.034	−0.112		0.098	0.038
PC	P-Value			0.464			0.280
	R^2			0.000			0.001
	Const.			−0.011			−0.018
	Slope			−0.078			−0.061
(c) 6-Step-Ahead Forecast							
AO	P-Value	0.475	0.494	0.653	0.521	0.415	0.137*
	R^2	0.052	0.014	0.007	−0.008	0.031	0.005
	Const.	−0.236	−0.195	−0.237	−0.015	−0.160	−0.231
	Slope	0.416*	0.274	0.246	−0.019	0.177	0.151
IMA	P-Value		0.620	0.523		0.537	0.155
	R^2		0.000	−0.001		0.039	0.010
	Const.		0.041	−0.001		−0.146	−0.216
	Slope		−0.142	−0.170		0.196	0.170
PC	P-Value			0.855			0.343
	R^2			−0.007			−0.009
	Const.			−0.042			−0.071
	Slope			−0.028			−0.026
(d) 8-Step-Ahead Forecast							
AO	P-Value	0.411	0.504	0.765	0.624	0.387	0.307
	R^2	0.049	0.005	0.001	0.027	−0.006	0.036
	Const.	−0.236*	−0.199	−0.227	−0.082	−0.141	−0.410*
	Slope	0.429*	0.223	0.232	0.150	0.071	0.422
IMA	P-Value		0.207	0.564		0.233	0.295
	R^2		0.005	−0.002		−0.005	0.012
	Const.		0.037	0.009		−0.059	−0.328*
	Slope		−0.206*	−0.197		−0.079	0.272
PC	P-Value			0.987			0.390
	R^2			−0.008			0.033
	Const.			−0.028			−0.269*
	Slope			0.009			0.350**

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 model listed in the column. *(**) indicate statistical significance at the 15% (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

		75Q3−10Q3			84Q1−10Q3		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
(a) 2-Step-Ahead Forecast							
AO	P-Value	0.828	0.909	0.968	0.533	0.066*	0.282
	R^2	−0.002	−0.006	−0.006	0.001	−0.008	−0.004
	Const.	−0.012	−0.033	−0.010	−0.020	−0.075*	−0.081*
	Slope	0.144	0.062	0.071	−0.083	−0.041	0.072
IMA	P-Value		0.645	0.906		0.287	0.297
	R^2		−0.004	−0.005		−0.008	0.013
	Const.		−0.021	0.002		−0.055*	−0.062
	Slope		−0.082	−0.073		0.043	0.155
PC	P-Value			0.847			0.211
	R^2			−0.007			0.030
	Const.			0.023			−0.006
	Slope			0.009			0.112*
(b) 4-Step-Ahead Forecast							
AO	P-Value	0.639	0.673	0.983	0.357	0.510	0.158
	R^2	0.003	0.001	−0.007	0.001	−0.009	−0.007
	Const.	0.022	0.035	−0.003	−0.027	−0.058	−0.104
	Slope	0.273	0.232	0.078	−0.079	0.006	−0.075
IMA	P-Value		0.955	0.653		0.714	0.478
	R^2		−0.007	−0.001		−0.003	−0.010
	Const.		0.014	−0.024		−0.032	−0.078
	Slope		−0.041	−0.195		0.085	0.004
PC	P-Value			0.450			0.194
	R^2			0.003			−0.001
	Const.			−0.038			−0.046
	Slope			−0.154			−0.081
(c) 6-Step-Ahead Forecast							
AO	P-Value	0.446	0.568	0.702	0.475	0.514	0.325
	R^2	0.009	0.002	−0.001	−0.009	0.008	−0.004
	Const.	0.021	−0.043	−0.098	−0.025	−0.065	−0.149*
	Slope	0.358	0.305	0.264	−0.019	0.166	0.138
IMA	P-Value		0.655	0.582		0.602	0.327
	R^2		−0.007	−0.007		0.011	−0.001
	Const.		−0.064	−0.119		−0.041	−0.124*
	Slope		−0.053	−0.094		0.184	0.157
PC	P-Value			0.846			0.387
	R^2			−0.007			−0.009
	Const.			−0.055			−0.084
	Slope			−0.041			−0.027
(d) 8-Step-Ahead Forecast							
AO	P-Value	0.352	0.652	0.651	0.572	0.398	0.229
	R^2	0.010	−0.002	0.004	0.023	−0.006	0.044
	Const.	0.023	−0.072	−0.140	−0.014	−0.113	−0.240*
	Slope	0.395*	0.234	0.428	0.208	0.116	0.674
IMA	P-Value		0.253	0.657		0.241	0.253
	R^2		−0.005	−0.008		−0.007	0.020
	Const.		−0.094	−0.163		−0.099	−0.225*
	Slope		−0.162	0.033		−0.092	0.466
PC	P-Value			0.873			0.419
	R^2			−0.005			0.040
	Const.			−0.069			−0.127
	Slope			0.194			0.558

Notes: See notes to Table 3. The threshold dummy takes 1 when the absolute value of real-time gap is larger than 1.19 and otherwise takes 0.

Table 5: GW Conditional Test: Recession Dummy

		75Q3−10Q3			84Q1−10Q3		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
(a) 2-Step-Ahead Forecast							
AO	P-Value	0.266	0.978	0.300	0.092*	0.079*	0.038**
	R^2	0.002	−0.007	0.011	0.015	−0.009	0.013
	Const.	−0.008	−0.018	−0.043	−0.054*	−0.085**	−0.087*
	Slope	0.269*	0.014	0.412*	0.169**	0.019	0.202**
IMA	P-Value		0.560	0.602		0.190	0.796
	R^2		0.013	−0.003		0.003	−0.009
	Const.		−0.011	−0.036		−0.031	−0.033
	Slope		−0.255	0.143		−0.149*	0.034
PC	P-Value			0.169			0.230
	R^2			0.041			0.050
	Const.			−0.025			−0.002
	Slope			0.398**			0.183**
(b) 4-Step-Ahead Forecast							
AO	P-Value	0.426	0.713	0.895	0.234	0.514	0.195
	R^2	−0.003	−0.007	−0.005	−0.001	−0.008	−0.008
	Const.	0.063	0.095	−0.006	−0.053*	−0.064	−0.126*
	Slope	0.225	0.009	0.180	0.094	0.064	0.061
IMA	P-Value		0.653	0.740		0.794	0.414
	R^2		0.002	−0.007		−0.009	−0.009
	Const.		0.031	−0.069		−0.011	−0.073
	Slope		−0.216	−0.046		−0.030	−0.033
PC	P-Value			0.348			0.369
	R^2			0.000			−0.010
	Const.			−0.100			−0.062
	Slope			0.170			−0.003
(c) 6-Step-Ahead Forecast							
AO	P-Value	0.457	0.965	0.809	0.290	0.158	0.230
	R^2	0.002	−0.006	−0.001	0.004	0.050	0.018
	Const.	0.067	0.015	−0.077	−0.040*	−0.076	−0.164*
	Slope	0.346	0.148	0.358	0.105	0.396*	0.384
IMA	P-Value		0.655	0.581		0.427	0.328
	R^2		−0.004	−0.007		0.022	0.006
	Const.		−0.052	−0.144		−0.035	−0.123
	Slope		−0.198	0.012		0.291	0.279
PC	P-Value			0.705			0.383
	R^2			−0.004			−0.009
	Const.			−0.091			−0.088
	Slope			0.210			−0.012
(d) 8-Step-Ahead Forecast							
AO	P-Value	0.479	0.696	0.446	0.181	0.103*	0.089*
	R^2	0.008	0.001	0.025	0.136	0.052	0.114
	Const.	0.063	−0.060	−0.147	−0.034	−0.155*	−0.244*
	Slope	0.504*	0.411	0.999	0.573*	0.597*	1.331**
IMA	P-Value		0.246	0.526		0.209	0.133*
	R^2		−0.007	−0.001		−0.009	0.037
	Const.		−0.124	−0.210		−0.121*	−0.210*
	Slope		−0.093	0.495		0.024	0.758*
PC	P-Value			0.261			0.221
	R^2			0.007			0.041
	Const.			−0.087			−0.089
	Slope			0.588**			0.734*

Notes: See notes to Table 3.

Table 6: GW Conditional Test: SPF Downturn Probability

		75Q3−10Q3			84Q1−10Q3		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
(a) 2-Step-Ahead Forecast							
AO	P-Value	0.151	0.355	0.188	0.249	0.082*	0.174
	R^2	0.041	0.017	0.050	0.006	−0.002	−0.009
	Const.	−0.143*	−0.140*	−0.197*	−0.070*	−0.057*	−0.074*
	Slope	0.009**	0.007	0.011*	0.002*	−0.002	0.000
IMA	P-Value		0.615	0.741		0.243	0.718
	R^2		0.001	−0.004		0.023	−0.004
	Const.		0.003	−0.054		0.013	−0.004
	Slope		−0.002	0.002		−0.004**	−0.002
PC	P-Value			0.279			0.455
	R^2			0.020			0.023
	Const.			−0.057			−0.017
	Slope			0.004*			0.002*
(b) 4-Step-Ahead Forecast							
AO	P-Value	0.269	0.465	0.382	0.283	0.465	0.207
	R^2	0.042	0.043	0.030	−0.007	−0.005	−0.006
	Const.	−0.126	−0.112	−0.183	−0.056*	−0.082	−0.092
	Slope	0.011*	0.011	0.011	0.001	0.002	−0.002
IMA	P-Value		0.983	0.746		0.875	0.448
	R^2		−0.007	−0.007		−0.008	−0.001
	Const.		0.014	−0.058		−0.026	−0.036
	Slope		−0.001	−0.001		0.001	−0.003
PC	P-Value			0.512			0.357
	R^2			−0.007			0.023
	Const.			−0.072			−0.010
	Slope			0.000			−0.003
(c) 6-Step-Ahead Forecast							
AO	P-Value	0.349	0.644	0.566	0.309	0.129*	0.520
	R^2	0.046	0.018	0.006	−0.002	0.020	−0.008
	Const.	−0.123	−0.145	−0.173	−0.049*	−0.103	−0.097
	Slope	0.012*	0.009	0.007	0.001	0.004	−0.001
IMA	P-Value		0.654	0.518		0.083*	0.612
	R^2		−0.004	−0.002		0.005	−0.004
	Const.		−0.022	−0.050		−0.054	−0.049
	Slope		−0.003	−0.005		0.003*	−0.003
PC	P-Value			0.858			0.342
	R^2			−0.006			0.037
	Const.			−0.028			0.006
	Slope			−0.002			−0.006
(d) 8-Step-Ahead Forecast							
AO	P-Value	0.437	0.401	0.262	0.226	0.051*	0.038**
	R^2	0.049	0.020	0.013	0.046	0.043	0.028
	Const.	−0.138	−0.209	−0.245	−0.065*	−0.237**	−0.294*
	Slope	0.014*	0.011	0.012*	0.006*	0.010*	0.013**
IMA	P-Value		0.211	0.637		0.192	0.063*
	R^2		−0.004	−0.007		−0.002	0.002
	Const.		−0.071	−0.107		−0.172*	−0.229*
	Slope		−0.004	−0.003		0.003	0.007*
PC	P-Value			0.964			0.558
	R^2			−0.007			−0.006
	Const.			−0.036			−0.057
	Slope			0.001			0.003

Notes: See notes to Table 3.

Table 7: GW Conditional Test: Real-Time Gap

		75Q3−10Q3			84Q1−10Q3		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
(a) 2-Step-Ahead Forecast							
AO	P-Value	0.560	0.229	0.318	0.538	0.069*	0.064*
	R^2	0.017	0.018	0.023	0.004	0.004	0.029
	Const.	0.009	−0.035	−0.012	−0.046	−0.073**	−0.048
	Slope	0.121	0.126	0.150	−0.039	0.044	0.083**
IMA	P-Value		0.561	0.767		0.123*	0.096*
	R^2		−0.007	−0.005		0.031	0.070
	Const.		−0.044	−0.022		−0.028	−0.002
	Slope		0.005	0.029		0.083	0.123*
PC	P-Value			0.867			0.568
	R^2			−0.005			0.018
	Const.			0.022			0.026
	Slope			0.023			0.039
(b) 4-Step-Ahead Forecast							
AO	P-Value	0.483	0.306	0.383	0.341	0.134*	0.130*
	R^2	0.047	0.079	0.040	−0.001	0.057	0.026
	Const.	0.069	0.068	−0.006	−0.050	−0.027	−0.092
	Slope	0.243	0.287*	0.241	−0.031	0.128*	0.116
IMA	P-Value		0.783	0.737		0.162	0.122*
	R^2		−0.003	−0.007		0.106	0.053
	Const.		−0.001	−0.075		0.024	−0.042
	Slope		0.044	−0.002		0.158**	0.146**
PC	P-Value			0.506			0.292
	R^2			−0.001			−0.009
	Const.			−0.074			−0.066*
	Slope			−0.047			−0.012
(c) 6-Step-Ahead Forecast							
AO	P-Value	0.356	0.120*	0.310	0.349	0.112*	0.082*
	R^2	0.058	0.039	0.031	−0.007	0.220	0.121
	Const.	0.094	0.018	−0.047	−0.025	0.034	−0.050
	Slope	0.284*	0.266*	0.264	0.017	0.272**	0.295**
IMA	P-Value		0.604	0.548		0.188	0.092*
	R^2		−0.007	−0.007		0.189	0.117
	Const.		−0.076	−0.141		0.059	−0.025
	Slope		−0.018	−0.020		0.255**	0.278**
PC	P-Value			0.843			0.385
	R^2			−0.007			−0.008
	Const.			−0.065			−0.084
	Slope			−0.002			0.023
(d) 8-Step-Ahead Forecast							
AO	P-Value	0.294	0.078*	0.362	0.291	0.065*	0.076*
	R^2	0.055	0.034	0.021	0.066	0.067	0.107
	Const.	0.107	−0.026	−0.047	0.061	−0.038	0.002
	Slope	0.295*	0.261*	0.271	0.139*	0.224*	0.437**
IMA	P-Value		0.235	0.650		0.129*	0.086*
	R^2		−0.007	−0.007		0.003	0.054
	Const.		−0.133*	−0.154		−0.099	−0.059
	Slope		−0.034	−0.024		0.086	0.298**
PC	P-Value			0.976			0.329
	R^2			−0.008			0.028
	Const.			−0.021			0.040
	Slope			0.010			0.212*

Notes: See notes to Table 3.

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

		75Q3−10Q3			84Q1−10Q3		
		IMA	PC	PC-TAR	IMA	PC	PC-TAR
(a) 2-Step-Ahead Forecast							
AO	P-Value	0.239	0.278	0.097*	0.073*	0.069*	0.022**
	R^2	0.071	0.028	0.099	0.019	−0.006	0.026
	Const.	0.004	−0.032	−0.021	−0.037	−0.083**	−0.066*
	Slope	0.951*	0.649	1.233*	0.224**	0.085	0.316**
IMA	P-Value		0.505	0.309		0.348	0.678
	R^2		0.011	0.004		−0.002	−0.007
	Const.		−0.036	−0.024		−0.047	−0.030
	Slope		−0.302	0.282		−0.139	0.091
PC	P-Value			0.042**			0.222
	R^2			0.062			0.052
	Const.			0.011			0.017
	Slope			0.584**			0.230**
(b) 4-Step-Ahead Forecast							
AO	P-Value	0.398	0.345	0.266	0.283	0.042**	0.083*
	R^2	0.072	0.082	0.088	−0.008	0.024	−0.002
	Const.	0.071	0.074	−0.007	−0.043	−0.055	−0.119*
	Slope	1.248*	1.244*	1.447*	0.052	0.354*	0.207
IMA	P-Value		0.999	0.695		0.084*	0.285
	R^2		−0.007	−0.005		0.018	−0.005
	Const.		0.003	−0.079		−0.012	−0.076
	Slope		−0.004	0.200		0.302**	0.155
PC	P-Value			0.414			0.360
	R^2			−0.001			0.000
	Const.			−0.082			−0.064
	Slope			0.204			−0.147
(c) 6-Step-Ahead Forecast							
AO	P-Value	0.379	0.239	0.300	0.072*	0.049**	0.075*
	R^2	0.078	0.053	0.075	0.009	0.251	0.062
	Const.	0.104	0.027	−0.040	−0.026	−0.017	−0.110
	Slope	1.427*	1.338**	1.700*	0.183**	1.199**	0.907
IMA	P-Value		0.603	0.516		0.099*	0.136*
	R^2		−0.007	−0.005		0.173	0.040
	Const.		−0.077	−0.143		0.009	−0.084
	Slope		−0.090	0.273		1.016**	0.724
PC	P-Value			0.643			0.387
	R^2			−0.002			0.004
	Const.			−0.067			−0.093
	Slope			0.363			−0.292
(d) 8-Step-Ahead Forecast							
AO	P-Value	0.386	0.139*	0.165	0.129*	0.087*	0.147*
	R^2	0.068	0.105	0.174	0.078	0.220	0.364
	Const.	0.126	−0.006	−0.022	0.041	−0.062	−0.042
	Slope	1.491*	1.967**	3.125**	0.672**	1.744**	3.501**
IMA	P-Value		0.196	0.335		0.139*	0.190
	R^2		0.000	0.034		0.084	0.273
	Const.		−0.132*	−0.148		−0.102	−0.082
	Slope		0.476	1.634*		1.072*	2.829**
PC	P-Value			0.281			0.279
	R^2			0.025			0.117
	Const.			−0.016			0.020
	Slope			1.158**			1.757**

Notes: See notes to Table 3.

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

Reference Model	Alternative Model	Forecast Horizon	Sample	Squared Error Diff. $D = 0$	Squared Error Diff. $D = 1$
AO	PC-TAR	2	Post 84	-0.087	0.115
AO	PC	8	Post 84	-0.155	0.442
AO	PC-TAR	8	Post 84	-0.244	1.087
IMA	PC-TAR	8	Post 84	-0.210	0.548

Notes: This table considers the cases in which both constant and slope terms are statistically significant in the GW regression (see Table 5) within the comparisons between each of the two Phillips curve models and either 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 Recession Probability

Reference Model	Alternative Model	Forecast Horizon	Sample	Cutoff Values (%)
AO	PC-TAR	2	Full	18.0
IMA	PC	6	Post 84	18.0
AO	PC-TAR	8	Full	20.4
AO	PC	8	Post 84	23.7
AO	PC-TAR	8	Post 84	22.6
IMA	PC-TAR	8	Post 84	32.7

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 AO or IMA models.

Table 11: Cutoff Values of Real-Time Gap

Reference Model	Alternative Model	Forecast Horizon	Sample	Cutoff Value
AO	PC-TAR	2	Post 84	0.578
IMA	PC-TAR	2	Post 84	0.016
AO	PC	4	Full	−0.237
AO	PC	4	Post 84	0.211
IMA	PC	4	Post 84	−0.152
IMA	PC-TAR	4	Post 84	0.288
AO	PC	6	Full	−0.068
AO	PC	6	Post 84	−0.125
AO	PC-TAR	6	Post 84	0.169
IMA	PC	6	Post 84	−0.231
IMA	PC-TAR	6	Post 84	0.090
AO	PC	8	Full	0.100
AO	PC	8	Post 84	0.170
AO	PC-TAR	8	Post 84	−0.005
IMA	PC-TAR	8	Post 84	0.198

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 the dark and middle shadings in Table 7 within the comparisons between each of the two Phillips curve models and either AO or IMA models. The cutoff values are in the units of the unemployment gap (expressed in percentage points).

Table 12: Cutoff Values of the Four-Quarter Change in Real-Time Gap

Reference Model	Alternative Model	Forecast Horizon	Sample	Cut-off Value
AO	PC-TAR	2	Full	0.017
AO	PC-TAR	2	Post 84	0.209
AO	PC	4	Full	−0.059
AO	PC-TAR	4	Full	0.005
AO	PC	4	Post 84	0.155
IMA	PC	4	Post 84	0.040
AO	PC	6	Full	−0.020
AO	PC-TAR	6	Full	0.024
AO	PC	6	Post 84	0.014
IMA	PC	6	Post 84	−0.009
AO	PC	8	Full	0.003
AO	PC-TAR	8	Full	0.007
AO	PC	8	Post 84	0.036
AO	PC-TAR	8	Post 84	0.012
IMA	PC-TAR	8	Full	0.091
IMA	PC	8	Post 84	0.095
IMA	PC-TAR	8	Post 84	0.029

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

Forecast horizon	75Q3−10Q3 Real Time	75Q3−10Q3 Final	84Q1−10Q3 Real-Time	84Q1−10Q3 Final
	(a) Mean Absolute Errors			
2	0.552	0.552	0.466	0.450
4	0.506	0.504	0.411	0.403
6	0.573	0.547	0.434	0.433
8	0.659	0.638	0.516	0.520
	(b) Root-Mean-Square Errors			
2	0.727	0.737	0.612	0.611
4	0.690	0.696	0.547	0.552
6	0.806	0.767	0.557	0.573
8	0.917	0.906	0.684	0.705

Table 14: GW Test Results Using Final Revised Gap: Phillips Curve vs. AO and IMA

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	0.739	0.046**	0.728	0.133*	0.449	0.135*	0.266	0.042**	0.147*	0.079*
	R^2	0.000	0.000	−0.004	−0.008	0.007	−0.006	0.019	0.031	0.043	−0.003
	Const.	−0.031	−0.082**	−0.054	−0.086*	−0.131	−0.061*	−0.039	−0.072*	−0.048	−0.082**
	Slope	–	–	0.179	0.045	0.005	−0.001	0.137	0.077*	0.816*	0.125
IMA	P-Value	0.233	0.300	0.491	0.221	0.273	0.255	0.277	0.065*	0.492	0.463
	R^2	0.000	0.000	−0.005	−0.002	0.008	0.014	−0.001	0.057	−0.006	−0.006
	Const.	−0.058	−0.045	−0.046	−0.032	0.012	0.009	−0.060	−0.032	−0.056	−0.045
	Slope	–	–	−0.090	−0.124	−0.004	−0.003**	0.043	0.103*	−0.082	−0.094
(b) 4-Step-Ahead Forecast											
AO	P-Value	0.424	0.274	0.704	0.527	0.380	0.542	0.381	0.169	0.374	0.054*
	R^2	0.000	0.000	−0.007	−0.009	0.044	−0.008	0.041	0.047	0.074	0.018
	Const.	0.087	−0.063	0.080	−0.065	−0.111	−0.081	0.084	−0.047	0.072	−0.061
	Slope	–	–	0.055	0.021	0.010*	0.001	0.203*	0.117*	1.099*	0.352*
IMA	P-Value	0.926	0.708	0.475	0.626	0.831	0.932	0.935	0.129*	0.909	0.119*
	R^2	0.000	0.000	−0.002	−0.007	−0.006	−0.009	−0.006	0.078	−0.007	0.015
	Const.	−0.006	−0.020	0.017	−0.012	0.015	−0.024	−0.006	−0.002	−0.005	−0.019
	Slope	–	–	−0.170	−0.072	−0.001	0.000	0.022	0.133**	−0.061	0.300**
(c) 6-Step-Ahead Forecast											
AO	P-Value	0.437	0.541	0.716	0.617	0.485	0.223	0.221	0.081*	0.270	0.039**
	R^2	0.000	0.000	−0.001	−0.003	0.064	0.015	0.049	0.116	0.070	0.146
	Const.	0.094	−0.049	0.063	−0.065	−0.140	−0.120	0.100	−0.025	0.093	−0.040
	Slope	–	–	0.247	0.142	0.012	0.004*	0.231**	0.185**	1.172**	0.966**
IMA	P-Value	0.830	0.763	0.861	0.934	0.977	0.240	0.974	0.114*	0.893	0.077*
	R^2	0.000	0.000	−0.006	−0.009	−0.007	0.003	−0.007	0.082	−0.006	0.093
	Const.	−0.017	−0.020	−0.004	−0.025	−0.017	−0.071	−0.017	0.000	−0.017	−0.013
	Slope	–	–	−0.099	0.037	0.000	0.003*	−0.001	0.158**	−0.177	0.781**
(d) 8-Step-Ahead Forecast											
AO	P-Value	0.962	0.293	0.605	0.355	0.488	0.110*	0.606	0.265	0.239	0.102*
	R^2	0.000	0.000	0.015	0.000	0.025	0.035	0.010	0.006	0.079	0.095
	Const.	0.006	−0.117	−0.058	−0.145	−0.172	−0.266*	0.013	−0.105	0.019	−0.101
	Slope	–	–	0.567	0.249	0.010	0.010*	0.145	0.097	1.457**	1.235**
IMA	P-Value	0.358	0.132*	0.654	0.281	0.492	0.245	0.511	0.241	0.638	0.209
	R^2	0.000	0.000	−0.007	0.007	−0.003	−0.003	0.000	−0.009	−0.008	0.016
	Const.	−0.114	−0.147*	−0.121	−0.111	−0.034	−0.201*	−0.119	−0.148	−0.114	−0.139
	Slope	–	–	0.064	−0.323	−0.004	0.003	−0.111	−0.011	0.016	0.586