7-2012

Do Forecasters Know the Slope of the Phillips Curve?

David Thomas Fisher

College of William and Mary

Follow this and additional works at: https://scholarworks.wm.edu/honorstheses

Recommended Citation

https://scholarworks.wm.edu/honorstheses/530

This Honors Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
Do Forecasters Know the Slope of the Phillips Curve?

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Arts in Economics from The College of William and Mary

by

David Thomas Fisher

Accepted for Honors

/s/ Olivier Coibion
Olivier Coibion, Director

/s/ Till Schreiber
Till Schreiber

/s/ John F. Boschen
John Boschen

Williamsburg, VA
April 20, 2012
Abstract: Despite the fact that professional forecasters have significant resources to observe the economy and that the Phillips Curve has been observed and recognized in nearly all literature and economic models, the Survey of Professional Forecasters has shown a consistent under-recognition of the Phillips Curve’s effects. This effect is observed across forecast horizons and in many other inflation forecasts, including the Livingston survey, Michigan Survey of Consumers, and the Federal Reserve Board’s Greenbook forecast. By examining changes in the mis-estimation of the slope of the Phillips Curve over time, we can observe changes in macroeconomic thought, as well as in the case of Greenbook forecasts explain some policy decisions of the Fed.

Table of Contents

1. Introduction .......................................................... 2
2. Are SPF Forecasters Underestimating the Slope of the Phillips Curve .......... 5
   Results ........................................................................... 8
   An Aside: Extension to Test Rationality .................................. 10
3. Variance over the Forecasting Horizon .......................................... 11
4. Other Forecasters and Other Baselines ............................................. 13
   SPF—Consumer Price Index ................................................. 13
   Livingston—CPI .............................................................. 16
   Livingston—GDP ............................................................. 20
5. Consumer Surveys ................................................................ 24
   Michigan Survey versus CPI ............................................... 25
   Michigan Survey versus Personal Consumption Expenditure ......... 26
6. Federal Reserve Greenbook Surveys .................................................. 29
7. Summary of Regressions to Date ................................................. 34
   Examining Overall Accuracy of Forecasts (via MSE) ....................... 36
8. Examining Rolling Samples .......................................................... 37
10. Conclusion ........................................................................... 46
11. Bibliography ........................................................................... 48

1. Introduction

Inflation forecasting is one of the core activities of all economic agents.

Ultimately, it stands as a fundamental consideration in intertemporal decision-making.

Ang et al (2007) ascribe inflation forecasting’s importance in three key areas, (1.) in
analyzing and making policy decisions, (2.) for investors’ ability to hedge, and (3.) in labor contracts, but these examples just touch the surface of the economic decisions for which actors must base decisions off predictions of inflation levels at varying points in the future. As an abstract concept, it is measured in innumerable forms: the Consumer Price Index and components, the Gross Domestic Product deflator and chain-type index and variants for e.g. GNP and sub-components, Personal Consumption Expenditure price index and derivatives, Producer Price Index, and many others. They consistently and imperfectly measure the core ideal, that inflation is a change in the price level.

Since Irving Fisher’s realized a link in 1926 for US data and A. W. Phillips published in 1958 his hypothesis on the relation between unemployment and inflation in the UK, the Phillips Curve has developed into what a former Vice President of the US Federal Reserve’s Board of Governors called “the ‘clean little secret’ of empirical macroeconomics” (Stock and Watson 1999b). The relationship outlined in the Phillips Curve can take many different forms. It is typically shown by relating inflation to real economic activity, iconically and most commonly in the form of the unemployment rate’s variation from NAIRU (the Non-Accelerating Inflation Rate of Unemployment), which in most specifications is allowed to—and does—change over time (Stock and Watson 1999a).

The series used to represent real economic activity in various Phillips Curve models has come from many sources. The Federal Reserve in its modeling of the economy systemically estimate a direct measure of the output gap, actually three estimates of it, as part of a large dynamic-stochastic-general-equilibrium model (Kiley 2010). Many others have used other proxies for real economic activity. Extending this
layout for the model, Stock and Watson (1999b) show that housing starts, capacity utilization, manufacturing growth, and trades sales, for example, each provide significant forecasting accuracy.

Further Stock and Watson (1999b) establish that inflation forecasters should take more into account than just unemployment. They find that de novo a wide breadth of economic variables (specifically those summarizing aggregate activity) produces the best possible predictive model of inflation over various human forecasters. Yet the model contrived by Stock and Watson (1999b) produces noticeably subpar accuracy if it fails to include a Phillips Curve specification as the authors discuss. These results are slightly at odds with Ang et al (2007) who find that both a time series ARIMA-based model they created as well as a financial-market sensitive model (optimized econometrically from various market variables) are outperformed by the Survey of Professional Forecasters (SPF has a lower error).

Mankiw, Reiw, & Wolfers (2003) examined rationality for many of the inflation series we examine later in this paper before delving into a pricing expectations discussion. They found evidence of bias in two series they examined: a significant, yet minimal constant error was found in the less sophisticated forecasting series. Further, when examining autocorrelation of inflation, they found rationality was violated for two series. Specifically, they found that the Michigan and Livingston surveys could have derived information from previous periods’ inflation-forecast errors in order to make current forecasts more accurate.

As to the importance of these surveys, the Federal Reserve obviously gets some benefit from analyzing the various forecasts submitted to it in their public policy roll,
even their forecasting role, before they release them to the public and the markets at large. Most importantly, Carroll (2003) establishes a dissemination pathway for professional forecasts to reach and be absorbed by households (via an expectational mechanism), thus impacting the expectations of the consumers who engage in economic activity and are samples to respond to the Michigan Survey of Consumers. Yet Coibion and Gorodnichenko (2011) have found evidence to dispute these precise results.

2. Are SPF Forecasters Underestimating the Slope of the Phillips Curve

The Survey of Professional Forecasters (hereafter SPF or Survey) was founded by the American Statistical Association and the National Bureau of Economic Research in 1968 and is now conducted quarterly by the Federal Reserve Bank of Philadelphia. The SPF draws from senior, private sector economists—professional forecasters—who “produce regular forecasts of economic variables as part of their jobs in the business world or on Wall Street.” Most respondents are from Wall Street financial firms, but the survey also includes significant contingents from banks, academia, consulting firms, and chief economists at Fortune 500 companies. (Croushore 1993) Currently, the SPF has about 50 respondents from various sectors of the economy, mostly working in various parts of the financial sector.

The respondents to the SPF are those who develop economic forecasts as a primary component of their job description. The depth of the survey and the sheer number of forecasts required to complete the mailed survey (i.e. the wide range of economic variables requested by FRB Philadelphia) mean that respondents are both serious and educated. As such, relative to other surveys available it is viewed as the most accurate. (Croushore 1998) It represents a very mature set of respondents who are likely
to make some of the most informed predictions available. Indeed Ang et al (2007) found that the Survey of Professional Forecasters outperforms very slightly the Livingston survey, and to a notable extent, consumer’s expectations in the Michigan Survey. They further find the SPF consistently outperforms the aforementioned (in Section 1) other time-series and financial-market models in predicting components of CPI, though not Personal Consumption Expenditure (PCE) price index inflation (which is best forecasted by a methodology explained in Ang et al (2007) but is far beyond the scope of this analysis).

Beyond the credibility of the survey respondents, the SPF is available on a consistent quarterly basis. Further, each observation corresponds to a well-defined and consistently measured financial market variable, which we will assume prima facie that it is uniquely predicting. In this case, we will examine the forecasters’ median prediction of the GDP-deflator based inflation rate from each period. The series has in fact gone through several incarnations that are accounted for in the development of the data set but will be ignored beyond this paragraph for notational simplicity. The series started with GNP inflation, beginning in 1992 Survey participants were asked to forecast GDP inflation, and beginning in 1996, the GDP chain-type price index. GDP deflator data, unlike that for CPI, was collected in the earliest samples of the survey. For this very reason, Croushore (1998) choose it for his primary analysis, as one, early example, as we will follow here.

Having sufficiently extolled the Survey respondents for their eminent, educated position in the market, we clearly expect them to be rational actors. Mankiw et al (2003) establish a long list of rationality criteria for forecast data as mentioned previously in
Section 1. First, that it should be noted that the SPF GDP inflation data do not exhibit a mean error different from 0 (where $\pi_t - E_{t-12}\pi_t = a$ and $a = 0$ is not rejected). Second, no additional information is available in the forecast itself to predict inflation (where $\pi_t - E_{t-12}\pi_t = a + bE_{t-12}\pi_t$, and $a = b = 0$ is not rejected). Yet when looking at the rationality of expectations, we should see actors/forecasters adjust their forecasts based on past performance (the existence and magnitude of past error). Their third criterion requires that no information be contained in previously observed ex-post inflation error that can aid forecast accuracy (that $\pi_t - E_{t-12}\pi_t = a + b\pi_{t-12} - E_{t-24}\pi_{t-12}$ and $a = b = 0$). In this case, they found there was a significant deviation from the null hypothesis in $b = 0.640$. Expanding to see where this variation came from, they conclude that macroeconomic variables have additional predictive power beyond what SPF forecasts already include. They find

$$\pi_t - E_{t-12}\pi_t = a + b(E_{t-12}\pi_t) + c(E_{t-13}) + d(TBILL_{t-13}) + e(un_{t-13})$$

leaves $c = d = e = 0$ rejected, suggesting that both the Treasury Bill rate of return and unemployment at the time the forecast is made both contain information that SPF forecasters both could and should have used to create a more accurate forecast.

It is not hard to move from one of these tests and develop a model to judge whether the Phillips Curve contains more (or less) information than the Survey forecasters are giving it credit for, whether they are properly estimating the Phillips Curve’s slope. When examining from the basis of $\pi_t - E_{t-12}\pi_t = 0$, if forecasters are properly estimating the slope of the Phillips Curve, we expect to see $b = 0$ when $\pi_t - E_{t-12}\pi_t = a + b(un_{t-12})$. This rational expectations model forms the backbone of the analysis we conduct throughout this paper.
Because of the timing of the SPF collection (slightly over a month before the end of the quarter), we will estimate the equation $\pi_t - E_{t-13}\pi_t = a + b(un_{t-13})$, which represents the unemployment information and economic reality roughly available to forecasters at the time they made their predictions. Adjusting $un_{t-13}$ to $un_{t-12}$ (in order to reflect the unemployment situation at the end of the quarter in which forecasts are made) does not in fact change the outcome of tests throughout this paper in any appreciable manner. However, it does compare forecasters more on par with the information they had at the time they made their predictions, excluding any information contained in any potential unexpected change in unemployment between periods $(t-12)$ and $(t-13)$.

The coefficient $b$ (for $un_{t-13}$) allows us to fully determine whether the Phillips Curve can provide information from which we can accurately predict inflation above and beyond the Survey forecasters’ estimates as presently constituted. The alternate hypothesis, that $b \neq 0$, suggests that forecasters’ are improperly estimating the slope of the Phillips Curve: i.e. that a stronger or weaker relationship between unemployment and inflation than presently hypothesized by forecasters would better reflect observed reality.

**Results**

In fact we find:

$$\pi_t - E_{t-13}\pi_t = 1.752*** - 0.256*** (un_{t-13}).$$

$$(0.566) \quad (0.079)$$

For notational purposes throughout the remainder of this paper, the author designates $*** = $ significance at a 1% level of confidence, $** = $ at 5%, and $* = $ at 10%. For this and all subsequent analyses, Newey-West standard errors are used with lag lengths of up to 4 periods allowed.
The relationship between ex-post inflation error and the unemployment rate is clearly shown in Figure 1 (below) as well. We are able to reject the null $b = 0$ at the 1% significance level. For analysis of SPF GDP inflation, we use here the maximal sample for which both ex-ante data is available from the Philadelphia Federal Reserve on SPF forecasts (2011) and ex-post actual series are available from the St. Louis Federal Reserve Bank’s FRED database (2012). In this case, the sample begins in 1968 quarter 4 (documented hereafter as 1968q4) i.e. a prediction made in the middle of 1968q4 (represented $t-13$) about inflation in the period $\pi_{t-12,t}$ starting at the end of 1968q4 ($t-12$) and ending 1969q4 (in this case $t$).

Figure 1. SPF GDP Deflator Ex-Post Error versus Unemployment at 1 Year.

In Figure 1 above, we clearly observe the negative coefficient seen in the regression on the unemployment series and the observed negative relationship in the data.
An Aside: Extension to Test Rationality

If we take into account the expansions of Mankiw et al (2003), we see this result holds up as we adjust the definition of rationality they use. To see if last year’s forecast has any information,

\[
\pi_t - E_{t-13}\pi_t = 1.415** - 0.313*** (un_{t-13}) + 0.185* E_{t-13}\pi_t
\]

shows us fundamentally the same result, that a stronger Phillips Curve relationship should be taken into account. To see if lagged error reaches 0,

\[
\pi_t - E_{t-13}\pi_t = 1.245** - 0.193**(un_{t-13}) + 0.526*** (\pi_{t-12} - E_{t-25}\pi_{t-12})
\]

autocorrelation shows that information about last year’s forecast error isn’t being taken into account in generating this year’s forecast. Unfortunately, once we take into account the additional power of the Phillips Curve in these rationality exercises, we lose some of the original power Mankiw et al noticed in the SPF data. In the first equation, by adding a Phillips Curve adjustment and scaling back the forecaster’s initial projection to \((1 - c)\), we see improved prediction power. The second is more powerful, that assuming forecasters make adjustments for the Phillips Curve, there is still a significant amount of autocorrelated error—implying forecasters aren’t taking last year’s forecasting errors into account beyond what would have been a result of the error caused by not including the error caused by their improper estimation of the Phillips Curve.

In order to establish the generality of the original regression, we should exclude the peak of the Great Inflation and the Volcker Disinflation. As such, we exclude predictions about 1973q1-1974q4 and 1978q1-1984q4 (thus predictions made the year before). We see over the remaining sample that the relationship still significantly holds.
Specifically, that \[ \pi_t - E_{t-13}\pi_{t-13} = 1.618*** - 0.253*** (un_{t-13}) \]

\[
(0.575) \quad (0.083)
\]

At this point, we have recognized in the updated data set a problem that is identified in passing in Mankiw et al (2003). In summary, within the Survey of Professional Forecasters history, we see a vast systematic relation of unemployment to forecast errors indicating an underestimation of the strength of the Phillips Curve. This statistically significant, systemic bias among professional forecasters also has economic significance with a 1-point change in inflation forecast for an approximately 4-point change in unemployment. Arising from this, we will first discuss how far this observation holds and then discuss possible sources and implications.

3. Variance over the Forecasting Horizon

Again expanding on our previous results while remaining within the Survey of Professional Forecasters, we will now investigate whether the Survey forecasters still underestimate the Phillips Curve’s slope over varying forecast time horizons, other than the previously analyzed 1-year inflation period, \( E_{t-13}\pi_{t-12,t} \). Specifically, we will look at the inflation rates for the present quarter (from the previous one, price level from periods t-15 to t-12), annualized quarter-to-quarter forward inflation rates, and the year-long period overlapping all four future quarters that we have already examined in Section 2:

\[
E_{t-13}\pi_{t-15,t-12}, \quad E_{t-13}\pi_{t-12,t-9}, \quad E_{t-13}\pi_{t-9,t-6}, \quad E_{t-13}\pi_{t-6,t-3}, \quad \text{and} \quad E_{t-13}\pi_{t-3,t}
\]

(put in a different basis that will not be referenced or presented further in the paper: \( E_{t-13}\pi_{t-3,t}, \quad E_{t-13}\pi_{t-3,t}, \quad E_{t-13}\pi_{t-3,t}, \quad \text{and} \quad E_{t-13}\pi_{t-3,t} \).
In order to accomplish this, we estimate \( \pi_{t-(i+3) t-i} - E_{t-13} \pi_{t-(i+3) t-i} = a + b(un_{t-13}) \)

for \( i = 0, 3, 6, 9 \) in addition to the model examined in the previous section. Table 1 and Figure 2 summarize the results we see over the entire available sample, 1968q4-2011q4.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>( a ) (c)</th>
<th>( b ) (un)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-15,t-12</td>
<td>0.958**</td>
<td>-0.139**</td>
</tr>
<tr>
<td></td>
<td>(0.377)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>t-12,t-9</td>
<td>1.823***</td>
<td>-0.269***</td>
</tr>
<tr>
<td></td>
<td>(0.645)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>t-9,t-6</td>
<td>2.194***</td>
<td>-0.334***</td>
</tr>
<tr>
<td></td>
<td>(0.804)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>t-6,t-3</td>
<td>2.484***</td>
<td>-0.381***</td>
</tr>
<tr>
<td></td>
<td>(0.912)</td>
<td>(0.132)</td>
</tr>
<tr>
<td>t-3,t</td>
<td>2.133***</td>
<td>-0.321***</td>
</tr>
<tr>
<td></td>
<td>(0.756)</td>
<td>(0.107)</td>
</tr>
<tr>
<td>t-12,t</td>
<td>1.752***</td>
<td>-0.256***</td>
</tr>
<tr>
<td></td>
<td>(0.566)</td>
<td>(0.079)</td>
</tr>
</tbody>
</table>

Figure 2. SPF GDP Inflation over All Time Horizons
From these results, it is clear that the Survey forecasters systemically underestimate the Phillips Curve over all forecasting time horizons in the sample. Also, the coefficients as the time horizon extends move in a more extreme direction generally as the time horizon becomes further away. Standard error of the predictor also is at its maximum when concerning inflation during the period 3 quarters away from the forecast. The specifics of third quarter estimation suggest it is a difficult distance over which to estimate the effect of the Phillips Curve (the error of which is reflected in the coefficient’s existence) likely with difficult arising from other effects as well (from the standard error). And as we are dealing with “forward” inflation rates, the prediction for the 3-quarter inflation does not consider the earlier quarters and/or the certainty of the predictions made about them: it relies solely on forecasters’ analysis of the change in price level from the end of quarter 2 to the end of quarter 3 independent of all else.

4. Other Forecasters and Other Baselines

*SPF—Consumer Price Index*

More broadly, the analyses discussed so far are based on GDP price index inflation. In general these figures represent a more holistic view of prices throughout the entire economy generally. Not only do they reflect consumer goods in the proportion they are consumed, the GDP deflator method also contains a significant amount of information about the prices of investment goods, producer (pre-customer) exchanges, and government activity, all in the proportion they make up of GDP. On the other hand, the most popular/visible measure of inflation is the Consumer Price Index (CPI) calculated monthly by the Bureau of Labor Statistics. While the Survey of Professional Forecasters first used GNP calculated inflation, it started collecting projections for CPI
indexed inflation in 1981 as the survey developed. To expand on the analysis done
previously we should see whether the same effect is seen in SPF’s CPI forecasting errors
as was seen in the GDP inflation series.

Over the same forecasting horizon, we will examine the same quarterly and
forward rates. However, we are now limited to the data for which there are SPF CPI
inflation forecasts, the timeframe 1981q3-2011q4. Again, we will estimate the regression
\[ \pi_{t-(i+3),t-i} - E_{t-i}(\pi_{t-(i+3),t-i}) = a + b(\text{un}_{t-i}) \text{ for } i = 0, 3, 6, 9 \]
over median forecasts from each period. In the regressions, we will also be using Newey-West standard errors. Notice
again in the data (Figure 3) that there is an evident negative relationship, one we will
elucidate in the regressions summarized in Table 2.

Figure 3. SPF CPI Inflation Forecast Error versus Unemployment at 1 Year.
Table 2. SPF CPI Inflation over All Time Horizons (and compared to GDP over consistent time sample—1981q3-2011q4 or latest).

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Panel A. SPF CPI Inflation</th>
<th>Panel B. GDP Inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$ (c)</td>
<td>$b$ (un)</td>
</tr>
<tr>
<td>t-15,t-12</td>
<td>0.571</td>
<td>-0.103</td>
</tr>
<tr>
<td></td>
<td>(0.633)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>t-12,t-9</td>
<td>0.877</td>
<td>-0.169</td>
</tr>
<tr>
<td></td>
<td>(0.878)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>t-9,t-6</td>
<td>0.749</td>
<td>-0.163</td>
</tr>
<tr>
<td></td>
<td>(0.968)</td>
<td>(0.156)</td>
</tr>
<tr>
<td>t-6,t-3</td>
<td>0.721</td>
<td>-0.172</td>
</tr>
<tr>
<td></td>
<td>(1.046)</td>
<td>(0.169)</td>
</tr>
<tr>
<td>t-3,t</td>
<td>0.933</td>
<td>-0.227</td>
</tr>
<tr>
<td></td>
<td>(1.086)</td>
<td>(0.178)</td>
</tr>
<tr>
<td>t-12,t</td>
<td>0.874</td>
<td>-0.185</td>
</tr>
<tr>
<td></td>
<td>(0.879)</td>
<td>(0.146)</td>
</tr>
</tbody>
</table>

Figure 4. SPF CPI Inflation over All Time Horizons

While the CPI set doesn’t have the same level of significance as the SPF GDP inflation series, it should be made very clear that it has similar (even slightly larger) magnitude of coefficients. It is the standard errors of the $b$ coefficient that are nearly
double the SPF GDP series. While we cannot say with any certainty that there is a significant effect here still, the lack of significance coming from added standard error does not in any way dismiss our conclusion that SPF forecasters still underestimated the slope of the Phillips Curve. For example, a source of variance that would be especially prominent in CPI data moreso than GDP inflation would be fuel prices, which (1.) make up a large portion of CPI and (2.) are notoriously volatile (and consequently hard to predict).

We should note, that this analysis in Table 2 is from a notably shorter sample, the time effect could have paid a role. Specifically with regards to the sample, we have conducted an analysis over the largest possible common time period available over all data sets we will introduce throughout. In this scenario (see Table 9), the Survey of Professional Forecasters’ forecast of CPI was highly significant (always surpassing 5% significance level). For reference, the only difference in the original sample and the revised one is that the new set cuts off the last five years, 2006-2011. In Table 2 Panel B when we apply the time restrictions on the maximal sample of the SPF CPI series onto the SPF GDP inflation series, we see that the significance level of the SPF GDP series decreases, but with the same characterization that the coefficients have not notably changed (albeit slightly decreased), merely standard errors have ballooned. As such, we suggest that the time sample does have an effect on the regression and its statistical significance (a theory which we will fully explicate in Section 8).

Livingston—CPI

The Livingston survey was started in 1946 by newspaper columnist Joseph Livingston. It began as an outgrowth of his biannual newspaper column for which he would ask his business friends’ opinions on the state of the economy. The database
became popular for researchers, so the Philadelphia Federal Reserve took on the responsibility to shepherd and fund the project, entirely so after Livingston’s 1989 death. Its characteristic semiannual structure makes it unique among the surveys analyzed here. (Croushore 1997)

Unlike the specificity of the Survey of Professional Forecasters, the Livingston survey is more general in that the participants it draws from come from companies with a much broader scope of disciplines and from individuals for whom making forecasts about future economic states may not be their prime responsibility. There are generally anywhere from 50-70 participants per survey. (Mankiw et al 2003) To reiterate, Ang et al (2007) found the Livingston survey one of the best predictors (just behind the SPF) of the CPI. It serves as a omnipresent force in expectations research for many years and has been used in countless academic works. The only complication to the Livingston survey is that CPI is not seasonally adjusted until 2004: an interesting historical choice. Livingston is generally regarded as a survey of business’ (firm’s) general expectations, in contrast to SPF being more focused (as from financial elite).

Again, we will estimate the regression $\pi_{t-(i-6),t-i} = E_{t-12} \pi_{t-(i-6),t-i} = a + b(\text{un}_{t-13})$ for $i = 0, 6$ to determine the degree to which Livingston forecasters underestimate the slope of the Phillips Curve as expressed through CPI. We are cautioned in Mankiw et al (2003) that “early [Livingston] data is unreliable,” a problem they fixed by ignoring all forecast data collected before 1954. Before following their direction and dropping all data points before 1954s1, we had a sample 1948s2-2011s2. Unlike the SPF survey, Livingston has a longer running series of the CPI than a series that can be derived for GDP inflation. In order to denote the semiannual data, we use e.g. 1950s1 to denote
quarters 1 and 2 in 1950 (1950s2 for quarters 3 and 4). For consistency in our specification of all equations regarding semiannual data, we use months consistently so it matches SPF’s order (a half-year/semiannual period being equal to 6 months).

Figure 5. Livingston (LVS) CPI Inflation Forecast Error versus Unemployment at 1 Year.

![Figure 5. Livingston (LVS) CPI Inflation Forecast Error versus Unemployment at 1 Year.](image)

Table 3. LVS CPI Inflation over All Time Horizons.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>(a) (c)</th>
<th>(b) (un)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-12,t-6</td>
<td>3.229***</td>
<td>-0.555***</td>
</tr>
<tr>
<td></td>
<td>(1.078)</td>
<td>(0.198)</td>
</tr>
<tr>
<td>t-6,t</td>
<td>2.891***</td>
<td>-0.412**</td>
</tr>
<tr>
<td></td>
<td>(0.978)</td>
<td>(0.168)</td>
</tr>
<tr>
<td>t-12,t</td>
<td>3.134***</td>
<td>-0.499***</td>
</tr>
<tr>
<td></td>
<td>(0.995)</td>
<td>(0.177)</td>
</tr>
</tbody>
</table>
Again here, we see over a different survey that business economists are underestimating the slope of the Phillips Curve in their inflation forecasts. It shows a very similar story to the one told about the Survey of Professional Forecasters. However, the coefficients here are stronger (though not statistically significantly so). This would suggest that participants in the Livingston survey mis-estimate the slope of the Phillips Curve to a greater degree than do SPF participants. Assuming this error is an indicator of or is correlated positively with other errors that are present in the forecasting process, this result follows and supports the narratives shown in the literature, specifically Croushore’s (1998) work and Ang et al’s (2007) empirical observations that the SPF is an ever so slightly more reliable and accurate measure.
At this point we will continue our analysis to the Livingston survey’s GDP inflation series. Like the Livingston CPI forecasts, it is semiannually observed. And similar to the SPF GDP series there is a break in the underlying inflation series forecasted in 1992: in all periods before 1992 real and nominal GNP are forecasted and after only GDP. It should be noted that there was no move to a chain-type price index as occurred in the Survey of Professional Forecasts. Again, no more reference will be made of this, and we will refer categorically to GDP (price index) inflation etc. to represent this series.

Unfortunately, the Livingston survey does not make a direct estimate of the GDP deflator. It is necessary to obtain each forecaster’s prediction for inflation from the equation \( E_{t-13}\pi_{t-12,t} = \Delta\%NGDP_{t-12,t} - \Delta\%RGDP_{t-12,t} \). We have a significantly curtailed time sample to consider because the first forecasts of real GDP are included in the Livingston survey in 1971 (forecasting nominal GDP was part of the survey from its inception). In order to calculate an individual forecaster’s inflation forecast for one period, it requires four separate pieces of information: beginning and ending nominal GDP, to calculate percent change in nominal GDP; and beginning and ending real GDP, to calculate percent change in real GDP. Both ending figures are a forecast for the given period. Once we have combined the two, percent-change figures, we can move to the forecaster’s inflation forecast, based on the relationship featured earlier in this paragraph. After reaching this level, we summarize the data to come to a median forecast for the period (and each horizon).
Figure 7. LVS GDP Inflation Forecast Error versus Unemployment at 6 Months.

Figure 8. LVS GDP Inflation Forecast Error versus Unemployment at 6 Months forward.
Figure 9. LVS GDP Inflation Forecast Error versus Unemployment at 1 Year.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Parameter a (c)</th>
<th>Parameter b (un)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-12,t-6</td>
<td>0.271</td>
<td>-0.253</td>
</tr>
<tr>
<td></td>
<td>(1.194)</td>
<td>(0.182)</td>
</tr>
<tr>
<td>t-6,t</td>
<td>1.451*</td>
<td>-0.225*</td>
</tr>
<tr>
<td></td>
<td>(0.793)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>t-12,t</td>
<td>0.919</td>
<td>-0.262**</td>
</tr>
<tr>
<td></td>
<td>(0.787)</td>
<td>(0.129)</td>
</tr>
</tbody>
</table>
Yet again, we see confirmation of a systemic underestimation of the Phillips Curve on the part of Livingston survey participants. That the two, half-year time step coefficients aren’t significant could be a result of the exact methodology used to come to a median inflation forecast. A broader, more sound (less varying / lower standard error) estimate could come from performing the median procedure (as discussed earlier in this sub-section) at an earlier stage in the procedure, so as to find median real GDP change and median nominal GDP change and subsequently coming to an inflation forecast (from those two medians). In previously conducted analyses not included herein, the author found a materially positive (if slight) change in significance level if the median procedure was conducted in the revised procedure (the one presented in this paragraph). However, we feel the former method represents the most faithful parallel of all the other series. In
either case, we see coefficients in line with what has been present in the Survey of Professional Forecaster’s coefficients of between -0.40 and -0.25 for GDP inflation, but higher than previously estimated standard errors. These standard errors are generally more than double what can be viewed in Table 1.

5. Consumer Surveys

The Michigan Survey of Consumers has been collecting information on consumer sentiment on a quarterly basis since 1960. They call over 500 households in the middle month of each quarter. From one of the questions asked and an especially devised procedure, the questioners are able to compute a mean inflation expectation—note the fact that this series uses the mean differing from all others to date. Specifically the question demanded of respondents during the survey is: “By about what percent do you expect prices to go up on average during the next 12 months?”

Unfortunately, the question does not directly peg against a measure of inflation as has been the case in all other cases/series examined in this paper. This leaves us with a level of vagary in determining how to judge first the correctness of (calculate ex-post errors for) the forecast and subsequently whether consumers—participants in the survey—are correctly judging the Phillips Curve. Nonetheless, we see it is likely they are not, as even professional forecasters with sophisticated modeling techniques are unable to accurately do. Further, Ang et al (2007) and Mankiw et al (2003) both show that Michigan estimates have significantly less predictive power when compared to SPF and Livingston, which assuming Phillips Curve error coincides with general error, confirms their analyses. Roberts (1997) confirms the conclusions here: that the Michigan survey does not conform to rationality in that consumers have not fully taken into account the
current state of many macroeconomic variables, including unemployment, while Mankiw et al (2003) go far further establishing that the Michigan Survey of Consumers fails an even more basic level of rationality as well.

Coibion and Gorodnichenko (2011) point out three significant limitations to the Michigan Survey. First, there is the lack of an index against which to peg forecasts to calculate errors as has been previously mentioned. Second, the lack of an index means consumers could be making forecasts of fundamentally different price indices—presumably they are making an estimate of inflation of either their unique basket of goods and services or the products to which they are exposed—which regardless introduces error. Finally, the methodology of the phone survey could introduce measurement errors.

*Michigan Survey versus CPI*

Instead of assuming which measure Question 19 from the Michigan survey is intended to parallel, we will explore it in relation to CPI, the typical measure, and the Personal Consumption Expenditures (PCE) deflator. As a point of reference, Ang et al (2007), Mankiw et al (2003), and Roberts (1997) all compare the Michigan survey to the CPI to the exclusion of inflation from PCE or any other measure. The specification for CPI is the same as always: $\pi_{t-12,t} - E_{t-13} \pi_{t-12,t} = a + b(\text{un}_t_{-13})$. With this, we find similar results to past CPI analyses in this paper.
Here we see fairly strong support for the hypothesis that regular citizens underestimate the slope of the Phillips Curve. Though it should be noted that the economic significance of this underestimation is markedly lower than for other CPI measures. Other forecasters see nearly double the coefficient on unemployment, exhibiting nearly double the unexplained effect from unemployment information a result of a much stronger error in their estimation of the Phillips Curve. (In addition, the constant term is insignificant in the Michigan Survey analysis in Table 5 above.)

**Michigan Survey versus Personal Consumption Expenditure**

Personal Consumption Expenditure is a price index formed from the consumption portion of the GDP in a similar method to the GDP deflator. As such it should serve as a
good measure of inflation with one of the key benefit of the GDP deflator: it represents the price change, not specifically for a basket of goods for urban consumers, but in what the market (of private citizens purchasing for themselves throughout the entire country) clears as a whole (or on average, conceptually equivalent). In any case it represents another good proxy to use for the change in the price level facing consumers, a good analog for both the question posed and inflation generally (specifically how it affects consumers). Using the same model specification as analyzing MSC with regards to CPI, we come to the following results.

Figure 12. MSC PCE Inflation Forecast Error versus Unemployment.

Table 6. MSC versus PCE Inflation.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>$a$ (c)</th>
<th>$b$ (un)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-12,t</td>
<td>-0.417</td>
<td>-0.088</td>
</tr>
<tr>
<td></td>
<td>(0.514)</td>
<td>(0.070)</td>
</tr>
</tbody>
</table>

Here we see an economically weaker coefficient on unemployment relative both to the one comparing the Michigan Survey to CPI (in Table 5), which was already weaker
than that for all other CPI surveys. One possibility is that it means that if Michigan Survey of Consumers respondents are making inflation forecasts based on PCE, that they are fairly closely estimating the Phillips Curve. Otherwise, we could just not have enough statistical power to make this assessment (variance in the underlying PCE might be too high). There is already evidence to suggest extensive disagreements between respondents to the Michigan Survey and their variance with varying levels of inflation (Mankiw et al 2003). Similarly, if PCE does not match what MSC participants are estimating, we would not notice this relationship even if they are failing to properly estimate the Phillips Curve.

One piece of information that could be helpful in determining which inflation series consumers are forecasting is which ex-post inflation series the predictions most accurately track. Taking this outcome at face value has obvious, insurmountable failings, but is a starting point. In Table 7, we have calculated the mean square error (MSE) for the two inflation-error series, ex-ante Michigan Survey of Consumers when compared to the observed CPI for the predicted period and the same compared to the observed PCE inflation for the period forecasted.

| Table 7. Comparison of Michigan Survey Forecast Errors by Inflation Series Forecasted. |
|---------------------------------|-----------------|-----------------|----------------|
| Tracking | CPI | PCE | SPF GDP |
| MSE      | 3.409 | 2.755 | 1.168 |

From this analysis, we see consumers are most accurately predicting the Personal Consumption Expenditure series in the Michigan Survey. This suggests that consumers in responding to the survey are most likely looking at the more nuanced PCE number that reflects the changing-weight basket of goods practically consumed in the economy by consumers like themselves. As to the strength of the Michigan Survey as a whole, it
could be argued that the fact that the consumer survey represents a sample of micro-level actors in economic activity, it might have added accuracy: Carroll (2003) states this “consensus among economists” that “macroeconomic outcomes depend critically on agent’s [those directly sampled in the Michigan survey] expectations.”

But in fact we see from this simplistic analysis that compared to the MSE of the 1-year ahead GDP inflation forecast of the Survey of Professional Forecast (included over its maximal sample available in Table 7), that professionals are far better at anticipating inflation. It seems Carroll’s (2003) assertion about agent’s expectations is not supported: consumers aren’t able to make their inflation expectations come true, or else there is remarkable forecasting error. The inaccuracy of Michigan Survey forecasts found here confirms a well-established body of literature.

Regardless of the outcome of the previous analysis, because of the wide usage of CPI as a counterpart to the Michigan Survey of Consumers in the broader literature as discussed previously, we will use the CPI as the appropriate comparison throughout the remainder of the paper.

6. Federal Reserve Greenbook Surveys

One of the most interesting questions inspired by the conclusions we have drawn thus far is whether the underestimation of the slope of the Phillips Curve extends to the highly regarded professionals at the Federal Reserve. We expect that the Federal Reserve should have enough institutional focus and resources to adequately make these forecasts correctly and be able to adapt best. In any case, they are the strongest standard-bearer, holding the bar highest for the remainder of our forecasters, and are arguably the least likely to be wrong or surprised.
Before every Federal Open Market Committee (FOMC) meeting, the Federal Reserve Board staff prepare forecasts of GDP inflation and many other macroeconomic variables that could be useful to the policy makers who will gather there as they make monetary policy decisions that are so integrally related to inflation. Five years after the meeting, the Federal Reserve Bank of Philadelphia compiles the data, so as not to disclose the Fed’s uncannily accurate predictions while they might have economic value and to protect any non-public sources (e.g. internal commercial bank statistics) they may have drawn from in the remainder of the Greenbook pamphlet (which contains a vast array of forecasts, analyses, and sensitivity projections that the FOMC may find pertinent). Specifically, FRB Philadelphia chooses the FOMC meeting closest to the center of the quarter (to facilitate comparison to the SPF and Livingston surveys).

In this Section of the paper, we will examine the Federal Reserve’s projection of inflation based on the GDP deflator. Again, as for the Survey of Professional Forecasters and the Livingston survey the predicted series changes such that pre-1992 the Greenbook predicts GNP deflator based inflation, 1992 to 1996 GDP deflator based inflation, and after 1996 GDP chain-weighted price index inflation.

Romer and Romer (2000) have well established the preeminence of the Federal Reserve Board staff’s Greenbook forecast of inflation, markedly so over commercial forecasts (specifically the Blue Chip and DRI series, proprietary data sets to which we do not have access, in addition to the SPF that we examine here). They hypothesize that the Greenbook’s accuracy comes from two sources: first a potential information advantage and second to the sheer amount of resources they use. They find the Federal Reserve has very limited access to new statistics: the Chairman only gets official releases the night
before they come out. In their qualitative analysis, they do ignore the bank statistics wing of the FRS and the proprietary, non-public information the Federal Reserve Banks extract from financial institutions, which are presumably available as Federal Reserve staff prepares the Greenbook forecast. Secondly, at longer horizons, the Federal Reserve could be using inside information about the Fed’s long-term commitment to various policies.

However, Romer and Romer (2000) clearly indicate in their analysis that they believe the advantage the Greenbook forecasts have evinced over professional forecasters comes from “the Federal Reserve commit[ting] far more resources to forecasting than even the largest commercial forecasters. …It has the same ‘technology’ as commercial forecasters… [i]t simply chooses to use more [technology, labor, and data] inputs than any commercial forecasters find profitable.” They conclude that overall the Federal Reserve is by far the most accurate type of forecaster of inflation.

Romer and Romer (2008) go on to briefly analyze the FOMC members’ forecasts. In making their forecasts, FOMC members have access to the Greenbook projections. They go back to their districts and prepare a forecast uniquely their own. Yet these new forecasts still produce subpar results. Romer and Romer’s (2008) analysis clearly show that someone projecting inflation with access to both Greenbook and FOMC members’ forecasts should entirely disregard the FOMC members’ forecast: they have no value relative to the Greenbook forecasts. Before rushing to the conclusion here that the individual members are inept, one should be aware of a slight complication, FOMC members are asked to prepare forecasts under “ideal monetary policy,” a direction given that differs from the Greekbook’s preparation procedure.
Due to ease of availability, ubiquity, and accuracy of the Greenbook forecast, in this analysis we will only look to the Greenbook’s (which we will occasionally abbreviate GBK) GDP inflation projections, which are available generally on all forecasting horizons starting in 1968 (though we begin some analysis in Section 8 in 1957 because of many missing forecasts at the one-year horizon. As previously mentioned, Greenbook data are released only after a 5-year delay: at the time of this analysis, projections from at latest 2005q4 were available (predictions made in the 4th quarter of 2005, the latest of which corresponds to inflation during the 4th quarter of 2006). Following a similar specification to all other series and using unemployment information from the month the forecast was made, we analyze the maximal sample of the Greenbook GDP inflation forecast below in Table 8.

Figures 13. GBK GDP Inflation Forecast Error versus Unemployment at 1 Year.
Table 8. GBK GDP Inflation over All Time Horizons.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>( a ) (c)</th>
<th>( b ) (un)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-15,t-12</td>
<td>1.108***</td>
<td>-0.199***</td>
</tr>
<tr>
<td></td>
<td>(0.302)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>t-12,t-9</td>
<td>1.712***</td>
<td>-0.277***</td>
</tr>
<tr>
<td></td>
<td>(0.506)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>t-9,t-6</td>
<td>2.119***</td>
<td>-0.330***</td>
</tr>
<tr>
<td></td>
<td>(0.736)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>t-6,t-3</td>
<td>1.998**</td>
<td>-0.306**</td>
</tr>
<tr>
<td></td>
<td>(0.920)</td>
<td>(0.130)</td>
</tr>
<tr>
<td>t-3,t</td>
<td>1.722*</td>
<td>-0.280*</td>
</tr>
<tr>
<td></td>
<td>(0.971)</td>
<td>(0.142)</td>
</tr>
<tr>
<td>t-12,t</td>
<td>1.363*</td>
<td>-0.218**</td>
</tr>
<tr>
<td></td>
<td>(0.734)</td>
<td>(0.103)</td>
</tr>
</tbody>
</table>

Figure 14. GBK GDP Inflation over All Time Horizons

We see that even the Federal Reserve is not adequately forecasting the slope of the Phillips Curve over the sample as a whole: the negative coefficient is significant at all horizons. First that we are able to find a systemic bias in the Federal Reserve’s Greenbook inflation forecasting is significant. Secondarily, this could have a significant
information content when retroactively analyzing policy decisions that the bank has made in light of the Phillips Curve knowing that it has been inaccurately estimated.

7. Summary of Regressions to Date

Before jumping in to an analysis of how the coefficients on unemployment change over time, it will be useful to combine all the regressions performed thus far into a consistent table. Table 9 contains Panel B, the regression information reported heretofore in the paper. Panel A brings the regressions over the largest possible sample that would contain 100% of data points in all series—notably it is cut off in 1981 at the start by SPF’s CPI projections and in 2005 by the Greenbook’s five-year release timetable. While we still have some 25 years of data to examine in our remaining, consistent sample, it does decrease the power of our observations, though some are clearly able to withstand the power loss.

Curiouser is the sudden significance of the SPF CPI error series, which suggests the last 5 years of that sample were in a different direction from what we have seen throughout the rest of the sample: this only increases the motivation to perform rolling regressions on each of the surveys, as we will examine in Section 8. Note SPF CPI is the series for which a 5-year change in sample size would represent the largest proportional change (i.e. it is the shortest sample) thus would be most susceptible to a change. Other than SPF CPI and the Michigan Survey versus CPI, the coefficients remained largely similar in significance level to the original results.
Table 9. Comparison of \( \alpha_{t-13} \) Coefficients over All Samples and Time Horizons.

### Panel A. Over Common Sample: 1981q3-2005q4 (& 1981s2-2005s2)

<table>
<thead>
<tr>
<th>Horizon</th>
<th>SPF GDP</th>
<th>SPF CPI</th>
<th>LVS CPI</th>
<th>LVS GDP</th>
<th>MSC CPI</th>
<th>GBK GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-15,t-12</td>
<td>-0.165***</td>
<td>-0.218**</td>
<td></td>
<td></td>
<td></td>
<td>-0.138***</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.097)</td>
<td></td>
<td></td>
<td></td>
<td>(0.042)</td>
</tr>
<tr>
<td>t-12,t-9</td>
<td>-0.271***</td>
<td>-0.350***</td>
<td></td>
<td></td>
<td>-0.514***</td>
<td>-0.148**</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.128)</td>
<td></td>
<td></td>
<td>(0.070)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>t-9,t-6</td>
<td>-0.329***</td>
<td>-0.358**</td>
<td>-0.181**</td>
<td>0.121**</td>
<td></td>
<td>-0.160**</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.160)</td>
<td></td>
<td></td>
<td></td>
<td>(0.070)</td>
</tr>
<tr>
<td>t-6,t-3</td>
<td>-0.381***</td>
<td>-0.371**</td>
<td></td>
<td></td>
<td>-0.150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.173)</td>
<td></td>
<td></td>
<td></td>
<td>(0.100)</td>
</tr>
<tr>
<td>t-3,t</td>
<td>-0.302**</td>
<td>-0.427**</td>
<td>0.156</td>
<td>0.113</td>
<td></td>
<td>-0.211*</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.176)</td>
<td></td>
<td></td>
<td></td>
<td>(0.116)</td>
</tr>
<tr>
<td>t-12,t</td>
<td>-0.226***</td>
<td>-0.367**</td>
<td>-0.556***</td>
<td>-0.436***</td>
<td>-0.160</td>
<td>-0.158**</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.144)</td>
<td></td>
<td></td>
<td></td>
<td>(0.079)</td>
</tr>
</tbody>
</table>

### Panel B. Over Maximal Samples

<table>
<thead>
<tr>
<th>Horizon</th>
<th>SPF GDP</th>
<th>SPF CPI</th>
<th>LVS CPI</th>
<th>LVS GDP</th>
<th>MSC CPI</th>
<th>GBK GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-15,t-12</td>
<td>1968q4</td>
<td>1981q2</td>
<td>1948s2</td>
<td>1971s1</td>
<td>1960q1</td>
<td>1965q4</td>
</tr>
<tr>
<td>t-12,t-9</td>
<td>2011q4</td>
<td>2011q4</td>
<td>2011s2</td>
<td>2011s2</td>
<td>2010q4</td>
<td>2005q4</td>
</tr>
<tr>
<td>t-12,t-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-9,t-6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-6,t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-3,t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-12,t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If at a minimum to investigate the change in SPF’s CPI forecast errors based on the Phillips Curve, we will now continue (in Section 8) to examine whether the sampling period affects the reported coefficient. Thus we are analyzing if forecasters are improving their estimation of the slope of the Phillips Curve over time (or becoming more inaccurate). There is already significant evidence to suggest that the Phillips Curve shifts over time (Stock and Watson 1999a). With these shifts, there is a potential for accuracy of forecasting to change. More importantly, shifts in perceptions about
Macroeconomics may impact the views of forecasters. Changes in these political views are evident in inaccurate estimation of the Phillips Curve and as our observed coefficient changes, we can analyze these changes.

*Examining Overall Accuracy of Forecasts (via MSE)*

Extending the analysis we conducted at the end of Section 5 in Table 7, we now examine all of our series versus their respective mean square error (MSE). MSE is a simple, strong means of observing the difference between a number and its true value very similar theoretically to standard deviation. In our case, it is a quick analysis to verify the results of Romer and Romer (2000). This will give us a simple answer about the relative error present in the various series.

<table>
<thead>
<tr>
<th>Table 10. Comparison of Mean Square Error for All Series at 1-Year Horizon.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizon</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td><strong>Panel B. Over Maximal Samples: see Table 9 for list</strong></td>
</tr>
<tr>
<td>Horizon</td>
</tr>
<tr>
<td>MSE</td>
</tr>
</tbody>
</table>

Quite surprisingly in Table 10 both in Panels A and B, we observe that the Survey of Professional Forecasters’ prediction of GDP inflation outperforms all other forecasts. The main surprise here is the contradiction with Romer and Romer (2000) showing that the Greenbook is outperformed by the SPF GDP series, if only slightly. This represents a significant departure from the existing, recognized body of literature, but as seen starting in Faust et al (2004). One possible explanation for this could merely be an increased time sample: we should at this point have 12 more years of Greenbook inflation forecasts than Romer and Romer had at the time—data from 1968q4 to 1991q4. However, after
changing our sample the Greenbook’s MSE is 2.035 while SPF’s GDP series’ MSE is 1.777, which does not change the outcome of our analysis.

Two further caveats that should be made about this analysis in comparison to Romer and Romer (2000). First, Romer and Romer (2000) use a far more complex forecast timing analysis; in our analysis we only examine one FOMC meeting’s Greenbook forecast per quarter (so as to match with the SPF survey). Second, Romer and Romer (2000) use a different methodology to determine which forecast is most reliable: the model they use to compare Greenbook to SPF forecasts is

$$\pi = a(SP) + b(GBK) + c.$$  By their empirical results, the year-ahead inflation rate (as compared to the data we have presented show $a = -1.08 (0.38)$, $b = 1.93 (0.35)$, and $c = 1.09 (0.53)$, indicating someone with access to the SPF next-year inflation forecast would double count the Greenbook results and then remove the information SPF GNP/GDP forecasts once (subtract them out). Obviously, the result we have achieved—of two competitive surveys one slightly outperforming the other—does not match their results any case.

8. Examining Rolling Samples

The graphs we hereafter present—Figures 15-20—reflect 10-year rolling regressions paralleling those specified earlier and throughout the paper in Sections 2-6. As pictured in any of the following, the year (on the x-axis) reflects the beginning quarter (or half-year) of the beginning of the 10-year sample. This analysis was, for consideration brevity’s sake, conducted only on 1-year inflation error. Plotted on the other axis is the coefficient of the regression for that sample, along with ±2 standard errors of the coefficient for that observation. The 10-year sample was considered a
minimum length given the power loss and fickleness of the coefficient that may be associated with any further decrease in the sample size. Nonetheless, we still see significant changes over the time giving us a significant amount of information about the questions at hand.

Figure 15. SPF GDP over 1-Year Horizon.

In the original sample of the Survey of Professional Forecasters’ prediction of the GDP deflator, we saw a clear negative relation. What we see depicted in Figure 15 is a (generally speaking) statistically significant underestimation of the slope of the Phillips Curve until a massive jump in the 1990s—which was statistically significant for a brief time in the other direction. This positive jump in the data would cover the early to mid-2000s. After this time, we see inflation error consistently close to 0 in this series, bounded nearly entirely within 0±0.5%. The modern sample (post-2000) seems to be fairly consistent with an accurate estimation of the Phillips Curve. Yet the pre-1980s
sample is remarkably below 0: the 1970s-era Phillips Curve estimation by professional forecasters significantly underestimated the slope of the Phillips Curve.

Figure 16. SPF CPI over 1-Year Horizon.

To refresh, we had seen in Tables 2 and 9 that there is insufficient evidence to suggest that SPF forecasters overall were systemically underestimating the Phillips Curve for the CPI series. The coefficients observed in the CPI inflation series were, however, strikingly similar to those in the GDP series. We see in Figure 16 a mirrored and equivalent story: coefficients in each time period are similar (to Figure 15) but standard errors are significantly greater—in this case that they alter the scale of the axis in Figure 16 (making direct comparison more difficult and opaque). Yet we also notice in the SPF GDP series pre-1981 (Figure 15) that, as previously discussed, the underestimation of the Phillips Curve in the 1970s was remarkably large. Given that that segment of the sample is not observed here, we have another explanation for a differing level of significance: a
period that would affect the mean in the direction of the hypothesized effect. Note both of these variances move in the same direction, lessening the statistical significance of the SPF CPI regression(s).

Figure 17. LVS CPI over 1-Year Horizon.

![Graph](image.png)

This Livingston sample (CPI) shows an interesting aberration at the beginning of our graph. At this moment, we recall Mankiw et al (2003) who vaguely state about the survey (started in 1948) that “the early data is unreliable.” Generally throughout Figure 17, we see the 1960-1990 period show a significantly negative coefficient. Especially if we account for the extra loss of power coming from the fact that a 10-year sample has half the observations for the biannual, Livingston survey when compared to the other surveys which are all quarterly elsewhere considered in this paper. Late in the sample, we see a similar jump as we have seen in all series beginning in the early 1990s.
As Table 4 would lead us to believe, there is only a weakly significant underestimate the slope of the Phillips Curve. Yet once we exclude the outlying, early observations, as Panel A of Table 9 does, and as we see beyond 1950 in Figure 17, we see a noteworthy, negative coefficient emerge. This is especially true when ignoring the post 1990s samples in Figure 17.

Figure 18. LVS GDP over 1-Year Horizon.

Before the samples starting in the 1980s begin, we see a negative coefficient colored by very high standard errors. Afterwards, there is a cyclical pattern—one that in the samples beginning in the 1980s does not match the other series—that moves around 0 but does not significantly deviate from 0 for any remarkable period of time. This is unexpected, given the results of the broader regressions in Table 9, which show a significant, negative coefficient. The problems mentioned previously with the weakness of the test over a semiannual, 10-year sample likely are further confusing the information
above. Though the point observations (i.e. ignoring the ±2se confidence interval) do contain some information about a local slope. However, the 1908s period positive coefficient is the only difference point, though it is never significantly positive and bordered on either side by a period of a negative coefficient, leaving us questioning whether it legitimately exists.

Figure 19. MSC CPI over 1-Year Horizon.

Within the Michigan Survey of Consumers, we see the 1970s establish the beginning of a clear negative coefficient on the rolling samples. Again, we see the positive jump in the samples beginning in the mid-1990s. As we noted in Section 5 though, standard errors are remarkably high meaning we have a very small period of a significant negative coefficient. Though the value of the coefficient is consistently negative for a long period of time (ignoring the confidence interval) and is much stronger pre-1980 than other forecasting methods and about on par thereafter. That it was only
during the Great Inflation that consumers were unable to accurately predict inflation, leads us to suspect this may only be a temporary aberration. There is likely to be significant information in the comparison of CPI to PCE inflation errors to spur some future research on the Michigan Survey’s inflation error over time.

Figure 20. Greenbook GDP over 1-Year Horizon. Samples begin 1975q3.

The forecasts of the Federal Reserve’s Greenbook are a different story. We wholeheartedly expect them to exhibit the highest degree of accuracy and rationality, even with respect to revising their estimate of the Phillips Curve as the world changes. Indeed, after samples beginning in roughly 1982, there is not a large deviation from 0, certainly not a significant one. Yet what is wholly remarkable is the strong, negative coefficient at the beginning of our sample. Certainly, -0.999 (the value in 1975q3 graphed in Figure 20) is the remarkably strong. This suggests the Fed underestimated the slope of the Phillips curve during this rolling sample period to a large degree. It should
be noted, that we had to curtail the rolling samples used to create Figure 20 because forecasts in periods before 1975q3 were few and far between at the year forecasting horizon. Unfortunately, we have no information about the Fed’s performance during the current crisis, solely because of the 5-year waiting period before the Fed can release its internal forecasts to the public.


Primiceri (2006) establishes a strong theory to explain the Great Inflation of the 1960s-1970s and the subsequent Volcker Disinflation of the early 1980s. To begin, he excoriates several existing theories that (1.) an abnormally high number of exogenous, non-policy shocks, (2.) that policy makers were not committed to low inflation, or (3.) that policy makers during the 1960s-1970s were more inept than previous and subsequent generations. None of those, he concludes, explain the full scope of the problems.

Instead, Primiceri (2006) looks to policy-makers’ beliefs about the coefficients of the Phillips Curve. He isolates the history of the Great Inflation into two distinct periods. The 1960s and the early 1970s were characterized by a belief in a low natural (NAIRU) rate of unemployment, a finding well established in the literature. As the Federal Reserve adjusted monetary policy to reach their overly optimistic unemployment target, the economy overheated resulting in a high rate of inflation. The fact that this high level of inflation was not linked to a lowering rate of unemployment caused much consternation, with some scholars advancing abandonment of the typically constructed Phillips Curve (McNees 1978). During the early 1970s however, policy-makers began to correctly estimate the natural rate of unemployment, yet Primiceri’s (2006) model suggests the Federal Reserve revised their perception of the slope of the Phillips Curve (lesser than
true estimate of the relationship between inflation and unemployment) in the early to mid-1970s, lessening the degree to which they would respond to historical inflation because of the high cost it would have in terms of employment.

It took the Volcker Disinflation before “policy-makers had a model of the economy that was approximately correct.” Primiceri’s explanation that the Federal Reserve feared the high cost of a lowering of inflation is confirmed more anecdotally by Okun (1978), who conducted a survey of the literature surrounding the era finding “all of the Philips Curves he examined] point to a very costly short-run tradeoff” between inflation and unemployment.

Yet the analysis conducted in throughout this paper gives us a unique insight not considered in Primiceri’s (2006) work. Using Greenbook data, we can take an inside look at the Federal Reserve’s thinking and predictions of inflation. As has been explicated throughout the earlier part of this paper, we find that Greenbook data has systemically underestimated the slope of the Phillips Curve. Specifically, in Section 8 and Figure 20, we found that the period for which the strength of the Phillips curve is underestimated to a level of statistical significance only for those 10-year samples beginning before 1978. This coincides with Primiceri’s “Overpessimism” with a Phillips Curve with a mis-estimated slope indicating, again, an overly high tradeoff between inflation and unemployment.

Beginning in 1982, the Federal Reserve staff’s estimation error of the Phillips Curve is much closer to zero and gradually continues to approach said no-bias mark. This matches Primiceri’s analysis as well. Thus we can conclude via Greenbook forecasts, that the Federal Reserve staff systemically underestimated the slope of the
Phillips Curve during the late 1970s, resulting in an overestimate of the cost of disinflation, confirming analyses both in Primiceri (2006) and Okun (1978). Thus, a dual-mandate-minded cost-benefit analysis mistakenly led the Federal Reserve to logically and rationally select high inflationary policy as a reasonable and appropriate policy path. Indeed as estimates become more accurate, the Fed became more willing to engage in disinflationary policy, notably the Volcker Disinflation of the 1980s.

10. Conclusion

Throughout many inflation forecast surveys we have seen a significant, systemic error with respect to the Phillips Curve. Our regressions have shown a consistent negative coefficient on unemployment showing that forecasters are consistently underestimating the Phillips Curve’s strength. This is surprising given the rigor of many of the forecasts, especially the Federal Reserve and the Survey of Professional Forecasters. As we learn in Section 8, much of the mischaracterization is a result of pre-1990s forecasting error. Since the 1990s, in private surveys, we have seen an overestimation of the Phillips Curve’s strength, but overall a much more accurate forecasting ability over all forecasters. The Federal Reserve has shown a fairly rational, bias-free forecasting system of late, which is in line with the wide literature that suggests Federal Reserve Greenbook forecasts are the best available forecasts of inflation. However, the gradual improvement of all forecasts could just be due to the lessening of inflation and volatility generally during the Great Moderation. In all, we are surprised, like Mankiw et al (2003) and others, at the broad inaccuracy of forecasts on such predictable bases as previous period error and well-recognized macroeconomic theories on available data, like the Phillips Curve.
Possible extensions of this research include extensions to additional forecasting series, including those used by Romer and Romer (2000), the financial market implicit forecasts created by the likes of Ang et al. (2007), implicit series from CBOE derivatives (which began limited trading in 2004), and longer-term forecasts from TIPS securities and the surveys included. As mentioned previously, a closer look at the existing forecasts’ performance during the recent crisis would be especially interesting. Other options considered, are breaking down forecast error by industry—especially relevant in the Livingston survey, where economists’ predictions are drawn from a wide array of finance, manufacturing, public sector, and other types of entities—where the predictions are given.
11. Bibliography


