Internet Search Behavior as an Economic Forecasting Tool:

The Case of Inflation Expectations[1]

Giselle Guzmán, Ph.D., FRM

Economic Alchemy LLC

Draft, Preliminary and Incomplete

This Version: November 29, 2011

[1] A previous version of this paper circulated under the title "An Inflation Expectations Horserace," and was submitted in partial fulfillment of the Ph.D. degree in Finance and Economics at Columbia University. Contact information: info@economicalchemy.com. I would like to thank Lawrence Klein, Joseph Stiglitz, Hal Varian, Marshall Blume, Marc Giannoni, Charles Jones, and Lawrence Glosten for their insights and guidance. Thanks also to Randy Moore, William Dunkelberg, Richard Curtin, Lynn Franco, and Dennis Jacobe for generously providing data. All equations estimated with EViews Version 6 for Windows PC by Quantitative Micro Software, LLC. Any errors are solely my own.

Internet Search Behavior as an Economic Forecasting Tool:

The Case of Inflation Expectations

Abstract:

This paper proposes a measure of real-time inflation expectations based on metadata, i.e., data about data, constructed from internet search queries performed on the search engine Google. The forecasting performance of the Google Inflation Search Index (GISI) is assessed relative to 37 other indicators of inflation expectations – 36 survey measures and the TIPS spread. For decades, the academic literature has focused on three measures of inflation expectations: the Livingston Survey, Survey of Professional Forecasters, and the Michigan Survey. While useful in developing models of forecasting inflation, these low frequency measures appear anachronistic in the modern era of higher frequency and real-time data. I demonstrate that higher frequency measures tend to outperform lower frequency measures in tests of accuracy, predictive power, and rationality. Furthermore, Granger Causality tests indicate that the GISI metadata indicator anticipates the inflation rate by 12 months, and out-of-sample forecasts show that the GISI has the lowest forecast error of all the inflation expectations indicators tested.

Keywords: Inflation, expectations, metadata, search, surveys, TIPS, Google, forecasting.

Internet Search Behavior as an Economic Forecasting Tool:

The Case of Inflation Expectations

I. Introduction

I examine the predictive power of various measures of inflation expectations, and introduce a novel measure of expectations revealed by internet search queries performed on the search engine Google. The main questions I seek to answer are the following: do higher frequency measures of inflation expectations perform better in tests of accuracy, predictive power, and rationality, compared to the standard lower frequency measures that economists have relied upon for decades? Moreover, can metadata, i.e., data about data, provide signals about the revealed expectations of economic agents, and do these signals have any predictive power in forecasting economic quantities?

According to the Pew Internet & American Life Project's May 2010 survey, 79% of American adults use the internet, and of these, 87% use a search engine to find information.[2] With the ubiquity of internet searches, websites, social media platforms, and blogs, we now have more data about ourselves, our opinions and our expectations than ever before in the history of humankind. Data about this data may contain important clues about human behavior that were unobservable just a decade or so ago. I examine a novel measure of expectations based on Google Trends, which provides near real-time tracking data on the volume of internet search queries related to a certain topic and users' propensity to search for these keywords on a relative

[2] http://www.pewinternet.org/

basis.[3] The topic in this case is inflation, as measured by changes in the number of searches performed on the keyword "inflation" in the US.

Internet search behavior can be interpreted as a measure of revealed expectations. Presumably, people search for information on topics they want to learn more about, or about things that are causing them concern. For example, if someone is not feeling well, they might search for information about the health symptoms they are experiencing. Similarly, if someone is concerned about managing their household expenses and believes the general price level is rising or that prices may soon rise, they might search for information about inflation. If someone is not worried about the prospect of rising prices, then they probably would not search for information about inflation, since most people are more fearful of inflation than of deflation. In an economy where wages are more sticky than prices, it is understandable that consumers would harbor anxiety about inflation. Therefore, we can surmise that if search query volumes on Google for the term "inflation" are increasing, then the public is feeling increasingly anxious about the prospect of rising prices. Hence, changes in the volume of search queries about inflation can be interpreted as a measure of the general public's revealed expectations for future changes in the price level.[4]

Early applications of analyzing metadata from internet search queries and social media have been in the field of epidemiology and symptom surveillance of diseases. Eysenbach, G. (2006) demonstrates a high correlation between epidemiological data from the 2004/2005 flu season in Canada and the number of clicks on a keyword triggered link in Google Adsense. Polgreen et al

[3] Google Trends provides weekly data that is updated daily.

[4] In this case the expectations are revealed expectations, in contrast to survey measures, which are stated expectations or bond spreads which are market-implied expectations.

(2008) examine Yahoo search queries and find a high correlation of the words "flu" and "influenza" with virologic and mortality surveillance data over several years. Ginsberg et. al. (2009) demonstrate that Google search query data can predict actual influenza rates 1-2 weeks in advance of the publication of the US Center for Disease Control's Influenza Sentinel Provider Surveillance Network. Ritterman et al (2009) use data mined from Twitter posts to demonstrate that the public's expectations regarding a future event (i.e., a swine flu pandemic) can be extracted from social media data and those public perceptions can be used to reduce the forecast error of implied probabilities obtained from prediction markets where participants wager on the likelihood of a future event.

Ground-breaking work by Choi and Varian (2009a) examines the use of Google Trends data to nowcast economic quantities, i.e., to estimate present economic activity that is subject to data reporting lags. They demonstrate how search query data can be used to estimate retail sales, automotive sales, home sales, and travel patterns. In subsequent work (2009b), the authors show that Google Trends can be used to predict initial claims for unemployment benefits. Similarly, other authors demonstrate the predictive power of Google job search query data in forecasting unemployment in the US (D'Amuri and Marcucci 2010), Germany (Askitas and Zimmermann 2009), Israel (Suhoy 2009), Norway (Anvik and Gjelstad 2010) and Italy (D'Amuri 2009). Google Trends data also has predictive power in nowcasting private consumption for Israel (Suhoy 2010) and in forecasting consumption for the US (Vosen and Schmidt 2011), with the authors finding that Google search query data outperforms survey-based indicators. Other researchers demonstrate the predictive power of search query data with applications to housing prices (Kulkarni et al 2009), politics and social trends (Constant and Zimmermann 2009), tourism (Song et al 2009), and entertainment (Goel et al 2010).

This paper is the first to examine inflation expectations derived from internet search query data. The importance of inflation expectations, for the real economy as well as for financial markets, cannot be overstated. Inflation expectations play a critical role in the Federal Reserve's determination of monetary policy and in establishing the Fed's credibility among market participants. Expectations of inflation are embedded in the investment and financing decisions of firms, the labor contract negotiations of managers and employees, and the consumption, investment, and savings decisions of individuals. For decades, economists have relied on a standard set of three survey measures of expected inflation, namely the semiannual Livingston Survey, the quarterly Survey of Professional Forecasters, and the quarterly Michigan Surveys of Consumers.[5] These three survey measures have been instrumental in developing models of inflation expectations formation, and in testing the rational expectations hypothesis.[6]

However, given the importance of inflation expectations, and the considerable attention the subject has received in the academic literature, it is somewhat surprising that economists have not made significant efforts to look beyond the standard set of three surveys to develop a more comprehensive set of measures to gauge inflation expectations. In particular, it seems odd that in a world driven by real-time information, economists are still relying on quarterly and semi-annual measures of inflation expectations, when higher frequency measures exist and are readily available. Furthermore, the internet and social media provide an unprecedented glimpse into the public psyche, providing researchers with an exciting and novel way to measure expectations

[5] Data were quarterly prior to 1976 and monthly thereafter, but most researchers use only the quarterly data in models of inflation forecasting.

[6] Term-structure models, ARIMA time series models, and Phillips Curve motivated models of inflation expectations are important tools as well, but those are not emphasized here since the focus is mainly on survey expectations.

revealed by human behavior in real-time. As the Federal Reserve's controversial "quantitative easing" policy response to the 2008 financial crisis continues to stir vigorous debate, it behooves researchers to devote significant efforts to identifying and testing new sources of inflation expectations data.

In this paper, I examine a collection of 38 measures of inflation expectations, with the aim of comparing the performance of higher frequency and lower frequency measures. Of particular interest is how the metadata inflation expectations index constructed from Google search query data performs relative to the 36 semi-annual, quarterly, and monthly survey measures and the market-implied TIPS spread. I run a horserace between all the measures and compare their accuracy, predictive content, and rationality using a battery of tests set forth in Thomas (1999), Grant and Thomas (1999), and Mehra (2002). I examine three types of measures – numerical forecasts of the inflation rate (survey-based and market-implied), diffusion indexes of the expected direction of the price level (survey-based), and a metadata index of search queries.[7] The predictive power of each type of measure is assessed with a test for Granger causality. Rationality for all measures is evaluated with tests for unbiasedness and efficiency. Out-of-sample forecast tests are also conducted. The analysis is extended for the case of numerical forecasts, which are tested for accuracy over the entire sample period by comparing summary statistics of the forecast errors. Finally, a test of equal forecast accuracy is performed to evaluate competing numerical forecasts.

[7] Diffusion indexes are useful in determining expected changes in the direction of the price level, either used as stand-alone measures or by extracting a common dynamic factor from a group of diffusion indexes, as in Stock and Watson (2002) and Guzmán (2003) and (2009).

The paper is organized as follows: Section II provides a brief review of related literature, Section III contains a description of the inflation expectations measures and indicator construction, Section IV describes the methodology and presents the results, and Section V concludes.

II. Literature Review

For decades, the academic literature has devoted significant efforts to developing and evaluating methods of forecasting inflation. In addition to other methods of forecasting inflation, a large body of literature has evolved on the subject of survey-based inflation expectations, with researchers debating and discussing the rationality, accuracy, and predictive power of these measures. The vast majority of these studies focus on three surveys: the Livingston Survey, the Michigan Survey, and the Survey of Professional Forecasters (SPF). Thomas (1999) examines consensus forecasts of economists from the Livingston Survey and households from the Michigan Survey, and finds that these surveys outperform benchmark forecasts generated by a naïve model of lagged inflation and by the Fisher relation. In addition, households outperform economists in tests of accuracy and unbiasedness. Grant and Thomas (1999) provide evidence that the Livingston and Michigan survey measures of inflation expectations are cointegrated with inflation realizations, supporting weak-form rationality of these survey respondents. Mehra (2002) examines the accuracy, predictive content, and rationality of the Livingston, Michigan, and SPF surveys, and reports that Michigan outperforms Livingston and SPF.

The Phillips curve has long been a standard tool for economists in forecasting inflation. Stock and Watson (1999) present an authoritative study of Phillips curve models and find that inflation forecasts generated by the Phillips curve produce the most accurate and reliable forecasts over the 1970-1996 period, compared with inflation forecasts using other macroeconomic variables and economic indicators. In addition, the authors find that the best-performing Phillips curve specification is one that uses a new composite index of aggregate economic activity comprising 168 individual activity measures, including surveys.

Indeed, an extensive literature has evolved on empirical factor models that exploit information from large data sets to predict key economic quantities such as inflation. Stock and Watson (2002) show that, when compared to standard benchmark models such as autoregressive, leading indicator, Phillips curve, and vector autoregressive models, the best forecast of inflation is obtained from a model employing lagged inflation and a single composite factor, constructed from a large set of indicators, including surveys. Other researchers, such as Guzmán (2003) have demonstrated that composite factors extracted from large data sets that include surveys along with other macroeconomic indicators can be effectively used to forecast aggregate stock returns. Guzmán (2009) shows how a composite factor constructed from a collection of surveys can improve both nowcasts and forecasts of aggregate stock returns as well as GDP growth. Similarly, Giannone, Reichlin, and Small (2008) show that composite factors obtained from high-frequency macroeconomic indicators and soft information such as surveys can significantly improve both nowcasts and forecasts of GDP growth. Indeed, surveys are gaining credibility as an important economic forecasting tool.

Economists currently rely on four primary methods of forecasting inflation: time series ARIMA models, forecasting regressions using variables motivated by the Phillips curve, term structure models, and inflation expectations derived from surveys of households and economists. Presumably, those economists participating as survey respondents are using some variation of the three non-survey methods to forecast inflation. Ang, Bekaert, and Wei (2007) compare and contrast these four methods of inflation forecasting and find that surveys outperform the other three methods. Adjustments to account for linear and non-linear bias in the survey data produce worse out-of-sample forecasting results than using the unadjusted survey median forecasts. In addition, the authors investigate models of combined forecasts and find that surveys outperform other model combinations, and when combined with other forecasts, the data tend to overweight survey forecasts and underweight the other forecasting methods.

However, Ang et. al. (2007), like others before them, examine only the three standard low frequency surveys – the quarterly Michigan Survey, the quarterly Survey of Professional Forecasters, and the semiannual Livingston Survey. Because of the long tradition these surveys have in the academic literature, many researchers are mistakenly under the impression that these three surveys are the only available surveys containing data on inflation expectations. In fact, there are many different survey measures of U.S. inflation expectations covering a wide range of respondent universes – households, businesses, economists, investors, manufacturers, and retailers.[8]

Of particular interest here is whether revealed expectations obtained from internet search behavior can help predict inflation better than the existing tools, namely surveys and yield curve measures. Yield curve measures of inflation expectations can be tainted by liquidity premium,

[8] Some of the survey expectations data have higher frequencies but shorter histories, and others relate to inflation expectations for categories such as vehicles, housing, retail sales and wages. While some may argue that these are not directly comparable to the aggregate price level, they are evaluated in this context because they are important components of the information set available to economic agents, and frequently ignored in the academic literature.

preferred habitat, demand forces, supply constraints, and other factors. Survey-derived measures of inflation expectations have a long tradition in the academic literature, but survey-derived expectations are frequently criticized because the data are considered subjective, and respondents may not have any incentive to answer truthfully. Much of the criticism surrounding survey data is overblown, and countless studies have demonstrated the value and predictive power of survey data. However, internet search behavior data may possess a relative advantage over survey data in that search behavior is an objective measure of expectations, unfettered by the possibility of framing effects associated with the manner that survey questions are posed, or potentially dubious motives of survey respondents.

III.a. Description of Inflation Expectations Measures

I examine a total of 38 measures of inflation expectations, obtained from 15 data sources. There are 36 survey measures, one market-implied yield curve measure, and one metadata measure. Table 1 contains the complete list of indicators. A description of each data source is given in Appendix A.

III.b. Indicator Construction

[Insert Equations 1-3 here]

IV. Methodology and Results

IV.a. Predictive Power

The predictive content of each of the inflation expectations indicators is measured by a test of Granger Causality. This test evaluates the possibility that inflation expectations and inflation realizations may be cointegrated, in the sense of Engle and Granger (1987). One would expect that actual inflation rates might influence inflation expectations. But, if inflation expectations influence the future rate of inflation, this would be of significant interest to policymakers, as it implies a bilateral feedback effect between inflation expectations and inflation realizations.

The tests for Granger Causality are specified as follows:

[Insert Equations 4-5 here]

Table 2 shows that, at 3 lags, the actual inflation rate influences most of the measures of inflation expectations, and this is not a surprise, as one would expect agents to form expectations based in part on recent past experience. What is intriguing is that several of the measures of inflation expectations influence the future actual inflation rate. In this case the null hypothesis for the absence of Granger Causality is rejected. Significant predictive power is demonstrated by the following measures of inflation expectations: NFIB Small Business 3-month Price Plans, Richmond Fed 6-month Retail Prices, Michigan Vehicles Price Conditions, Blue Chip 1-year CPI forecast, Survey of Professional Forecasters 1-year CPI forecast, Livingston 1-year Median CPI forecast and the Michigan 5-year mean inflation forecast. Since many of these indicators are available at a monthly frequency, there is a clear advantage to using monthly indicators in addition to or instead of the quarterly and semi-annual frequency measures.

Table 3 shows that, at 12 lags, the actual inflation rate once again influences many of the measures of inflation expectations. In addition, several measures of inflation expectations demonstrate predictive power over the actual future inflation rate. The Google Inflation Search Index, the Livingston 6-month mean PPI forecast, the Blue Chip 1-year CPI forecast, the Survey of Professional Forecasters 1-year CPI forecast and the Michigan median 1-year inflation forecast all demonstrate a statistically significant ability to anticipate the future actual inflation rate.

It is interesting to note that only the Google Inflation Search Index anticipates the actual inflation rate without bilateral feedback. This indicates that inflation expectations derived from internet search queries are not influenced by the actual rate of inflation, but are useful in forecasting future inflation. Given that the GISI is a real-time measure of inflation expectations, it could be a useful tool in developing a real-time inflation forecasting model.

The results of the Granger Causality tests lend support to some alternative theoretical macroeconomic models. For instance, the finding that inflation expectations of businesses and retailers Granger cause future inflation rates makes sense to the extent that there may exist strategic complementarities between the price-setting decisions of manufacturers or suppliers of different goods, in the sense suggested by Calvo (1983). This theory of pricing can justify an aggregate supply relation that takes the form of an expectations-augmented Phillips curve, where the location of the short-run Phillips curve is determined by expectations regarding future inflation.

Indeed, in many macroeconomic models of the New Keynesian variety, current inflation is mainly determined by current expectations of future inflation. This is because price-setters will optimally adjust their prices such that current prices reflect a mark-up above their expected average nominal marginal costs for the duration that prices are expected to remain fixed.

Therefore, expected future inflation will affect current inflation because current prices are aligned with average expected future nominal marginal costs. Thus, inflation expectations can lead to self-fulfilling deflations or inflations, i.e., there is convergence to a rational-expectations equilibrium as a result of adaptive learning dynamics.[9]

Alternatively, the results could be explained by a sticky information model as proposed by Mankiw and Reis (2002), rather than the sticky prices underlying the New Keynesian models. In the sticky information model, current inflation is determined by past expectations of current inflation.[10] Some researchers, most notably Carroll (2003) and Lanne, Luoma, and Luoto (2009), argue that the inflation expectations data from the Michigan Survey is consistent with a sticky information model and that agents are slow to update their beliefs, thus providing the microfoundations for the model proposed by Mankiw and Reis.

Finally, another alternative explanation for the Granger Causality results could be that the apparent cointegration could be explained by a common shock affecting both current and future inflation. For example, even if actual inflation and expected inflation are unrelated, a commodity price shock could induce a revision of current expectations for inflation one year hence, and also affect inflation every month from then on. While this explanation is possible, it is not probable here due to the fact that most of the sample period occurs over a time span during which there was no major commodity price shock.

---

[9] Woodford (2003)

[10] In a sense, the sticky information model is like a Phillips curve with backward-looking expectations instead or forward-looking expectations.

IV.b. Rationality – Unbiasedness

Thomas (1999) states: "If inflation expectations are fully rational, they should exhibit two fundamental characteristics. First, they should be unbiased – that is, agents should forecast inflation correctly on average. Second, forecasts should be efficient – that is, agents should employ all relevant information for which the marginal benefit of gathering and utilizing the information exceeds the marginal cost."

The test for unbiasedness is estimated by OLS and specified as follows:[11]

[Insert Equation 6 here]

The equation is estimated by regressing the actual inflation rate π_t on the previously made forecast of inflation π^e_t and testing the joint null hypothesis that $\alpha=0$ and $\beta=1$. Forecasts are considered unbiased if the null hypothesis cannot be rejected. The joint null hypothesis is tested with a Chi-squared test. The Chi-squared test only applies to the numerical forecasts, since the hypothesis that $\beta=1$ would be meaningless for a diffusion or metadata index.

Table 4 contains the results of the test for unbiasedness, where the joint null hypothesis $\alpha=0$ and $\beta=1$ is tested for Equation (6), for the 22 numerical forecasts of inflation expectations.[12] The results of the Chi-squared tests indicate that the null hypothesis is decisively rejected for 21 of the 22 numerical forecasts. This means that each of these 21 indicators systematically either overestimate or underestimate the actual inflation rate. The only measure of inflation expectations where the null hypothesis fails to be rejected is the Blue Chip Indicators Survey

[11] Model is estimated by OLS with Newey-West HAC standard errors with lag truncation parameter set to equal forecast horizon in order to avoid overlapping standard errors.

[12] The Chi-squared test is not applicable to the diffusion indexes or the metadata index.

one-year forecast for CPI inflation. In this case, the Chi-squared p-value is 0.3093, and we fail to reject the joint null hypothesis $\alpha=0$ and $\beta=1$.

IV.c. Rationality – Efficiency

The test for efficiency is estimated by OLS and specified as follows:[13]

[Insert Equation 7 here]

The equation is estimated by regressing the forecast error e_t on the information set I_t either individually or jointly. The information set includes those variables that are pertinent to a comprehensive model of inflation. The variables are tested for significance first individually, then jointly. If any or all of the variables in the information set are significantly negatively correlated with the forecast error, this implies that agents failed to take all relevant information into account when forming their inflation expectations. Weak-form efficiency implies that agents have taken into consideration only the information contained in past inflation rates, while strong-form efficiency implies that agents have considered information about all variables that are germane to forecasting inflation.

Following Thomas (1999) and Mehra (2002), the variables employed in the information set I_t are: the lagged 12-month rate of CPI inflation, a measure for the output gap, M1 and M2 growth, and a measure for oil price inflation. Since most of the data have a monthly frequency, the unemployment rate is used as a proxy for the output gap, with this substitution following

[13] Model is estimated by OLS with Newey-West HAC standard errors with lag truncation parameter set to equal forecast horizon in order to avoid overlapping standard errors.

Gramlich (1983). The measure for oil price inflation is the lagged 12-month rate of change for the producer price index for fuels. A description of the economic variables and data sources is contained in Appendix B.

Tables 5 – 10 present results for the tests for efficiency, to find out if agents employed relevant information in forming inflation expectations. In this test, forecast errors are regressed on inflation-related variables to determine if there is a correlation. The variables are first tested separately and then together in a joint specification. Testing whether agents used knowledge of lagged inflation in forming expectations is a test of weak-form efficiency. To test for strong-form efficiency, four variables are tested: the unemployment rate (a substitute measure for the output gap), the lagged 12-month growth rate of the narrow (M1) and broad (M2) monetary aggregates, and a measure for energy price inflation (the 12-month rate of change of the producer price index for fuels). In each case, the independent variable is defined so that failure of agents to take account of the variable in the manner suggested by conventional economic theory would result in a negative and statistically significant coefficient on the variable.[14] That is, if agents fail to account for past inflation, money growth, etc., they would underestimate inflation and have a negative forecasting error, resulting in a negative sign on the coefficient for the variable. Conversely, if agents account too much for past inflation, money growth, etc., they would overestimate inflation and have a positive forecasting error, resulting in a positive sign on the coefficient for the variable.

Table 5 contains the results for the efficiency test with respect to the most recent 12-month rate of CPI inflation known to agents at the time the inflation expectations are measured. The table indicates that most agents adequately took into account the past rate of CPI inflation, but

[14] Thomas (1999)

respondents to some surveys did not. Specifically, NFIB Small Business Price Plans, the Philadelphia Fed's, Dallas Fed's, New York Fed's, and Kansas City Fed's expectations for Prices Paid, and the Livingston Mean and Median CPI forecasts all failed to consider adequately the lagged inflation rate in forming inflation expectations. Due to the insufficient use of information concerning the past inflation rate, weak-form efficiency can be rejected for these measures of inflation expectations.

Conversely, Table 5 indicates that some measures of inflation expectations attributed too much influence to the past CPI inflation rate, resulting in a positive forecasting error. The Richmond Fed's survey expectations for 6-month prices paid, prices received, retail prices, non-retail prices, and services prices all have a positive and statistically significant coefficient on lagged CPI. Similarly, the Blue Chip 1-year GDP deflator and 1-year CPI forecasts, the Michigan 1-year median and 5-year mean and median inflation forecasts, the Survey of Professional Forecasters 1-year CPI, and the Livingston survey's 6-month mean and median CPI and PPI forecasts all have forecast errors that are positively correlated with the lagged inflation rate. This indicates that respondents to these surveys overestimated the impact of past inflation when forming their expectations for future inflation.

Table 5 indicates that weak-form efficiency is supported for the majority of the higher frequency inflation expectations measures. The Google Inflation Search Indicator, Philadelphia Fed 6-month prices received, Richmond Fed 6-month wages, Kansas City Fed 6-months prices received, New York Fed 6-months prices received, Dallas Fed 6-months prices received, and Dallas Fed 6-month wages, are all measures of inflation expectations that adequately took account of the lagged CPI inflation rate. The same is true for the Michigan Survey's price conditions for durable goods, vehicles, and housing. In addition, the UBS/Gallup 1-year inflation forecast, the Michigan 1-year mean inflation forecasts, the TIPS spread, the Conference Board 1-year inflation forecast, and the Livingston survey's 1-year mean and median PPI forecasts are also weak-form efficient measures of inflation expectations.

Table 6 presents the results of the efficiency test with respect to the lagged 12-month growth rate of the narrow monetary aggregate (M1). Expectations from the Blue Chip 1-year GDP deflator, Michigan 1-year median inflation and Livingston 1-year mean and median CPI forecasts all have negative and statistically significant coefficients, meaning that they fail to take sufficient account of M1 growth. Because these survey measures failed to take adequate account of M1 growth, strong-form efficiency can be rejected for these measures of inflation expectations.

Conversely, the Kansas City Fed's 6-month prices received, New York Fed's 6-month prices paid, and the Michigan 5-year mean and median inflation forecasts all have forecast errors that are positively correlated with lagged M1 growth, indicating that respondents to these surveys overestimated the influence of lagged M1 growth when forming their inflation expectations.

Strong-form efficiency with respect to lagged M1 growth is supported for several measures of inflation expectations. The Google Inflation Search Index, the NFIB Small Business 3-month price plans, the Philadelphia Fed's 6-month prices paid and prices received, the Richmond Fed's 6-month prices paid, prices received, retail prices, non-retail prices, services prices, and wages all take into account the lagged growth rate of the narrow monetary aggregate. The Kansas City Fed's 6-month prices paid, the New York Fed's 6-month prices received, the Dallas Fed's 6-month prices paid, prices received, and wages, and the Livingston Survey's 6-month mean and median CPI and PPI also efficiently incorporate information about lagged M1 growth, as do the Michigan Survey's price conditions for durable goods, vehicles, and housing. The UBS/Gallup

1-year inflation forecast, the Michigan 1-year mean inflation, the Blue Chip 1-year CPI forecast, the TIPS spread, the Conference Board 1-year inflation forecast, the SPF 1-year CPI forecast, and the Livingston survey mean and median 1-year PPI forecasts adequately take M1 growth into account as well. Strong-form efficiency is supported for all these measures of inflation expectations.

Table 7 contains the results of the efficiency test with respect to the lagged 12-month growth rate of the broad monetary aggregate (M2). The Richmond Fed Survey's expectations for 6-month prices paid and prices received, the Livingston Survey's 1-year mean and median CPI expectations, and the Michigan 5-year mean and median inflation forecasts all fail to take proper account of the lagged 12-month growth rate of the broad monetary aggregate, as indicated by the significant negative correlation between the forecasting error and lagged M2 growth. Because these survey measures failed to take adequate account of M2 growth, strong-form efficiency can be rejected for these measures of inflation expectations.

Conversely, the Philadelphia Federal Reserve's 6-month prices received and the Richmond Fed's 6-month Retail prices, as well as the Michigan Survey's durable goods and housing price conditions, and the UBS/Gallup 1-year inflation forecast are measures of inflation expectations with forecast errors that are positively significantly correlated with M2 growth, suggesting that forecasters attributed too much influence of M2 growth on the future inflation rate.

Strong-form efficiency with respect to M2 growth is supported for several of the higher frequency measures of inflation expectations. The Google Inflation Search Index, NFIB Small Business price plans, Philadelphia Fed's 6-month prices paid, Richmond Fed's 6-month non-retail prices, services prices, and wages, Kansas City Fed's 6-month prices paid and prices received, New York Fed's 6-month prices paid and prices received, and the Dallas Fed's 6-month prices paid, prices received, and wages all efficiently incorporate information about M2 growth, thus exhibiting strong-form efficiency. Similarly, the Michigan Survey's price conditions for vehicles, the Blue Chip Survey's 1-year forecast for the GDP deflator and CPI, the Michigan Survey 1-year mean and median inflation forecast, the TIPS spread, the Conference Board Survey's 1-year inflation forecast, the Survey of Professional Forecasters 1-year CPI forecast, and the Livingston Survey's mean and median 6-month CPI and PPI, and mean and median 1-year PPI forecasts are also strong-form efficient with respect to M2 growth.

The results for the efficiency test with respect to oil price inflation are displayed in Table 8. Survey expectations for 6-month prices paid from neither the Philadelphia Fed, nor the Kansas City Fed, nor the Dallas Fed adequately took into account the lagged oil price inflation, as indicated by the negative and statistically significant coefficient. Due to the inadequate use of information concerning energy price inflation, strong-form efficiency can be rejected for these survey measures of inflation expectations.

Conversely, a positive correlation between oil price inflation and the forecast error is noted for the Richmond Fed's 6-month prices paid, prices received, retail, non-retail, and services prices, the New York Fed's 6-month prices received, and the Livingston Survey's 6-month mean and median CPI forecasts, indicating that these measures of inflation expectations attributed too much importance to oil price inflation in forming expectations for future inflation.

Several of the higher frequency measures demonstrate strong-form efficiency with respect to oil price inflation. The Google Inflation Search Index, NFIB Small Business price plans, Philadelphia Fed's 6-month prices received, Richmond Fed's 6-month wage expectations, Kansas City Fed's 6-month prices received, New York Fed's 6-month prices paid, and the Dallas Fed's 6-month prices received and wage expectations all adequately took account of oil price inflation in forming expectations for future inflation. The Michigan Survey's price conditions for durable goods, vehicles, and housing prices, as well as the mean and median 1-year and 5-year inflation forecasts, also sufficiently incorporate information regarding oil price inflation, thereby exhibiting strong-form efficiency. The Blue Chip 1-year GDP deflator and CPI forecasts, the UBS/Gallup 1-year inflation forecast, the TIPS spread, the Conference Board 1-year inflation forecast, the SPF 1-year CPI forecast, and the Livingston Survey's mean and median 6-month PPI, and 1-year PPI and CPI forecasts are also strong-form efficient with respect to oil price inflation.

Table 9 presents the results for the efficiency test with respect to the lagged unemployment rate, a proxy for the output gap. The table indicates that none of the Richmond Fed's measures of inflation expectations for the services sector effectively incorporates information about the unemployment rate. The Richmond Federal Reserve Surveys of expectations for retail prices, non-retail prices, and service sector prices all have a negative and statistically significant coefficient on the lagged unemployment rate. Due to the inadequate use of information concerning the unemployment rate, strong-form efficiency can be rejected for the Richmond Fed's service sector surveys.

Conversely, the NFIB Small Business price plans, the Blue Chip 1-year CPI forecast, the Survey of Professional Forecasters 1-year CPI forecast, and the Michigan 5-year mean and median inflation forecasts, all have forecast errors that are positively correlated with the unemployment rate, suggesting that forecasters attributed too much influence from the unemployment rate on their forecasts of future inflation.

Strong-form efficiency with respect to the unemployment rate is indicated for several of the higher frequency measures. The Google Inflation Search Index, the Philadelphia, Kansas City, and New York Fed's 6-month prices paid and prices received, and the Dallas and Richmond Fed's 6-month prices paid, prices received, and wages all efficiently incorporated information about the unemployment rate in forming inflation expectations. Additionally, the Michigan survey's price conditions for durable goods, vehicles, housing, 1-year mean and median inflation forecasts, the Blue Chip 1-year GDP deflator forecast, the UBS/Gallup 1-year inflation forecast, the TIPS spread, the Conference Board 1-year inflation forecast, and the Livingston Survey's mean and median 6-month and 1-year CPI and PPI forecasts also display strong-form efficiency with respect to the lagged unemployment rate.

Table 10 presents the results for the efficiency test using the joint specification, with the lagged CPI, M1 and M2 growth, oil price inflation, and unemployment rate tested together in one equation. The table indicates that most of the measures do not efficiently incorporate information from all of these variables simultaneously, refuting strong-form efficiency for 36 of the 38 indicators tested. Note that only the Conference Board Survey 1-year inflation expectations and the Michigan Survey median 1-year inflation expectations pass the joint specification test with statistical significance, indicating strong-form efficiency for these two survey measures.

IV.d. Out-of-Sample Forecasts

Out-of-sample forecasts of all the inflation expectations indicators are performed using a basic predictive model for actual inflation regressed on expected inflation and past inflation. Due to the high serial correlation in the rate of inflation, the model is specified to test whether the survey forecasts have any predictive power for the future rate of inflation beyond the information contained in past inflation data. The model is estimated by OLS as follows:[15]

[Insert Equation 8 here]

A forecast is produced by estimating parameters using data available through December 2005, for the particular measure of inflation best predicted by the expectations measure over the corresponding horizon. The estimated parameters are then used to fit the equation in forecasts over the out-of-sample period, January 2006 to October 2008.[16] The forecasts are then evaluated by comparing the Root Mean Squared Forecast Errors (RMSFE) to determine the accuracy of the forecasts over the respective horizon corresponding to the frequency of the indicator.

Table 11 presents results for out-of-sample forecasts using a basic predictive model for actual inflation regressed on expected inflation and past inflation. The most accurate out-of-sample forecast is the Google Inflation Search Index, with a RMSFE 0.2929 in forecasting PCE inflation. The next best is the Philadelphia Fed's 6-month expectations for prices received, with a RMSFE of 1.1989. The standard economists' data set does not perform as well, with the SPF 1-year CPI forecast registering a RMSFE of 1.5430, the Michigan Mean 1-year inflation forecast registering a RMSFE of 2.5405 and the Livingston mean 1-year CPI forecast registering a

[15] Model is estimated by OLS with Newey-West HAC standard errors with lag truncation parameter set to equal forecast horizon in order to avoid overlapping standard errors.

[16] Out-of-sample forecasts are not performed for the Michigan 5-year inflation forecasts due to insufficient data.

RMSFE of 3.4935. Most of the higher frequency monthly measures of inflation expectations significantly outperform the standard lower frequency quarterly and semiannual survey measures, indicating that there are benefits to using higher frequency data. In particular, there are significant benefits to using a metadata index constructed from internet search queries on the keyword "inflation".

IV.e. Accuracy of Numerical Forecasts

[Insert Section IV.e. here]

IV.f. Numerical Forecast Comparison Tests

With measures of numerical inflation forecasts from so many different sources, it is inevitable that there will be apparent differences in forecast accuracy within the sample. This raises the question as to whether the outcome is due to pure chance. A test of equal predictive accuracy is performed to determine whether these observed differences in numerical forecasts are statistically significant or not.

Since the information set is limited, i.e., available data only include a set of forecasts and actual values of the predictand, a model-free test is appropriate. I employ a variant of the Morgan-Granger-Newbold (1977) (MGN) test, proposed by Harvey, Leybourne, and Newbold (1997)

(HLN). The test will allow an objective evaluation of the forecast accuracy of each of the numerical forecasts and determine whether the observed differences are due to chance or due to superior forecasting ability. The methodology is described as follows.

Let y_t represent the actual values of inflation observed for period $t = 1, 2, 3, \ldots T$, and let $\hat{y}_{it}$ represent the forecast of indicator i for $i = 1, 2$. Then the forecast error is defined as:

$$e_{it} = \hat{y}_{it} - y_t \tag{12}$$

The loss depends on forecast and actual values only through the forecast error:

$$g(y_t, \hat{y}_{it}) = g(\hat{y}_{it} - y_t) = g(e_{it}) \tag{13}$$

And the loss differential between the two forecasts is given by:

$$d(t) = g(e_{1t}) - g(e_{2t}) \tag{14}$$

Two forecasts have equal accuracy if and only if the loss differential has zero expectation for all t. Therefore, we can test the null hypothesis:

H_0: $E(d_t) = 0$ *for all* t

versus the alternative hypothesis:

H_1: $E(d_t) = \mu \neq 0$.[17]

[17] Mariano and Preve (2008).

The MGN test assumes that: A(1) the loss is quadratic; and A(2) forecast errors are (a) zero mean, (b) Gaussian, and (c) serially uncorrelated. Then, the null hypothesis of equality of forecast mean squared errors is equivalent to equality of forecast error variances.

Let:

$$x_t = e_{1t} - e_{2t} \tag{15}$$

$$z_t = e_{1t} + e_{2t} \tag{16}$$

The MGN test statistic is given by:

$$S = [(1-r^2)/(n-1)]^{-1/2}\, r \tag{17}$$

where

$$r = [\Sigma x_t^2 \Sigma z_t^2]^{-1/2}\, \Sigma x_t z_t$$

The HLN test casts the MGN test in a regression framework:

$$z_t = \beta x_t + \varepsilon_t \tag{18}$$

such that (18) is identical to the null hypothesis that $\beta = 0$, i.e.:

$$S = [\hat{\sigma}^2/\Sigma x_t^2]^{-1/2}\hat{\beta} \tag{19}$$

where

$$\hat{\beta} = \Sigma x_t z_t / \Sigma x_t^2 \qquad \text{and} \qquad \hat{\sigma}^2 = (n-1)^{-1}\, \Sigma(z_t - \hat{\beta} x_t)^2$$

and $\hat{\sigma}^2$ estimates the variance σ^2 of ε_t.[18]

I perform the HLN test of equal forecast accuracy by designating a benchmark forecast for each of the 12-month CPI, PPI, and PCE tests.[19] The benchmark selection rule is as follows: the monthly indicator with the median RMSE is chosen, and in the case of a tie, i.e., if there is an even number of indicators, then the series with the longest available history is chosen. The benchmark forecast is then converted to either quarterly or semiannual frequency as needed to perform the regression test.

The HLN (1997) variation of the MGN (1977) test is performed to test for equal forecasting accuracy, i.e., equality of forecast error variances. Table 15 presents results of the HLN test of numerical forecasts for the CPI over a 12-month horizon. The benchmark selection rule indicates that the Michigan Survey's 1-year median inflation forecast is the benchmark measure. The null hypothesis of $\beta = 0$ can be decisively rejected for all the measures of inflation expectations except for the Survey of Professional Forecasters. This means that the null hypothesis of equal forecasting accuracy is rejected for the majority of the measures.

Table 16 presents results of the HLN test of numerical forecasts for the PCE over a 12-month horizon. The benchmark selection rule indicates that the Michigan Survey's 1-year median inflation forecast is again the benchmark measure. The null hypothesis of $\beta = 0$ is rejected for the Blue Chip GDP Deflator, the Blue Chip CPI, the TIPS Spread, Michigan 1-year mean, and

[18] Harvey, Leybourne, and Newbold (1997).

[19] The HLN test is not performed for the 6-month or 5-year numerical forecasts due to insufficient observations.

the Conference Board 1-year inflation forecasts. The null hypothesis fails to be rejected for the SPF 1-year CPI forecast and all of the Livingston forecasts.

Table 17 presents results of the HLN test of numerical forecasts for the PPI over a 12-month horizon. The benchmark selection rule indicates that the Blue Chip 1-year CPI inflation forecast is the benchmark measure. The null hypothesis of $\beta = 0$ is rejected for the Blue Chip GDP Deflator, the Michigan 1-year median, Michigan 1-year mean, and the Conference Board 1-year inflation forecasts. The null hypothesis fails to be rejected for the SPF 1-year CPI forecast, the TIPS Spread, and all of the Livingston forecasts.

V. Conclusion

I have shown that higher frequency measures of inflation expectations tend to outperform the standard three lower frequency surveys – the quarterly Michigan Survey, the quarterly Survey of Professional Forecasters, and the semiannual Livingston Survey – in terms of accuracy, predictive power, rationality, and out-of-sample forecasts. While there is no single winner that consistently outperforms all of the other measures on the complete battery of tests, the results indicate that several of the monthly surveys conducted by the regional Federal Reserve banks perform well, as do the NFIB Small Business Survey, the Conference Board Survey, and the Blue Chip Survey. It is worth noting that the Blue Chip Survey is the only indicator that passes the test for unbiasedness. Particularly disappointing is the performance of the TIPS spread, given the degree to which many financial markets professionals rely upon it. It must be emphasized that TIPS have a managed supply curve, therefore the TIPS spread is not a clean measure of inflation expectations and cannot be treated as one.

What is interesting is that many of the surveys that are not typically used in the academic literature perform better relative to those that are typically used. Other authors have found that inflation forecasts from the standard three low frequency surveys outperform inflation forecasts generated by time series ARIMA models, regression models using Phillips curve-derived real activity measures, and term-structure models. Since the higher frequency surveys examined in this paper outperform the standard three lower frequency surveys, it is likely that the higher frequency surveys would outperform inflation forecasts generated from the aforementioned alternative methods as well.

The most intriguing and impressive performance was given by the Google Inflation Search Index. In addition to significantly anticipating the rate of PCE inflation by 12 months and having the lowest out-of-sample forecast error of the indicators tested, the GISI passes the weak-form and strong-form efficiency tests for all variables when tested separately. While the measure does not pass the efficiency test using the joint specification, it is important to recognize that that the benefit of near real-time availability may outweigh the lack of strong-form efficiency in this respect. The measure also does not pass the unbiasedness test, suggesting that it may not be rational. Then again, no measure passes all of the rationality tests, and rationality of economic agents is arguably a dubious assumption.

The GISI indicator has two main advantages compared to the other measures of inflation expectations. The first advantage is timeliness, as Google Trends is updated on a daily basis, and could theoretically be made available in real-time via an API feed. Not only is the timelier data more advantageous for economic forecasting, but the improved availability resolves the timing issue associated with survey data releases. The second advantage is flexibility, since the data can easily be converted to lower frequency series for the purposes of traditional economic analysis, such as monthly frequency for estimating an expectations-augmented Phillips Curve relationship, or quarterly frequency for estimating the Taylor rule. Indeed, there are many potential applications in economics and finance for this type of metadata analysis.[20]

There is no doubt that the internet and social media are the vox populi of the modern era. These new channels of communication are the mirrors that reflect our collective social mood. People learn what others are thinking and express what they themselves are thinking by posting and promoting their thoughts and opinions on websites like Twitter, Facebook, and on blogs. People reveal what is on their minds by querying topics on search engines such as Google, Yahoo, and Bing. These activities together create a learning process and a continuous feedback loop, whereby people form, express, and reveal expectations in real-time.

If the revealed expectations derived from internet search query data have predictive power for future outcomes, then the symbiotic nature of social media and internet search behavior suggests that these channels may facilitate self-fulfilling prophecies. If the predicted outcomes relate to economic events, then there are potentially significant ramifications for investors, economists, and policy-makers alike. The internet and social media can become important tools for the Federal Reserve to gauge and manage inflation expectations towards achieving policy objectives.

[20] Preliminary tests using various metadata indicators to nowcast and forecast other economic quantities, as well as predict returns in the commodities and equities markets have yielded statistically significant results. Clearly, more research is warranted in this area.

Appendix A. Inflation Expectations Data - Description and Sources

[Insert Appendix A here]

Appendix B. Economic Data - Description and Sources

[Insert Appendix B here]

References

[1] Ang, A., Bekaert, G., and Wei, M., Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics* 54 (2007), 1163-1212.

[2] Anvik, C. and Gjelstad, K., "Just Google it." Forecasting Norwegian unemployment figures with web queries, *Center for Research in Economics and Management CREAM Publication* 11 (2010).

[3] Askitas, N. and Zimmerman, K., Google econometrics and unemployment forecasting, *Applied Economics Quarterly* 55 (2) (2009), 107-120.

[4] Calvo, G., Staggered prices in a utility-maximizing framework, *Journal of Monetary Economics* 12 (1983), 383-398.

[5] Carroll, C., Macroeconomic expectations of households and professional forecasters, *The Quarterly Journal of Economics* 118 (1) (2003), 269-298.

[6] Choi, H., and Varian, H., Predicting the present with Google Trends, www.google.com/googleblogs/.../google_predicting_the_present.pdf (2009a).

[7] Choi, H., and Varian, H., Predicting initial claims for unemployment benefits, research.google.com/archive/papers/initialclaimsUS.pdf (2009b).

[8] Constant, A. and Zimmermann, K., Face to face with the financial crisis: The U.S. presidential election from a transnational perspective, *German Institute for Economic Research Weekly Report* 16/2008 (4) (2009).

[9] D'Amuri, F., Predicting unemployment in short samples with internet job search query data, *MPRA Paper* 18403 (2009).

[10] D'Amuri, F. and Marcucci, J., "Google it!" Forecasting the US unemployment rate with a Google job search index, *Fondazione Eni Enrico Mattei Working Papers* 421 (2010).

[11] Engle, R. F., and Granger, C.W.J., Co-integration and error correction: Representation, estimation, and testing, *Econometrica* 55 (1987), 251-276.

[12] Eysenbach, G., Infodemiology: Tracking flu-related searches on the web for syndromic surveillance, *AMIA 2006 Symposium Proceedings* (2006), 244-248.

[13] Giannone, D., Reichlin, L., and Small, D., Nowcasting: The real-time informational content of macroeconomic data, *Journal of Monetary Economics* 55 (2008), 665-676.

[14] Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M., and Brilliant, L., Detecting influenza epidemics using search engine query data, *Nature* 457 (19) (2009).

[15] Goel, S., Hofman, J., Lahaie, S. Pennock, D., and Watts, D., Predicting consumer behavior with web search, *PNAS Early Edition* (2010).

[16] Gramlich, E.M., Models of inflation expectations formation: A comparison of household and economist forecasts, *Journal of Money, Credit, and Banking* 15 (2) (1983), 155-173.

[17] Granger, C.W.J., and Newbold, P., *Forecasting Economic Time Series*, Academic Press, Florida, 1977.

[18] Grant, A.P. and Thomas, L.B., Inflationary expectations and rationality revisited, *Economics Letters* 62 (1999), 331-338.

[19] Guzmán, G., *GDP Growth Signals, Investor Psychology, and Hedging Pressure: A Multivariate Approach to Forecasting Returns on the S&P500 Index*, Masters Degree Independent Study, The Wharton School, University of Pennsylvania, 2003.

[20] Guzmán, G., Using sentiment surveys to predict GDP growth and stock returns, in: *The Making of National Economic Forecasts*, L. R. Klein, ed., Edward Elgar Publishing Ltd., London, 2009, pp. 319-351.

[21] Harvey, D., Leybourne, S., and Newbold, P., Testing the equality of prediction mean squared errors, *International Journal of Forecasting* 13 (1997), 281-291.

[22] Kulkarni, R., Haynes, K., Stough, R., and Paelinck, J., Forecasting housing prices with Google econometrics, *George Mason University School of Public Policy Research Paper* 10 (2009).

[23] Lanne, M., Luoma, A. and Luoto, J., A naïve sticky information model of households' inflation expectations, *Journal of Economic Dynamics & Control* 33 (2009), 1332-1344.

[24] Mankiw, G. and Reis, R., Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips Curve, *The Quarterly Journal of Economics* 117 (4) (2002), 1295-1328.

[25] Mariano, R. and Preve, D., *Statistical Tests for Multiple Forecast Comparison,* Presentation, T.W. Anderson Conference, Stanford University, 2008.

[26] Mehra, Y.P., Survey measures of expected inflation: Revisiting the issues of predictive content and rationality, *Federal Reserve Bank of Richmond Economic Quarterly* 88/3 (2002), 17-35.

[27] Morgan, W.A., A test for significance of the difference between the two variances in a sample from a normal bivariate population, *Biometrika* 31 (1939), 13-19.

[28] Polgreen, P.M., Chen, Y., Pennock, D.M., & Forrest, N.D., Using internet searches for influenza surveillance, *Clinical Infectious Diseases* 47 (2008), 1443-1448.

[29] Ritterman, J., Osborne, M., and Klein, E., Using prediction markets and Twitter to predict a swine flu pandemic, *Proceedings of the 1st International Workshop on Mining Social Media* (2009).

[30] Song, H., Pan, B., and Ng, D., Forecasting demand for hotel rooms with search engine query volume data, *College of Charleston Working Paper* (2009).

[31] Stock, J.H. and Watson, M.W., Forecasting inflation, *Journal of Monetary Economics* 44 (1999), 293-335.

[32] Stock, J.H. and Watson, M.W., Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* 20 (2) (2002), 147-162.

[33] Suhoy, T., Query indices and a 2008 downturn: Israeli data, *Bank of Israel Discussion Paper* (2009).

[34] Suhoy, T., Monthly assessments of private consumption, *Bank of Israel Discussion Paper* (2010).

[35] Thomas, L.B., Survey measures of expected U.S. inflation, *The Journal of Economic Perspectives* 13 (4) (1999), 125-144.

[36] Vosen, S. and Schmidt, T., Forecasting private consumption: survey-based indicators vs. Google Trends, *Journal of Forecasting* doi: 10.1002/for.1213 (2011).

[37] Woodford, M., *Interest and Prices*, Princeton University Press, Princeton, 2003.