Extracting Information from Different Expectations

Andrew B. Martinez*

April 15, 2021

Abstract

Long-term expectations are believed to play a role in driving future inflation and guiding monetary policy responses. However, expectations are not directly observed and available measures can present a wide range of values. To understand what drives these differences, we examine the evolution of survey and market-based consumer price inflation expectations in the United States between 2003-2019. We show that inflation forecasts can be improved by incorporating the differential between survey and market-based measures. Next, we decompose and extract the differentials in rigidity and information. While both play a role, the information differential is most important and explains the overall forecast improvements. Using machine learning methods, we find that up to half of the information differential is explained by real-time changes in measures of liquidity. This information helps predict the divergence in long-term inflation expectations in 2020 as well as their increase in early 2021.

Keywords: Breakeven inflation, error correction, forecast encompassing, model selection

JEL classifications: C32, C52, C53, D84, E31, E37

*Office of Macroeconomic Analysis, U.S. Department of the Treasury; H.O. Stekler Research Program on Forecasting, The George Washington University; Climate Econometrics, Nuffield College. The views expressed here are my own and do not necessarily represent those of the Treasury Department or the U.S. Government. Thanks to Constantin Burgi, Jennifer L. Castle, James A. Girola, David T. Griffiths, David F. Hendry, Alex Schiboula, Tara Sinclair, Eileen Tipoe and to participants of the 21st IWH-GW-CIREQ Macroeconometric Workshop, the 2nd Vienna Workshop on Economic Forecasting, and the 32nd Dynamic Econometrics Conference for helpful comments and discussions. All results were obtained using Ox 8.1 and PcGive 15; see Doornik (2018) and Doornik and Hendry (2018). All errors are my own. Contact: Andrew.Martinez@treasury.gov.

1 Introduction

Long-term inflation expectations are believed to play an important role in driving and predicting future inflation. However, expectations are difficult to measure and there are many ways to do so. Alternative methods, which include surveys (of consumers, businesses, forecasters, etc.), financial-markets, and models thereof, produce a wide range of results. For example, since 2014 the Federal Reserve Bank of New York has conducted surveys of primary dealers' and market participants' expectations about 5-year, 5-year forward CPI inflation. Market participants' median expectations have never exceeded and on average are 0.12 percentage points lower than primary dealers (with a t-statistic greater than 10). More generally, differences were also pronounced during the spread of COVID-19 in 2020 when market-based measures declined sharply, surveys of professional forecasters were broadly unchanged, and surveys of consumers increased.

The dispersion between survey and market-based measures of inflation expectations raises important questions about their relative informativeness and whether they can be used to predict future inflation. In this paper, we analyze the differences between survey and market-based measures, their informativeness, and what might be driving them. We use a forecast-encompassing framework to understand whether the differential between surveys and markets can improve inflation forecasts. Next, we decompose and identify the information and rigidity differentials and then test for which channel matters for improving inflation forecasts. Finally, we use machine learning methods to select variables from multiple sources that best explain the information differential and then assess whether the selected information captures the recent evolution of the expectations differentials in 2020 and early 2021.

We find that the differential between survey and market-based expectations can be used to improve inflation forecasts. While the differential is driven by both the rigidity and information channels, the information differential closely captures most of the discrepancies between alternative measures of expectations and explains the overall forecast improvements due to differences between survey and market-based measures. The information differential is explained by measures of real-time changes in liquidity including money, stocks, and bank reserves. We show that differences in how measures respond to changes in liquidity captures most of the divergence and recovery in expectations during the spread of COVID-19.

Our findings relate to the literature comparing alternative measures of inflation expectations, which was sparked by Bernanke (2007) when he asked "which measure or combination of measures should central bankers focus to assess inflation developments?".[1] The initial response was in favor of surveys of forecasts from professional economists (see Ang et al., 2007) which have since been incorporated into many models

[1]See also Thomas (1999) for an earlier comparison of survey-based measures which favors professional forecasters.

to improve their forecasts; e.g. see Clark (2011), Kozicki and Tinsley (2012), Faust and Wright (2013), and Chan et al. (2018). However, recent findings support using measures of short-term consumer (Coibion et al., 2018 and Chen, 2019) or market-based (Kliesen, 2015, Grothe and Meyler, 2018 and de Mendonça et al., 2020) inflation expectations.[2] This coincides with a decline in the usefulness of survey-based inflation expectations (see Trehan, 2015 and Berge, 2018) and as a result, policy-makers pay attention to several measures of inflation expectations albeit with different emphasis (see Bullard, 2016, Yellen, 2017, Böninghausen et al., 2018, and Clarida, 2020). We build on this literature by embedding alternative measures within Faust and Wright (2013)'s autoregressive gap model of inflation and show that long-term market-based measures can improve upon professional survey-based forecasts.

The analysis is also related to the literature on deviations from full information rational expectations, which considers the role of rigidities (Mankiw et al., 2003) and/or noisy information (Sims, 2003 and Woodford, 2003). Unlike previous studies, that test for evidence of these channels in individual and/or aggregate professional forecasters (see Coibion and Gorodnichenko 2015a, Coibion et al. 2018, Bordalo et al. 2018 and Angeletos et al. 2020), we consider the differential between aggregate survey and market-based measures. We allow for different information and rigidities across measures and then decompose them to show that the information differential matters most for the overall differential. Our approach differs from the decomposition in Reis (2020) by focusing on the disagreement across measures of expectations rather than the disagreement within measures due to heterogeneous agents.

Finally, our analysis also extends the literature on drivers of expectations differentials. Previous studies find that inflation expectations respond differently to news (Carroll, 2003), food and oil prices (Coibion and Gorodnichenko, 2015b), macroeconomic data releases (Bauer, 2015), financial volatility (Stillwagon, 2018), and belief shocks (Candia et al., 2020). We take a different approach by extracting the information differential and searching across many potential information sources. We find that the information differential is well explained by real-time changes in liquidity including money, reserves, and stocks.

The rest of the paper follows. The next section describes the data and measures of expectations used. Section 3 describes the analytical methods while section 4 presents the results. Finally, section 5 concludes.

2 Measures of Inflation and Inflation Expectations

There are multiple measures of inflation expectations derived from surveys in the United States. We focus on measures of long-term expectations that extend at least five-years-ahead. These include the Blue Chip

[2] In contrast, Bauer and McCarthy (2015) finds short-term market expectations perform worse than professional forecasters.

Economic Indicators (BCEI), the Livingston Survey (LIV), the Survey of Professional Forecasters (SPF), and the University of Michigan's Survey of Consumers (MSC).[3] The long-term BCEI forecasts are updated twice a year in March and October, the LIV forecasts are updated in June and December, the SPF forecasts are updated quarterly and the MSC is updated monthly.

Inflation expectations derived from financial markets are often based on the difference between Treasury Inflation-Protected Securities (TIPS) and nominal Treasury Securities and are referred to as Treasury Breakeven Inflation.[4] There are, however, several concerns associated with this measure which complicate its interpretation as a measure of expectations. In particular, breakeven inflation includes a possibly time-varying inflation risk premia as well as a liquidity bias due to differences in the respective markets.

The U.S. Treasury produces a measure of Treasury Breakeven Inflation (TBI) which directly addresses the liquidity bias. The main features of this measure are (1) it is based on off-the-run Treasury securities which substantially reduces differences in liquidity across the nominal Treasury and TIPS markets; and (2) it is calculated using spot rates instead of yields to ensure consistency with inflation rates; see Girola (2019). This measure extends back to 2003 and is updated on a monthly basis using daily data. Church (2019) shows that TBI correlates well with future non-seasonally adjusted CPI.

Alternatively, models are used to extract expectations from breakeven inflation. One approach comes from D'Amico et al. (2018, DKW), who use a no-arbitrage pricing model to extract inflation expectations from nominal yields, TIPS yields, and inflation. Their measure is updated monthly by the Federal Reserve Board; see Kim et al. (2019). We also consider two other model-based measures of long-term inflation expectations. Haubrich et al. (2012, HPR) use a term-structure model to extract monthly inflation expectations from surveys of professional forecasters, nominal Treasury yields, and inflation swaps. Their measure is available on a monthly basis since 1982 and is maintained by the Federal Reserve Bank of Cleveland.[5] Aruoba (2020, ARU) produces a term-structure of inflation expectations based on information from multiple surveys of professional forecasters. This measure is available on a monthly basis since 1998 and is maintained by the Federal Reserve Bank of Philadelphia.

We measure U.S. inflation using the October 2020 vintage of the seasonally adjusted consumer price inflation (CPI-U). We transform this series into a quarterly annualized inflation rate by averaging the price level over the quarter and then taking the log difference multiplied by 400. The sample of interest is from 2003 Q1 - 2019 Q4 over which the measures of long-term inflation expectations are available.

[3] Long-term measures of expectations are also available from Blue Chip Financial Forecasters and Consensus Economics, the New York Federal Reserve's Surveys of Primary Dealers and Market Participants as well as the Atlanta Federal Reserve's Business Inflation Expectations.

[4] TIPS are based on non-seasonally adjusted CPI. Inflation swaps are also available starting in 2004.

[5] Ajello et al. (2019) and Williams (2020) also propose model-based measures of inflation expectations derived from TIPS.

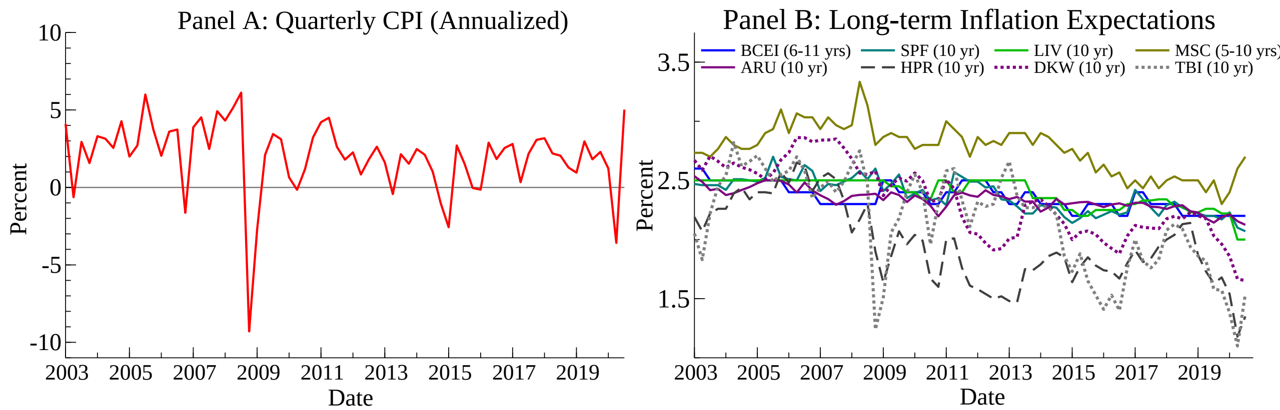


Figure 2.1: Inflation and Long-term Inflation Expectations (2003 Q1 - 2020 Q3)

Figure 2.1 plots inflation in Panel A and the measures of long-term inflation expectations in Panel B. Each measure is presented on a CPI basis with the exception of the Michigan survey which is not tied to a specific measure of inflation. The measures are plotted quarterly from 2003 through 2020 Q3 where higher- (lower-) frequency measures are averaged (extended) across the full quarter.[6] At the end of 2019, long-term inflation expectations ranged between 1.6 and 2.3 percentage points. There was a clear difference between survey-based measures (indicated by solid lines) and measures derived primarily from financial markets (indicated by dotted or dashed lines). In recent years, survey-based measures were all higher than financial market-based measures. This gap became even more apparent during the spread of COVID-19 in 2020.

Table 2.1: Correlation Across Measures of Expectations (2003-2019)

	BCEI	SPF	LIV	MSC	ARU	HPR	DKW	TBI
BCEI	1.00	0.73	0.67	0.37	0.77	0.29	0.41	0.44
SPF		1.00	0.81	0.60	0.79	0.60	0.74	0.60
LIV			1.00	0.72	0.68	0.41	0.63	0.69
MSC				1.00	0.54	0.42	0.62	0.68
ARU					1.00	0.53	0.54	0.50
HPR						1.00	0.87	0.53
DKW							1.00	0.59
TBI								1.00
PC	0.61	0.85	0.82	0.77	0.74	0.80	0.89	0.84

Notes: Pearson correlation coefficients are presented. PC: First Principal Component of inflation expectations.

We can examine the correlations between measures of inflation expectations to understand their similarities. Table 2.1 presents the correlations across measures and with their first principle component. The measures of expectations derived from professional forecasters are highly correlated with one another and

[6]See Appendix Figure A.2 for an illustrative timeline of the release dates for each measure in 2020 Q1.

with the ARU measure. Model-based measures of financial market expectations, HPR and DKW, are highly correlated with one-another despite using different information but are not particularly similar to TBI. Thus, while some measures are similar, there are substantial differences between them, which suggests they may contain unique information. On the other hand, the first principal component of all the measures correlates well with each measure individually. This suggests that there is a common component or trend. While similarities can be extracted from multiple measures of expectations, e.g. see Ahn and Fulton (2020) and Williams (2020), we focus instead on the differences between measures.

3 Methods

There are several competing explanations for what drives differences in expectations. For example, Mankiw et al. (2003) propose a sticky-information model where gaps in expectations are explained by differences in rigidities. So, for example, market participants might update their expectations more quickly to the same set of information. An alternative explanation comes from a model of noisy-information, see Sims (2003) and Woodford (2003), where agents may have access to different kinds of information. In this context, the gap is explained by access to (or a focus on) different information. Thus, market participants may be responding to a different, perhaps better or worse, information set.[7]

To assess the relevance of these explanations, it is necessary to establish a framework that can be used to disentangle them. We start with Faust and Wright (2013)'s autoregressive model of the gap between inflation and long-run expectations. The h-step-ahead forecasts from this model can be written as

$$\tilde{\pi}_{t+h|t} = \tilde{\mu}_{t+h|t} + \rho^{h} (\pi_t - \mu_t), \tag{1}$$

where π_t is inflation and μ_t represents long-term expectations, generated at time t, about future inflation. In the simplest version of (1), Faust and Wright (2013) fix the autoregressive parameter $\rho \equiv 0.46$.[8] Future inflation expectations follow a random walk so that $\tilde{\mu}_{t+h|t} = \mu_t$. This implies that (1) has a time-varying long-run mean (referred to as a time-varying trend). Other studies formulate variants of (1) using alternative long-run survey-based professional forecasts or statistical measures.[9] We are interested if market-based measures can add value to survey-based measures and so we appeal to the forecast-encompassing literature.

[7]The Federal Reserve Bank of New York's October 2015 Survey of Market Participants indicates that both explanations matter.

[8]The parameter for more recent periods is likely smaller than 0.46 as discussed by Chen (2019). We exclude a Phillips curve relationship to adhere with Faust and Wright (2013).

[9]Backwards looking statistical measures were proposed by Atkeson and Ohanian (2001), Stock and Watson (2007) and Ball and Mazumder (2011). Martinez et al. (2021) illustrates why backwards looking measures perform well in some settings.

Suppose for simplicity of exposition that there are only two different measures of expectations whose relative information content are of interest: survey-based expectations $\mu_{t,Survey}$ and financial market-based expectations $\mu_{t,Markets}$. Furthermore, suppose that these measures are used as different time-varying long-run means in (1) to generate alternative forecasts of inflation. Then it is possible to formulate a forecast-encompassing regression as in Chong and Hendry (1986) and Fair and Shiller (1989, 1990):

$$\pi_{t+h} = \beta_{0,h} + \beta_{S,h}\tilde{\pi}_{t+h|t,Survey} + \beta_{M,h}\tilde{\pi}_{t+h|t,Markets} + u_{t+h}, \tag{2}$$

where $\beta_{0,h}$ is the bias and u_{t+h} is the unexplained residual.[10] We can test for the uniqueness of information available from each of the forecasts in this equation.[11] For example, the joint hypothesis of $\beta_{S,h} = 1$ and $\beta_{M,h} = 0$ implies that the survey-based forecast sufficiently explains inflation at horizon h so that the market-based forecast provides no additional value. Alternatively, the joint hypothesis of $\beta_{S,h} = 0$ and $\beta_{M,h} = 1$ implies that the survey-based forecast is not informative beyond the market-based forecast.[12]

The equation can be reinterpreted in the context of forecast combinations as originally discussed by Bates and Granger (1969). In fact, as illustrated by Granger and Ramanathan (1984), the population-optimal weights for combining the forecasts can be obtained from (2) when a homogeneity restriction $\beta_{S,h} + \beta_{M,h} \equiv 1$ is imposed. In the context of forecast-encompassing, Ericsson (1993) shows that the homogeneity restriction implies that (2) can be rewritten by subtracting $\tilde{\pi}_{t+h|t,Survey}$ from both sides so that

$$e_{t+h,Survey} = \beta_{0,h} + \beta_{M,h}\left(\tilde{\pi}_{t+h|t,Markets} - \tilde{\pi}_{t+h|t,Survey}\right) + \tilde{u}_{t+h}, \tag{3}$$

where $e_{t+h,Survey} = \pi_{t+h} - \tilde{\pi}_{t+h|t,Survey}$ is the survey-based forecast error. The importance of the differential between the market-based forecast and the survey-based forecast is determined by the optimal weight on the market-based forecasts $\beta_{M,h}$ in (3). When both forecasts are generated from (1) with fixed ρ, the forecast differential is the horizon weighted expectations differential

$$\left(\tilde{\pi}_{t+h|t,Markets} - \tilde{\pi}_{t+h|t,Survey}\right) = (1-\rho^h)\left(\mu_{t,Markets} - \mu_{t,Survey}\right), \tag{4}$$

where greater weight is given to the expectations differential at longer horizons. Therefore, the focus on the forecast differential is a formalization of the implicit framework by Clarida (2020). In this context,

[10] The bias term captures expectation differentials that are constant over time; e.g. see Bürgi (2020).

[11] The forecasts are non-nested and parameters are fixed so standard t-tests are used (Harvey et al., 1998) without correcting for parameter estimation uncertainty (West, 2001) or nested models (Clark and McCracken, 2001; Hansen and Timmermann, 2015).

[12] Romer and Romer (2000) spurred a separate literature by referring to these as information advantage hypotheses.

the null hypothesis $\beta_{M,h} = 0$ means that the expectations differential provides no explanatory power for the survey-based forecast errors. Alternatively, the null hypothesis that $\beta_{M,h} = 1$ means that the expectations differential completely explains the survey-based forecast errors. Together, both null hypotheses assess whether the differential between market and survey-based measures improves survey-based forecasts.

This framework can be extended in multiple ways. First, it can be used to evaluate two survey-based forecasts. Although the direction of information flow becomes less clear, the approach can still be used to assess whether the expectations differential helps explain the forecast errors. Second, it can be augmented by conditioning on additional information. For example, the unemployment gap can be added to (3) to evaluate if omitting the Phillips curve from the forecasts matters. Third, the forecast-encompassing equation can be extended to test multiple horizons (see Hungnes, 2018) and/or more than two forecasts (see Martinez, 2015 and Ericsson and Martinez, 2019) jointly. Finally, the tests can also be adapted to allow for time variation and instabilities using fluctuation tests; see Rossi and Sekhposyan (2016).

While forecast-encompassing tests evaluate the general difference in information content between the forecasts, they can also be reinterpreted as tests of the combined information and rigidity differentials between measures. To see this, suppose that measures of expectations are described as

$$\mu_{t,m} = \gamma_m f_{t,m} + (1 - \gamma_m)\, \mu_{t-1,m}, \tag{5}$$

where $f_{t,m}$ is a measure-specific function of information available at time t and γ_m denotes how quickly expectations are updated by this information (i.e. rigidity).[13] Then the differential between market-based and survey-based expectations can be iteratively decomposed as:

$$\begin{aligned}(\mu_{t,Markets} - \mu_{t,Survey}) = (\gamma_{Markets} - \gamma_{Survey}) &\sum_{j=0}^{t-1} (1 - \gamma_{Markets})^j \left[\frac{\Delta\mu_{t-j,Survey}}{\gamma_{Survey}}\right] \\ &+ \gamma_{Markets} \sum_{j=0}^{t-1} (1 - \gamma_{Markets})^j \left[f_{t-j,Markets} - f_{t-j,Survey}\right] \\ &+ (1 - \gamma_{Markets})^t \left(\mu_{0,Markets} - \mu_{0,Survey}\right),\end{aligned} \tag{6}$$

where the first term is the rigidity differential ($\nabla\gamma_{M,t}$), the second term is the discounted information differential ($\nabla f_{M,t}$) and the third term captures the initial conditions ($\nabla\mu_{0,M,t}$). Thus, market-based expectations can be both less rigid and have access to different information such that a combination of the information and rigidity differentials drives the total differential. To better understand which of these channels matters

[13] This is consistent with the earlier assumption that $\mu_{t+h|t}$ is a random walk when $f_{s,t}$ is also treated as a random walk.

most for forecast performance we can combine (6) with (3) and (4) to get

$$e_{t+h,Survey} = \tilde{\beta}_{0,h} + \beta_{M,h,1}\tilde{\nabla}_h \gamma_{M,t} + \beta_{M,h,2}\tilde{\nabla}_h f_{M,t} + \beta_{M,h,3}\tilde{\nabla}_h \mu_{0,M,t} + \bar{u}_{t+h}, \tag{7}$$

where $\tilde{\nabla}_h = (1-\rho^h)\nabla$. This equation nests (3), which is a restricted version of (7) implying that $\beta_{M,h,1} \equiv \beta_{M,h,2} \equiv \beta_{M,h,3}$. It can be used to test the importance of each channel and to assess which channel or comibnation of channels matters most for improving the suvey-based forecast performance.

However, we first need to identify γ_m for each measure of expectations. To do this, we impose a common function of information across each measure (m) of expectations as a biased and noisy prediction of current inflation $f_{t,m} = \eta_m + \pi_t + \varepsilon_{t,m}$ so that (5) becomes

$$\mu_{t,m} = \gamma_m(\pi_t + \eta_m) + (1-\gamma_m)\mu_{t-1,m} + \gamma_m \varepsilon_{t,m}, \tag{8}$$

which is similar to the approaches by Coibion and Gorodnichenko (2015a) and Jorgensen and Lansing (2019) among others except that we allow for measure specific intercepts and information beyond current inflation to impact long-term expectations formation. For example, Bauer (2015) finds that both market-based and survey-based measures of expectations respond to inflation and macroeconomic news releases.

Considering (1) and (8) together implies that expectations and inflation are determined simultaneously when $\gamma_m \neq 0$. This means that γ_m cannot be identified from (8) alone. Therefore, we combine (1) and (8) as a system of simultaneous equations

$$\begin{pmatrix} 1 & -1 \\ -\gamma_m & 1 \end{pmatrix} \begin{pmatrix} \pi_t \\ \mu_{t,m} \end{pmatrix} = \begin{pmatrix} \eta_{\pi,m} \\ \gamma_m \eta_{\mu,m} \end{pmatrix} + \begin{pmatrix} \rho & -\rho \\ 0 & 1-\gamma_m \end{pmatrix} \begin{pmatrix} \pi_{t-1} \\ \mu_{t-1,m} \end{pmatrix} + \begin{pmatrix} v_t \\ \gamma_m \varepsilon_{t,m} \end{pmatrix}, \tag{9}$$

where we assume that the shocks v_t and $\varepsilon_{t,m}$ are mean zero with a general covariance structure. If $\gamma_m \neq 1$, then we can rewrite (9) as a system of reduced form equations

$$\begin{pmatrix} \pi_t \\ \mu_{t,m} \end{pmatrix} = \frac{\gamma_m}{1-\gamma_m} \begin{pmatrix} \frac{1}{\gamma_m}\eta_{\pi,m} + \eta_{\mu,m} \\ \eta_{\pi,m} + \eta_{\mu,m} \end{pmatrix} + \begin{pmatrix} \frac{\rho}{1-\gamma_m} & 1 - \frac{\rho}{1-\gamma_m} \\ \frac{\gamma_m \rho}{1-\gamma_m} & 1 - \frac{\gamma_m \rho}{1-\gamma_m} \end{pmatrix} \begin{pmatrix} \pi_{t-1} \\ \mu_{t-1,m} \end{pmatrix} + \frac{\gamma_m}{1-\gamma_m} \begin{pmatrix} \frac{1}{\gamma_m} v_t + \varepsilon_{t,m} \\ v_t + \varepsilon_{t,m} \end{pmatrix}, \tag{10}$$

which can also be expressed in vector equilibrium-correction form

$$\begin{pmatrix} \Delta\pi_t \\ \Delta\mu_{t,m} \end{pmatrix} = \frac{\gamma_m}{1-\gamma_m} \begin{pmatrix} \frac{1}{\gamma_m}\eta_{\pi,m} + \eta_{\mu,m} \\ \eta_{\pi,m} + \eta_{\mu,m} \end{pmatrix} + \begin{pmatrix} \frac{\rho}{1-\gamma_m} - 1 \\ \frac{\gamma_m \rho}{1-\gamma_m} \end{pmatrix} (\pi_{t-1} - \mu_{t-1,m}) + \frac{\gamma_m}{1-\gamma_m} \begin{pmatrix} \frac{1}{\gamma_m} v_t + \varepsilon_{t,m} \\ v_t + \varepsilon_{t,m} \end{pmatrix}, \tag{11}$$

where when $\gamma_m \neq 0$ weak exogeneity is violated such that there is a reduced-rank restriction that pins down both equations and implies that the system in levels is a non-stationary I(1) process; see Hendry (1995).

We can estimate the bivariate vector equilibrium-correction model in (11) as

$$\begin{pmatrix} \Delta\hat{\pi}_t \\ \Delta\hat{\mu}_{t,m} \end{pmatrix} = \begin{pmatrix} \hat{c}_{\pi,m} \\ \hat{c}_{\mu,m} \end{pmatrix} + \begin{pmatrix} \hat{\alpha}_{1,m} \\ \hat{\alpha}_{2,m} \end{pmatrix} (\pi_{t-1} - \mu_{t-1,m}), \tag{12}$$

where $\hat{\alpha}_{1,m}$ and $\hat{\alpha}_{2,m}$ are the estimates of $\alpha_{1,m} = \frac{\rho}{1-\gamma_m} - 1$ and $\alpha_{2,m} = \frac{\gamma_m \rho}{1-\gamma_m}$ and where $\hat{c}_{i,m}$ are the variable specific intercepts. To identify γ_m from (12), assuming ρ is known, we impose the restriction that $\hat{\alpha}_{2,m} = \hat{\alpha}_{1,m} + 1 - \rho$. Then $\hat{\gamma}_m = 1 - \frac{\rho}{\hat{\alpha}_{1,m}+1}$. Assuming the residuals are i.i.d. normal then, when applying the multivariate delta method (Casella and Berger, 2002), the coefficient standard error is $\hat{\sigma}_{\hat{\gamma}_m} = \frac{\rho\hat{\sigma}_{\hat{\alpha}_{1,m}}}{(\hat{\alpha}_{1,m}+1)^2}$.[14]

Once we identify the degree of information rigidity for each measure of expectations, it is possible to decompose differences in expectations using (6) and to test their relative informativeness using (7). We can also perform a counterfactual exercise by shutting down the differences in rigidities ($\gamma_{Markets} \equiv \gamma_{Survey}$) while conditioning on the original set of information or by shutting down the information channel ($f_{t,Markets} \equiv f_{t,Survey}$) to illustrate the roles that focusing on alternative information or having different rigidities/persistence can have in generating the wedge between alternative measures of expectations.

We also examine if economic variables can explain the information differential. To do this, we formulate a model based on the extracted information differential for each measure of expectations

$$\nabla f_{t,m} = (f_{t,m} - f_{t,BCEI}) = \beta_{0,m} + \sum_{i=1}^{N} \beta_{i,m} y_{i,t} + o_{t,m}, \tag{13}$$

where $y_{i,t}$ represents a source of information at time t and $o_{t,m}$ is the unexplained residual. When $N > T$, it is not possible to estimate (13) using traditional methods. We select over all N variables using the general-to-specific automatic model selection procedure implemented in 'Autometrics'; see Doornik (2009). Autometrics performs a tree search over subsets of variables and uses F-tests to eliminate them in groups. It then checks the selected model against the starting point to see if the user-specified amount of information loss is exceeded. This reduces concerns about highly correlated variables that plague other model selection procedures; see Doornik (2008) and Hendry and Doornik (2014). Once we isolate the most important sources of information for each differential, we can use them to predict the information differential, understand whether they drive the overall differential, and ascertain to what extent they drive historical forecast improvements.

[14]The i.i.d. assumption can be relaxed by estimating $\hat{\sigma}_{\hat{\alpha}_{1,m}}$ using HAC methods; see Andrews (1991).

4 Results

We start by testing the overall information differences in individual forecast pairs. The forecast-encompassing test results are presented in Table 4.1 where major columns represent different horizons and major rows represent different measures of expectations. For simplicity we choose BCEI as the baseline and focus on the MSC and market-based measures as alternatives. In each block of cells, the first entry presents the estimate of $\beta_{M,h}$ from (3). The second entry is the p-value associated with the null hypothesis that $\beta_{M,h} = 1$. The third entry is the p-value associated with the null hypothesis that $\beta_{M,h} = 0$.

Table 4.1: Forecast-Encompassing Coefficients and Probabilities Relative to BCEI

Horizon:	1	2	3	4	5	6	7	8	Joint
MSC	3.12	-0.91	-1.30	-0.11	0.87	0.13	-1.50	-1.27	-1.71
	⟦0.261⟧	⟦0.390⟧	⟦0.234⟧	⟦0.442⟧	⟦0.916⟧	⟦0.515⟧	⟦0.075⟧*	⟦0.063⟧*	⟦0.111⟧
	{0.100}	{0.681}	{0.499}	{0.939}	{0.470}	{0.925}	{0.282}	{0.294}	{0.417}
HPR	0.83	0.72	1.73	2.49	2.59	1.97	1.57	1.46	1.05
	⟦0.872⟧	⟦0.794⟧	⟦0.350⟧	⟦0.033⟧**	⟦0.028⟧**	⟦0.271⟧	⟦0.585⟧	⟦0.650⟧	⟦0.690⟧
	{0.421}	{0.494}	{0.029}**	{0.001}***	{0.001}***	{0.028}**	{0.137}	{0.155}	{0.008}***
DKW	2.00	1.96	2.20	2.49	2.35	1.82	1.62	1.30	0.85
	⟦0.361⟧	⟦0.322⟧	⟦0.162⟧	⟦0.076⟧*	⟦0.189⟧	⟦0.449⟧	⟦0.582⟧	⟦0.778⟧	⟦0.983⟧
	{0.070}*	{0.045}**	{0.012}**	{0.004}***	{0.017}**	{0.095}*	{0.152}	{0.225}	{0.053}*
TBI	0.10	-0.52	-0.20	0.58	1.25	0.62	-0.42	-0.99	-0.33
	⟦0.372⟧	⟦0.139⟧	⟦0.169⟧	⟦0.508⟧	⟦0.677⟧	⟦0.519⟧	⟦0.020⟧**	⟦0.006⟧***	⟦0.114⟧
	{0.918}	{0.608}	{0.816}	{0.356}	{0.039}**	{0.285}	{0.484}	{0.158}	{0.436}

Notes: Values in each block are (1) the estimated coefficient, (2) p-value associated with the null-hypothesis that the coefficient is equal to unity in square brackets and (3) p-value associated with the null hypothesis that the coefficient is equal to zero in squigly brackets. Joint estimates and tests follow Hungnes (2018). All equations include a dummy variable for 2008 Q4. Tests use HAC estimates from Andrews (1991). *$p<0.1$ **$p<0.05$ ***$p<0.01$

For example, in the third major row and the first major column, the estimate for the optimal weight on DKW at $h = 1$ is $\hat{\beta}_{DKW,1} = 2.00$. The homogeneity restriction implies that the optimal weight on BCEI is therefore $\hat{\beta}_{BCEI,1} = -1.00$. The null hypothesis that $\beta_{DKW,1} = 1$ cannot be rejected with a probability of 36.1 percent. However, the null hypothesis that $\beta_{DKW,1} = 0$ is weakly rejected with a probability of 7.0 percent. Together, these results suggest that the expectations differential between DKW and BCEI explains the BCEI-based forecast-errors at this horizon.

The forecast-encompassing results are generally supportive of market-based measures of expectations. Both the HPR and the DKW differentials are informative across several horizons with DKW doing particularly well up through six-quarters-ahead. Furthermore, both the HPR and DKW differentials are informative for BCEI when considering all horizons jointly. The evidence is mixed for the TBI differential in that it is informative at the 1-year-ahead forecast horizon but is not informative at longer horizons. Appendix Table

A.4 illustrates that these results are robust to the use of alternative survey-based measures as a baseline and are broadly robust to controlling for the unemployment gap as a measure of the Phillips curve. Furthermore, following Ericsson (1992), Appendix A.1 shows that the forecast encompassing results are mirrored by the relative RMSE rankings. Appendix A.2 shows that the results are somewhat sensitive to time-varying instabilities, especially for MSC and TBI, with the most significant information differences around 2010-12.

Overall, the results for HPR and DKW indicate that the differential between the survey (BCEI) and market-based (HPR and DKW) measures of expectations can help (and in some cases completely) explain survey-based forecast errors. This supports the hypothesis that market-based measures have access to better information and/or respond more quickly than survey-based measures to the same information. However, the overall results are unable to shed light on which of the channels is most important for why the differential improves forecast performance. We turn to this question after identifying the measure-specific rigidities.

4.1 Identifying Rigidity

We can decompose the differences between alternative measures of expectations in order to better understand what is driving them and where the relative value stems from. To do this, we start by estimating the degree of rigidity in each measure of expectations using (12).

The estimates of γ_m are shown in Table 4.2. The first row presents the baseline estimates assuming the model is correctly specified and without allowing for any instabilities. Most estimates are not significantly different from zero, which implies that expectations are weakly exogenous with respect to inflation and so follow a random walk process. However, the HPR measure has an estimate which is positive and significantly different from zero. This implies that it is equilibrium-correcting such that it adjusts to reduce the gap with past inflation. In particular, the estimate of 0.025 for HPR implies that it adjusts back to the equilibrium relationship within 10 years. If the BCEI and TBI estimates were statistically significant, the negative values would imply that they are equilibrium diverging such that they adjust to increase the gap with inflation.

We can assess the robustness of these estimates by allowing for general forms of model instability and model misspecification, which can impact how expectations interact with inflation; see Castle et al. (2014). We allow for any number of outliers and shifts at any point in time by estimating (12) with Impulse Indicator Saturation (IIS) and Differenced-IIS (DIIS) using a target gauge of 0.1% (see Hendry et al., 2008 and Castle et al., 2011). Table 4.2 shows that when accounting for instabilities, the estimate for BCEI goes from being weakly exogenous to equilibrium-correcting. There is a concern that the BCEI's lower frequency update schedule could affect these estimates. However, Appendix Table A.5 shows similar estimates for other higher frequency surveys of professional forecasters.

Table 4.2: Estimates of Rigidity for Alternative Measures of Long-Term Expectations

	BCEI	MSC	HPR	DKW	TBI
No Model Instability:	-0.005	0.005	0.025*	0.013	-0.023
	(0.007)	(0.012)	(0.015)	(0.010)	(0.025)
Model Instability / Misspecification:	0.019**	0.007	0.025*	0.013	-0.031
	(0.008)	(0.011)	(0.015)	(0.010)	(0.024)

Notes: The estimation sample is 2003Q2 - 2019Q4 (67 observations). Allow for model instability/misspecification by selecting over and retaining outliers and shifts using Autometrics with Impulse Indicator Saturation (IIS) and Differenced-IIS (DIIS). The target gauge is 0.1% *so that under the null we expect 0.2 irrelevant indicators to be retained on average. See Appendix Table A.6 for details. Estimated standard errors are in parentheses.* *$p<0.1$ **$p<0.05$ ***$p<0.01$

The estimates of γ_m imply that the inertia in long-run expectations was very high over this period regardless of which measure is used. This finding is consistent with other studies that show persistence has increased in the past few decades; e.g. see Jorgensen and Lansing (2019). Note that our estimates of γ_m are sensitive to the choice of ρ. For example, if inflation persistence has declined to $\rho = 0.2$, as suggested by Chen (2019), then the estimates of γ_m for the survey-based measures would be slightly larger, whereas the estimates for the market-based measures would be somewhat smaller. However, these differences are small enough such that they do not have meaningful differences on the resulting analysis.

4.2 Which Channel Matters?

Once we identify the estimate of rigidity for each measure of expectations, we can decompose the overall differential between measures into the rigidity channel, the information channel and the initial conditions channel using the decomposition in (6). We can then test which channel matters most for improving forecast accuracy using (7). For the purpose of this exercise we focus on one channel of information at a time by imposing that the other channels are uninformative.

The results of this exercise are presented in Table 4.3 where each major row represents an alternative measure of expectations and each major column represents a channel of information. The first major column presents the combined "Total" of all the channels where we impose that $\beta_{M,h,1} \equiv \beta_{M,h,2} \equiv \beta_{M,h,3}$ in (7) such that the results are identical to those presented in Table 4.1. Each of the other major columns focuses on a specific channel. For example, in the third major column labeled "Information" we impose the restriction that $\beta_{M,h,1} \equiv \beta_{M,h,3} \equiv 0$ and focus on the information channel by testing restrictions on $\beta_{M,h,2}$. In the third major row of the first minor column of this set the estimate for $\beta_{DKW,1,2} = 1.98$, which is very similar to the estimate of 2.00 that we obtain for the total information result in the first minor column of the first major column. The p-values associated with the two forecast-encompassing tests are also similar such that we find evidence that DKW's information channel improves the BCEI-based forecast.

Table 4.3: Decomposed Forecast-Encompassing Tests (Relative to BCEI)

	Total		**Rigidity**		**Information**		**Initial Conditions**	
Horizon:	1	Joint	1	Joint	1	Joint	1	Joint
MSC	3.12	-2.31	5.20	-6.65	2.95	-2.60	30.03	52.08
	⟦0.261⟧	⟦0.249⟧	⟦0.460⟧	⟦0.173⟧	⟦0.347⟧	⟦0.255⟧	⟦0.178⟧	⟦0.419⟧
	{0.100}	{0.417}	{0.360}	{0.199}	{0.156}	{0.470}	{0.164}	{0.393}
HPR	0.83	1.67	-21.13	3.34	0.94	1.55	-4.08	-7.96
	⟦0.872⟧	⟦0.310⟧	⟦0.069⟧*	⟦0.213⟧	⟦0.944⟧	⟦0.633⟧	⟦0.143⟧	⟦0.082⟧*
	{0.421}	{0.008}***	{0.082}*	{0.229}	{0.256}	{0.015}**	{0.238}	{0.167}
DKW	2.00	1.45	17.10	-10.88	1.98	1.57	35.54	63.72
	⟦0.361⟧	⟦0.904⟧	⟦0.231⟧	⟦0.199⟧	⟦0.402⟧	⟦0.918⟧	⟦0.196⟧	⟦0.328⟧
	{0.070}*	{0.053}*	{0.204}	{0.203}	{0.092}*	{0.087}*	{0.184}	{0.308}
TBI	0.10	-0.49	3.67	-6.00	0.10	-0.06	-9.54	-16.41
	⟦0.372⟧	⟦0.338⟧	⟦0.563⟧	⟦0.166⟧	⟦0.262⟧	⟦0.655⟧	⟦0.120⟧	⟦0.341⟧
	{0.918}	{0.608}	{0.428}	{0.201}	{0.901}	{0.699}	{0.159}	{0.420}

Notes: Values in each block are (1) the estimated coefficient, (2) p-value associated with the null-hypothesis that the coefficient is equal to unity in square brackets and (3) p-value associated with the null hypothesis that the coefficient is equal to zero in squigly brackets. Joint estimates and tests follow Hungnes (2018). Equations include a dummy variable for 2008 Q4. Tests use HAC estimates from Andrews (1991). *$p<0.1$ **$p<0.05$ ***$p<0.01$

Looking more broadly at the results in Table 4.3 we see clear evidence that the information channel drives the forecast improvements. For both HPR and DWK we find evidence that the information channel encompasses the overall SPF forecast performance while the other channels do not. These results hold across all forecast horizons In fact, for HPR, there is evidence in the initial conditions channel that it has a worse starting point, which makes it even more evident that HPR has an important information advantage relative to BCEI. On the other hand, we do not see any strong evidence for any channel that MSC or TBI are more or less informative than BCEI. Overall, these results illustrate that the information channel drives the improvements in forecast accuracy relative to professional forecasters.

We can illustrate the importance of the information channel another way by performing a counterfactual exercise that indicates what the expectations differential would be if measures had the same rigidity or information. The contributions from the various components of the expectations differentials are plotted in Figure 4.1 for each of our four measures. This is based on the decomposition in (6) using the robust estimates for γ_m when allowing for model instabilities in Table 4.2.[15] We start by plotting the overall differences between BCEI and four other measures. Next, we perform a counterfactual exercise by shutting down the differences in the information channel by imposing that, $f_{t,m} \equiv f_{t,BCEI}$, which implies that only the differences in rigidity and the initial conditions are allowed to operate. Figure 4.1 shows that this explains very little of the overall differentials, with the possible exceptions of MSC and TBI.

[15]Note that we set $\gamma_{TBI} = 0.005$ in line with the smallest other estimate to ensure that its estimated value is non-negative.

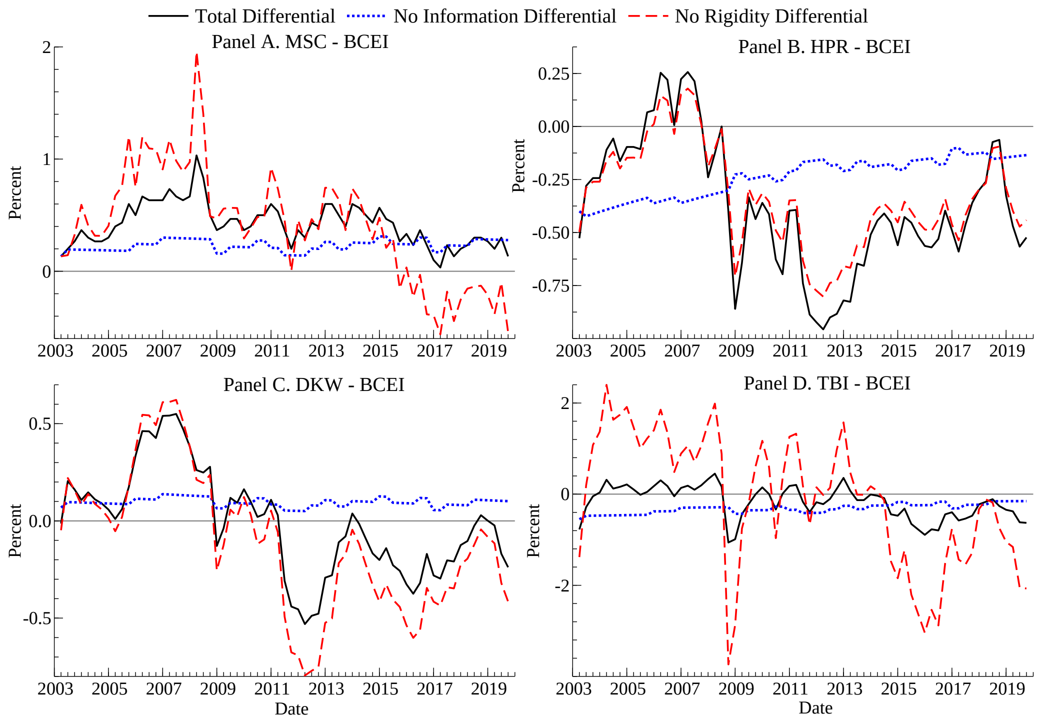


Figure 4.1: Sources of Expectations Differentials by Measure (2003 Q2 - 2019 Q4)

We perform another counterfactual exercise by closing the rigidity differential, such that $\gamma_m \equiv \gamma_{BCEI}$, which means that only the information differential and the initial conditions are allowed to operate. Figure 4.1 shows that the information differential captures most of the variation in the overall differential, especially for DKW and HPR. This is consistent with and reinforces the decomposed encompassing results in Table 4.3 that the information channel matters most for the overall differential between measures of expectations.

The counterfactual exercises illustrate some other interesting measure-specific features. For example, the total differential between MSC and BCEI and the information differential follow each other closely until 2015. After 2015, the information differential falls and the rigidity differential becomes more important. This is consistent with other findings that recent changes in consumer expectations are not driven exclusively by information (Vergbrugge and Binder, 2016) and may be explained by changing survey demographics as older participants leave the sample; e.g. see Malmendier and Nagel (2016) and Binder and Makridis (2020).

Overall the results provide fairly consistent evidence that the information channel is what matters most for driving overall differences between expectations and forecast improvements. For both DKW and HPR, differences in information with BCEI explain most of the overall differential and explain the forecast improvements across horizons. This suggests that they are focusing on / responding to better information than the BCEI or other measures. In the next subsection we explore what information might be driving these advantages and how stable these relationships were during the COVID-19 pandemic in 2020.

4.3 What Explains the Information Differential?

Previous research has found that inflation expectations respond to a variety of information sources including news (Carroll, 2003), food and oil prices (Coibion and Gorodnichenko, 2015b), gasoline price shocks (Kilian and Zhou, 2020), macroeconomic data releases (Bauer, 2015) and financial volatility (Stillwagon, 2018). However, previous studies have focused on a relatively small subset of information and on overall expectations differentials. We extend the literature by focusing exclusively on the information differentials after removing other aspects of expectations and by searching over a much more general set of information.

Our analysis draws from two major sources of information. First, we consider every non-interest or exchange rate series available in real-time vintages of FRED-MD between 2003-2019 as potential sources of information that could produce a wedge between measures of expectations; see McCracken and Ng (2016). Second, we consider macroeconomic data announcement surprises based on the aggregated difference between data releases with available Bloomberg forecasts; see Altavilla et al. (2017). By selecting over both sources of information individually we can assess whether previous findings are corroborated and potentially discover additional sources of information that have not yet been found to be important.

For the FRED-MD database, we aggregate the available data over the quarter and compute the quarterly percent change for each vintage of the database that was available at the end of each quarter from 2003 Q1 through 2019 Q4. We start by formulating a general unrestricted model as in (13) where $y_{i,t} = \Delta x_{i,t}$ and $\Delta x_{i,t}$ represents the quarterly percent change information source i at data vintage t. Since there are 103 variables in the FRED-MD database that satisfy our criteria, we choose a conservative target gauge of 0.1% so that under the null on average we expect to retain less than 0.1 irrelevant variables by chance. This ensures that we only retain those variables that matter most for explaining the information differential. Following Coibion and Gorodnichenko (2015b), we force oil prices into the model so that they are always retained.[16]

For the Bloomberg macro announcements, we aggregate surprises over the quarter for each quarter from 2003 Q1 through 2019 Q4. We then formulate another general unrestricted model as in (13) where $y_{i,t} = \sum_{j=1}^{J} \left(x_{i,t_j} - \tilde{x}_{i,t_j} \right)$ and $\tilde{x}_{i,t_j}$ represents the available Bloomberg forecast for macroeconomic release i at day j of quarter t. Since there are 50 sources of Bloomberg macro surprises that satisfy our criteria, we choose a conservative target gauge of 0.2% so that under the null on average we expect to retain roughly 0.1 irrelevant variables by chance. We force surprises about the Institute for Supply Management's prices paid in to the model since these are often considered to be a leading indicator of price movements. Once we have selected information from both FRED-MD and the Bloomberg surprises individually, we combine the selected information into a single model of selected information sources.

[16] We include the percent change rather than the levels, which circumvents the concerns raised in Kilian and Zhou (2020).

Table 4.4: Selected Information Sources (2003-2019)

Source	$\nabla f_{MSC,t}$	$\nabla f_{HPR,t}$	$\nabla f_{DKW,t}$	$\nabla f_{TBI,t}$
Real-time FRED-MD Changes:				
	Oil Prices	**Oil Prices**	**Oil Prices**	**Oil Prices**
	Govt Transfers	M2 Multiplier	Industrial Stock Gap	Bank Reserves
	Prices of Raw Materials			
Bloomberg Macro Surprises:				
	ISM Prices Paid	**ISM Prices Paid**	**ISM Prices Paid**	**ISM Prices Paid**
	NAHB Housing Index	Building Permits	Empire Manufacturing	Fed Funds Rate
		Consumer Confidence		Import Prices
		ISM Manufacturing		
		Leading Index		
R^2**:**	0.41	0.47	0.31	0.61

Notes: All equations estimated and selected between 2003-2019. Bold values indicate variables that were retained regardless of selection.

The selected information from each source and for each of the expectations differentials is listed in Table 4.4. This illustrates that a handful of variables explain between a third and two thirds of the variation in information differentials. Aside from the forced variables, there is relatively limited overlap between the retained information sources across measures. We interpret this as capturing the idiosyncratic differences after removing common information. For example, the idiosyncratic information differential between MSC and BCEI is associated with changes in oil prices, government transfers (i.e. the net difference between personal income and personal income excluding transfers), the prices of crude materials and housing price surprises. This indicates that long-term consumer expectations are more responsive to supply shocks and the business cycle, which is consistent with the results in Candia et al. (2020).

The selected information indicates that the market-based measures respond more than professional forecasters to changes in measures of liquidity. The HPR differential is related to changes in the M2 money multiplier while the DKW differential responds to changes in industrial stocks, and the TBI differential is associated with changes in bank reserves and interest rate surprises. The relationship between measures of liquidity and the market-based differentials indicates that markets are more sensitive to changes in these measures than professional forecasters. Beyond that we also find that the DKW measure is more responsive to information about industrial and manufacturing related measures.

The selection and estimation sample is dominated by the 2008-09 financial crisis during which there were dramatic changes to the economy along with a large-scale expansion of the money supply through the Federal Reserve's credit and Quantitative Easing programs. Therefore the selected information is likely driven by the large changes that occurred during this period. It is thus important to assess whether the selected information is robust to other periods or is specific to the 2008-09 financial crisis.

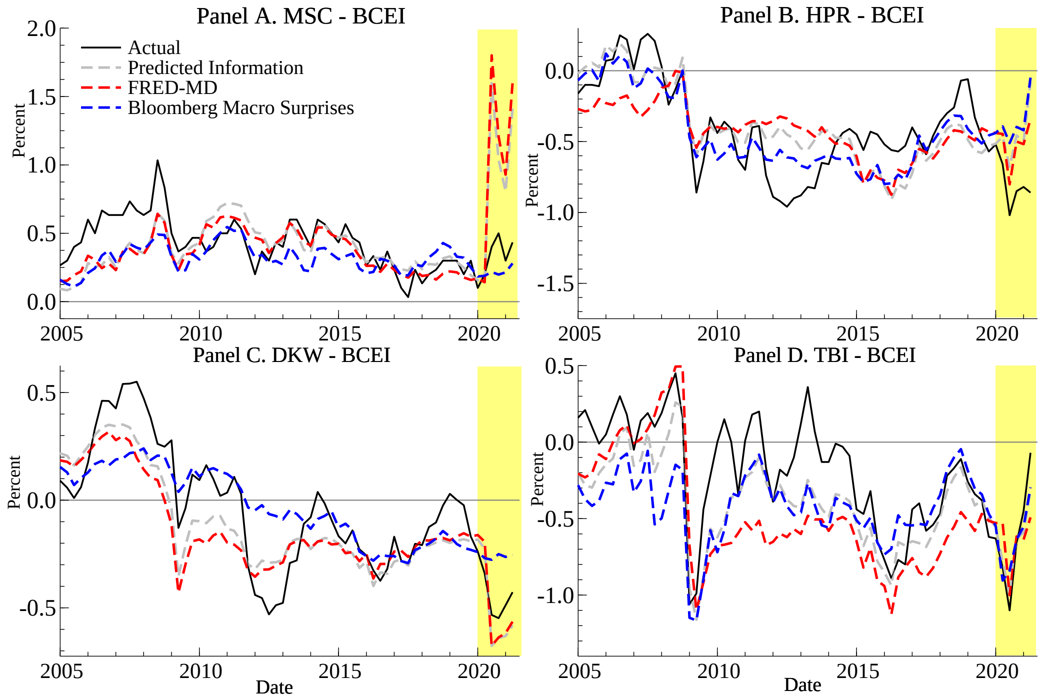


Figure 4.2: Actual and Predicted Expectations Differentials (2005 Q1 - 2021 Q1)

To assess the robustness of our selection results we project the information differential using previously selected information through 2021 Q1 and then feed the projected differential through (6) to obtain a prediction of the total expectations differential. Figure 4.2 presents the results for the total selected information as well as for individual information sources. It illustrates that there are important differences across the two information sources as well as across individual measures of expectations differentials.

The projection of the MSC expectations differential drastically overshoots the relative increase in consumer expectations. This is driven by the information selected from the FRED-MD database and is largely associated with the increase in government transfers during the COVID-19 pandemic. This illustrates that if consumers responded to changes in fiscal policy during the COVID-19 pandemic the same way as they did during the Great Recession then their inflation expectations would have been much higher. Thus, this indicates that consumers have adapted their expectations in light of the pandemic.

Projections of the market-based expectations differentials align with actual differentials. The selected information from the FRED-MD database captures the sharp decline and recovery in market-based expectations much more closely than it does for consumers. However, the selected information from the Bloomberg macroeconomic surprises does not capture this downturn with the exception of TBI. This indicates that the selected measures of liquidity capture the changes in market-based measures and are broadly robust to the massive changes in the economy that occurred since the start of the COVID-19 pandemic.

Table 4.5: Testing for Information Instability

	MSC			**HPR**			**DKW**			**TBI**		
Info Source:	Both	FMD	BBG	Both	FMD	BBG	Both	FMD	BBG	Both	FMD	BBG
2003-19:	0.18	0.16	0.19	0.21	0.28	0.18	0.15	0.18	0.18	0.29	0.40	0.36
2020-21:	0.77	0.92	0.18	0.45	0.34	0.52	0.13	0.13	0.22	0.14	0.25	0.17
Chow Test	17.7^{***}	28.7^{***}	0.83	4.00^{***}	1.33	7.87^{***}	0.72	0.52	1.42	0.20	0.36	0.20
	[0.000]	[0.000]	[0.535]	[0.003]	[0.263]	[0.000]	[0.608]	[0.763]	[0.230]	[0.960]	[0.871]	[0.960]
	F(5,62)	F(5,64)	F(5,65)	F(5,60)	F(5,65)	F(5,62)	F(5,62)	F(5,65)	F(5,64)	F(5,63)	F(5,65)	F(5,65)

Notes: 2003-2019 is the in-sample estimation period 2020-2021 is the forecast period. The values for both periods are represented by the root mean square errors. FMD: FRED-MD. BBG: Bloomberg Surprises. P-values are in square brackets. *$p<0.1$ **$p<0.05$ ***$p<0.01$

We can also assess the stability of the selected information using Chow (1960) tests of parameter instability. Table 4.5 presents the results from this exercise. We find clear evidence of instability for the Michigan Survey of Consumers for both the overall and FRED-MD information sources. Although the information from the Bloomberg surprises only gradually captures the increase in the MSC-BCEI differential in 2020-21, it has not performed substantially worse than the earlier period and so does not indicate instability.

Another measure that shows signs of information instability is the HPR differential for both the overall and Bloomberg surprises. The Bloomberg surprises information fails to capture the sharp and persistent decline in the differential. However, this instability may stem in part from the fact that the HPR measure suffers from a zero-lower bound problem. Overall, the results are consistent with the view that measures of liquidity capture the movements in market-based measures of inflation expectations and that these relationships are robust to the large changes that have occurred since the start of the COVID-19 pandemic.

5 Conclusions

Long-term inflation expectations are believed to play an important role in driving and predicting future inflation. However, inflation expectations are not directly observed and the various measures derived from surveys of professional forecasters, consumers, and financial markets present a wide range of values. This raises questions as to whether the difference between survey and market-based measures contains information that can be used to forecast inflation.

In this paper we seek to understand some of these questions by decomposing these differences to extract the information. We start with a forecast-encompassing framework to understand whether the differential between survey and market-based measures is informative for forecasts. Next, we decompose and identify the information and rigidity channels and test which of these matters for improving inflation forecasts. Finally, we use machine learning methods to select over many possible sources of information and assess if the selected information captures the recent divergence in expectations.

Applying the methods to multiple measures of inflation expectations in the United States since 2003, we find that the differential between survey and market-based measures adds value to forecasts derived from survey-based measures. We identify the degree of rigidity in each measure of expectations using a constrained bivariate equilibrium correction model and use these estimates to decompose the overall differential into its relative contributions. We find that although the rigidity and information differentials both play a role, the information differential closely captures the overall differential and helps to explain the overall forecast improvements.

The information differential for market-based measures is explained by a handful of variables which correspond with broad measures of liquidity. We show that these changes in liquidity explain much of the historical variation in the information differential and can help predict the divergence in market-based measures of expectations in 2020 . Overall, our findings illustrate that market-based measures of expectations include unique information and that this information can be used to improve forecast performance.

References

Ahn, H. J. and Fulton, C. (2020). Index of Common Inflation Expectations. FEDS Notes No. 2020-09-02, Board of Governors of the Federal Reserve System, Washington, D.C.

Ajello, A., Benzoni, L., and Chyruk, O. (2019). Core and 'Crust': Consumer prices and the term structure of interest rates. *The Review of Financial Studies*, 33(8):3719–3765.

Altavilla, C., Giannone, D., and Modungo, M. (2017). Low frequency effects of macroeconomic news on government bond yields. *Journal of Monetary Economics*, 92:31–46.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

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

Angeletos, G.-M., Huo, Z., and Sastry, K. (2020). Imperfect Macroeconomic Expectations: Evidence and Theory. In Eichenbaum, M. S. and Hurst, E., editors, *NBER Macroeconomics Annual*, volume 35. Chicago: University of Chicago Press.

Aruoba, B. S. (2020). Term Structures of Inflation Expectations and Real Interest Rates. *Journal of Business & Economic Statistics*, 38(3):542–553.

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

Ball, L. and Mazumder, S. (2011). Inflation dynamics and the great recession. *Brookings Papers on Economic Activity*, 2011(1):337–381.

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

Bauer, M. D. (2015). Inflation expectations and the news. *International Journal of Central Banking*, 11(2).

Bauer, M. D. and McCarthy, E. (2015). Can we rely on market-based inflation forecasts? *FRBSF Economic Letter*, 30.

Berge, T. J. (2018). Understanding survey-based inflation expectations. *International Journal of Forecasting*, 34(4):788–801.

Bernanke, B. S. (2007). Inflation Expectations and Inflation Forecasting. Remarks by Chairman Ben S. Bernanke At the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts. Available Online At: `https://www.federalreserve.gov/newsevents/speech/bernanke20070710a.htm` [Last Accessed: 4th Jan, 2020], Board of Governors of the Federal Reserve System, Washington, D.C.

Binder, C. and Makridis, C. (2020). Stuck in the seventies: Gas prices and consumer sentiment. *Review of Economics and Statistics*, pages 1–42.

Böninghausen, B., Kidd, G., and de Vincent-Humphreys, R. (2018). Interpreting recent developments in market based indicators of longer term inflation expectations. ECB Economic Bulletin Issue 6/2018, European Central Bank, Frankurt, Germany.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2018). Over-reaction in macroeconomic expectations. NBER Working Paper No. 24932, National Bureau of Economic Research, Cambridge, MA.

Bullard, J. (2016). Inflation expectations are important to central bankers, too. The Regional Economist, Presidents Message, Federal Reserve Bank of St. Louis, St. Louis, MO.

Bürgi, C. (2020). Consumer inflation expectations and household weights. RPF Working Paper No. 2020-002, GWU Research Program on Forecasting, Washington, D.C.

Candia, B., Coibion, O., and Gorodnichenko, Y. (2020). Communication and the beliefs of economic agents. NBER Working Paper No. 27800, National Bureau of Economic Research, Cambridge, MA.

Carroll, C. D. (2003). Macroeconomic Expectations of Households and Professional Forecasters. *Quarterly Journal of Economics*, 118(1):269–298.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series. Pacific Grove, CA: Duxbury, second edition.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, 3(1).

Castle, J. L., Doornik, J. A., Hendry, D. F., and Nymoen, R. (2014). Misspecification testing: Non-invariance of expectations models of inflation. *Econometric Reviews*, 33(5-6):553–574.

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

Chen, Y. G. (2019). Inflation, inflation expectations, and the phillips curve. CBO Working Paper 2019-07, Congressional Budget Office, Washington, D.C.

Chong, Y. Y. and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, 53(4):671–690.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, pages 591–605.

Church, J. D. (2019). Inflation expectations and inflation realities: a comparison of the Treasury Breakeven Inflation curve and the Consumer Price Index before, during, and after the Great Recession. BLS Monthly Labor Review, U.S. Bureau of Labor Statistics, Washington, D.C.

Clarida, R. H. (2020). Financial Markets and Monetary Policy: Is There a Hall of Mirrors Problem? Remarks by Vice Chair Richard H. Clarida at the 2020 U.S. Monetary Policy Forum, sponsored by the Initiative on Global Markets at the University of Chicago Booth School of Business, New York, New York. Available Online At: `https://www.federalreserve.gov/newsevents/speech/files/clarida20200221a.pdf` [Last Accessed: 26th Feb, 2020], Board of Governors of the Federal Reserve System, Washington, D.C.

Clark, T. E. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341.

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nesting models. *Journal of Econometrics*, 105(1):85–110.

Clements, M. P. and Hendry, D. F. (1993). On the limitations of comparing mean square forecast errors. *Journal of Forecasting*, 12(8):617–637.

Coibion, O. and Gorodnichenko, Y. (2015a). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–78.

Coibion, O. and Gorodnichenko, Y. (2015b). Is the phillips curve alive and well after all? inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics*, 7(1):197–232.

Coibion, O., Gorodnichenko, Y., and Kamdar, R. (2018). The formation of expectations, inflation, and the phillips curve. *Journal of Economic Literature*, 56(4):1447–1491.

D'Amico, S., Kim, D. H., and Wei, M. (2018). Tips from TIPS: the informational content of Treasury Inflation-Protected Security prices. *Journal of Financial and Quantitative Analysis*, 53(1):395–436.

de Mendonça, H. F., Garcia, P. M., and Vicente, J. V. M. (2020). Rationality and anchoring of inflation expectations: An assessment from survey-based and market-based measures. *Journal of Forecasting*.

Diebold, F. X. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics*, 70:915–925.

Doornik, J. A. (2009). Autometrics. In Shephard, N. and Castle, J. L., editors, *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pages 88–121. Oxford, UK: Oxford University Press.

Doornik, J. A. (2018). *Ox 8: An Object-Oriented Matrix Programming Language*. Richmond, UK: Timberlake Consultants Ltd.

Doornik, J. A. and Hendry, D. F. (2018). *Empirical Econometric Modelling*, volume I of *PcGive 15*. Richmond, UK: Timberlake Consultants Ltd.

Ericsson, N. R. (1992). Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration. *Journal of Policy Modeling*, 14(4):465–495.

Ericsson, N. R. (1993). On the limitations of comparing mean square forecast errors: Clarifications and extensions. *Journal of Forecasting*, 12(8):644–651.

Ericsson, N. R. and Martinez, A. B. (2019). Evaluating government budget forecasts. In Williams, D. and Calabrese, T., editors, *The Palgrave Handbook of Government Budget Forecasting*, chapter 3, pages 37–69. Springer.

Fair, R. C. and Shiller, R. J. (1989). The informational content of ex ante forecasts. *Review of Economics and Statistics*, pages 325–331.

Fair, R. C. and Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *The American Economic Review*, pages 375–389.

Faust, J. and Wright, J. H. (2013). Forecasting Inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2A, chapter 1, pages 3–56. Amsterdam: Elsevier.

Fulton, C. and Hubrich, K. (2021). Forecasting US inflation in real time. Finance and Economics Discussion Series 2021-014, Board of Governors of the Federal Reserve System, Washington, D.C.

Girola, J. A. (2019). The Treasury Breakeven Inflation Curve. Presentation by James A. Girola, Washington, D.C. Available Online At: `https://www.treasury.gov/system/files/226/bri_mar2019_acc.pdf` [Last Accessed: 23rd Jan, 2020], U.S. Department of the Treasury, Office of Economic Policy, Washington, D.C.

Granger, C. W. J. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204.

Grothe, M. and Meyler, A. (2018). Inflation forecasts: Are market-based and survey-based measures informative? *International Journal of Financial Research*, 9(1):171–188.

Hansen, P. R. and Timmermann, A. (2015). Equivalence between out-of-sample forecast comparisons and wald statistics. *Econometrica*, 83(6):2485–2505.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business & Economic Statistics*, 16(2):254–259.

Haubrich, J., Pennacchi, G., and Ritchken, P. (2012). Inflation expectations, real rates, and risk premia: Evidence from inflation swaps. *The Review of Financial Studies*, 25(5):1588–1629.

Hendry, D. F. (1995). *Dynamic Econometrics*. Advanced Texts in Econometrics. Oxford, UK: Oxford University Press.

Hendry, D. F. and Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*. Arne Rynde Memorial Lectures. Cambridge, MA: MIT Press.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, 23(2):337–339.

Hillebrand, E., Huang, H., Lee, T.-H., and Li, C. (2018). Using the entire yield curve in forecasting output and inflation. *Econometrics*, 6(3):40.

Hoesch, L., Rossi, B., and Sekhposyan, T. (2020). Has the information channel of monetary policy disappeared? Revisiting the empirical evidence. Federal Reserve Bank of San Francisco Working Paper No. 2020-08, Federal Reserve Bank of San Francisco, San Francsico, CA.

Hungnes, H. (2018). Encompassing tests for evaluating multi-step system forecasts invariant to linear transformations. Discussion Paper No. 871, Statistisk Sentralbyrå, Oslo, Norway.

Jiao, X., Pretis, F., and Schwarz, M. (2021). Testing for coefficient distortion due to outliers with an application to the economic impacts of climate change.

Jorgensen, P. L. and Lansing, K. J. (2019). Anchored inflation expectations and the flatter phillips curve. Federal Reserve Bank of San Francisco Working Paper 2019-27, Federal Reserve Bank of San Francisco, San Francsico, CA.

Kilian, L. and Zhou, X. (2020). Oil prices, gasoline prices and inflation expectations: A new model and new facts. Federal Reserve Bank of Dallas Working Paper No. 2025, Federal Reserve Bank of Dallas, Dallas, TX.

Kim, D. H., Walsh, C., and Wei, M. (2019). Tips from TIPS: Update and Discussions. FEDS Notes No. 2019-5-21-3, Board of Governors of the Federal Reserve System, Washington, D.C.

Kliesen, K. L. (2015). How accurate are measures of long-term inflation expectations? Economic Synopses 9, Federal Reserve Bank of St. Louis, St. Louis, MO.

Kozicki, S. and Tinsley, P. A. (2012). Effective use of survey information in estimating the evolution of expected inflation. *Journal of Money, Credit and Banking*, 44(1):145–169.

Malmendier, U. and Nagel, S. (2016). Learning from inflation experiences. *The Quarterly Journal of Economics*, 131(1):53–87.

Mankiw, N. G., Reis, R., and Wolfers, J. (2003). Disagreement about inflation expectations. *NBER Macroeconomics Annual*, 18:209–248.

Martinez, A. B. (2015). How good are US government forecasts of the federal debt? *International Journal of Forecasting*, 31(2):312–324.

Martinez, A. B. (2017). Testing for differences in path forecast accuracy: Forecast-error dynamics matter. Working Paper No. 17-17, Federal Reserve Bank of Cleveland, Cleveland, OH.

Martinez, A. B., Castle, J. L., and Hendry, D. F. (2021). Smooth robust multi-horizon forecasts. *Advances in Econometrics*, Forthcoming.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Reis, R. (2020). The people versus the markets: A parsimonious model of inflation expectations. mimeo.

Romer, C. D. and Romer, D. H. (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, 90(3):429–457.

Rossi, B. and Sekhposyan, T. (2016). Forecast rationality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. *Journal of Applied Econometrics*, 31(3):507–532.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Stillwagon, J. R. (2018). TIPS and the VIX: Spillovers from financial panic to breakeven inflation in an automated, nonlinear modeling framework. *Oxford Bulletin of Economics and Statistics*, 80(2):218–235.

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

Thomas, L. B. (1999). Survey Measures of Expected U.S. Inflation. *Journal of Economic Perspectives*, 13(4):125–144.

Trehan, B. (2015). Survey measures of expected inflation and the inflation process. *Journal of Money, Credit and Banking*, 47(1):207–222.

Vergbrugge, R. J. and Binder, C. C. (2016). Digging into the downward trend in consumer inflation expectations. Economic Commentary 2016-11, Federal Reserve Bank of Cleveland, Cleveland, OH.

West, K. D. (2001). Tests for forecast encompassing when forecasts depend on estimated regression parameters. *Journal of Business & Economic Statistics*, 19(1):29–33.

Williams, P. D. (2020). Inflation Expectations in the U.S.: Linking Markets, Households, and Businesses. IMF Working Paper WP/20/240, International Monetary Fund, Washington, D.C.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In Aghion, P., Frydman, R., Stiglitz, J. E., and Woodford, M., editors, *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, chapter 1, page 25. Princeton, NJ: Princeton University Press.

Yellen, J. L. (2017). Inflation, Uncertainty, and Monetary Policy. Remarks by Chair Janet L. Yellen at the "Prospects for Growth: Reassessing the Fundamentals", 59th Annual Meeting of the National Association for Business Economics, Cleveland, Ohio. Available Online At: `https://www.federalreserve.gov/newsevents/speech/files/yellen20170926a.pdf` [Last Accessed: 4th Nov, 2020], Board of Governors of the Federal Reserve System, Washington, D.C.

A Appendix

A.1 MSE's

Ericsson (1992) notes that a smaller MSFE is a necessary but not sufficient condition to ensure that one forecast encompasses the other. The full sample forecast performance presented in Table A.1 illustrates that the differences between BCEI and most measures are generally not very large. MSC and LIV perform significantly worse, particularly at longer horizons. HPR and DKW perform slightly better although only HPR is significantly better around 1-year-ahead. TBI is inconsistent in that it performs significantly worse at the shortest and longest horizons but is significantly better around 1-year-ahead. ARU also performs significantly better at 1-year-ahead.

Table A.1: Relative Forecast Performance of Alternative Measures of Expectations

	RMSE	Relative RMSE (in %)								
h	BCEI	SPF	LIV	MSC	ARU	HPR	DKW	TBI	AVE	PC-F
1	2.34	100.2	100.1	101.8	100.2	100.3	99.9	100.9	100.2	100.2
2	2.44	100.5	100.5	104.6	100.3	99.8	99.6	102.7	100.6	100.6
3	2.35	100.4	100.6	105.3	100.2	97.5	99.4	102.2	100.2	100.2
4	2.35	99.8	100.4	105.0	99.7	96.7*	99.2	100.0	99.5	99.6
5	2.34	99.8	100.6	104.8	99.4*	97.1	99.6	98.0	99.3	99.4
6	2.33	100.2	101.3	105.7	99.9	98.3	100.2	99.6	100.0	100.1
7	2.32	100.4	101.4	106.5*	100.4	98.9	100.5	102.1	100.7	100.6
8	2.33	100.8	101.1	106.5*	100.2	98.2	100.4	103.1	100.7	100.7
Joint	2.34	100.2	100.6*	103.6*	99.9	98.0	100.0	100.8	100.2	100.2

Notes: h represents the number of quarters-ahead that are being forecast. AVE is the average of the different forecasts while PC-F is the First Principal Component of the forecasts following Hillebrand et al. (2018). The joint metric is a generalized version of the RMSE; see Clements and Hendry (1993). Tests of equal predictive accuracy are conducted using Diebold and Mariano (1995) and for joint horizons using Martinez (2017) where stars indicate a rejection of the null hypothesis of equal accuracy with the following probabilities: *$p<0.1$ **$p<0.05$ ***$p<0.01$

The differences in relative forecast performance are larger when focusing on the last ten years. This is illustrated in Table A.2 which shows that MSC performs significantly worse whereas the HPR performs better across all horizons. The path measures also suggest that SPF, ARU, HPR and DKW outperform relative to BCEI whereas LIV and MSC underperform relative to BCEI. Overall HPR outperforms across most metrics while MSC under-performs. However, MSC's underperformance is most likely due to inconsistencies in the measure of inflation which the MSC targets; see Bürgi (2020). The performance of measures such as SPF, ARU, DKW and TBI is less consistent across horizons.

Table A.2: Relative Forecast Performance of Alternative Measures of Expectations since 2010

	RMSE	Relative RMSE (in %)								
h	BCEI	SPF	LIV	MSC	ARU	HPR	DKW	TBI	AVE	PC-F
1	1.37	99.8	99.6	104.7	99.7	99.5	99.0	101.1	99.7	99.7
2	1.52	100.1	100.7	109.4*	100.2	96.7	98.5	101.6	99.9	99.9
3	1.51	99.8	101.1**	113.0*	99.6	93.3	97.6	99.1	99.2	99.3
4	1.51	98.6	100.6*	113.5*	98.6	89.9	96.1*	95.7	97.6**	97.8**
5	1.52	97.9	100.8	113.4*	98.1*	88.7	95.9*	94.5	97.0**	97.2**
6	1.51	99.7	103.1*	115.4*	99.2	89.3	96.9	98.4	98.6	98.7
7	1.50	101.4	102.9	118.2*	100.2	90.8	97.3	105.7	100.4	100.4
8	1.51	101.6	102.7*	118.5**	100.3	92.7	98.3	111.5*	101.5	101.5
Joint	1.36	99.7*	100.5*	103.9	99.7	95.4	98.4	100.3	99.5	99.5

Notes: h represents the number of quarters-ahead that are being forecast. AVE is the average of the different forecasts while PC-F is the First Principal Component of the forecasts following Hillebrand et al. (2018). The joint metric is a generalized version of the RMSE; see Clements and Hendry (1993). Tests of equal predictive accuracy are conducted using Diebold and Mariano (1995) and for the joint horizons using Martinez (2017) where stars indicate a rejection of the null hypothesis of equal accuracy with the following probabilities: *$p < 0.1$ **$p < 0.05$ ***$p < 0.01$

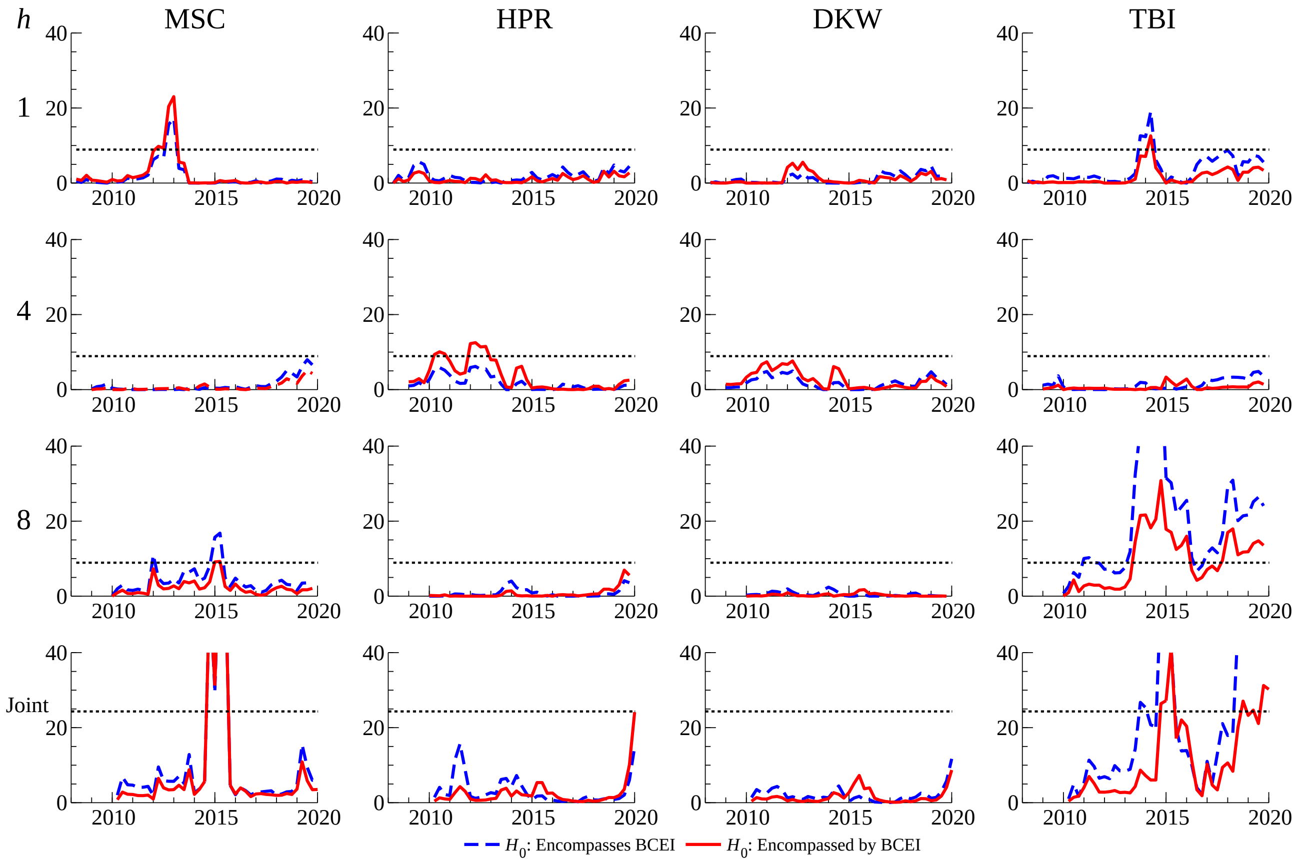
A.2 Fluctuation Encompassing Tests

We can assess the stability of the encompassing tests over time by re-estimating the encompassing test statistics based on (3) using a 20 quarter (five-year) rolling window. Since this is a restricted version of Hoesch et al. (2020)'s information-advantage fluctuation regression, the critical values from Rossi and Sekhposyan (2016) can be used.[17] We focus on four horizons: $h = \{1, 4, 8, \text{Joint}\}$ and four measures of expectations: MSC, HPR, DKW and TBI. The recursive test statistics are presented in Figure A.1.

There is strong evidence of instability in the encompassing test results for MSC and for TBI. This is especially true for the 1-quarter-ahead, 2-years-ahead, and the joint forecast horizons. For MSC, this occurs around the middle of the sample around 2012-2015 which coincides with the start of the post-2008 sample and could indicate underlying instabilities in MSC at that time. The instability of TBI is associated with the latter half of the sample around 2015 and 2018-19 where both hypotheses are strongly rejected.

The results for HPR and DKW are relatively stable with neither hypothesis rejected except for a brief period around 2010-12 at the one-year-ahead horizon where both HPR and DKW are more likely to encompass BCEI. There is also evidence towards the end of the sample that HPR is more informative than BCEI for all forecast horizons jointly.

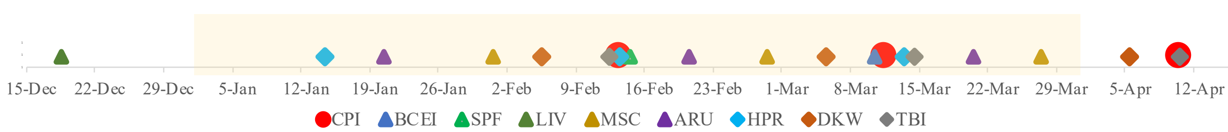
[17] We use the model-free critical values. Rossi and Sekhposyan (2016) argue that these critical values are valid across all horizons.



Notes: Each regression tests a single restriction (8 restrictions for the joint) and the estimation window of 20 quarters represents about 30% of the overall sample. All statistics are computed using HAC estimates from Andrews (1991). The dotted black line represents the critical value at which the null hypothesis is rejected at a 5% confidence level.

Figure A.1: Fluctuation Encompassing Tests

A.3 Additional Figures and Tables



Notes: DKW and TBI only contain information up through the end of the month prior to which they are released. The shaded area represents Q1.

Figure A.2: Illustrative Timeline of Release Dates for 2020 Q1

Table A.3: Forecast-Encompassing Coefficients and Probabilities Relative to BCEI

Horizon:	1	2	3	4	5	6	7	8	Joint
SPF	4.77	2.54	1.93	4.33	4.04	2.19	0.61	0.09	0.22
	⟦0.298⟧	⟦0.604⟧	⟦0.732⟧	⟦0.219⟧	⟦0.321⟧	⟦0.647⟧	⟦0.878⟧	⟦0.667⟧	⟦0.860⟧
	{0.189}	{0.392}	{0.478}	{0.111}	{0.044}**	{0.401}	{0.808}	{0.965}	{0.754}
LIV	10.00	2.03	1.47	3.00	2.05	-3.30	-3.58	-3.12	-2.86
	⟦0.023⟧**	⟦0.776⟧	⟦0.896⟧	⟦0.532⟧	⟦0.696⟧	⟦0.122⟧	⟦0.144⟧	⟦0.060⟧*	⟦0.000⟧***
	{0.012}**	{0.575}	{0.681}	{0.349}	{0.447}	{0.233}	{0.252}	{0.152}	{0.001}***
ARU	0.82	-2.33	-1.32	3.30	5.76	1.56	-1.66	-2.29	-1.26
	⟦0.969⟧	⟦0.367⟧	⟦0.524⟧	⟦0.493⟧	⟦0.159⟧	⟦0.876⟧	⟦0.499⟧	⟦0.286⟧	⟦0.107⟧
	{0.862}	{0.527}	{0.716}	{0.327}	{0.090}*	{0.663}	{0.673}	{0.457}	{0.154}

Notes: Horizon is number of quarters-ahead that are being forecast. The values in each block are 1): the estimated coefficients from equation (3) with a dummy variable for 2008 Q4, 2): The p-value associated with the null-hypothesis that the coefficient is equal to unity in the square brackets; and 3): The p-value associated with the null hypothesis that the coefficient is equal to zero in the squigly brackets. All tests use HAC estimates from Andrews (1991). *$p<0.1$ **$p<0.05$ ***$p<0.01$

Table A.4: Additional Joint Forecast-Encompassing Tests

	MSC		HPR		DKW		TBI	
	(1): Base	(2): (1)$\vert\nabla U$	(3): Base	(4): (3)$\vert\nabla U$	(5): Base	(6): (5)$\vert\nabla U$	(7): Base	(8): (7)$\vert\nabla U$
BCEI	-1.71	-1.71	1.05	1.34	0.85	0.86	-0.33	-0.33
	⟦0.111⟧	⟦0.132⟧	⟦0.690⟧	⟦0.588⟧	⟦0.983⟧	⟦0.988⟧	⟦0.114⟧	⟦0.134⟧
	{0.417}	{0.436}	{0.008}***	{0.024}**	{0.053}*	{0.136}	{0.436}	{0.442}
SPF	-2.63	-2.63	1.27	1.75	1.06	1.08	-0.43	-0.43
	⟦0.025⟧**	⟦0.035⟧**	⟦0.612⟧	⟦0.561⟧	⟦0.964⟧	⟦0.976⟧	⟦0.035⟧**	⟦0.050⟧**
	{0.114}	{0.134}	{0.014}**	{0.023}**	{0.080}*	{0.155}	{0.287}	{0.299}
LIV	-2.11	-2.11	1.34	1.74	1.28	1.30	-0.29	-0.29
	⟦0.123⟧	⟦0.162⟧	⟦0.571⟧	⟦0.534⟧	⟦0.964⟧	⟦0.966⟧	⟦0.102⟧	⟦0.131⟧
	{0.473}	{0.525}	{0.005}***	{0.016}**	{0.127}	{0.188}	{0.455}	{0.482}
ARU	-1.99	-2.09	1.21	1.48	0.98	0.99	-0.25	-0.25
	⟦0.065⟧*	⟦0.075⟧*	⟦0.466⟧	⟦0.527⟧	⟦0.976⟧	⟦0.980⟧	⟦0.175⟧	⟦0.238⟧
	{0.351}	{0.351}	{0.002}***	{0.013}**	{0.093}*	{0.134}	{0.597}	{0.634}

Notes: The values are the estimated coefficients from equation (3) with a dummy variable for 2008Q4. The estimates and tests follow as a special case from Hungnes (2018). ∇U represents the unemployment gap constructed using using quarterly unemployment and real-time estimates of the NAIRU from CBO and the Federal Reserve Board; see Fulton and Hubrich (2021). Measures along the rows represent the base forecasts. The p-value associated with the null-hypothesis that the coefficient is equal to unity is in the square brackets. The p-value associated with the null hypothesis that the coefficient is equal to zero is in the squigly brackets. *$p<0.1$ **$p<0.05$ ***$p<0.01$

Table A.5: Estimates of γ_m for Alternative Measures of Expectations

	BCEI	MSC	HPR	DKW	TBI	SPF	LIV	ARU
No Model Instability:	-0.005	0.005	0.025*	0.013	-0.023	0.025***	0.002	0.011**
	(0.007)	(0.012)	(0.015)	(0.010)	(0.025)	(0.009)	(0.005)	(0.005)
Model Instability:	0.019**	0.007	0.025*	0.013	-0.031	0.022**	0.001	0.011**
	(0.008)	(0.011)	(0.015)	(0.010)	(0.024)	(0.009)	(0.003)	(0.005)
Outlier Distotion Test:	719.7***	2.270	0.000	0.002	9.159***	5.941**	11.65***	0.025
	[0.000]	[0.132]	[0.996]	[0.968]	[0.002]	[0.015]	[0.001]	[0.874]

Notes: The estimation sample is 2003Q2 - 2019Q4 (67 observations). Allow for model instability/misspecification by selecting over and retaining outliers and shifts using Autometrics with Impulse Indicator Saturation (IIS) and Differenced-IIS (DIIS). The target gauge is 0.1% so that under the null we expect 0.2 irrelevant indicators to be retained on average. See Appendix Table A.6 for details on which outliers were retained. Outlier Distortion test is from Jiao et al. (2021). Estimated standard errors are in parentheses and p-values are in square brackets. *$p<0.1$ **$p<0.05$ ***$p<0.01$

Table A.6: Dates of Detected Outliers and Differenced Outliers

Date	BCEI	SPF	LIV	MSC	ARU	HPR	DKW	TBI
2006 Q4		DI						DI
2008 Q2				DI				
2008 Q4	I	I	I	I	I	I	I	I
2009 Q1	I							
2010 Q3				DI				DI
2010 Q4			I					
2011 Q4			I					
2013 Q4			I					
2014 Q4			I					

Notes: I = Impulse and DI = Differenced Impulse.