On the external validity of experimental inflation forecasts: A comparison with five categories of field expectations*

Camille Cornand•   Paul Hubert♦

August 26, 2019

Abstract

Establishing the external validity of experimental inflation forecasts is essential if laboratory experiments are to be used as decision-making tools for monetary policy. Our contribution is to document whether different measures of inflation expectations, based on various categories of agents (participants in experiments, households, industry forecasters, professional forecasters, financial market participants and central bankers), share common patterns. We do so by analysing the forecasting performance of these different categories of data, their deviations from full information rational expectations, and the variables that enter the determination of these expectations. Overall, the different categories of forecasts exhibit common features: forecast errors are comparably large and autocorrelated, and forecast errors and forecast revisions are predictable from past information, suggesting the presence of some form of bounded rationality or information imperfections. Finally, lagged inflation positively affects the determination of inflation expectations. While experimental forecasts are relatively comparable to survey and financial market data, more heterogeneity is observed compared to central bank forecasts.

Keywords: inflation expectations, experimental forecasts, survey forecasts, market-based forecasts, central bank forecasts.

JEL Classification: E3, E5, E7.

---

* We would like to thank Klaus Adam, Cars Hommes, Domenico Massaro, Damjan Pfajfar, Matthias Weber and Blaz Žakelj for sharing their experimental data. We also thank Christophe Blot and Pauline Gandré for detailed feedbacks, two anonymous referees, the editors, Herbert Dawid and Jan Tuinstra, and participants to the 2018 BEAM-ABEE workshop, in particular John Duffy, Cars Hommes, Simas Kucinskas, Tomasz Makarewicz, and Baptiste Massenot for valuable comments. This research was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program ''Investissements d'Avenir'' (ANR-11-IDEX-007) operated by the French National Research Agency (ANR).
• Univ Lyon, CNRS, GATE L-SE UMR 5824, F-69130 Ecully, France; email: cornand@gate.cnrs.fr.
♦ Sciences Po - OFCE, 10 place de Catalogne, 75014 Paris, France; email: paul.hubert@sciencespo.fr.

1. Introduction

Understanding the formation of economic agents' inflation expectations is crucial for the conduct of monetary policy. Recently, a growing macro-experimental literature has focused on inflation expectation formation in the laboratory.[1] Laboratory experiments—particularly Learning-to-Forecast Experiments (LtFE)[2] (see Hommes (2011) for a survey)—have been used to validate expectation hypotheses and learning models. These experiments can also serve as important tools for central bankers, providing a test bed for competing policy actions or monetary policy rules (Cornand and Heinemann, 2014, 2019). The external validity of expectations is critical for policy initiatives to be valid outside the laboratory (Duffy, 2008).[3]

The question of how agents form inflation expectations is generally studied using survey data on expectations of future inflation. Survey data present the advantages (over experimental data) of providing "natural" expectations that are, in principle, more representatively sampled and that should be subject to more external validity. However, they present the inconvenience of paying a fixed reward, implying that there is no incentive for economic agents to provide an answer that is as accurate as possible.[4] Extracting expectations from financial instruments (such as inflation swaps) provides a response to the lack of incentives. However, there are no direct observations of expectations, and their extrapolation can interfere with biases or other uncontrolled elements (yielding a problem of testing joint hypotheses). While survey measures are not directly linked to financial decisions, market-based measures incorporate liquidity or risk premia. In contrast, laboratory experiments provide data on expectations that respond to incentives. They also offer data, the generating process of which is perfectly known by the experimenter and provides the experimenter with a large number of independent observations. Overall, all inflation expectation data (experimental data, market-based data, survey data and central bank data) have some relative advantages and drawbacks, and they provide an overview of how economic agents form inflation expectations, how successful they are in forming these forecasts, the extent to which they deviate from the benchmark of full information rational expectations, and the variables that enter their determination.

---

[1] More generally, experimental investigations of macroeconomic phenomena have started to become important, especially in the field of monetary economics. Among the various studies using experimental methods in macroeconomics, see, e.g., Arifovic and Sargent (2003), Engle-Warnick and Turdaliev (2010), Luhan and Scharler (2014), Lambsdorff et al. (2013), Kryvtsov and Petersen (2013) and Noussair et al. (2015). For a nice and rather comprehensive overview of macroexperiments, see Duffy (2016). Behavioural macroeconomic theory papers have also recently built upon experiments, e.g., De Grauwe and Macchiarelli (2015), Cole and Milani (2017), and Bertasiute et al. (2018), to mention but a few.

[2] In LtFE, participants in the experiment play the role of professional forecasters. Their task is to provide their expectations about an economic variable (for example, the market price or, in the case in which we are interested, inflation). Their payoffs depend negatively on their forecast error. The expectations that are formed by subjects in the laboratory are aggregated (either by mean or median), and this summary statistic is introduced into the theoretical model as the aggregate expectation of agents. Most recent experiments on DSGE models have used variants of the standard three equations: IS curve, Phillips curve, and policy rule. This model is directly implemented via a computer program, except for the expectations determined by participants in the experiment. The computer program derives the current values of variables conditional on the model parameters, which yield, period after period, time series of the main variables.

[3] External validity is particularly important for macro-experimentalists since they necessarily have recourse to small-scale laboratory evidence. A complementary line of research consists of testing the robustness of macro-experiments when considering large group sizes (100 participants) instead of the standard group size of 6 to 10 participants. See Hommes et al. (2018) and Bao et al. (2019).

[4] In fact, some surveys conducted among professionals have introduced incentives to respond as accurately as possible. For example, the publisher of Blue Chip Economic Indicators organizes an annual dinner, during which the most accurate forecaster of the previous year is honoured and identified in later issues of the newsletter. Something similar is organized by the central bank of Brazil for professional inflation forecasts.

The debate on the external validity of experimental inflation forecasts echoes the issue—noted by Carroll (2003), among others—of the heterogeneity between the inflation expectations of various categories of agents—professional forecasters, industrial forecasters, central bankers, households, financial market participants—and how each of these groups forms expectations. While the literature has provided some comparisons of survey measures of inflation expectations (see, e.g., Thomas (1999) for such a study of US data), a comparison across *types* of measures (survey, market based, central banks, and experimental), based on various categories of agents (including households, industry, professional forecasters, financial market participants, central bankers, and participants in laboratory experiments), has not yet been performed. Our contribution is therefore to document whether different types of measures share common patterns by:

- Assessing the characteristics of forecast errors

To this end, we analyse for each of our different samples (experimental, survey, financial market and central bank data) whether forecast errors are significant to evaluate whether economic agents (participants in laboratory experiments, households, industry, professional forecasters, financial market participants, central bank staff and policymakers) make systematic mistakes, i.e., form biased inflation expectations. We also measure the magnitude of absolute forecast errors to establish the forecasting quality of our different samples. To do so, we rely on the methodology used by Diebold and Mariano (1995), Romer and Romer (2000), and Ang, Bekaert and Wei (2007) to test the forecasting performance of different types of data.

- Capturing deviations from the full information rational expectation benchmark

To capture these deviations (either due to informational friction or because economic agents use simple rules of thumb), we study for each of our different samples for whether forecast errors are autocorrelated, whether they are correlated with forecast revisions and whether forecast revisions depend on past forecast revisions. Here, we follow the methodology of Andolfatto, Hendry and Moran (2007), Coibion and Gorodnichenko (2012, 2015), and Andrade and Le Bihan (2013), who noted the role of information rigidities.

- Evaluating the usual determinants of inflation forecasts

We study how lagged inflation and output gaps affect inflation forecasts to evaluate whether the usual determinants of inflation are used by the different considered categories of agents. The formation processes of expectations and the variables entering this process have been notably studied by Mankiw and Reis (2002), Sims (2003), Lanne, Luoma and Luoto (2009), Pfajfar and Santoro (2010), Fendel, Lis and Rülke (2011), and Dräger, Lamla and Pfajfar (2016). We follow their methodology.

While these analyses have been conducted separately and have excluded experimental data, our paper precisely aims at conducting these analyses in parallel to provide a large set of characteristic comparisons and to enlarge the data sets by including experimental inflation expectations. Using standard measures of forecast characteristics, determinants of forecast errors and expectation formation determination, these comparisons are intended to determine whether the various sets of inflation expectations exhibit heterogeneous or common patterns and thus to examine the external validity of experimental inflation forecasts. While establishing the external validity of experimental inflation expectations represents a key step in the development of macro-experiments, especially those addressing monetary policy issues, to our knowledge, there is no available study relying on a sample of experimental data on inflation expectations, confronting it with field data.[5] From this perspective, our aim is also to advise the policymaker about the informativeness of the different types of data. Indeed, *all* of these sources are used by central banks in their policy decisions.

Despite the considerable heterogeneity among our six different categories of data, we find that the different data sets exhibit various common features: forecast errors are comparably large; autocorrelations of forecast errors are positive and significant; forecast errors and forecast revisions are very often predictable; and the standard lagged inflation determinant of inflation expectations is robust to the data sets. There is nevertheless some heterogeneity among the six different sets. If experimental forecasts are relatively comparable to survey and financial market data, central banks' forecasts seem superior since they do not exhibit systematic bias, are less autocorrelated and are hardly predictable. Excluding central bank forecasts, we conclude that experimental data are comparable to other data sets in the sense that forecast errors exhibit the same type of bias (except for industry forecasts), and lagged forecast revisions significantly predict forecast revisions.

The paper is structured as follows. Section 2 presents the data. Section 3 describes the empirical results. Section 4 concludes the paper.

2. Data

We collect inflation expectation forecast data for four different types of measures of inflation expectations (experimental data, survey data, financial market data, and central bank data), corresponding to six categories of agents (participants in experiments, households, industry, professional forecasters, financial market participants, central bank's staffs and policymakers). We also collect macroeconomic data. Descriptive statistics of the different series are provided in Table A in the Appendix.

We acknowledge the heterogeneity of the different data sets with respect to their forecasting horizons, their frequency, and the sample period considered. Regarding the horizon and frequency, while they might be different from one set to the other, it is worth emphasizing that, for all categories of agents, they correspond to the relevant horizon and frequency for their respective usual economic decisions, while the forecasting horizon and frequency in experimental forecasts are abstract. Regarding the sample period for field data, we focus our empirical analysis on a relatively recent sample period, from 1987 to 2017, for comparability purposes between types of agents, as well as macroeconomic and structural environments. Two exceptions are inflation swaps that start in 2004 and Greenbook forecasts that end in 2012.

2.1. Data from laboratory experiments *(from various published research papers)*

We collect a sample of macro-experimental data on inflation expectation formation from published papers.[6] The Learning-to-Forecast (LtF) design, based on the New Keynesian (NK) reduced-form model or some variant, offers the incentives to induce accurate inflation forecasts.[7]

[5] Some recent papers have focused on establishing the external validity of experiments on expectation formation in a different manner. In particular, Armantier et al. (2015) presented a study in which they compared consumers' survey data on inflation expectations to the behaviour of the same subjects in a financially incentivized investment experiment. They showed that stated beliefs in the survey and experimental decisions were highly correlated and conformed to theoretical predictions. Armantier et al. (2017) mixed experimental and survey methods to investigate how consumers' inflation expectations respond to new information. They randomly provided a subset of agents with factual information (i.e., either past-year average food price inflation or the average forecast of next-year overall inflation in the Survey of Professional Forecasters). They were thus able to identify the causal effects of new information on agents' expectations. Our methodology and aim are different: we investigate whether experimental data share the same pattern as field data.

The first four considered experimental papers implement variants of the standard NK three equation model: IS curve, Phillips curve, and policy rule.

$$y_t = E_t y_{t+1} - \varphi(i_t - E_t \pi_{t+1}) + g_t$$
$$\pi_t = \lambda y_t + \rho E_t \pi_{t+1} + u_t$$
$$i_t = \bar{\pi} + \phi_\pi(\pi_t - \bar{\pi}) + \phi_y(y_t - \bar{y})$$

where π_t and y_t are the inflation rate and output gap in period t, $\bar{\pi}$ and $\bar{y}$ are their steady state values, i_t is the nominal interest rate, g_t and u_t are exogenous disturbances, $E_t \pi_{t+1}$ is the average expected inflation, $E_t y_{t+1}$ is the average expected output gap, φ, λ, ρ, ϕ_π, and ϕ_y are positive parameters. The economy is qualitatively described to participants in experiments. Instructions include an explanation of the mechanisms that govern model equations (in particular, monetary policy is described as central bank intervention reducing interest rates to increase inflation and conversely).

The first paper is that of Pfajfar and Žakelj (2018) (henceforth PZ), which presented an LtFE (conducted at the University Pompeu Fabra in Spain and the University of Tilburg in the Netherlands) based on the above-presented reduced form of the NK model. They asked subjects to form inflation expectations (more precisely, a prediction of the *t*+1 period inflation and the 95% confidence interval of their inflation prediction) but no output gap expectations. Instead, a computer program feeds the model with naïve output gap expectations: $E_t y_{t+1} = y_{t-1}$ in the above model specification. The parameters values are standard: $\rho = 0.99$, $\lambda = 0.3$, $\varphi = 0.164$, and $\bar{\pi} = 3$. Since they investigated the targeting rule that best stabilizes the economy, they considered four different treatments, corresponding to four different policy rules: inflation forecast targeting, with three different degrees of monetary policy aggressiveness: $\phi_y = 0$ and $\phi_\pi = 1.5$ or 1.35 or 4; and contemporaneous inflation targeting, with an intermediate degree of monetary policy aggressiveness: π_t is replaced by $E_t \pi_{t+1}$ in the monetary policy rule with $\phi_\pi = 1.5\%$. The disturbances g_t and u_t follow an AR-1 process. Subjects observe the history of macroeconomic variables: at each period *t*, participants observe inflation, the output gap and the interest rate up to period *t*-1. There are 70 periods, and each period corresponds to one quarter. The number of observations amounts to 24 independent groups.

The second paper is Cornand and M'baye (2018a) (henceforth CMB1), who focused on a design very close to that of PZ. Their experiment relied on the same model, but the considered parameter values were slightly different: $\rho = 0.99$, $\lambda = 0.3$, $\varphi = 1$, $\bar{\pi} = 5$ (to avoid the focal 2% value). As in PZ, CMB1 asked subjects to state inflation expectations but not output expectations, for which they assumed again $E_t y_{t+1} = y_{t-1}$. CMB1 studied the role of the central bank's inflation target communication by comparing treatments in which the central bank explicitly announces its inflation target to treatments in which the central bank does not announce the target. They considered four treatments differing by the type of inflation-targeting procedure. More precisely, under strict inflation targeting, the sole objective of the central bank is to stabilize inflation ($\phi_\pi = 1.5$, $\phi_y = 0$). Under the explicit strict inflation target treatment, the central bank announces its 5% target, while under the implicit

[6] Other closely related LtFEs include Assenza et al. (2013), Arifovic and Petersen (2017) and Mauersberger (2018).

[7] While a full model (e.g., in Noussair et al. (2015)) generates payoffs in line with microfoundations and economic agents' roles in reality (in so far as subjects represent consumers and producers who directly interact on the product and labour markets), inflation forecasts are complicated to infer from the sole observation of subjects' decisions in markets. It is however possible to simultaneously elicit expectations and decisions in markets as in Petersen (2015).

strict inflation target treatment, there is no announcement about the target value. Under a flexible inflation target, the central bank has both an inflation objective and an output gap stabilization objective ($\phi_\pi = 1.5$, $\phi_y = 0.5$). Again, depending here on whether the flexible inflation target is made explicit or not, the central bank communicates its target for inflation or not. An important point is that, in contrast to the other considered experiments, subjects can be provided with information about the target depending on the treatment. Additionally, in contrast to PZ, g_t and u_t are independently and normally distributed shocks with a standard deviation of 0.1. For each treatment, CMB1 has 4 sessions, and each session lasted 50 periods.

The third paper is Cornand and M'baye (2018b) (henceforth CMB2), which is similar to CMB1 in terms of design. The parameter values are the same ($\rho = 0.99$, $\lambda = 0.3$, $\varphi = 1$, $\bar{\pi} = 5$), and they focus on the case in which the central bank stabilizes both inflation and the output gap ($\phi_\pi = 1.5$, $\phi_y = 0.5$). They considered four different treatments differing with respect to whether the central bank implements a band or point target and also by the size of shocks (g_t and u_t are independently and normally distributed shocks with standard deviations of 0.08 or 0.4 depending on the treatment). More precisely, the considered treatments were the following. In the band inflation targeting with small shock treatment, the central bank simply announces a band inflation target (interval [4% - 6%]) to the public, in a context in which shocks have low variance. In the point inflation targeting with small shock treatment, the central bank explicitly communicates its 5% numerical target with a tolerance band of +/-1% around its target in a context in which the variance of shocks is low. In the band inflation targeting with large shock treatment, the central bank simply announces the band inflation target ([4% - 6%]) to the public but in a context in which the variance of shocks is relatively high. In the point inflation targeting with large shock treatment, the central bank explicitly communicates its 5% numerical target with a tolerance band of +/-1% around its target but in a context in which the variance of shocks is relatively high. The authors had 4 sessions for each treatment, and each session lasted 60 periods. Both experiments by Cornand and M'baye were conducted at the GATE-Lab of the University of Lyon in France.

The fourth paper is Hommes et al. (2017) (henceforth HMW), which presented an LtFE (conducted at the CREED lab at the University of Amsterdam in the Netherlands) based again on the simple standard version of the New Keynesian model considered above. The parameter values are the same as in Cornand and M'baye: $\rho = 0.99$, $\lambda = 0.3$, and $\varphi = 1$, except for the target, which is lower, $\bar{\pi} = 3.5$. The main difference is that subjects' task consists of forming *both* inflation and output gap expectations in period *t* for period *t*+1.[8] They considered two different treatments, corresponding to the implementation of two different policy rules by the central bank: one in which the central bank reacts to inflation only ($\phi_\pi = 1.5$, $\phi_y = 0$); and one in which the central bank additionally reacts to the output gap ($\phi_\pi = 1.5$, $\phi_y = 0.5$). As in PZ, CMB1 and CMB2, the subjects observe the history of macroeconomic variables (all realizations of inflation, output gaps and interest rates) up to period *t*-1. As in CMB1 and CMB2, g_t and u_t are independently and normally distributed shocks with a standard deviation of 0.1. There are 50 periods; the number of observations amounts to 43 independent groups.

The fifth paper is Adam (2007). His experiment occurred at the University of Salerno in Italy and at the Goethe University of Frankfurt in Germany. The model for the experiment has a different

[8] Their aim was to test a theoretical behavioural model showing that the central bank's reaction to the output gap, beyond inflation, reduces inflation volatility.

structure than that of the other considered experiments. Current output and inflation are determined as functions of predetermined variables and agents' expectations of future endogenous variables according to:

$$\pi_t = -\bar{\pi} + \frac{1}{\bar{y}\varepsilon} y_{t-1} + \left(1 - \frac{1}{\bar{\pi}\varepsilon}\right) E_{t-1}\pi_t + E_{t-1}\pi_{t+1}$$

$$y_t = \left(1 - \frac{1}{\bar{\pi}}\right)\bar{y} + \frac{1}{\bar{\pi}}\left(1 - \frac{1}{\bar{\pi}\varepsilon}\right) y_{t-1} - \frac{\bar{y}}{\bar{\pi}^2}\left(1 - \frac{1}{\bar{\pi}\varepsilon}\right) E_{t-1}\pi_t - \frac{\bar{y}}{\bar{\pi}^2} E_{t-1}\pi_{t+1} + v_t.$$

In the experiment, the considered parameter values are: $\varepsilon = 2$ most of the time or $\varepsilon = 1$ depending on the low or high elasticity treatment, $\bar{\pi} = 1.04$, $\bar{y} = 100$, and $v_t \backsim ii\ U[-1,1]$. Participants play the role of firms that set prices one period in advance and must perform inflation forecasts 2 periods ahead (although prices are sticky for a single period only). The subjects do not know any features of the underlying economy. In any period *t*, the subjects observe the history of output and inflation up to period *t*-1, and they are asked to forecast inflation for periods *t* and *t*+1.[9] The experiment is between an LtFE and a Learning-to-Optimize (LtO) experiment since there is an additional optimization task. The economy is subject to shocks (mean zero white noise shock with small bounded support). There are 6 independent groups.

Table 1 summarizes the similarities and differences between the experimental designs of the considered papers. It shows that the data-generating processes, the precise tasks of subjects, the incentives, the information sets and the characteristics of subjects could differ from one experiment to the next.

Table 1 – Characteristics of experimental designs

<table>
<tr><th rowspan="2"></th><th colspan="3">Data generating process</th><th rowspan="2">Subjects' task</th><th rowspan="2">Incentives in ECU
f stands for the absolute forecast error</th><th rowspan="2" colspan="2">Information set at date *t:* history of variables up to period *t*-1</th><th rowspan="2">Subjects: undergraduate students</th></tr>
<tr><th>Model</th><th>Output gap expecta-tions</th><th>Shocks</th></tr>
<tr><td>PZ</td><td rowspan="4">Qualitative description of the same NK model structure with relatively standard and comparable parameter values</td><td rowspan="3">Pre-programmed naïve</td><td>AR-1 process</td><td rowspan="3">Inflation forecasts in *t* for *t*+1</td><td>$max\left\{\frac{100}{1+f} - 20{,}0\right\}$</td><td rowspan="4">Inflation, Output gap, Interest rate, profits from forecasts</td><td></td><td>Group size: 9
Majors: economics, business</td></tr>
<tr><td>CMB1</td><td rowspan="4">iid distributions

variances may differ</td><td rowspan="2">$max\left\{\frac{160}{1+f} - 40{,}0\right\}$</td><td rowspan="2">In most treatments, information about the inflation target</td><td>Group size: 6
Majors: economics, business</td></tr>
<tr><td>CMB2</td><td>Group size: 6
Majors: engineering, business</td></tr>
<tr><td>HMW</td><td>Part of subjects' task</td><td>Inflation and output gap forecasts in *t* for *t*+1</td><td>$\frac{100}{1+f}$</td><td></td><td>Group size: 6
Majors: economics, business</td></tr>
<tr><td>Adam</td><td>Variant of the model, no monetary policy Unknown to subjects</td><td>Pre-programmed naïve</td><td>Inflation forecasts for periods *t* and *t*+1, Price</td><td>$\frac{400}{1+f} - 100$</td><td colspan="2">Inflation, Output gap</td><td>Group size: 5
Majors: business, engineering</td></tr>
</table>

[9] Note that only average forecasts per group (and not individual ones) are available for this paper.

				setting			

The assumed data-generating process has an impact on the heuristics that subjects use. Whenever the model implemented in the experiment requires being fed with expectations not determined by the subjects, the issue of the form of expectation to be used arises. All of the considered experiments, except for HMW, implement a naïve form for output gap expectations. While there is survey and experimental evidence arguing in favour of such an expectational form, this assumption could have the consequence of reinforcing inertia in subjects' stated inflation expectations. The form and size of shocks also matter. With i.i.d. shocks, potential fluctuations in inflation are only endogenously driven by agents' expectations. Depending on the context one wants to reproduce, alternative ways of modelling shocks might have to be considered (e.g., AR-1, as in PZ). Finally, knowledge of the data-generating process or not has an impact on subjects' forecasting ability.

Some of the considered experiments required the participants to perform a single task (CMB1, CMB2, PZ), while others (Adam and HMW) required multiple tasks. In particular, Adam asked for the setting of prices, as well as the forming of two expectations. Bao et al. (2013) showed that combining LtF with LtO leads to larger errors in both. In addition, the incentives in the LtF design could be more or less strong. In Adam, these incentives were relatively weak compared to the other papers.

The information set of participants was relatively comparable across the considered experiments. Note, however, that the subjects were more informed in CMB1 (half of data) and CMB2, since they were explicitly informed about the quantitative central bank's inflation target. This information might have generated more forward-looking behaviour.

Finally, the characteristics of subjects participating in the experiment and the group size were again reasonably comparable. Regarding the characteristics of participants, all of the considered experiments invited undergraduate students in economics, business or engineering from various European universities, and these subjects thus represent a very homogenous, educated group.[10] Concerning group size, coordination is likely more difficult with a larger number of subjects, but considering groups greater than 4, as here, is sufficient to justify the use of the competitive equilibrium as equilibrium concept (see, e.g., Huck et al. (2004)).

2.2. Survey data

2.2.1. Households *(Michigan)*

The Michigan Survey of Consumer Attitudes and Behavior surveys a cross-section of the population about their expectations over the next year. Most papers using the Michigan survey cover only the period since 1978, during which these data have been collected monthly and on a quantitative basis: respondents were asked to state their precise quantitative inflation expectations. Before then, the Michigan survey was qualitative. It has been conducted quarterly since 1946, although for the first 20 years, the respondents were asked only whether they expected prices to rise, fall, or stay the same. Each month, a sample of approximately 500 households is interviewed, in which the sample is chosen to statistically represent households in the US, excluding Alaska and Hawaii. The monthly phone call survey focuses on respondents' perceptions and expectations regarding personal finances, business conditions and news regarding the economy in general, as well as macroeconomic aggregates, such as unemployment, interest rates and inflation. Furthermore, the survey collects individual and household socioeconomic characteristics.

[10] We do not have the precise demographic data for the different considered experiments. However, in this respect, LtFE participants probably fall in the category of the so-called WEIRD subjects (Heinrich et al., 2010a, b).

2.2.2. Industry *(Livingston)*

The Livingston Survey was started in 1946 by the late columnist Joseph Livingston. It is the oldest continuous survey of economists' expectations. It summarizes the forecasts of economists from industry, the government and academia in the US. The Federal Reserve Bank of Philadelphia took responsibility for the survey in 1990. The Livingston Survey covers analysts and economists working in industry. It is conducted twice per year, in June and December, so it has a semiannual frequency. It provides twelve-month Consumer Price Index (CPI) inflation forecasts from approximately 50 survey respondents.

2.2.3. Professional forecasters *(Survey of Professional Forecasters)*

The Survey of Professional Forecasters (SPF) is collected and published by the Federal Reserve Bank of Philadelphia. It focuses on professional forecasters mostly in the banking sector in the US. Surveys are sent to approximately 40 panellists at the end of the first month of the quarter, the deadline for submission is the second week of the second month of the quarter, and forecasts are published between the middle and end of February, May, August, and November. GDP price index forecasts (available since 1968) are fixed-horizon forecasts for the current and the next four quarters. They are provided as annualized quarter-over-quarter growth rates. We also perform our analysis with CPI forecasts provided since 1981. We consider the median of individual responses, rather than the mean, which could be affected by potential outliers.

2.3. Financial market instruments *(swap data)*

Market-based inflation expectations are derived from inflation swaps. These instruments are financial market contracts to transfer inflation risk from one counterparty to another. We consider instantaneous forwards at different maturities that measure expected inflation as of the date of maturity of the contract. In general, the advantage of financial market expectations over survey measures of expectations is that they are directly related to payoff decisions, so there is no strategic response bias or no difference between stated and actual beliefs. However, one disadvantage is that financial market expectations do not provide a direct measure of inflation expectations since they are affected by credit risk, liquidity and inflation risk premia. Swaps tend to be a better market measure for deriving inflation expectations than inflation-indexed bonds because they are generally less sensitive to liquidity and risk premia. Another advantage of market-based measures is that they are available daily. For comparison purposes, we also perform our analysis at the monthly frequency and calculate the average of all of the working day observations in each month. These data are available since October 2004 only for liquidity reasons.

2.4. Central bank (Fed)

2.4.1. Federal Open Market Committee (FOMC)

The FOMC has published forecasts for inflation and real GDP growth twice per year in the Monetary Policy Report to the Congress since 1979. Since October 2007, their publication has been quarterly. We consider forecasts of the GDP deflator until 1988, then the Consumer Price Index until 1999 and then the Personal Consumption Expenditures (PCE) measure of inflation following the focus of the FOMC. These forecasts are fourth quarter-over-fourth quarter growth rates for the current and next calendar years. Until 2005, the forecast for the next year was published only per year. While each FOMC member was required to submit a forecast, the Monetary Policy Reports provide only summary statistics for each variable. In particular, they report "central tendency" values, which show the highest and lowest forecasts after dropping the extremes (commonly defined as the three highest and three lowest values, although this practice is not consistently made clear in the reports) and the "range" of forecasts listing the highest and lowest values. We consider the midpoint of the "full range" of all individual FOMC members' forecasts, which should be more informative for all views in the FOMC than the central tendency. It should then also be more comparable with surveys or experimental data, which are not truncated. These FOMC forecasts are a mix of the model-based forecasts of the Greenbook (see below) and FOMC members' judgement.

2.4.2. Greenbook

The "Greenbook" contains the forecasts of the staff of the Federal Reserve. These forecasts are model-based forecasts, formed and provided to the FOMC members before FOMC meetings. They are made available to the public after a five-year embargo, and they forecast different measures of inflation and real GDP/GNP growth over different quarterly horizons up to 1 year ahead. They are available for all horizons since 1969Q4 and are measured as annualized quarter-over-quarter growth rates.

2.5. Macroeconomic data

Regarding experimental data, macroeconomic variables (inflation and output gap) are generated by a computer program that implements a model of the economy, conditional on the parameters and on the expectations and prices that participants to the experiment are asked for, depending on the design (inflation expectations for all experiments considered in this paper and prices for Adam only).

For observed macroeconomic data, we use the Consumer Price Index for All Urban Consumers (FRED mnemonic: CPIAUCSL), the Gross Domestic Product: Implicit Price Deflator Index (GDPDEF), the Personal Consumption Expenditures: Chain-type Price Index (PCECTPI) and the Real Gross Domestic Product, Billions of Chained 2009 Dollars (GDPC1).

2.6. Comparability of laboratory and field data

It is not clear which type of field expectations is closer to the expectations formed in a LtFE. We discuss the respects in which the various field expectations (households, industry, market participants, professional or central bank expectations) could be compared to experimental expectations, especially depending on the design features.

Intuitively, households seem most similar to laboratory subjects outside of the laboratory. However, participants in the considered laboratory experiments are usually students, who are thus more educated than average people in surveys. Instead, experimental expectations are comparable to industry's expectations due to a more comparable level of education between the students usually recruited to perform an experimental task and people working in industry.

The considered experimental data are close to survey data in the sense that both provide direct observations of expectations. However, in LtFE, tasks are incentivized, in contrast to surveys. A weak form of incentives might be more compatible to the survey data of households, while stronger forms might be more in line with survey of professional forecasters or central bankers' forecasting tasks, whose incentives to form accurate forecasts are either financial or reputational. Financial market participants' forecasts are not direct observations but are comparable to experimental forecasts in the sense that both categories of agents are directly financially incentivized to form accurate forecasts.

Experimental forecasts in LtFE are close to financial market participants' forecasts, professional forecasts and central bank forecasts due to the NK data-generating process motivating their forecasts (in the sense that they depend on inflation and economic slack) and the similarities among their information sets. Indeed, in LtFE, macroeconomic data information and the data-generating process are made salient to participants.[11]

Finally, regarding heterogeneity, at least for experimental (see Table 1), survey and central bank data, there is some heterogeneity within samples in the manner in which data are collected (type of survey or type of design and model underlying experiments), and the category of economic agents can vary. A priori, there is no reason for one sample to be more heterogeneous than another.

3. Empirical evidence

We first analyse the forecasting performance of our different types of data (inflation expectations from participants in laboratory experiments, households, industry, professional forecasters, financial market participants, central bankers). Then, following Coibion and Gorodnichenko (2012), we study the extent to which our different samples compare to each other in terms of deviation from the full information rational expectation benchmark. Third, we analyse in what respect the usual determinants enter our different categories of inflation expectations.

While survey data are available for a very long period of time, to have comparable samples (unbiased for potentially lower-quality forecasts in the past), we focus on the relatively recent period of 1987-2017.[12] For laboratory experiments, we rely on more than 38,000 observations.[13]

3.1. Forecast characteristics: bias and accuracy

What is the forecasting performance of our different categories of economic agents? Are forecast errors biased for the different categories of agents?[14] Is their accuracy comparable? Our strategy consists of analysing whether forecast errors exhibit systematic biases before focusing on absolute forecast errors to establish the overall quality of these forecast errors.

We denote by π_t the inflation rate in period *t* and $\overline{\pi_{\mathrm{t+1\backslash t}}}$ the mean forecast across agents made in period *t* for inflation in period *t*+1. The forecast error in period *t* for inflation in period *t*+*h* is defined by: $\mathrm{FE_{t,t+h}} \equiv \pi_{t+h} - \overline{\pi_{\mathrm{t+h\backslash t}}}$. We first test the following equation (following Romer and Romer (2000) and Ang, Bekaert, and Wei (2007)):

$$\mathrm{FE_{t,t+h}} = \beta_1 + \epsilon_t \qquad (1)$$

[11] The information set of participants is leverage that can be used to mimic different categories of agents. Poor information about the data-generating process, as well as the economic environment (as in Adam), is likely a design feature that can account for the average household's environment. Indeed, households are probably economic agents from the field that are the least informed about the working of the economy, as well as macroeconomic data. In contrast, a clearer data-generating process (as in the other considered experiments) might be more appropriate to generate expectations in line with those of analysts and economists from the industry (as is the case in the Livingstone Survey). Offering more salient macroeconomic data and information about the data-generating process is a way to account for the expectations of more informed categories of agents, such as financial market participants, professional forecasters and central bankers.

[12] Results for the full sample are provided in the Appendix (Tables B, C, D, E, F, and G).

[13] We use individual data when available since they provide more information. We nevertheless offer some robustness checks using the groups in Table M in the Appendix for comparability purposes with average survey data, the midpoint of central bank data and the market clearing price of swap data. In addition, Table O in the Appendix reports standard errors clustered at the individual or group level for experimental data. Standard errors are slightly higher, but the main messages remain unchanged.

[14] Note that we rely on an unconditional bias. An alternative methodology calculating a bias conditional on the perception of shocks is provided by Kucinskas and Peters (2018).

where β_1 is the estimated constant, ϵ_t is the error term and the null hypothesis is: the estimated constant β_1 is not significantly different from 0.

While forecast rationality (as defined, e.g., in Romer and Romer (2000)) implies that forecast errors should theoretically—consistent with the commonly maintained rational expectations assumption—be null on average, the literature provides some evidence that this may not be the case: economic agents are usually prone to making persistent forecast errors.[15]

Table 2 presents the results of the estimation of Equation (1) for our different types of measures and categories of agents. To render the results more easily comparable, Table 2 also provides forecast errors normalized by the standard deviation of the predicted variable (i.e., the different measures of inflation). A significant coefficient indicates that the forecast is biased. A positive coefficient indicates that economic agents underestimate inflation.

Forecast errors are almost systematically negative in our sample of laboratory experiments: participants in laboratory experiments tend to overestimate inflation. However, there is some heterogeneity. The data of HMW participants instead significantly underestimate inflation. In the sub-samples by Adam and CMB2, the error is not significant; however, as becomes clear below from the analysis of absolute forecast errors, this outcome is due to a compensation in errors.

The forecasts of financial market participants (extracted from inflation swaps) are always negative and significant. They increase with the horizon. Considering daily or monthly data does not affect these results. Similarly, forecast errors in the household sample are also negative and significant.

Forecast errors of professional forecasters exhibit a slightly less clear pattern: those regarding CPI for the next quarter (cpi1q) are not significant, while those for the next year (cpi4q) are. GDP price index forecasts for the current and next quarters (pgdp1q and pdgd4q) are also significant. We conclude that forecast errors of professional forecasters are most of the time negative and significant.

In contrast, in the industry sample, forecast errors are not significant, regardless of the horizon of the forecast (expectations at 6 months (living6m) and at 12 months (living12m)). Central bank forecasts are also not significant (except pce4q).

Table B in the Appendix performs the same analysis for the full field sample period. The results exhibit less clear patterns for survey and central bank forecasts. For a longer time period, the forecast errors of industry and central bank become significant, while those of households and professional forecasters become insignificant.

[15] For evidence see, e.g., Roberts (1997), Croushore (1997), Thomas (1999), and more recently, Mankiw et al. (2004) and Mehra (2002). Note, however, that some authors have pointed out that expectations are apparently biased errors serially correlated that cannot be construed as evidence against the rational expectations hypothesis. Andolfatto et al. (2007) argued that the hypothesis of unbiasedness tends to be rejected in particular in small samples but less often in larger samples and that it might be rational to be adaptive for agents when they cannot disentangle the effects of persistent and transitory shocks. Note finally that Romer and Romer (2000) observed that forecast rationality is insured after correcting for serial correlation for almost all of their data (except the Blue Chip sample, for which rationality is obtained when excluding Volcker disinflation). In our case, since some of the series exhibit forecast rationality, we do not control for serial correlation for comparability purposes to estimate the same regression specification in all cases.

Table 2 - Forecast errors

Experimental forecasts								
	(1)	(2)	(3)	(4)	(5)	(6)		
	All obs.	PZ	Adam	CMB1	CMB2	HMW		
β_1	-0.009*	-0.042***	-0.022	-0.116***	0.007	0.064***		
	[0.00]	[0.01]	[0.13]	[0.01]	[0.01]	[0.01]		
β_1 (normalised)	-0.005*	-0.026***	-0.013	-0.072***	0.004	0.040***		
	[0.00]	[0.01]	[0.08]	[0.01]	[0.00]	[0.00]		
N	38424	14904	510	4704	5664	12642		
Market-based forecasts								
	Daily			Monthly average				
	(1)	(2)	(3)	(4)	(5)	(6)		
	swap3y	swap5y	swap10y	swap3y	swap5y	swap10y		
β_1	-0.766***	-1.069***	-1.920***	-0.740***	-1.060***	-1.919***		
	[0.03]	[0.02]	[0.03]	[0.14]	[0.10]	[0.15]		
β_1 (normalised)	-0.526***	-0.735***	-1.320***	-0.508***	-0.728***	-1.318***		
	[0.02]	[0.01]	[0.02]	[0.10]	[0.07]	[0.10]		
N	2514	1992	687	116	92	32		
Survey forecasts								
	Livingston		Michigan	SPF				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q	
β_1	0.113	-0.063	-0.406***	-0.197**	-0.360***	0.008	-0.215**	
	[0.22]	[0.21]	[0.06]	[0.08]	[0.09]	[0.08]	[0.10]	
β_1 (normalised)	0.037	-0.021	-0.143***	-0.078**	-0.142***	0.003	-0.073**	
	[0.07]	[0.07]	[0.02]	[0.03]	[0.03]	[0.03]	[0.03]	
N	61	60	372	122	119	123	120	
Central bank forecasts								
	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_1	0.075	-0.051	0.032	0.035	0.200	0.263	0.286	0.558**
	[0.07]	[0.08]	[0.09]	[0.11]	[0.15]	[0.20]	[0.19]	[0.26]
β_1 (normalised)	0.038	-0.026	0.013	0.014	0.059	0.078	0.200	0.390**
	[0.03]	[0.04]	[0.04]	[0.04]	[0.04]	[0.06]	[0.14]	[0.18]
N	124	120	103	100	103	100	51	48

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (1) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Overall, comparing our different samples, SPF forecasts (except cpi1q), financial market participants' forecasts, households' and laboratory experiment forecasts are consistent and exhibit systematic errors: inflation is over-estimated. Forecasts obtained in laboratory experiments exhibit lower errors. One possible reason for the relatively lower forecast errors in experimental data is the relative simplicity of the data-generating process and more detailed information sets.[16] Central banks' and industry's forecasts exhibit no significant results. As becomes clear below, considering absolute forecast errors qualifies such a result.[17]

Second, we provide a robustness check of the results from Table 2 by estimating the absolute forecast error. Our aims are to evaluate whether positive and negative errors possibly compensated

[16] Adam has the largest average absolute forecast errors likely because his participants do not know anything about the data-generating process and had to perform multiple tasks.

[17] Table I in the Appendix reports measures of the dispersion and tails of forecast errors, showing that the different data distributions exhibit tails fatter than the tails from a normal distribution, except with market-based data.

for each other and to determine the quality of forecasts for our different categories of agents and types of measures.

The absolute forecast error in period *t* for inflation in period *t+h* is defined by: $\left|\mathrm{FE}_{\mathrm{t,t+h}}\right| \equiv \left|\pi_{t+h} - \overline{\pi_{\mathrm{t+h\backslash t}}}\right|$. Following Romer and Romer (2000)[18] and Ang, Bekaert, and Wei (2007) among others, we test the following equation:

$$\left|\mathrm{FE}_{\mathrm{t,t+h}}\right| = \beta_2 + \epsilon_t \tag{2}$$

where β_2 is the estimated constant, ϵ_t is the error term and the null hypothesis is: the estimated constant β_2 is not significantly different from 0.

Table 3 provides estimates of Equation (2) for our different samples. To render results more easily comparable, Table 3 also provides absolute forecast errors normalized by the standard deviation of the predicted variable (i.e., the different measures of inflation).

Table 3 shows that absolute forecast errors are all significantly different from zero. Experimental forecast errors all become significant, households' forecast errors become significant even with a long period sample, industry's forecast errors become significant even in the short 1990-2017 sub-period, professional forecast errors all become significant, but there seems to be an inconsistency since swap3y errors are larger than swap5y (for both daily and monthly data). Also for SPF, we see that the CPI provision is better than the GDP deflator at one year.[19] In terms of accuracy (as evaluated by the average magnitude of absolute forecast errors), forecast errors are comparable in our different data sets, especially when considering absolute forecast errors normalized by the standard deviation.[20] They seem to be more pronounced for market-based data and much less pronounced for experimental data (although the size of the sample might play a role).

Overall, the comparison between the analyses of forecast errors and absolute forecast errors enables us to formulate the following result.

Result 1: While the forecast errors of participants in laboratory experiments, financial market participants, households and professional forecasters are systematically biased (i.e., forecasts exhibit over-evaluation of inflation, on average), those from central banks and industry surveys are not. However, for each type of measure (experimental, survey, financial market and central bank data) and each category of economic agent, the forecast accuracy in terms of absolute forecast error is comparably large.

A few remarks are in order. First, this result confirms the superiority of central bank's forecasts already established in the literature.[21] Romer and Romer (2000) – who compared the forecasting performance of the Federal Reserve (using Greenbook data) and of commercial banks (using Blue

[18] More precisely, Romer and Romer (2000) use the mean squared errors to estimate forecast accuracy, rather than the absolute forecast error.

[19] These results are robust to the considered period, as shown in Table C in the Appendix.

[20] That forecast errors are also comparable within the sample of experimental data is interesting. Because the designs by CMB1 and CMB2 focus on the central bank's inflation target communication, while other experimental designs do not, we could have expected more heterogeneity in forecast errors. Indeed, Dräger et al. (2016) linked consistent expectations, central bank communication and forecast accuracy. More precisely, they interpreted consistency as effectiveness of central bank communication.

[21] Note, however, that this superiority should be considered with care because the number of observations is small. Nevertheless, it is not smaller than some of the other data sets (Livingston, SPF or monthly swaps, for instance) and not smaller than equivalent samples in the literature.

Chip Economic Indicators, Data Resources, Inc. (DRI) and SPF) – showed that the Federal Reserve forecasts inflation better than commercial forecasters.

Table 3 - Absolute forecast errors

Experimental forecasts

	(1) All obs.	(2) PZ	(3) Adam	(4) CMB1	(5) CMB2	(6) HMW
β_2	0.462***	0.573***	1.813***	0.329***	0.258***	0.418***
	[0.00]	[0.01]	[0.10]	[0.01]	[0.01]	[0.00]
β_2 (normalised)	0.288***	0.357***	1.130***	0.205***	0.161***	0.260***
	[0.00]	[0.01]	[0.06]	[0.01]	[0.00]	[0.00]
N	38424	14904	510	4704	5664	12642

Market-based forecasts

	Daily			Monthly average		
	(1) swap3y	(2) swap5y	(3) swap10y	(4) swap3y	(5) swap5y	(6) swap10y
β_2	1.347***	1.217***	1.920***	1.340***	1.199***	1.919***
	[0.02]	[0.02]	[0.03]	[0.10]	[0.08]	[0.15]
β_2 (normalised)	0.926***	0.837***	1.320***	0.920***	0.824***	1.318***
	[0.01]	[0.01]	[0.02]	[0.07]	[0.05]	[0.10]
N	2514	1992	687	116	92	32

Survey forecasts

	Livingston		Michigan	SPF			
	(1) living6m	(2) living12m	(3) mich1y	(4) pgdp1q	(5) pgdp4q	(6) cpi1q	(7) cpi4q
β_2	1.184***	1.159***	0.870***	0.695***	0.837***	0.674***	0.843***
	[0.16]	[0.15]	[0.04]	[0.05]	[0.05]	[0.05]	[0.06]
β_2 (normalised)	0.392***	0.384***	0.307***	0.275***	0.331***	0.229***	0.286***
	[0.05]	[0.05]	[0.01]	[0.02]	[0.02]	[0.02]	[0.02]
N	61	60	372	122	119	123	120

Central bank forecasts

	FOMC		Greenbook					
	(1) fomc_cy	(2) fomc_ny	(3) pgdp1q	(4) pgdp4q	(5) cpi1q	(6) cpi4q	(7) pce1q	(8) pce4q
β_2	0.572***	0.739***	0.750***	0.826***	1.041***	1.407***	1.042***	1.421***
	[0.04]	[0.05]	[0.06]	[0.07]	[0.11]	[0.14]	[0.13]	[0.18]
β_2 (normalised)	0.291***	0.376***	0.292***	0.322***	0.309***	0.417***	0.729***	0.994***
	[0.02]	[0.03]	[0.02]	[0.03]	[0.03]	[0.04]	[0.09]	[0.12]
N	124	120	103	100	103	100	51	48

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (2) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Second, our first result qualifies the superiority of professional forecasts. The literature has usually found that professional forecasters stand in a better position than many other economic agents in forecasting inflation (e.g., Carroll (2003)). Ang, Bekaert and Wei (2007) offered a comparison of four methods to study inflation forecasting: time series forecasts, forecasts based on the Phillips curve, forecasts from the yield curve, and surveys (Livingston, Michigan, and SPF). Their comparison also showed the superiority of survey forecasts in forecasting inflation. Nevertheless, commenting on such a result, they wrote (p. 1165), "*That the median Livingston and SPF survey forecasts do well is perhaps not surprising, because presumably many of the best analysts use time-series and Phillips curve models. However, even participants in the Michigan survey who are consumers, not professionals, produce accurate out-of-sample forecasts, which are only slightly worse than those of the professionals in the Livingston and SPF surveys.*" Our results are in line with this comment and extend it to experimental data.

Third, our first result emphasizes the performance comparability of laboratory data to various categories of field data: experimental forecasts are biased like other data and are not less accurate than other data.

3.2. Testing deviations from full information rational expectations

We evaluate our different categories of economic agents for whether forecast errors are autocorrelated, whether they depend on forecast revisions, and whether forecast revisions depend on past forecast revisions. This analysis follows Coibion and Gorodnichenko (2012). Theoretically, in a frictionless world and under rational expectations, forecast errors should not be correlated with previous forecast errors. However, two possible departures from this benchmark case can be considered. First, economic agents might not form rational expectations and can, for instance, use simple rules of thumb. Bounded rationality can give rise to autocorrelation in forecast errors. Second, although economic agents form rational expectations, information rigidities can generate autocorrelation in forecast errors. For instance, in the sticky information model (Mankiw and Reis, 2002), "*forecast errors depend both on the inflation process after the shock and on the degree of information rigidity. […] As the degree of information rigidity rises, conditional forecast errors will become increasingly persistent.*" (Coibion and Gorodnichenko, 2012, p. 122). There is much evidence in the literature that they are. For example, Diebold (1989) showed that inflation forecast errors are typically serially correlated and hence predictable. Romer and Romer (2000, p. 433) also showed that "*the serial correlation increases as the horizon for the forecasts becomes longer.*" However, more recently, evidence has been mixed. Coibion and Gorodnichenko (2012) showed that "*forecast errors are not predictable using lagged inflation conditional on lagged forecast errors.*"

First of all, to study the potential autocorrelation of forecast errors, we test the following equation:

$$\mathrm{FE}_{\mathrm{t,t+h}} = C_1 + \beta_3 \mathrm{FE}_{\mathrm{t-1,t+h-1}} + \epsilon_t \quad (3)$$

where β_3 is the estimated coefficient, C_1 is a constant, ϵ_t is the error term and the null hypothesis is: the estimated coefficient β_3 is not significantly different from 0.

Table 4 provides estimations of Equation (3) for our different samples and shows that the forecast errors are indeed autocorrelated: the forecast error is predictable owing to the former error, suggesting that economic agents do not form rational expectations. This characteristic is robust over almost all of our samples (with only a few exceptions: industry data (Livingston), Greenbook data (cpi1q and pce1q)). Note that Table D in the Appendix shows that, considering a longer period of time, autocorrelation of forecast errors on industry data also becomes significant. In terms of amplitude, market-based forecast errors seem to be more autocorrelated, while Greenbook data exhibit less autocorrelation in forecast errors (although one should be careful in interpretation because coefficients are biased by the frequency of the sample). Considering fixed effects for experimental data does not yield further insights, as shown in Table N in the Appendix.

Table 4 - Autocorrelation of forecast errors

Experimental forecasts

	(1) All obs.	(2) PZ	(3) Adam	(4) CMB1	(5) CMB2	(6) HMW
β_3	0.540***	0.680***	0.115**	0.370***	0.229***	0.461***
	[0.00]	[0.01]	[0.04]	[0.01]	[0.01]	[0.01]
constant	-0.007*	-0.010	-0.022	-0.070***	0.007	0.020***
	[0.00]	[0.01]	[0.13]	[0.01]	[0.01]	[0.01]
N	37748	14688	500	4608	5568	12384
R²	0.29	0.45	0.01	0.15	0.06	0.23

Market-based forecasts

	Daily			Monthly average		
	(1) swap3y	(2) swap5y	(3) swap10y	(4) swap3y	(5) swap5y	(6) swap10y
β_3	0.991***	0.993***	0.996***	0.912***	0.911***	0.929***
	[0.00]	[0.00]	[0.00]	[0.04]	[0.04]	[0.07]
constant	-0.008*	-0.008**	-0.008	-0.085	-0.096	-0.117
	[0.00]	[0.00]	[0.01]	[0.06]	[0.06]	[0.15]
N	2513	1991	686	115	91	31
R²	0.98	0.99	0.99	0.85	0.83	0.85

Survey forecasts

	Livingston		Michigan	SPF			
	(1) living6m	(2) living12m	(3) mich1y	(4) pgdp1q	(5) pgdp4q	(6) cpi1q	(7) cpi4q
β_3	-0.202	-0.102	0.898***	0.427***	0.548***	0.611***	0.744***
	[0.13]	[0.13]	[0.02]	[0.08]	[0.08]	[0.07]	[0.06]
constant	0.138	-0.071	-0.039	-0.112	-0.161**	0.009	-0.059
	[0.22]	[0.22]	[0.03]	[0.07]	[0.08]	[0.06]	[0.07]
N	61	60	372	122	119	123	120
R²	0.04	0.01	0.81	0.18	0.30	0.38	0.56

Central bank forecasts

	FOMC		Greenbook					
	(1) fomc_cy	(2) fomc_ny	(3) pgdp1q	(4) pgdp4q	(5) cpi1q	(6) cpi4q	(7) pce1q	(8) pce4q
β_3	0.666***	0.819***	0.469***	0.665***	0.109	0.236**	0.223	0.321**
	[0.07]	[0.05]	[0.09]	[0.08]	[0.10]	[0.10]	[0.14]	[0.14]
constant	0.027	-0.012	0.018	0.008	0.178	0.204	0.259	0.372
	[0.05]	[0.05]	[0.08]	[0.08]	[0.15]	[0.19]	[0.19]	[0.27]
N	124	120	103	100	103	100	50	47
R²	0.45	0.67	0.22	0.45	0.01	0.06	0.05	0.10

Note: Standard errors in brackets. * p < 0.10, ** p < 0.05, *** p < 0.01. Parameters are obtained by estimating equation (3) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Second, we analyse whether forecast errors are predictable owing to forecast revisions. The forecast revision in period *t* for inflation in period *t+h* is defined by: $\mathrm{FR}_{\mathrm{t,t+h}} \equiv \overline{\pi_{\mathrm{t+h\backslash t}}} - \overline{\pi_{\mathrm{t+h-1\backslash t-1}}}$. We estimate the following equation (equivalent to Equation (11) in Coibion and Gorodnichenko (2012)):

$$\mathrm{FE}_{\mathrm{t,t+h}} = C_2 + \beta_4 \mathrm{FR}_{\mathrm{t,t+h}} + \epsilon_t \quad (4)$$

where β_4 is the estimated coefficient, C_2 is a constant, ϵ_t is the error term and the null hypothesis is: the estimated coefficient β_4 is not significantly different from 0. If economic agents form full information rational expectations, forecast errors should theoretically not be correlated with forecast revisions.

Table 5 provides estimates of Equation (4) for our different samples. A significant coefficient indicates that the error is predictable owing to the revision.

Table 5 - Forecast errors on forecast revisions

Experimental forecasts

	(1) All obs.	(2) PZ	(3) Adam	(4) CMB1	(5) CMB2	(6) HMW
β_4	0.066***	0.231***	0.314***	-0.384***	-0.365***	-0.141***
	[0.01]	[0.01]	[0.07]	[0.01]	[0.01]	[0.01]
constant	-0.008*	-0.040***	-0.017	-0.116***	0.007	0.064***
	[0.00]	[0.01]	[0.12]	[0.01]	[0.01]	[0.01]
N	38423	14903	510	4704	5664	12642
R^2	0.00	0.05	0.04	0.24	0.18	0.02

Market-based forecasts

	Daily			Monthly average		
	(1) swap3y	(2) swap5y	(3) swap10y	(4) swap3y	(5) swap5y	(6) swap10y
β_4	-0.542***	-0.453*	-0.155	-0.740*	-0.354	0.408
	[0.18]	[0.23]	[0.68]	[0.43]	[0.40]	[1.17]
constant	-0.767***	-1.069***	-1.920***	-0.763***	-1.061***	-1.913***
	[0.03]	[0.02]	[0.03]	[0.14]	[0.10]	[0.16]
N	2513	1991	686	115	91	31
R^2	0.00	0.00	0.00	0.03	0.01	0.00

Survey forecasts

	Livingston		Michigan	SPF			
	(1) living6m	(2) living12m	(3) mich1y	(4) pgdp1q	(5) pgdp4q	(6) cpi1q	(7) cpi4q
β_4	-0.640*	-0.750	-0.260	-0.036	0.037	-0.232	-0.59
	[0.36]	[0.48]	[0.18]	[0.23]	[0.41]	[0.19]	[0.50]
constant	0.104	-0.08	-0.406***	-0.197**	-0.359***	0.007	-0.223**
	[0.22]	[0.21]	[0.06]	[0.08]	[0.09]	[0.08]	[0.10]
N	61	60	372	122	119	123	120
R^2	0.05	0.04	0.01	0.00	0.00	0.01	0.01

Central bank forecasts

	FOMC		Greenbook					
	(1) fomc_cy	(2) fomc_ny	(3) pgdp1q	(4) pgdp4q	(5) cpi1q	(6) cpi4q	(7) pce1q	(8) pce4q
β_4	-0.182	0.295	-0.212*	0.121	0.146	-0.651	0.193	-0.549
	[0.20]	[0.39]	[0.12]	[0.28]	[0.09]	[0.72]	[0.13]	[1.10]
constant	0.074	-0.049	0.029	0.037	0.203	0.248	0.329*	0.544**
	[0.07]	[0.08]	[0.09]	[0.11]	[0.15]	[0.20]	[0.19]	[0.27]
N	124	120	103	100	103	100	50	47
R^2	0.01	0.01	0.03	0.00	0.02	0.01	0.04	0.01

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (4) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Errors of professional forecasters and central bankers (except pdgp1q) are not predictable (as for financial market participants over long horizons) owing to forecast revision, meaning that they tend to use the information available to them. In contrast, coefficients are significant in the samples of laboratory participants, industry (except linving6m) and to some extent, for financial market participants (especially swap3y).[22] Data from laboratory experiments are thus relatively comparable to financial market participants' data in this respect (although the amplitude is even more pronounced for market-based data).

[22] It seems that both industry and household data are more reliable in recent periods than in the past since Table E in Appendix shows that, over a longer period, forecast errors on forecast revisions on industry and household data become significant when they were not for a shorter period or even more significant when they were already significant.

However, there is some heterogeneity among our samples and within the sample of experiments in terms of signs. When economic agents revise their expectations upward between *t*-1 and *t*, it reduces their forecast error in industry data (when significant, i.e., during the 1956-2017 period, as presented in Table E of the Appendix) and for the experiment by PZ and Adam, but it increases their forecast error in CMB1, CMB2 and in HMW,[23] in household data (over a long time period, as shown in Table E in the Appendix) and in financial market data over short horizons.[24] We can thus conclude that, in contrast to professional and central bankers' forecasts, for households and industry, forecast errors are predictable owing to forecast revisions (although less than for the other samples) and in the same direction as in the financial market participant sample, as well as in the same direction of some of the experimental sub-samples.

Third, we evaluate whether forecast revisions can be predicted owing to past forecast revisions. More precisely, the question that we answer is whether economic agents revise their expectations upwards if the former forecast is revised upwards. To this end, we test the following equation:

$$\mathrm{FR}_{\mathrm{t,t+h}} = C_3 + \beta_5 \mathrm{FR}_{\mathrm{t-1,t+h-1}} + \epsilon_t \quad (5)$$

where β_5 is the estimated coefficient, C_3 is a constant, ϵ_t is the error term and the null hypothesis is: the estimated coefficient β_5 is not significantly different from 0. Forecast revisions should theoretically not be correlated with lagged forecast revisions since there is no reason for economic agents to always revise their forecasts in the same direction.

Table 6 presents estimations of Equation (5) for our different samples. A significant and positive coefficient indicates that economic agents revise their expectations upwards if the former forecast is revised upwards. This table exhibits significant coefficients (except for households, swap10y for financial market participants with monthly data, pgdp4q for professional forecasters and most central bank's forecasts). The sign is generally negative (except for PZ's and HMW's sub-samples with experimental data). Table F in the Appendix overall confirms this analysis over longer periods, except for industry data that become non-significant and household data that become significant.[25]

Overall, we can state the following result:

Result 2: For each type of measure and each category of agent with the notable exception of central bank data, inflation forecasts depart from the full information rational expectation benchmark.

- *(a) Forecast errors are highly autocorrelated for all types of measures (experiments, surveys, financial markets, and central bank data) and for each category of agent, except for industry forecasters.*
- *(b) Forecast errors are predictable owing to forecast revisions except for central banks and professional forecasters.*
- *(c) Lagged forecast revisions enter significantly and usually negatively in forecast revisions for all types of measures and each category of agents—indicating that economic agents alternately revise their expectations upwards and downwards—except for central bank data.*

---

[23] Considering fixed effects does not alter the results regarding experimental data, as shown in Table N in the Appendix.

[24] Table H in the Appendix provides additional information to Table 5 by providing a decomposition between current and lagged inflation forecasts following Equation (12) in Coibion and Gorodnichenko (2012). In particular, in the case in which previous estimates of β_4 would be insignificant, one can see whether one of the two (current and lagged inflation forecasts) is significant. Overall, the main result of Table 5 is confirmed by Table H.

[25] Considering fixed effects does not alter the results regarding experimental data, as shown in Table N in the Appendix.

Table 6 - Forecast revisions on lagged forecast revisions

Experimental forecasts

	(1) All obs.	(2) PZ	(3) Adam	(4) CMB1	(5) CMB2	(6) HMW
β_5	0.078***	0.258***	-0.156***	-0.471***	-0.441***	0.051***
	[0.01]	[0.01]	[0.04]	[0.01]	[0.01]	[0.01]
constant	-0.001	-0.008	-0.011	-0.019	0.001	0.019***
	[0.00]	[0.01]	[0.08]	[0.01]	[0.01]	[0.01]
N	37748	14687	500	4608	5568	12385
R^2	0.01	0.07	0.02	0.25	0.21	0.00

Market-based forecasts

	Daily			Monthly average		
	(1) swap3y	(2) swap5y	(3) swap10y	(4) swap3y	(5) swap5y	(6) swap10y
β_5	-0.310***	-0.223***	-0.268***	-0.353***	-0.278***	-0.103
	[0.02]	[0.02]	[0.02]	[0.08]	[0.08]	[0.08]
constant	0.000	0.000	0.000	-0.009	-0.007	-0.006
	[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.01]
N	3295	3295	3295	150	150	150
R^2	0.10	0.05	0.07	0.12	0.08	0.01

Survey forecasts

	Livingston		Michigan	SPF			
	(1) living6m	(2) living12m	(3) mich1y	(4) pgdp1q	(5) pgdp4q	(6) cpi1q	(7) cpi4q
β_5	-0.466***	-0.432***	-0.075	-0.188**	-0.126	-0.230**	-0.176**
	[0.11]	[0.12]	[0.05]	[0.09]	[0.09]	[0.09]	[0.09]
constant	-0.025	-0.032	-0.001	-0.009	-0.014	-0.011	-0.015
	[0.07]	[0.05]	[0.02]	[0.03]	[0.02]	[0.04]	[0.02]
N	62	62	373	124	124	124	124
R^2	0.22	0.19	0.01	0.04	0.02	0.05	0.03

Central bank forecasts

	FOMC		Greenbook					
	(1) fomc_cy	(2) fomc_ny	(3) pgdp1q	(4) pgdp4q	(5) cpi1q	(6) cpi4q	(7) pce1q	(8) pce4q
β_5	0.004	-0.039	-0.408***	-0.173*	-0.237**	0.014	-0.218	0.122
	[0.09]	[0.09]	[0.09]	[0.10]	[0.10]	[0.10]	[0.14]	[0.14]
constant	-0.004	-0.007	-0.014	-0.015	-0.029	-0.021	-0.045	-0.013
	[0.03]	[0.02]	[0.07]	[0.04]	[0.15]	[0.03]	[0.20]	[0.03]
N	124	124	104	104	104	104	50	50
R^2	0.00	0.00	0.17	0.03	0.06	0.00	0.05	0.02

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (5) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Our second result calls for some comment. First, our second result confirms those of previous studies. In particular, it is in line with Andrade and Le Bihan (2013) who—relying on the European Central Bank Survey of Professional Forecasters—showed that forecasters fail to update their forecasts and have predictable forecast errors. As argued by Coibion and Gorodnichenko (2012, p. 136), "*because professional forecasters are some of the most informed economic agents, […] any evidence of information rigidity on their part [is] particularly notable*". Our result also goes in the direction of that of Romer and Romer (2000), emphasizing the idea that staff and policymakers from the Fed form better forecasts than professional (commercial) forecasters.

Second, despite differences in the economic contexts or in the structures of information flows and in the design regarding experimental data sets (for example, with vs. without communication about a forward-looking variable, such as the central bank target in CMB1 and CMB2 vs. PZ and HMW) that could have induced more or less information friction, our results are relatively consistent within samples. The most important heterogeneity in results is observed when regressing forecast errors on forecast revisions (Result 2 (b)) and can possibly be attributed to these differences in economic context and design.[26]

3.3. Forecast determination

Are inflation forecasts determined by the usual candidates, namely, lagged inflation and the lagged output gap for each of our different categories of economic agents? Based on insights from, e.g., Lanne et al. (2009), Fendel et al. (2010), and Dräger et al. (2016), we test the following equation:

$$\overline{\pi_{\mathrm{t+1\backslash t}}} = C_4 + \beta_6 \pi_{t-1} + \beta_7 y_{t-1} + \epsilon_t \tag{6}$$

where β_6 and β_7 are the estimated coefficients, y_{t-1} denotes the output gap in period *t-1*, C_4 is a constant, and ϵ_t is the error term. The output gap is the detrended measure of real GDP or industrial production (depending on data set frequencies) using the Christiano-Fitzgerald filter. Based on the standard Phillips curve relationship, we expect inflation forecasts to be significantly positively affected by both lagged inflation and the lagged output gap.

Table 7 reports results from the estimation of Equation (6) of inflation expectation determination; it also reports estimates of Equation (6) when considering lagged GDP, instead of the lagged output gap for field data.[27] We observe that all categories of economic agents significantly consider lagged inflation when forming their inflation expectations. In terms of the amplitude of the response of forecasts to lagged inflation, market-based data are those that respond less strongly. The response of forecasts to lagged inflation for industry data is also relatively low. In other samples, the responses seem comparable, although experimental data tend to exhibit a larger response. The latter could be the consequence of the focal role of inflation data in the incentivized coordination game described by LTFEs. The results in terms of amplitude should be considered cautiously, however, since the sample frequencies can impede comparisons.

Regarding lagged output gap, it always significantly enters the formation of inflation expectations for experimental data, while significance is less clear for field data. That experimental data differ from all other types of data in that the forecasts of experimental subjects significantly rely on lagged output gap in addition to lagged inflation, might arguably come from an experimenter demand effect,[28] precisely due to the observability of the lagged output gap in the laboratory. Because laboratory experiments offer a stylized context and select a few pieces of specific information to be disclosed to participants, experimental subjects might be particularly prone to such demand effects. In the experiments, the data of which are exploited in the present paper, the instructions and screens that the participants observed describe the main macroeconomic variables of the stylized economy. Lagged output gap is one of them, on the same level as lagged inflation.

[26] Over a longer time period, including data from a less transparent period of time, economic agents (household and industry) might have been more backward than forward looking, which could impede them in using potentially available information and thus making more accurate forecasts. The same type of reasoning can be applied to experimental data. Indeed, in contrast to participants in Adam's and PZ's experiments, participants in the experiments by CMB1 and CMB2, in which some treatments are provided with the inflation target of the central bank, possibly induced more forward-looking behaviour.

[27] Results for the full field sample period are presented in Table G reported in the Appendix.

[28] "*Experimenter demand effects […] refer to changes in behavior by experimental subjects due to cues about what constitutes appropriate behavior (behavior "demanded" from them)*" (Zizzo, p.75, 2010).

Table 7 - Forecast determination

Experimental forecasts	(1) All obs.	(2) PZ	(3) Adam	(4) CMB1	(5) CMB2	(6) HMW
β_6	0.964***	0.977***	0.584***	0.690***	0.371***	1.116***
	[0.00]	[0.00]	[0.01]	[0.02]	[0.02]	[0.01]
$\beta_{7\ OG}$	-0.241***	-0.161***	-0.235***	-0.385***	-0.194***	-0.200***
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]	[0.01]
constant	0.134***	0.059***	1.668***	1.613***	3.088***	-0.402***
	[0.01]	[0.01]	[0.05]	[0.12]	[0.07]	[0.02]
N	38424	14903	510	4704	5664	12643
R^2	0.85	0.89	0.84	0.32	0.12	0.69

Market-based forecasts	Daily			Monthly average		
	(1) swap3y	(2) swap5y	(3) swap10y	(4) swap3y	(5) swap5y	(6) swap10y
β_6	0.178***	0.086***	0.054***	0.190***	0.096***	0.075***
	[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.02]
$\beta_{7\ OG}$	-0.307	-0.432**	-0.413**	-0.008	-0.041	-0.057**
	[0.21]	[0.21]	[0.18]	[0.03]	[0.03]	[0.02]
constant	1.962***	2.389***	2.675***	1.924***	2.365***	2.628***
	[0.01]	[0.01]	[0.01]	[0.05]	[0.05]	[0.04]
N	3267	3267	3267	151	151	151
R^2	0.35	0.11	0.06	0.37	0.12	0.11
β_6	0.177***	0.085***	0.052***	0.186***	0.082***	0.057***
	[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.02]
$\beta_{7\ GDP}$	1.019***	1.114***	1.430***	0.050	0.040	0.046
	[0.19]	[0.19]	[0.16]	[0.04]	[0.04]	[0.03]
constant	1.962***	2.389***	2.675***	1.929***	2.390***	2.662***
	[0.01]	[0.01]	[0.01]	[0.05]	[0.05]	[0.04]
N	3266	3266	3266	150	150	150
R^2	0.35	0.12	0.09	0.37	0.11	0.08

(continued)

As in the case of experimental data, note that, when significant, lagged output gaps negatively enter the formation of market-based and professional inflation forecasts. It suggests that the latter perceive higher unemployment as stimulating future inflation. The rationale could be that these economic agents expect the central bank to stimulate the economy in a context of higher unemployment: a positive output gap can be interpreted as a signal about the endogenous response of the central bank. Regarding experimental forecasts, recall that, when monetary policy was implemented, subjects were told that the central bank would intervene by moving interest rates to stabilize inflation.

In contrast, when significant, lagged output gap positively enters the formation of household, industry and central bank forecasts. In line with the usual tradeoff between inflation and output gap as depicted by the Phillips curve, these economic agents perceive higher unemployment as likely to lower future inflation.

Lagged output gap is usually not observed in the field (its estimation can, however, differ across the different categories of agents), which might explain the lack of significance in our results. Instead, lagged GDP growth can be observed in the field. Table 7 shows that GDP growth positively enters the determination of inflation forecasts of field data, and significance is more often obtained than when considering lagged output gap. The positivity of coefficient β_7 is in line with the standard interpretation of the Phillips curve. That significance is not always reached can be attributed to the forming of inflation expectations being a task that can interfere with other tasks in the field and that can possibly be influenced by various kinds of quantitative and qualitative information in a noisy and complex field environment. This feature stands in contrast to the stylized environment provided by laboratory experiments.

Table 7 continued - Forecast determination

Survey forecasts	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_6	0.161***	0.180***	0.250***	0.485***	0.445***	0.480***	0.467***
	[0.06]	[0.06]	[0.02]	[0.04]	[0.05]	[0.05]	[0.05]
$\beta_{7\ OG}$	0.001**	0.001*	0.054**	-0.001**	-0.001***	-0.001	-0.002***
	[0.00]	[0.00]	[0.03]	[0.00]	[0.00]	[0.00]	[0.00]
constant	2.055***	2.159***	2.384***	1.069***	1.299***	1.359***	1.562***
	[0.18]	[0.18]	[0.05]	[0.13]	[0.13]	[0.15]	[0.14]
N	62	62	373	124	124	124	124
R²	0.24	0.24	0.38	0.52	0.44	0.43	0.45
β_6	0.177***	0.187***	0.262***	0.439***	0.378***	0.422***	0.391***
	[0.06]	[0.06]	[0.02]	[0.04]	[0.04]	[0.05]	[0.05]
$\beta_{7\ GDP}$	0.119*	0.118*	-0.005	0.050	0.059*	0.110***	0.075**
	[0.06]	[0.06]	[0.01]	[0.03]	[0.03]	[0.04]	[0.04]
constant	1.718***	1.841***	2.363***	1.069***	1.340***	1.238***	1.582***
	[0.22]	[0.21]	[0.05]	[0.13]	[0.14]	[0.15]	[0.15]
N	62	62	373	124	124	124	124
R²	0.22	0.24	0.371	0.51	0.41	0.46	0.43

Central bank forecasts	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_6	0.669***	0.541***	0.214***	0.214***	0.112	0.233***	-0.008	0.061***
	[0.05]	[0.05]	[0.04]	[0.04]	[0.07]	[0.05]	[0.07]	[0.02]
$\beta_{7\ OG}$	0.000	0.000	0.611***	0.531**	0.264	0.373	0.351	0.514***
	[0.00]	[0.00]	[0.21]	[0.21]	[0.37]	[0.25]	[0.33]	[0.09]
constant	0.718***	1.115***	1.637***	1.589***	2.350***	1.867***	1.791***	1.350***
	[0.13]	[0.12]	[0.15]	[0.14]	[0.26]	[0.18]	[0.24]	[0.06]
N	124	124	104	104	104	104	52	52
R²	0.63	0.55	0.28	0.27	0.03	0.20	0.02	0.50
β_6	0.663***	0.524***	0.229***	0.219***	0.106	0.231***	0.000	0.074***
	[0.05]	[0.04]	[0.04]	[0.04]	[0.07]	[0.05]	[0.07]	[0.03]
$\beta_{7\ GDP}$	0.016	0.067**	-0.010	0.027	0.047	0.039	-0.005	-0.014
	[0.03]	[0.03]	[0.03]	[0.03]	[0.06]	[0.04]	[0.06]	[0.02]
constant	0.691***	0.987***	1.608***	1.495***	2.238***	1.762***	1.774***	1.331***
	[0.14]	[0.13]	[0.17]	[0.16]	[0.28]	[0.19]	[0.25]	[0.09]
N	124	124	104	104	104	104	52	52
R²	0.63	0.56	0.22	0.23	0.03	0.19	0.00	0.15

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (6) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

These results from field data are in line with the literature (Mavroeidis et al., 2014), which has usually found no significant effect of output gap (or output growth gap) on inflation forecasts but a positive and significant effect of GDP growth.[29] Mavroeidis et al. (2014) found a large dispersion of the

[29] Table N confirms the analysis when fixed effects are considered for experimental data. Table G in the Appendix confirms the analysis presented in Table 7 over a longer period of time for surveys and central banks' forecasts. It nevertheless shows that the output gap significantly enters the forecast determination for industry data and, to some extent (short-term horizon), for professional forecasters data. Regarding GDP growth, Table G shows that the analysis remains the same for

estimated parameters for economic slack in their meta-analysis of NK Phillips curves. Our result additionally complements the empirical literature by providing a systematic analysis across different categories of economic agents.[30]

Result 3: For all categories of agents, inflation forecasts depend on lagged inflation. Regarding the dependence of inflation forecasts on real activity variables, output gap affects experimental inflation forecasts, while GDP growth affects all categories of field expectations.

4. Conclusion

While it is crucial for experimental inflation forecasts to be valid outside the laboratory to be useful for policymakers and policy initiatives, the issue of the external validity of laboratory experiments in terms of inflation expectation formation has barely been studied. The contribution of the present paper is twofold. First, it provides an overview of different categories of agents (participants in experiments, households, industry, professional forecasters, financial market participants and central bankers) on inflation expectations according to three dimensions: forecast accuracy, deviations from full information rational expectations, and the determinants of inflation forecasts. Beyond offering a systematic comparison between different categories of economic agents, this large picture of inflation expectations allows for evaluating the external validity of experimental data.

Overall, our different sets exhibit some common features, but they also present some heterogeneity. Regarding common patterns (for which experimental data are not an exception):

- forecast errors are large and forecast accuracy is comparable;
- forecast errors are significantly autocorrelated; and
- lagged inflation significantly determines inflation expectations.

In contrast, among features of dissent, we can note the specific status of central bank's forecasts. The latter exhibit superiority since they are not systematically biased (contrasting with all other data sets), they are less autocorrelated, forecast errors are not predictable owing to forecast revision (in contrast to all other data except those of professional forecasters), and forecast revisions are not predictable owing to past forecast revisions.

Comparing experimental data to every other set of data excluding central bank forecasts, we observe that

- forecast errors exhibit the same kind of bias (except for industry forecasts); and
- lagged forecast revisions significantly predict forecast revisions.

the full sample. Table J in the Appendix provides estimations of Equation (6) considering the current output gap instead of the lagged output gap. Tables K and L in the Appendix provide estimations of Equation (6) when considering the current or lagged unemployment gap, instead of the lagged output gap for field data. Using these alternative activity variables (for which measurement error is supposedly lower), Tables K and L show that the coefficient β_{7UG} is negative in most of our samples.

[30] The literature has indeed mainly focused on single (or less numerous) data sets. Using the Michigan survey data on inflation expectations, Lanne et al. (2009) showed that households use past releases of actual inflation (rather than forward-looking forecasts) to form their inflation expectations. Dräger et al. (2016) focused on consumers and professional forecasters and analysed the extent to which their expectations are in line with the Fisher equation, the Phillips curve, and the Taylor rule. Fendel et al. (2010) showed that professional forecasters use the expectations-augmented Phillips curve model when they forecast macroeconomic variables.

We also observe that experimental data might be closer to survey (households, professional forecasters) data in some respects (autocorrelation of forecast errors and predictability of forecast revision) and closer to data extracted from financial markets in other respects (forecast errors in forecast revisions).

We thus conclude that there is as much heterogeneity among the different sets once central bank forecasts are excluded. There is thus no reason to oppose experimental data to field data since the latter do not form a more homogenous group when we exclude experimental data.

Finally, our results raise possible avenues for research that we now discuss. One of the diverging features between experimental forecasts and other types of forecasts can possibly be attributed to an experimenter demand effect. While such a feature does not have large consequences in terms of forecast accuracy or deviations from full information rational expectations compared to other data sets, a potential consequence is that the design of experiments intending to elicit forecasts—particularly LtFEs—might have to be revisited. In particular, the way in which information about macroeconomic variables is presented might have to be presented differently.

Going one step further, this study could be used by experimenters to design experiments that mimic real-world features. More precisely, depending on the type of expectation (households, firms, financial market participants, etc.) that they intend to capture, experimenters should insure that participants reach, on average, the forecast properties found in field data. This assurance would help to reproduce stylized facts in the laboratory as a precondition for simulating the impact of alternative policy measures in a cost-effective manner in the laboratory. While we do not expect usual undergraduate student participants in experiments to achieve the same forecasting performance as highly qualified professional central bankers, a way to mimic their performance in the laboratory could be to provide appropriate training to participants (for instance, Petersen (2014) showed that making forecast errors salient helps participants to reduce these errors), beyond providing the correct incentives.

References

Adam K. (2007). "Experimental Evidence on the Persistence of Output and Inflation", *Economic Journal*, 117, 603-635.

Andolfatto D., S. Hendry and K. Moran (2007). "Are Inflation Expectations Rational?", *Journal of Monetary Economics*, 55(2), 406-422.

Andrade, P. and H. Le Bihan (2013). "Inattentive Professional Forecasters", *Journal of Monetary Economics*, 60(8), 967-982.

Ang, A., G. Bekaert and M. Wei (2007). "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?", *Journal of Monetary Economics*, 54(4), 1163-1212.

Arifovic J. and L. Petersen (2017). "Stabilizing expectations at the zero lower bound: experimental evidence", *Journal of Economic Dynamics and Control*, 82, 21-43.

Arifovic J. and T. J. Sargent (2003). "Laboratory experiments with an expectational Phillips curve", *Evolution and procedures in central banking*, 23-55.

Armantier O., W. Bruine de Bruin, G. Topa, W. van der Klaauw and B. Zafar (2015). "Inflation Expectations and Behavior: Do Survey Respondents Act on Their Beliefs?", *International Economic Review*, 56, 505-536.

Armantier O., W. van der Klauw, G. Topa and B. Zafar (2017). "The Price is Right: Updating Inflation Expectations in a Randomized Price Information Experiment", forthcoming in *Review of Economics and Statistics*.

Assenza T., P. Heemeijer, C. Hommes and D. Massaro (2013). "Individual expectations and aggregate macro behavior", *CeNDEF Working Paper*, University of Amsterdam.

Bao T., M. Hennequin, C. Hommes and D. Massaro (2019). "Coordination on bubbles in large-group asset pricing experiments", *Journal of Economic Dynamics and Control*, forthcoming.

Bao T., J. Duffy and C. Hommes (2013). "Learning, forecasting and optimizing: An experimental study", *European Economic Review*, 61, 186-204.

Bertasiute A., D. Massaro and M. Weber (2018). "The behavioral economics of currency unions: Economic integration and monetary policy", *Bank of Lithuania Working Paper*.

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

Coibion O. and Y. Gorodnichenko (2012). "What Can Survey Forecasts Tell Us About Informational Rigidities?", *Journal of Political Economy*, 120, 116-159.

Coibion O. and Y. Gorodnichenko (2015). "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts", *American Economic Review*, 105(8), 2644-78.

Cole S. J. and F. Milani (2017). "The misspecification of expectations in New Keynesian models: A DSGE-VAR approach", *Macroeconomic Dynamics*, 1-34, doi:10.1017/S1365100517000104.

Cornand C. and F. Heinemann (2014). "Experiments on monetary policy and central banking", in “Experiments in Macroeconomics”, John Duffy (eds.), Vol. 17, *Research in Experimental Economics*, Emerald Group Publishing.

Cornand C. and F. Heinemann (2019). "Experiments on macroeconomics: methods and applications", forthcoming in the *Handbook of Research Methods and Applications in Experimental Economics*, A. Schram and A. Ule (eds).

Cornand C. and C.K. M'baye (2018a). "Does inflation targeting matter? An experimental investigation", *Macroeconomic Dynamics*, 22(2), 362-401.

Cornand C. and C.K. M'baye (2018b). "Band or Point Target? An Experimental Study", *Journal of Economic Interaction and Coordination*, 13(2): 283-309.

Croushore D. (1997). "The Livingstone Survey: Still useful after all these years", Federal Reserve Bank of Philadelphia, *Business Review* March/April, 15-27.

De Grauwe P. and C. Macchiarelli (2015). "Animal spirits and credit cycles", *Journal of Economic Dynamics and Control*, *59*, 95-117.

Diebold F. (1989). "Forecast combination and encompassing: reconciling two divergent literatures", *International Journal of Forecasting*, 5, 589-92.

Diebold F. and R. Mariano (1995). "Comparing Predictive Accuracy", *Journal of Economic and Business Statistics*, 13, 253-63.

Dräger L., M. Lamla and D. Pfajfar (2016). "Are Survey Expectations Theory-Consistent? The Role of Central Bank Communication and News", *European Economic Review*, 85, 84-111.

Duffy J. (2008). "Experimental Macroeconomics", in: S. Durlauf and L. Blume (Eds.), *New Palgrave Dictionary of Economics*, 2nd Ed., New York: Palgrave Macmillan.

Duffy J. (2016). "Macroeconomics: A Survey of Laboratory Research", in J.H. Kagel and A.E. Roth (Eds.), *Handbook of Experimental Economics*, 2, Princeton: Princeton University Press, 1-90.

Engle-Warnick J. and N. Turdaliev (2010). "An experimental test of Taylor-type rules with inexperienced central bankers. *Experimental Economics*, *13*(2), 146-166.

Fendel R., E. Lis and J.-C. Rülke (2011). "Do professional forecasters believe in the Phillips curve? Evidence from G7 countries", *Journal of Forecasting*, 30(2), 268-87.

Henrich J., S. J. Heine and A. Norenzayan (2010). "The weirdest people in the world?", *Behavioral and Brain Sciences*, 33, 61-83.

Henrich J., S. J. Heine and A. Norenzayan (2010). "Most people are not WEIRD", *Nature*, 466(7302), 29.

Hommes C.H. (2011). "The Heterogeneous Expectations Hypothesis: Some Evidence from the Lab", *Journal of Economic Dynamics and Control*, 35(1), 1-24.

Hommes C.H., A. Kopányi-Peuker and J. Sonnemans (2018). "Bubbles, crashes and information contagion in large-group asset market experiments", mimeo.

Hommes C.H., D. Massaro and M. Weber (2017). "Monetary Policy under Behavioral Expectations: Theory and Experiment", *Bank of Lithuania Working Paper series*, 42.

Huck S., H. Normann and J. Oechssler (2004). "Two are few and four are many: number effects in experimental oligopolies", *Journal of Economic Behavior and Organization*, 53, 435-446.

Kucinskas S. and F.S. Peters (2018). "Measuring Biases in Expectation Formation", mimeo.

Kryvtsov O. and L. Petersen (2013). "Expectations and monetary policy: experimental evidence", Bank of Canada Working Paper No. 2013-44.

Lambsdorff J. G., M. Schubert and M. Giamattei (2013). "On the role of heuristics—Experimental evidence on inflation dynamics", *Journal of Economic Dynamics and Control*, *37*(6), 1213-1229.

Lanne M., A. Luoma and J. Luoto (2009). "A naïve sticky information model of households’ inflation expectations", *Journal of Economic Dynamics and Control*, 33(6), 1332-44.

Luhan W. J. and J. Scharler (2014). "Inflation illusion and the Taylor principle: An experimental study", *Journal of Economic Dynamics and Control*, *45*, 94-110.

Mankiw G. and R. Reis (2002). "Sticky information versus sticky prices: A proposal to replace the new Keynesian Phillips curve", *Quarterly Journal of Economics*, 117(4), 1295-328.

Mankiw G., R. Reis, and J. Wolfers (2004). "Disagreement about Inflation Expectations", In *NBER Macroeconomics Annual 2003*, M. Gertler and K. Rogoff (eds.), 209–48, Cambridge, MA: MIT Press.

Mauersberger F. (2018). "Monetary policy rules in a non-rational world: A macroeconomic experiment", *University of Columbia Working Paper*.

Mavroeidis S., M. Plagborg-Moller and J. H. Stock (2014). "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve", *Journal of Economic* Literature, 52(1), 124-88.

Mehra Y. P. (2002). "Survey Measures of Expected Inflation: Revisiting the Issues of Predictive Content and Rationality". *Federal Reserve Bank of Richmond Economic Quarterly,* 88 (Summer), 17-36.

Noussair C. N., D. Pfajfar and J. Zsiros (2015). "Pricing decisions in an experimental dynamic stochastic general equilibrium economy", *Journal of Economic Behavior and Organization*, 109, 188-202.

Petersen L. (2014). "Forecast Error Information and Heterogenous Expectations in Learning to Forecast Macroeconomic Experiments", in “Experiments in Macroeconomics”, John Duffy (eds.), Vol. 17, *Research in Experimental Economics*, Emerald Group Publishing.

Petersen L. (2015). "Do expectations and decisions respond to monetary policy?", *Journal of Economic Studies*, 42(6), 972-1004.

Pfajfar D. and E. Santoro (2010). "Heterogeneity, learning and information stickiness in inflation expectations”, *Journal of Economic Behavior & Organization*, 75(3), 426-444.

Pfajfar D., and Žakelj B. (2018). "Inflation Expectations and Monetary Policy Design: Evidence from the Laboratory", *Macroeconomic Dynamics*, 22(4), 1035-75.

Roberts J.M. (1997). "Is Inflation Sticky?" Journal of *Monetary Economics*, 39 (January), 173-96.

Romer C. and D. Romer (2000). "Federal Reserve Information and the Behavior of Interest Rates", *American Economic Review*, 90, 429–57.

Sims C. (2003). "Implications of Rational Inattention", *Journal of Monetary Economics*, 50, 665-90.

Thomas L. B. (1999). "Survey measures of expected US inflation", *Journal of Economic Perspectives*, 13(4), 125-44.

Zizzo D. J. (2010), "Experimenter demand effects in economic experiments", *Experimental Economics*, 13(1), 75-98.

APPENDIX

Table A - Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Experimental forecasts					
All obs.	39 100	3.734	1.732	-13.9	52
PZ	15 120	3.057	2.204	-13.9	52
Adam	520	4.058	1.437	0.1	10.93
CMB1	4 800	5.296	0.999	3	16
CMB2	5 760	4.875	0.522	0	16
HMW	12 900	3.424	0.834	-5	12
Market-based forecasts					
Daily *(November 2004 - June 2017)*					
swap3y	3 297	2.332	0.443	0.285	4.65
swap5y	3 297	2.568	0.373	1.139	3.277
swap10y	3 297	2.789	0.315	1.773	3.46
Monthly *(November 2004 - June 2017)*					
swap3y	152	2.318	0.452	0.547	3.177
swap5y	152	2.564	0.382	1.266	3.229
swap10y	152	2.785	0.315	1.972	3.368
Survey forecasts					
Livingston *(June 1956 - December 2017)*					
living6m	124	3.081	2.103	0.06	10.67
living12m	124	3.24	2.054	0.25	10.27
Michigan *(January 1978 - December 2017)*					
mich1y	481	3.605	1.706	0.4	10.4
SPF *(October 1968 - December 2017)*					
pgdp3q	197	3.518	2.067	0.734	9.942
pgdp6q	192	3.49	1.803	1.457	8.69
cpi3q	146	2.918	1.22	0.604	7.928
cpi6q	146	3.14	1.213	1.847	7.926
Central bank forecasts					
FOMC *(July 1979 - December 2017)*					
fomc_cy	154	3.061	2.097	0.4	10.25
fomc_ny	154	3.14	1.903	1.05	9.5
Greenbook *(October 1969 - December 2012)*					
pgdpf1q	173	3.642	2.384	0.1	11.5
pgdpf4q	165	3.396	2.059	0.8	9.5
cpif1q	133	3.531	2.824	-3.2	15.1
cpif4q	133	3.234	1.893	0.9	9.8
pcef1q	52	1.763	1.179	-1.8	5.9
pcef4q	52	1.49	0.44	0.7	2.4

Note: Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. For experimental data: PZ indicates data from Pfajfar and Zakelj (2018), Adam those from Adam (2008), CMB1 those from Cornand and M'baye (2018), CMB2 those from Cornand and M'baye (2016) , and HMW those from Hommes, Massaro and Weber (2017).

Table B - Forecast errors - Full samples

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_1	0.616***	0.450**	-0.024	0.018	-0.056	-0.073	-0.424***
	[0.18]	[0.20]	[0.07]	[0.09]	[0.13]	[0.08]	[0.10]
N	123	122	480	195	187	145	142

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_1	-0.115*	-0.391***	0.092	0.164	-0.080	-0.04	0.286	0.558**
	[0.07]	[0.10]	[0.09]	[0.12]	[0.15]	[0.18]	[0.19]	[0.26]
N	154	150	172	161	132	129	51	48

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (1) with OLS for the full sample period for survey and central bank forecasts. Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table C - Absolute forecast errors - Full samples

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_2	1.468***	1.543***	1.039***	0.974***	1.312***	0.692***	0.976***
	[0.14]	[0.15]	[0.05]	[0.06]	[0.09]	[0.05]	[0.07]
N	123	122	480	195	187	145	142

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_2	0.641***	0.949***	0.924***	1.137***	1.190***	1.530***	1.042***	1.421***
	[0.04]	[0.07]	[0.06]	[0.09]	[0.11]	[0.13]	[0.13]	[0.18]
N	154	150	172	161	132	129	51	48

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (2) with OLS for the full sample period for survey and central bank forecasts. Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table D - Autocorrelation of forecast errors - Full samples

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_3	0.311***	0.496***	0.921***	0.533***	0.779***	0.596***	0.769***
	[0.09]	[0.08]	[0.02]	[0.06]	[0.05]	[0.07]	[0.05]
constant	0.403**	0.208	-0.005	0.003	-0.028	-0.042	-0.086
	[0.18]	[0.18]	[0.03]	[0.08]	[0.08]	[0.06]	[0.07]
N	122	121	479	194	183	144	141
R^2	0.10	0.25	0.85	0.29	0.61	0.37	0.60

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_3	0.689***	0.845***	0.437***	0.783***	0.157*	0.318***	0.223	0.321**
	[0.06]	[0.04]	[0.07]	[0.05]	[0.09]	[0.08]	[0.14]	[0.14]
constant	-0.026	-0.064	0.042	0.023	-0.09	-0.045	0.259	0.372
	[0.05]	[0.06]	[0.08]	[0.08]	[0.15]	[0.18]	[0.19]	[0.27]
N	153	149	171	153	131	128	50	47
R^2	0.49	0.72	0.19	0.58	0.03	0.10	0.05	0.10

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (3) with OLS for the full sample period for survey and central bank forecasts. Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table E - Forecast errors on forecast revisions - Full samples

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_4	0.577**	1.232***	-0.498***	0.275	0.555*	-0.259	0.009
	[0.26]	[0.35]	[0.15]	[0.18]	[0.31]	[0.16]	[0.34]
constant	0.584***	0.411**	-0.028	0.014	-0.089	-0.095	-0.413***
	[0.18]	[0.19]	[0.06]	[0.09]	[0.13]	[0.07]	[0.10]
N	122	121	479	194	183	144	141
R^2	0.04	0.09	0.02	0.01	0.02	0.02	0.00

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_4	0.029	0.425	-0.043	0.525**	0.153*	-0.008	0.193	-0.549
	[0.17]	[0.29]	[0.10]	[0.22]	[0.09]	[0.49]	[0.13]	[1.10]
constant	-0.103	-0.374***	0.081	0.085	-0.09	-0.062	0.329*	0.544**
	[0.07]	[0.10]	[0.09]	[0.12]	[0.15]	[0.19]	[0.19]	[0.27]
N	153	149	171	153	131	128	50	47
R^2	0.00	0.01	0.00	0.04	0.02	0.00	0.04	0.01

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (4) with OLS for the full sample period for survey and central bank forecasts. Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For surveys, the forecasting horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table F - Forecast revisions on lagged forecast revisions - Full samples

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_5	0.028	0.113	-0.236***	-0.104	-0.253***	-0.194**	-0.11
	[0.09]	[0.09]	[0.04]	[0.07]	[0.07]	[0.08]	[0.08]
constant	0.005	0.008	-0.009	-0.004	-0.011	-0.046	-0.044*
	[0.06]	[0.05]	[0.02]	[0.04]	[0.03]	[0.04]	[0.02]
N	122	122	479	195	186	144	144
R^2	0.00	0.01	0.06	0.01	0.07	0.04	0.01

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_5	-0.018	-0.035	-0.354***	-0.063	-0.287***	0.055	-0.218	0.122
	[0.08]	[0.08]	[0.07]	[0.08]	[0.08]	[0.09]	[0.14]	[0.14]
constant	-0.056*	-0.051*	-0.02	-0.028	-0.126	-0.049	-0.045	-0.013
	[0.03]	[0.03]	[0.07]	[0.04]	[0.14]	[0.03]	[0.20]	[0.03]
N	152	152	171	154	131	131	50	50
R^2	0.00	0.00	0.13	0.00	0.08	0.00	0.05	0.02

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (5) with OLS for the full sample period for survey and central bank forecasts. Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For surveys, the forecasting horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table G - Forecast determination - Full samples

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_6	0.576***	0.563***	0.545***	0.649***	0.534***	0.562***	0.557***
	[0.04]	[0.04]	[0.01]	[0.02]	[0.02]	[0.04]	[0.04]
$\beta_{7\ OG}$	0.000	0.000	-0.02	0.000	-0.001*	-0.001*	-0.001***
	[0.00]	[0.00]	[0.04]	[0.00]	[0.00]	[0.00]	[0.00]
constant	0.973***	1.180***	1.643***	0.848***	1.324***	1.249***	1.482***
	[0.17]	[0.17]	[0.05]	[0.10]	[0.11]	[0.12]	[0.12]
N	123	123	480	196	191	146	146
R^2	0.69	0.68	0.82	0.85	0.75	0.63	0.62
β_6	0.576***	0.558***	0.545***	0.657***	0.542***	0.539***	0.527***
	[0.04]	[0.03]	[0.01]	[0.02]	[0.02]	[0.03]	[0.04]
$\beta_{7\ GDP}$	0.029	0.033	-0.01	0.129***	0.161***	0.092***	0.096***
	[0.05]	[0.05]	[0.01]	[0.03]	[0.03]	[0.03]	[0.03]
constant	0.888***	1.099***	1.662***	0.458***	0.850***	1.077***	1.322***
	[0.22]	[0.22]	[0.06]	[0.12]	[0.14]	[0.14]	[0.14]
N	123	123	480	196	191	146	146
R^2	0.69	0.68	0.821	0.86	0.78	0.64	0.63

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_6	0.950***	0.833***	0.585***	0.481***	0.620***	0.449***	-0.008	0.061***
	[0.03]	[0.03]	[0.03]	[0.03]	[0.06]	[0.04]	[0.07]	[0.02]
$\beta_{7\ OG}$	0.000	0.000	0.36	0.211	-0.222	0.181	0.351	0.514***
	[0.00]	[0.00]	[0.26]	[0.25]	[0.47]	[0.29]	[0.33]	[0.09]
constant	0.206*	0.636***	1.071***	1.305***	1.282***	1.607***	1.791***	1.350***
	[0.11]	[0.12]	[0.17]	[0.16]	[0.28]	[0.17]	[0.24]	[0.06]
N	153	153	172	164	133	133	52	52
R^2	0.86	0.80	0.68	0.63	0.46	0.54	0.02	0.50
β_6	0.957***	0.851***	0.582***	0.483***	0.620***	0.450***	0.000	0.074***
	[0.03]	[0.03]	[0.03]	[0.03]	[0.06]	[0.04]	[0.07]	[0.03]
$\beta_{7\ GDP}$	0.044	0.099***	-0.013	0.056**	-0.04	0.019	-0.005	-0.014
	[0.03]	[0.03]	[0.03]	[0.03]	[0.06]	[0.04]	[0.06]	[0.02]
constant	0.072	0.327**	1.115***	1.134***	1.391***	1.553***	1.774***	1.331***
	[0.15]	[0.16]	[0.19]	[0.18]	[0.32]	[0.20]	[0.25]	[0.09]
N	153	153	172	164	133	133	52	52
R^2	0.86	0.81	0.68	0.64	0.46	0.54	0.00	0.15

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (6) with OLS for the full sample period for survey and central bank forecasts. Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For surveys, the forecasting horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table H - Forecast errors on current and past inflation forecasts

Experimental forecasts

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_{H1}	-0.062***	0.098***	-0.624***	-0.692***	-0.642***	-0.473***
	[0.00]	[0.01]	[0.06]	[0.01]	[0.01]	[0.01]
β_{H2}	-0.198***	-0.371***	-1.237***	0.077***	0.087***	-0.182***
	[0.00]	[0.01]	[0.06]	[0.01]	[0.01]	[0.01]
constant	0.964***	0.796***	7.537***	3.148***	2.709***	2.307***
	[0.01]	[0.02]	[0.31]	[0.04]	[0.05]	[0.02]
N	38423	14903	510	4704	5664	12642
	0.2	0.273	0.581	0.728	0.481	0.541

Market-based forecasts

	Daily			Monthly average		
	(1)	(2)	(3)	(4)	(5)	(6)
	swap3y	swap5y	swap10y	swap3y	swap5y	swap10y
β_{H1}	-1.133***	-0.757***	-0.332	-1.386***	-0.576	0.235
	[0.18]	[0.24]	[0.68]	[0.45]	[0.48]	[1.36]
β_{H2}	-0.046	0.149	-0.018	0.107	0.134	-0.526
	[0.18]	[0.24]	[0.68]	[0.45]	[0.48]	[1.27]
constant	2.165***	0.598*	-0.901	2.388**	0.146	-1.067
	[0.20]	[0.31]	[0.60]	[0.94]	[1.38]	[3.21]
N	2513	1991	686	115	91	31
	0.082	0.017	0.004	0.117	0.017	0.007

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_{H1}	-0.771**	-0.795	0.020	-0.194	-0.215	-0.190	-0.632
	[0.38]	[0.51]	[0.18]	[0.23]	[0.39]	[0.20]	[0.51]
β_{H2}	0.515	0.711	0.541***	-0.114	-0.214	0.272	0.561
	[0.38]	[0.50]	[0.18]	[0.23]	[0.39]	[0.20]	[0.50]
constant	0.745	0.144	-2.113***	0.532**	0.716***	-0.208	-0.023
	[0.69]	[0.70]	[0.30]	[0.22]	[0.26]	[0.25]	[0.33]
N	61	60	372	122	119	123	120
	0.067	0.042	0.086	0.094	0.142	0.019	0.015

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_{H1}	-0.212	0.223	-0.443***	-0.165	0.012	-0.811	-0.030	-1.103
	[0.20]	[0.39]	[0.12]	[0.24]	[0.11]	[0.71]	[0.16]	[1.11]
β_{H2}	0.151	-0.368	-0.023	-0.430*	-0.278**	0.434	-0.404**	-0.004
	[0.21]	[0.39]	[0.12]	[0.24]	[0.11]	[0.71]	[0.16]	[1.11]
constant	0.216	0.304	1.083***	1.366***	0.920***	1.227**	1.092***	2.200**
	[0.17]	[0.26]	[0.23]	[0.23]	[0.35]	[0.49]	[0.40]	[0.93]
N	124	120	103	100	103	100	50	47
	0.013	0.022	0.222	0.287	0.072	0.055	0.13	0.078

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating a modified version equation (4) with OLS. Forecast revisions can be decomposed between current and past inflation forecasts. We replace forecast revisions by these two variables: β_{H1} is associated to current inflation forecasts and β_{H2} is associated to current inflation forecasts. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample.

Table I - Dispersion and tails of forecast errors

Experimental forecasts

	(1) All obs.	(2) PZ	(3) Adam	(4) CMB1	(5) CMB2	(6) HMW
Variance	1.09	1.00	8.21	0.05	0.06	0.27
Kurtosis	40.81	24.46	7.22	5.21	5.02	7.97

Market-based forecasts

	Daily			Monthly average		
	(1) swap3y	(2) swap5y	(3) swap10y	(4) swap3y	(5) swap5y	(6) swap10y
Variance	2.31	0.89	0.70	2.38	0.88	0.71
Kurtosis	3.28	2.84	2.39	3.29	2.80	2.39

Survey forecasts

	Livingston		Michigan	SPF			
	(1) living6m	(2) living12m	(3) mich1y	(4) pgdp1q	(5) pgdp4q	(6) cpi1q	(7) cpi4q
Variance	4.12	4.99	2.06	1.64	3.35	0.83	1.45
Kurtosis	6.55	6.11	5.09	4.25	7.37	4.15	3.60

Central bank forecasts

	FOMC		Greenbook					
	(1) fomc_cy	(2) fomc_ny	(3) pgdp1q	(4) pgdp4q	(5) cpi1q	(6) cpi4q	(7) pce1q	(8) pce4q
Variance	0.66	1.43	1.39	2.47	2.99	4.40	1.91	3.26
Kurtosis	3.03	4.03	3.72	6.67	7.24	7.35	10.72	9.38

Note: Variance and kurtosis are computed at the group level for experimental forecasts for comparability with other forecast data. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table J - Forecast determination - Current output gap

Experimental forecasts

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_6	0.968***	1.005***	0.363***	0.644***	0.328***	0.995***
	[0.00]	[0.00]	[0.01]	[0.03]	[0.02]	[0.01]
$\beta_{7\ OG}$	-0.129***	-0.021	-0.234***	-0.322***	-0.171***	-0.281***
	[0.01]	[0.02]	[0.01]	[0.03]	[0.02]	[0.01]
constant	0.124***	0.015	2.562***	1.872***	3.293***	0.04
	[0.01]	[0.01]	[0.05]	[0.15]	[0.07]	[0.03]
N	39100	15120	520	4800	5760	12900
R^2	0.85	0.89	0.84	0.29	0.11	0.69

Market-based forecasts

	Daily			Monthly average		
	(1)	(2)	(3)	(4)	(5)	(6)
	swap3y	swap5y	swap10y	swap3y	swap5y	swap10y
β_6	0.178***	0.086***	0.054***	0.183***	0.090***	0.072***
	[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.02]
$\beta_{7\ OG}$	-0.386*	-0.327	-0.388**	0.019	-0.02	-0.047**
	[0.21]	[0.21]	[0.18]	[0.03]	[0.03]	[0.02]
constant	1.962***	2.389***	2.675***	1.939***	2.378***	2.635***
	[0.01]	[0.01]	[0.01]	[0.05]	[0.05]	[0.04]
N	3267	3267	3267	151	151	151
R^2	0.35	0.11	0.06	0.37	0.11	0.10

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_6	0.164***	0.181***	0.250***	0.478***	0.432***	0.468***	0.454***
	[0.06]	[0.06]	[0.02]	[0.04]	[0.05]	[0.05]	[0.05]
$\beta_{7\ OG}$	0.001**	0.001*	0.057**	-0.001*	-0.001***	0.000	-0.001***
	[0.00]	[0.00]	[0.03]	[0.00]	[0.00]	[0.00]	[0.00]
constant	2.041***	2.150***	2.382***	1.089***	1.338***	1.394***	1.598***
	[0.18]	[0.18]	[0.05]	[0.13]	[0.13]	[0.15]	[0.14]
N	62	62	373	124	124	124	124
R^2	0.24	0.24	0.38	0.52	0.43	0.42	0.44

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_6	0.675***	0.547***	0.202***	0.205***	0.107	0.229***	-0.016	0.049***
	[0.05]	[0.05]	[0.04]	[0.04]	[0.07]	[0.05]	[0.07]	[0.02]
$\beta_{7\ OG}$	0.000	0.000	0.637***	0.516**	0.269	0.309	0.382	0.519***
	[0.00]	[0.00]	[0.21]	[0.21]	[0.36]	[0.25]	[0.32]	[0.08]
constant	0.702***	1.100***	1.666***	1.610***	2.361***	1.875***	1.805***	1.367***
	[0.13]	[0.12]	[0.15]	[0.15]	[0.26]	[0.18]	[0.24]	[0.06]
N	124	124	104	104	104	104	52	52
R^2	0.63	0.55	0.28	0.27	0.03	0.20	0.03	0.52

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (6) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table K - Forecast determination - Lagged unemployment gap

Market-based forecasts

	Daily			Monthly average		
	(1)	(2)	(3)	(4)	(5)	(6)
	swap3y	swap5y	swap10y	swap3y	swap5y	swap10y
β_6	0.178***	0.086***	0.053***	0.207***	0.101***	0.069***
	[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.02]
$\beta_{7\ UG}$	-0.659	-0.544	-0.501	0.536***	0.488***	0.293**
	[0.96]	[0.95]	[0.82]	[0.16]	[0.16]	[0.14]
constant	1.962***	2.389***	2.675***	1.888***	2.353***	2.639***
	[0.01]	[0.01]	[0.01]	[0.05]	[0.05]	[0.04]
N	3267	3267	3267	151	151	151
R^2	0.35	0.11	0.06	0.42	0.16	0.10

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_6	0.176***	0.191***	0.251***	0.474***	0.433***	0.476***	0.454***
	[0.06]	[0.06]	[0.02]	[0.04]	[0.05]	[0.05]	[0.05]
$\beta_{7\ UG}$	-0.214**	-0.175*	-0.359**	0.143	0.270**	0.165	0.289***
	[0.10]	[0.10]	[0.16]	[0.10]	[0.11]	[0.12]	[0.11]
constant	2.011***	2.125***	2.380***	1.104***	1.339***	1.373***	1.604***
	[0.18]	[0.18]	[0.05]	[0.13]	[0.13]	[0.15]	[0.14]
N	62	62	373	124	124	124	124
R^2	0.23	0.24	0.38	0.51	0.43	0.43	0.44

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_6	0.660***	0.535***	0.212***	0.217***	0.105	0.237***	-0.026	0.047**
	[0.05]	[0.04]	[0.04]	[0.04]	[0.07]	[0.05]	[0.07]	[0.02]
$\beta_{7\ UG}$	-0.100	-0.030	-0.248*	-0.132	-0.217	-0.062	-0.321	-0.296***
	[0.09]	[0.09]	[0.13]	[0.13]	[0.23]	[0.16]	[0.20]	[0.06]
constant	0.741***	1.131***	1.639***	1.574***	2.373***	1.850***	1.822***	1.367***
	[0.12]	[0.12]	[0.15]	[0.15]	[0.26]	[0.18]	[0.24]	[0.07]
N	124	124	104	104	104	104	52	52
R^2	0.63	0.55	0.24	0.23	0.03	0.19	0.05	0.44

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (6) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table L - Forecast determination - Current unemployment gap

Market-based forecasts

	Daily			Monthly average		
	(1)	(2)	(3)	(4)	(5)	(6)
	swap3y	swap5y	swap10y	swap3y	swap5y	swap10y
β_6	0.178***	0.086***	0.053***	0.205***	0.104***	0.075***
	[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.02]
$\beta_{7\ UG}$	0.305	0.188	0.142	0.404**	0.471***	0.370***
	[0.96]	[0.95]	[0.82]	[0.16]	[0.17]	[0.14]
constant	1.962***	2.389***	2.675***	1.893***	2.348***	2.629***
	[0.01]	[0.01]	[0.01]	[0.05]	[0.05]	[0.04]
N	3267	3267	3267	151	151	151
R^2	0.35	0.11	0.06	0.40	0.15	0.12

Survey forecasts

	Livingston		Michigan	SPF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	living6m	living12m	mich1y	pgdp1q	pgdp4q	cpi1q	cpi4q
β_6	0.172***	0.187***	0.249***	0.474***	0.435***	0.471***	0.457***
	[0.06]	[0.06]	[0.02]	[0.04]	[0.05]	[0.05]	[0.05]
$\beta_{7\ UG}$	-0.226**	-0.188*	-0.465***	0.139	0.272**	0.121	0.299***
	[0.10]	[0.10]	[0.16]	[0.10]	[0.11]	[0.12]	[0.11]
constant	2.019***	2.131***	2.387***	1.103***	1.335***	1.389***	1.596***
	[0.18]	[0.18]	[0.05]	[0.13]	[0.13]	[0.15]	[0.14]
N	62	62	373	124	124	124	124
R^2	0.24	0.24	0.39	0.51	0.43	0.42	0.44

Central bank forecasts

	FOMC		Greenbook					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	fomc_cy	fomc_ny	pgdp1q	pgdp4q	cpi1q	cpi4q	pce1q	pce4q
β_6	0.662***	0.536***	0.210***	0.216***	0.097	0.241***	-0.05	0.033
	[0.05]	[0.05]	[0.04]	[0.04]	[0.07]	[0.05]	[0.07]	[0.02]
$\beta_{7\ UG}$	-0.055	-0.001	-0.203	-0.106	-0.256	-0.003	-0.425**	-0.321***
	[0.10]	[0.09]	[0.14]	[0.13]	[0.23]	[0.16]	[0.20]	[0.06]
constant	0.736***	1.127***	1.641***	1.574***	2.394***	1.838***	1.876***	1.399***
	[0.13]	[0.12]	[0.15]	[0.15]	[0.26]	[0.18]	[0.24]	[0.07]
N	124	124	104	104	104	104	52	52
R^2	0.63	0.55	0.23	0.23	0.04	0.18	0.08	0.48

Note: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equation (6) with OLS. Market-based forecasts are considered at a daily or monthly frequency, Livingston has a semiannual frequency, Michigan monthly, SPF quarterly. FOMC and Greenbook are taken at a quarterly frequency. N is the size of each sample. For market-based forecasts, the forecasting horizon is 3, 5 and 10 years. For surveys, the horizon for Livingston is 6 or 12 months, for Michigan 1-year, and for SPF 1-quarter and 4-quarter. For central bank forecasts, the horizon is the current and next calendar years for FOMC and 1-quarter and 4-quarter for Greenbook.

Table M - Experimental data - By group

	(1)	(2)	(3)	(4)	(5)	(6)
Forecast errors - Equation (1)						
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_1	-0.005	-0.042*	-0.022	-0.116***	0.007	0.064***
	[0.01]	[0.02]	[0.13]	[0.01]	[0.01]	[0.01]
N	6001	1656	510	784	944	2107
Absolute forecast errors - Equation (2)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_2	0.542***	0.573***	1.813***	0.329***	0.258***	0.418***
	[0.01]	[0.02]	[0.10]	[0.01]	[0.01]	[0.01]
N	6001	1656	510	784	944	2107
Autocorrelation of forecast errors - Equation (3)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_3	0.341***	0.850***	0.115**	0.626***	-0.019	0.553***
	[0.01]	[0.01]	[0.04]	[0.03]	[0.03]	[0.02]
constant	-0.007	-0.001	-0.022	-0.040***	0.009	0.014
	[0.01]	[0.01]	[0.13]	[0.01]	[0.01]	[0.01]
N	5892	1632	500	768	928	2064
R^2	0.116	0.696	0.013	0.419	0	0.319
Forecast errors on forecast revisions - Equation (4)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_4	0.472***	0.848***	0.314***	-0.051***	-0.034	0.273***
	[0.02]	[0.02]	[0.07]	[0.02]	[0.03]	[0.02]
constant	-0.003	-0.034**	-0.017	-0.116***	0.007	0.064***
	[0.01]	[0.02]	[0.12]	[0.01]	[0.01]	[0.01]
N	6001	1656	510	784	944	2107
R^2	0.127	0.574	0.039	0.012	0.002	0.063
Forecast revisions on lagged forecast revisions - Equation (5)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_5	0.221***	0.777***	-0.156***	-0.440***	-0.611***	0.418***
	[0.01]	[0.02]	[0.04]	[0.03]	[0.03]	[0.02]
constant	0.001	0.002	-0.011	-0.019	0.002	0.019**
	[0.01]	[0.01]	[0.08]	[0.01]	[0.01]	[0.01]
N	5893	1632	500	768	928	2065
R^2	0.051	0.619	0.024	0.272	0.383	0.196
Forecast determination - Equation (6)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_6	0.841***	0.948***	0.157***	0.704***	0.625***	1.117***
	[0.01]	[0.02]	[0.03]	[0.04]	[0.02]	[0.02]
$\beta_{7\ OG}$	0.229***	0.009	0.103***	-0.244***	-0.047***	0.279***
	[0.01]	[0.05]	[0.03]	[0.03]	[0.02]	[0.02]
constant	0.622***	0.188***	3.435***	1.580***	1.843***	-0.533***
	[0.03]	[0.04]	[0.13]	[0.18]	[0.08]	[0.06]
N	6002	1656	510	784	944	2108
R^2	0.692	0.86	0.06	0.615	0.621	0.74

Note: Standard errors in brackets. * p < 0.10, ** p < 0.05, *** p < 0.01. Parameters are obtained by estimating equations (1) to (6) with OLS. The individual data are averaged for each group in each experiment. N is the size of each sample.

Table N - Experimental data - Including fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Autocorrelation of forecast errors - Equation (3)						
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_3	0.523***	0.677***	0.106**	0.138***	0.189***	0.449***
	[0.00]	[0.01]	[0.05]	[0.01]	[0.01]	[0.01]
constant	0.191	0.016	-0.098	0.040	-0.247***	-0.016
	[0.13]	[0.11]	[0.45]	[0.09]	[0.06]	[0.08]
N	37748	14688	500	4608	5568	12384
R^2	0.3	0.449	0.021	0.293	0.091	0.242
Forecast errors on forecast revisions - Equation (4)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_4	0.065***	0.231***	0.314***	-0.384***	-0.365***	-0.141***
	[0.01]	[0.01]	[0.07]	[0.01]	[0.01]	[0.01]
constant	0.337**	0.120	0.339	-0.004	0.091*	0.017
	[0.15]	[0.14]	[0.43]	[0.08]	[0.05]	[0.10]
N	38423	14903	510	4704	5664	12642
R^2	0.042	0.062	0.048	0.511	0.232	0.045
Forecast revisions on lagged forecast revisions - Equation (5)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_5	0.077***	0.258***	-0.156***	-0.471***	-0.441***	0.051***
	[0.01]	[0.01]	[0.04]	[0.01]	[0.01]	[0.01]
constant	-0.062	0.006	-0.027	0.017	-0.01	0.101
	[0.14]	[0.14]	[0.28]	[0.12]	[0.06]	[0.10]
N	37748	14687	500	4608	5568	12385
R^2	0.008	0.069	0.024	0.25	0.209	0.005
Forecast determination - Equation (6)						
	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_6	0.928***	0.930***	0.573***	0.643***	0.110***	1.133***
	[0.00]	[0.01]	[0.01]	[0.02]	[0.02]	[0.01]
$\beta_{7\ OG}$	-0.312***	-0.389***	-0.214***	-0.394***	-0.011	-0.207***
	[0.01]	[0.02]	[0.01]	[0.02]	[0.02]	[0.01]
constant	0.355***	0.240***	1.438***	1.834***	3.964***	-0.358***
	[0.10]	[0.09]	[0.10]	[0.16]	[0.09]	[0.07]
N	38424	14903	510	4704	5664	12643
R^2	0.866	0.893	0.851	0.467	0.259	0.698

Note: Standard errors in brackets. * p < 0.10, ** p < 0.05, *** p < 0.01. Parameters are obtained by estimating equations (3) to (6) with OLS and fixed effects. N is the size of each sample.

Table O - Experimental data - Clustered Standard Errors

Forecast errors - Equation (1)

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_1	-0.009	-0.042***	-0.022	-0.116***	0.007	0.064***
	[0.01]	[0.01]	[0.09]	[0.04]	[0.01]	[0.01]
N	38424	14904	510	4704	5664	12642

Absolute forecast errors - Equation (2)

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_2	0.462***	0.573***	1.813***	0.329***	0.258***	0.418***
	[0.03]	[0.04]	[0.47]	[0.04]	[0.02]	[0.02]
N	38424	14904	510	4704	5664	12642

Autocorrelation of forecast errors - Equation (3)

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_3	0.540***	0.680***	0.115*	0.370***	0.229***	0.461***
	[0.06]	[0.07]	[0.06]	[0.09]	[0.05]	[0.02]
constant	-0.007**	-0.010*	-0.022	-0.070***	0.007	0.020***
	[0.00]	[0.01]	[0.08]	[0.02]	[0.01]	[0.00]
N	37748	14688	500	4608	5568	12384
R^2	0.29	0.45	0.01	0.15	0.06	0.23

Forecast errors on forecast revisions - Equation (4)

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_4	0.066	0.231*	0.314***	-0.384***	-0.365***	-0.141***
	[0.07]	[0.14]	[0.08]	[0.02]	[0.02]	[0.02]
constant	-0.008	-0.040***	-0.017	-0.116***	0.007	0.064***
	[0.01]	[0.01]	[0.09]	[0.04]	[0.01]	[0.01]
N	38423	14903	510	4704	5664	12642
R^2	0.00	0.05	0.04	0.24	0.18	0.02

Forecast revisions on lagged forecast revisions - Equation (5)

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_5	0.078	0.258*	-0.156***	-0.471***	-0.441***	0.051*
	[0.08]	[0.15]	[0.05]	[0.02]	[0.04]	[0.03]
constant	-0.001	-0.008**	-0.011*	-0.019***	0.001	0.019***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
N	37748	14687	500	4608	5568	12385
R^2	0.01	0.07	0.02	0.25	0.21	0.00

Forecast determination - Equation (6)

	(1)	(2)	(3)	(4)	(5)	(6)
	All obs.	PZ	Adam	CMB1	CMB2	HMW
β_6	0.964***	0.977***	0.584***	0.690***	0.371***	1.116***
	[0.02]	[0.03]	[0.01]	[0.07]	[0.06]	[0.03]
$\beta_{7\ OG}$	-0.241***	-0.161**	-0.235***	-0.385**	-0.194***	-0.200***
	[0.05]	[0.08]	[0.02]	[0.15]	[0.03]	[0.02]
constant	0.134*	0.059	1.668***	1.613***	3.088***	-0.402***
	[0.07]	[0.07]	[0.07]	[0.35]	[0.29]	[0.09]
N	38424	14903	510	4704	5664	12643
R^2	0.85	0.89	0.84	0.32	0.12	0.69

Note: Clustered standard errors in brackets, at the individual level for columns 1-2 and 4-6 and at the group level for column 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Parameters are obtained by estimating equations (1) to (6) with OLS. The individual data are average for each group in each experiment. N is the size of each sample.