Forecast Error Information and Heterogeneous Expectations in Learning-to-Forecast Experiments*

Luba Petersen
Simon Fraser University

Current Version: July 2014

Abstract

This paper explores the importance of accessible and focal information in influencing beliefs and attention in a learning-to-forecast laboratory experiment where subjects are incentivized to form accurate expectations about inflation and the output gap. We consider the effects of salient and accessible forecast error information and learning on subjects' forecasting accuracy and heuristics, and on aggregate stability. Experimental evidence indicates that, while there is considerable heterogeneity in heuristics used, subjects' forecasts can be best described by a constant-gain learning model where subjects respond to forecast errors. Salient forecast error information reduces subjects' overreaction to their errors and leads to greater forecast accuracy, coordination of expectations and macroeconomic stability. The benefits of this focal information are short-lived and diminish with learning.

JEL classifications: C92, E2, E52, D50, D91
Keywords: experimental macroeconomics, laboratory experiment, monetary policy, expectations, learning to forecast, availability heuristic, focal points, communication, rational inattention

*Corresponding author: Luba Petersen: lubap@sfu.ca. This paper has benefited greatly from the helpful comments and suggestions of Jasmina Arifovic, David Freeman, Oleksiy Kryvtsov, and an anonymous referee.

1 Introduction

Expectations are an important driver of economic activity. What households believe about the future state of the economy will shape their decisions on how much to consume, work, and invest. Firms' pricing decisions depend significantly on expectations of future demand and aggregate price levels. An understanding of how expectations are formed and evolve is key to managing expectations and promoting economic stability.

It can be challenging to identify the effects of information, policy, and disturbances on expectations and the overall economy using traditional empirical approaches without making assumptions about the underlying structure of expectation formation and the aggregate data generating process. As a result, laboratory experiment have become an increasing popular source of data on expectation formation. In a highly controlled, incentivized environment where the data-generating process of the economy is established by the experimenter, one can more cleanly identify how individuals form beliefs in response to different policies, shocks, or information. Numerous experimental macroeconomic papers have now explored the effects monetary policy rules, communication, and the structure of the economy have on expectation formation and aggregate outcomes.

As learning-to-forecast experiments become a more policy-relevant source of data, further research on how certain design decisions influence behaviour is warranted. For example, how we place information on subjects' screens may bias behaviour and create potentially unintended focal points. Placing historical information right next to where subjects submit their forecast has the potential to generate adaptive or trend-chasing behaviour, especially if the information is presented graphically. Locating that information elsewhere while making the current period shocks more salient may generate increased sensitivity to those shocks and a reduction in adaptive behaviour.

This paper seeks to begin the discussion on focal information in expectation-driven environments. We construct a macroeconomic environment where subjects' aggregated forecasts about future output and inflation influence the current state of the economy. We conduct an experiment to understand how focal forecast error information influences forecasting heuristics and accuracy. Improving the salience and availability of forecast error information may encourage subjects to utilize the information when forming their expectations, improve overall coordination of heuristics, and lead to greater economic stability. We also explore how forecasting strategies, accuracy, and economic stability change with learning. This paper addresses not only an important question for the design of experiments, but also for the design of central bank communication. We are among the first to investigate experimentally the role of information in the heterogeneity of expectations. Our laboratory experiment provides evidence of how the negative consequences of rational inattention can be ameliorated through public announcements.

Our main finding is that increasing the salience of forecast errors encourages inexperienced subjects to correct their forecasting behaviour and results in significantly smaller forecast errors. Moreover, the improved coordination of forecasting behaviour leads to greater output and inflation stability. Over time, this information become less useful as a coordinating device as subjects continue to increase their usage of and overreact to past forecast errors when forming their expectations, leading to larger forecast errors and increased volatility. The fact that expectations can be influenced by focal information, at least in the short-run, suggests that *what* policy makers emphasize when communicating to the public can be very important in influencing economic stability.

2 Related Literature

Dozens of experiments have been conducted to understand how expectations are formed as structural features of an environment or information sets change. Duffy (2012) provides an extensive literature review discussing the evolution of the experimental expectation-formation literature. This paper contributes most directly to the learning-to-forecast New Keynesian experiments pioneered by Adam (2007). These experiments involve subjects forming output gap and/or inflation expectations in a multivariate multi-equation linearized environment where current output and inflation depend on aggregate expectations. Subjects are paid based based on their forecast error, inducing an incentive to form accurate forecasts. Pfajfar and Zakelj (2013) vary the type of nominal interest rate rule to explore the relationship between expectation formation and monetary policy. They find that forward-looking rules tend to generate expectational cycles and higher inflation variability than contemporaneous rules. Within the set of forward-looking policy rule treatments, they vary the sensitivity of interest rates to future expected inflation and output, and find that more aggressive policy leads to greater economic stability. In a companion paper, Pfajfar and Zakelj (2014) utilize their experimental data to test the rational expectations hypothesis in 70-period temporally linked economies. They find that they cannot reject rationality for 40% of their sample, while many subjects' expectations can be modelled by some form of adaptive behaviour. Changes in aggregate variables influence the likelihood of switching. For example, subjects are more likely to switch their strategies during recessions. Assenza et al. (2013) also study heterogeneous expectations in a New Keynesian experiment where subjects only forecast inflation, while the output gap expectation is either set to the steady state, formed naively based on past realized values, or formed by another human subject in the group. The authors also vary the central bank's reaction function between passive and active monetary policy. Like Pfajfar and Zakelj, they observe subjects frequently switching between forecasting heuristics. They find that an estimated evolutionary switching model of heterogeneous expectations can better describe expectation formation than a homogeneous expectations model.

A related paper by Roos and Luhan (2013) investigates how subjects gather and utilize information in a combined forecasting and optimization experiment. Subjects played the roles of either workers or firms who were incentivized to form accurate forecasts of wages and prices and maximize their utility or profits, respectively. A 'rudimentary' description of the data generating process was provided to subjects at the beginning of the experiment. Each period, subjects could choose to purchase for a small cost market information presented either cross-sectionally or in time-series form. The authors observe very low demand for information and that the majority of the information requests come from a small subset of subjects. Information purchases lead firms to earn higher profits but does not improve forecast accuracy for either type. Average absolute forecast errors do diminish over time and is attributed to learning.

This experiment extends the experimental design of Kryvtsov and Petersen (2014), in which subjects interact in a learning-to-forecast New Keynesian economy similar to the ones developed by Pfajfar and Zakelj (2013, 2014) and Assenza et al. (2013). Kryvtsov and Petersen (KP henceforth) study how the strength of the expectations channel of monetary policy changes in response to increased persistence of shocks, more aggressive monetary policy, and central bank forward guidance. Among other things, they find that providing focal central bank forecasts of their own path of interest rates has mixed effects. Inexperienced subjects condition on the information which leads to improved economic stability. With learning however, the aggregate economies either strongly condition on the forward guidance resulting in high aggregate stability or it creates increased confusion and greater instability.

New Keynesian environments can become fraught with multiple equilibria when agents' expectations are heterogeneous. To successfully forecast, subjects in these learning-to-forecast experiments must coordinate their expectations, and more specifically, their forecasting rules. Schelling (1960) argues that information that focuses players' attention on one equilibrium can facilitate coordination and that it may be rational to condition one's decision on the focal information. Numerous laboratory experiments have since shown that, in games with multiple Nash equilibria, focal points can facilitate and improve coordination (Mehta et al. (1994) generate focal points through variation in labelling, Blume and Gneezy (2000) on endogenously generated focal points.) Nagel (1995) observes high levels of coordination on focal points in a Keynesian-inspired 'beauty contest' game where subjects are rewarded for guessing closest to p times the mean of all numbers submitted. Recent theoretical work by Demertzis and Viegi (2008, 2009) has shown that the communication of an inflation target can serve as an effective focal point at coordinating and stabilizing expectations in Morris-Shin (2002) environments where there is poor or ambiguous public information. Thus, it is reasonable to conclude that communicating focal forecast error information in learning-to-forecast experiments would generate improved coordination and improved payoffs.

Extensive survey and experimental evidence suggests that individuals use heuristics to form their beliefs about the economy. Pfajfar and Santoro (2010) use the University of Michigan's Survey of Consumer Attitudes and Behavior to study expectation formation. They observe considerable heterogeneity in forecasting behaviour, including rational forecasters, highly adaptive forecasters and constant gain learners. Milani (2009) shows that adding constant gain learning where agents update their forecasts based on previous forecast errors can improve the fit of monetary DSGE models with alternative expectation specifications, including those with rational expectations. On the other hand, Keane and Runkle (1990) find strong support for rational expectation formation using price forecast data from the ASA-NBER Survey of Professional Forecasters. Using experimental evidence, Pfajfar and Santoro (2014) find that 37% of their subject pool can be described as using a general model that employs all available information, while 38% extrapolate trends in some form. Another 9% form adaptive expectations while the remaining 16% exhibit behaviour consistent with sticky information and adaptive learning models. Finally, KP find strong support for an adaptive-lagged expectation formation rule where subjects condition both on current shocks and lagged realized values of inflation and output.

As forecasting experiments become a more policy-relevant source of data, further research needs to be conducted on how certain design decisions influence behaviour. For example, how we place information on subjects' screens may create unintended focal information that can bias forecasting behaviour. Placing historical information right next to where subjects submit their forecast has the potential to generate adaptive or trend-chasing behaviour, especially if the information is presented graphically. Locating that information elsewhere while making the current period shocks more salient may generate increased sensitivity to those shocks and a reduction in adaptive behaviour. Through a series of related experiments, Tversky and Kahneman (1973) demonstrate how individual behaviour can be biased by information that is easily accessible. In the context of this experiment, the *availability heuristic* would suggest that subjects would condition their expectations more on past forecast errors when this information is made more readily available.

3 Experimental Design

The experiments were conducted in Montreal, Quebec. Both non-student and student subjects were invited to participate in sessions that involved 30 minutes of instruction and 90 minutes of game participation. Each session consisted of 9 subjects interacting together as a single group. Earning, including a \$10 show up fee, ranged from \$18 to \$47, and average \$35.25 for two hours of participation.

The experiment took place within a simplified New Keynesian economy where households and firms make optimal decisions given their expectations. The theoretical framework is derived in Woodford (2003). The aggregate economy implemented in the experiment can be described by the following four equations, calibrated to match three moments in the Canadian data: standard deviation of inflation deviations (0.44 per cent), serial correlation of inflation deviations (0.4), and the ratio of standard deviations of output gap and inflation (4.4).

$$x_t = E_t^* x_{t+1} - (i_t - E_t^* \pi_{t+1} - r_t^n) \ , \tag{1}$$

$$\pi_t = 0.989 E_t^* \pi_{t+1} + 0.13 x_t \ , \tag{2}$$

$$i_t = 1.5E^*_{t-1}\pi_t + 0.5E^*_{t-1}x_t \ , \tag{3}$$

$$r^n_t = 0.57r^n_{t-1} + \epsilon_t \ . \tag{4}$$

Equation 1 is the Investment-Saving (IS) curve and describes how the output gap x_t, a measure of aggregate demand above its natural level, responds to current aggregate expectations of the future output gap and deviations of the real interest rate, $i_t - E^*_t\pi_{t+1}$, from the natural rate of interest, r^n_t. As the real interest rises above the natural rate of interest, contractionary pressures cause the output gap to decrease.

Equation 2 is the New Keynesian Phillips curve and describes the supply side of the economy. The equation is derived from monopolistically competitive firms' intertemporal optimization problem. As aggregate expectations of future inflation, $E^*_t\pi_{t+1}$, or aggregate demand increase, current inflation will increase. Firms are able to update their prices randomly, leading to sluggish adjustment in prices.

Equation 3 is the reaction function of the central bank and describes how the nominal interest rate is set. According to this specification, the central bank increases nominal interest rates in response to higher expected inflation and output gap formed in the previous period for the current period. This specification allows for a period t nominal interest rate to be provided to subjects when they form their period $t+1$ forecasts. Under rational expectations, this formulation is equivalent to a standard central bank function that targets current period realized output and inflation.

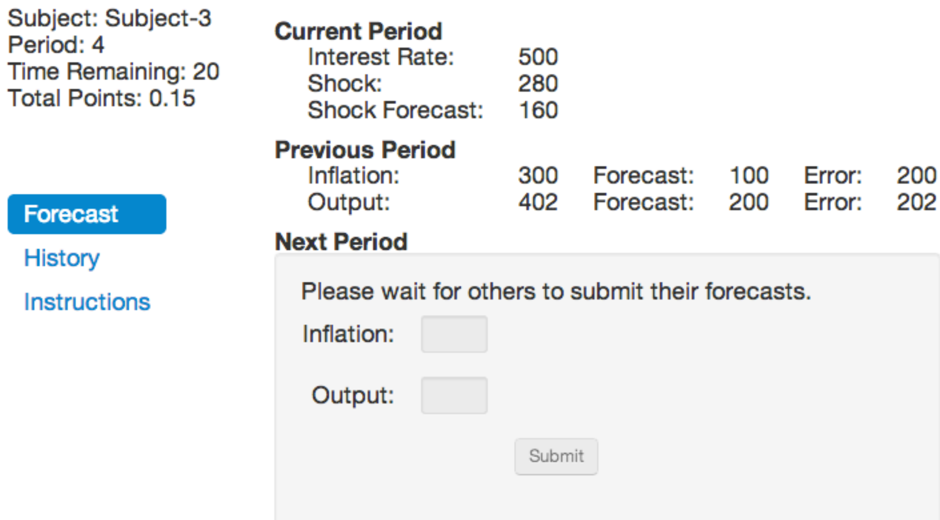
Finally, Equation 4 describes the stochastic process of the natural rate of interest as an AR(1) process where ϵ_t is assumed to be drawn randomly from a normal distribution with mean zero and variance σ^2_r, where $\sigma_r = 1.13$

Each experimental session consisted of two stationary repetitions, consisting of approximately 50 periods each. These repetitions were initialized at the long-run steady state of zero inflation, output gap, and nominal interest rate. Each period, subjects were provided information about the current period's interest rate, shock to the natural rate of interest, and the expected shock size in the following period. They were then asked to provide forecasts for next period's inflation and output gap in basis points (e.g. 1% would be submitted as 100 basis points). Subjects were allowed to submit positive or negative numbers, and there was no limit to the values they may submit. Each period lasted up to 1 minute in the first 10 periods of each repetition, and 45 seconds thereafter. If all subjects submitted their decisions before time elapsed, which was generally the case, the experiment immediately move on to the next period. Before moving onto the next period, the current period's inflation and output as well as the next period's nominal interest rate were computed using the median forecasts for inflation and output. The motivation for using the median, rather than the average forecast as done in similar experiments, was to minimize the ability of a single subject to manipulate the economy, and because the median tends to be a better measure of central tendency.

Two information treatments are considered in this experiment. We analyze behaviour in a benchmark environment (abbreviated as the "B" treatment) where subjects must actively obtain historical information and compare this to a treatment where subjects are provided with additional forecast error information on the main screen ("FEI" treatment). The purpose of this alternative information environment is to identify whether the focal forecast error information influences how inexperienced and experienced subjects form forecasts. Figure 1 is a screenshot of the main screen that subjects interacted with during the FEI sessions. In the B treatment, subjects did not have immediately accessible information on previous period forecast errors. Mehta et al. (1992) identify *closeness* or *proximity* as a feature that enhances the salience of a specific strategy and the usefulness of it as a coordinating device. In this experiment, forecast error information in the FEI treatment is made salient primarily because of its proximity to where subjects submit their forecasts, on the main screen and its relative ease in accessing. We also study the effect of experience on forecasting, disagreements, and macroeconomic stability. By resetting the environment and conducting a second stationary repetition, we can observe whether forecasting behaviour, forecast errors, and aggregate outcomes are significantly different with learning. We conducted 5 sessions of the Benchmark treatment and 4 sessions of the Forecast Error Information treatment.

The experimental design builds on Kryvtsov and Petersen (2014) and differs on a number of dimensions from the previous literature. First and most importantly, the experimental interface is considerably different. The only available information on the main screen is the current nominal interest rate, the natural rate of interest rate shock occurring in the current period, and a forecast of next period's shock. Historical information is placed on a secondary screen which subjects must actively click on in order to obtain past information about their past forecasts and realized aggregate variables. This differs from the interfaces of previous experiments that place all the current and historical information on a single screen. The purpose of this

Figure 1: Screenshot from Forecast Error Information treatment [I]



(I) This is the main screen of the interface subjects interacted with in the FEI treatment. "Previous Period" information was not included in the Benchmark treatment.

modification is to minimize the degree to which subjects are 'primed' to focus on past information when forming their forecasts and to create a more realistic environment where subjects must 'look up' past information if they are interested in utilizing it to make forecasts. As in Roos and Luhan (2012), the data generating process of the economy is provided to subjects in a supplementary technical instructions screen that subjects could access if they wanted more information. Providing subjects with the data generating process makes it easier to identify the set of information that may be used in forming forecasts. However, unlike Roos and Luhan, we do not charge subjects for this information. Instead, they must utilize their limited time to look up the information. Finally, subjects submitted forecasts for both output and inflation, where in the earlier literature, subjects either forecasted one of the two variables or forecasted inflation for one and two periods ahead.

Participants were presented with detailed instructions before the experiment began. We explained using non-technical language how the output gap, inflation, and nominal interest rate would evolve given their forecasts and exogenous shocks. Subjects were informed that their only task would be to submit forecasts for the following period's output gap and inflation, and that their score would depend on the accuracy of their forecast. Specifically, their score would be computed each period according to the following payoff function:

$$Score_t = 0.3(e^{-0.01|E^*_{t-1}\pi_t - \pi_t|} + e^{-0.01|E^*_{t-1}x_t - x_t|}) \ , \tag{5}$$

where $E^*_{t-1}\pi_t - \pi_t$ and $E^*_{t-1}x_t - x_t$ were the subject's forecast errors associated with forecasts submitted in period $t-1$ for period t variables. With more than 100 periods of play, a subject had the potential to earn over $70 by making accurate forecasts. This scoring rule incentivizes subjects to form accurate forecasts. This scoring rule is very similar to that used in the previous experimental literature in that scores decrease monotonically with the forecast errors and the minimum score a subject can earn in any period is zero. In the rules used by Assenza et al. (2013) and Pfajfar and Zakelj (2014), there is diminishing marginal loss from forecast errors while under our rule. Under our rule, the per-period score reduces by 50% for every 100 basis point forecast error for both inflation and output, continually incentivizing subjects to make as accurate forecasts as possible.

We also clearly explained that the median forecast for each of inflation and output formed each period would be used in the calculation of the output gap, inflation, and the nominal interest rate. Subjects never directly observed each others' forecasts or the median forecasts. We also explained to subjects how they could access detailed information about the economy in the technical instructions. Subjects were given a 4-period practice phase of approximately 10 minutes to learn the interface and better understand the timing of the game.

Forecasting Models

As a starting point, we begin with a rational expectations forecasting model. Given the parameterization of the environment, rational expectations solution does not depend on any endogenous state variables but only on exogenous state variables. The rational expectations solution for the output gap is simply a function of the current period shock and parameters of the model:

$$E_t x_{t+1} = \Phi r^n_t \ . \tag{6}$$

The rational expectations solution for inflation forecasts follows an identical structure.

We also consider a variety of alternative forecasting models. The simplest is the *naive expectations* model, where the agent would form their expectation of a variable based on its previous realized value. We consider an adaptive model of the form:

$$E_t x_{t+1} = \beta x_{t-1} \ . \tag{7}$$

It is not immediately obvious whether presenting subjects with forecast error information would increase the importance they place on past realized values when forming their expectations. On one hand, the forecast error information reduces the need to utilize historical information when forming forecasts, and should reduce the reliance on past values. On the other hand, past output and inflation are clearly presented on the main screen and also become more focal.

We consider the possibility that subjects' forecasts respond to trends in inflation and output, as in the model for output expectations below:

$$E_t x_{t+1} - x_{t-1} = \alpha + \eta(x_{t-1} - x_{t-2}) \ . \tag{8}$$

If the estimated $\hat{\eta} \geq 0$, agents expect that the previous upward or downward movements in the variable that they are forecasting will continue in the next period, i.e. the subjects are *trend-chasing*. If $\hat{\eta} < 0$, agents expect that the movement in the variable of interest will reverse its trend, and we describe this as *contrarian expectations*. In order to observe the trend, subjects would need to review the historical screen or else remember values from two periods prior. Given the additional information presented on the main screen in the FEI treatment, we would expect less time spent on the history screen and generally a reduction in trend-chasing behaviour compared to the B treatment.

Presenting salient forecast error information may prime subjects to condition on their forecast errors when forming expectations about the future. This type of behaviour for output forecasts can be described by a *constant gain adaptive expectations* model:

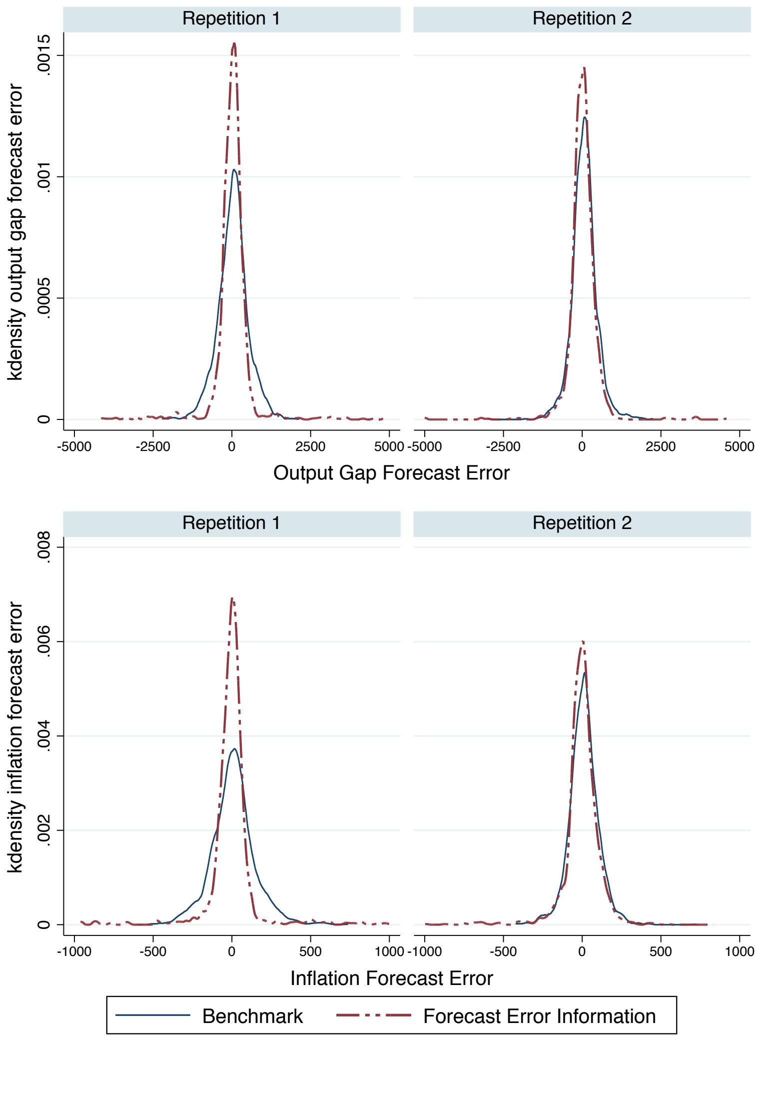
$$E_t x_{t+1} = E_{t-1} x_t + \gamma(E_{t-1} x_t - x_{t-1}) \ , \tag{9}$$

with a similarly structured model for inflation forecasts. The dependent variable in this estimation is the change in expectations, $E_t x_{t+1} - E_{t-1} x_t$. An estimated $\hat{\gamma} < 0$ suggests that when subjects over-forecasts a variable, they will correct their forecast downward next period. If focal forecast error information is important in influencing forecasting behaviour, we should expect the estimated $\hat{\gamma}$'s in the two treatments to be significantly different from one another.

4 Findings

Forecast Errors

Figure 2: Kernel densities of output gap and inflation forecast errors



Does making forecast errors more salient improve forecasting ability? In the Benchmark treatment, to identify one's forecast error, a subject would need to review the history screen to identify how accurate their forecasts were. This could be observed by comparing the distance of the time series graphs of realized variables to that of forecasted variables. This task was made simpler in the FEI treatment, where subjects forecast errors were presented on the main screen. If subjects were to condition on their past forecast errors and can successfully correct under- or over-forecasting, then we should expect to see smaller forecast errors in the FEI treatment.

The kernel densities of the forecast errors for each treatment are displayed in Figure 2 by repetition and summary statistics on the squared forecast errors are given in Table 1. We also report the effect size using Glass' Δ, which is the difference between the mean B forecast error and the mean FEI forecast error, and is measured in standard deviations.[1] The difference across treatments is stark in the first repetition. The density function for the FEI treatment is more heavily centered around zero. The median and mean absolute forecast errors in the FEI treatment are generally smaller for both output and inflation in both repetitions (the only exception is inflation forecasts, which are modestly higher in the second repetition of the FEI treatment).

These differences across treatments diminish with learning in the second repetition. Relative to the first repetition of the Benchmark treatment, median and mean forecast errors for both output and inflation decrease with learning, and the variance of forecast errors also declines. Forecast errors somewhat worsen with learning in the FEI treatment. For output forecast errors, the median and mean error increases but the standard deviation decrease. This suggests that the tails of the distribution of forecast errors are getting fatter but less extreme. For inflation forecast errors, the median error increases, but mean and standard deviation decrease. The changes in FEI forecast errors are negligible and statistically insignificant.

The null hypothesis that the forecast errors under the B and FEI treatments are drawn from the same distribution is rejected through a two-sample Kolmogorov Smirnov test ($p < 0.001$ for inflation and output forecasts in both repetitions). Comparing within a treatment, the distributions of output gap and inflation forecast errors are significantly different across repetitions in both the B treatment ($p < 0.001$) and the FEI treatment ($p < 0.01$).

Evaluation of Forecasting Models

We study the fit of the rational expectations, adaptive expectations, trend-chasing, and constant gain adaptive expectations models across repetitions. Each model is estimated as a fixed-effects regression with standard errors clustered at the session level. We consider the effects of information and learning in separate regressions. The results for inflation forecasts are presented in Tables 2 and 3, while output forecasts are

[1] A Glass Δ of 0.44 for output forecast errors in the first repetition implies that the mean B forecast error was 0.44 standard deviations larger than the mean FEI forecast error.

presented in Tables 4 and 4. As measures of fit, we compute R^2, Akaike Information Criterion (A.I.C.), and Bayesian Information Criterion (B.I.C.) statistics.

The rational expectations model is presented in columns (1) and (5) for all tables. Subjects significantly condition both their inflation and output forecasts on the current period shock in both repetitions. While subjects with forecast error information utilize the shock less in their forecasts, the differences across treatments is not statistically significant. With experience, subjects in both treatments learn to place more weight on the shock in their forecasts. The learning effect is statistically significant for subjects in the FEI treatment ($p < 0.05$).

Table 1: Absolute Forecast Errors for Output Gap and Inflation[I]

Treatment		Output Gap		Inflation	
		Rep. 1	Rep. 2	Rep 1.	Rep. 2
B					
	median	278	219	75	52
	mean	363.75	295.86	102.73	69.78
	stdev	319.32	278.67	98.85	64.12
	Glass Δ^a		0.21		0.33
			(0.15-0.27)		(0.27-0.39)
FEI					
	median	166	178	37	47
	mean	224.30	235.61	75.94	72.32
	stdev	475.62	235.52	272.11	224.44
	Glass Δ^b		-0.02		0.01
			(-0.09,0.04)		(-0.05,0.08)
	Glass Δ^c	0.44	0.22	0.27	-0.04
		(0.37-0.50)	(0.15-0.28)	(0.21-0.33)	(-0.10-0.02)

(I) Summary statistics for forecast errors
(a,b) Effect sizes associated with a Glass Δ test across repetitions.
(c) Effect sizes associated with a Glass Δ test across treatments. Values in brackets are the 95% confidence interval of the estimated effect size. The estimates are calculated using the Benchmark group's standard deviation.

The adaptive expectations model is presented next in columns (2) and (6). Lagged values of output and inflation play a quantitatively large and significant role in the forecasts made by subjects, across levels of experience and information. There are no significant differences in adaptive behavior across treatments when inexperienced subjects form their forecasts. In the second repetition, the experienced FEI subjects place significantly more weight than B subjects on past inflation when forming their inflation forecasts ($p < 0.05$). The role of past output levels in output forecasts does not significantly differ across treatments, however B subjects become significantly less adaptive in their output forecasts with experience.

Columns (3) and (7) presents the results from the trend-chasing expectations models. Generally, the model does not perform well at describing the variability in inflation and output expectations. Past trends in inflation do not generate large or significant trend-chasing behavior among inexperienced subjects. Experienced subjects do exhibit weakly significant contrarian expectations and there are no considerable differences across treatments. In forming their output forecasts, subjects in both treatments exhibit contrarian heuristics. This behavior is only statistically significant among those in the Benchmark treatment and does not change significantly with learning.

Finally, the constant gain expectations model addresses how subjects update their forecasts in response to the previous period's forecast errors. The results of this set of regressions are presented in columns (4) and (8). This model of forecasting fits the data the best according to all of our goodness of fit measures. Across all treatments and repetitions, subjects significantly respond to their past errors when forming their forecasts of future inflation and output. Inexperienced B subjects react significantly more to their inflation forecast errors when forming their forecasts than their FEI counterparts ($p < 0.01$). The salient forecast error information works to stabilize FEI subjects' responsiveness to their forecast errors. With experience, the FEI subjects increase their reaction to their inflation forecast errors by more than 65%, while the B subjects responsiveness is largely unchanged. Experienced FEI subjects significantly overreact to their errors relative to the B subjects. The weight that subjects place on past errors in their output forecast does not differ significantly across treatments or with learning. On average, inexperienced FEI subjects exhibit a larger aggressive reaction to their past forecast errors, however there is considerable heterogeneity among subjects and the differences between treatments is only significantly at the 15% level.

Table 2: Comparison of Estimated Expectation Models [I]

	Inflation Forecasts							
	Repetition 1				Repetition 2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$-π_{t-1}	$E_t\pi_{t+1}$-$E_{t-1}\pi_t$	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$-π_{t-1}	$E_t\pi_{t+1}$-$E_{t-1}\pi_t$
r_t^n	0.235**				0.251***			
	(0.05)				(0.03)			
r_t^n x FEI	-0.139				-0.035			
	(0.08)				(0.05)			
π_{t-1}		0.559***				0.580***		
		(0.04)				(0.02)		
π_{t-1} x FEI		0.207				0.275**		
		(0.21)				(0.06)		
$\pi_{t-1} - \pi_{t-2}$			-0.051				-0.055*	
			(0.02)				(0.02)	
$\pi_{t-1} - \pi_{t-2}$ x FEI			-0.045				0.188	
			(0.11)				(0.09)	
$E_{t-2}\pi_{t-1} - \pi_{t-1}$				-0.871***				-0.880***
				(0.02)				(0.04)
$E_{t-2}\pi_{t-1} - \pi_{t-1}$ x FEI				0.255***				-0.142**
				(0.05)				(0.04)
α	39.604***	21.224***	10.015***	6.861***	11.721***	3.395***	6.629***	6.044***
	(0.13)	(2.55)	(0.01)	(0.25)	(0.81)	(0.28)	(0.01)	(0.24)
N	3981	3880	3780	3767	3960	3862	3765	3755
R^2	0.0163	0.0672	0.000545	0.415	0.0333	0.121	0.00107	0.534
A.I.C.	53332.3	51856.5	50217.7	49464.8	51513.5	49938.4	48860.0	48730.9
B.I.C.	53344.9	51869.0	50230.2	49477.3	51526.1	49950.9	48872.5	48743.3

(I) Significance levels: $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Comparison of Estimated Expectation Models [I]

	Inflation Forecasts							
	Benchmark				FEI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$-π_{t-1}	$E_t\pi_{t+1}$-$E_{t-1}\pi_t$	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$	$E_t\pi_{t+1}$-π_{t-1}	$E_t\pi_{t+1}$-$E_{t-1}\pi_t$
r_t^n	0.235**				0.096**			
	(0.05)				(0.03)			
r_t^n X EXP	0.017				0.121**			
	(0.05)				(0.03)			
π_{t-1}		0.559***				0.766**		
		(0.04)				(0.20)		
π_{t-1} X EXP		0.021				0.089		
		(0.04)				(0.18)		
$\pi_{t-1} - \pi_{t-2}$			-0.051				-0.096	
			(0.02)				(0.10)	
$\pi_{t-1} - \pi_{t-2}$ x EXP			-0.004				0.229	
			(0.02)				(0.17)	
$E_{t-2}\pi_{t-1} - \pi_{t-1}$				-0.871***				-0.616***
				(0.02)				(0.04)
$E_{t-2}\pi_{t-1} - \pi_{t-1}$ X EXP				-0.009				-0.406***
				(0.03)				(0.04)
α	23.767***	12.239***	9.521***	8.345***	28.241***	12.453**	6.715***	3.917***
	(0.79)	(0.54)	(0.01)	(0.27)	(0.19)	(3.16)	(0.01)	(0.31)
N	4511	4421	4331	4308	3430	3321	3214	3214
R^2	0.0949	0.268	0.00192	0.539	0.00788	0.0509	0.000516	0.460
A.I.C	54019.3	52033.5	51868.0	51509.0	47619.3	46052.0	44353.5	44020.7
B.I.C	54032.1	52046.3	51880.7	51521.7	47631.6	46064.2	44365.7	44032.9

(I) Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Comparison of Estimated Expectation Models [I]

	Output Forecasts							
	Repetition 1				Repetition 2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	$E_t x_{t+1}$	$E_t x_{t+1}$	$E_t x_{t+1}$-x_{t-1}	$E_t x_{t+1}$-$E_{t-1} x_t$	$E_t x_{t+1}$	$E_t x_{t+1}$	$E_t x_{t+1}$-x_{t-1}	$E_t x_{t+1}$-$E_{t-1} x_t$
r_t^n	0.629**				0.727***			
	(0.16)				(0.08)			
r_t^n X FEI	-0.155				-0.020			
	(0.10)				(0.06)			
x_{t-1}		0.511***				0.465***		
		(0.03)				(0.02)		
x_{t-1} x FEI		-0.015				0.104		
		(0.15)				(0.08)		
$x_{t-1} - x_{t-2}$			-0.107*				-0.133**	
			(0.04)				(0.04)	
$x_{t-1} - x_{t-2}$ x FEI			-0.117				-0.027	
			(0.15)				(0.15)	
$E_{t-2} x_{t-1} - x_{t-1}$				-0.792***				-0.806***
				(0.04)				(0.05)
$E_{t-2} x_{t-1} - x_{t-1}$ x FEI				-0.152				0.012
				(0.08)				(0.14)
α	31.015***	31.518***	34.902***	30.143***	26.738***	20.602***	46.649***	37.263***
	(0.25)	(0.26)	(0.18)	(0.49)	(3.22)	(0.82)	(0.02)	(1.72)
N	3981	3880	3780	3767	3960	3862	3765	3755
R^2	0.0360	0.154	0.0127	0.568	0.130	0.335	0.0266	0.605
A.I.C.	58683.4	56778.1	55666.4	55373.9	54538.8	51742.8	51827.0	51420.6
B.I.C.	58696.0	56790.7	55678.9	55386.3	54551.3	51755.4	51839.5	51433.1

(I) Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Comparison of Estimated Expectation Models [I]

	Output Forecasts							
	Benchmark				FEI			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	$E_t x_{t+1}$	$E_t x_{t+1}$	$E_t x_{t+1}$-x_{t-1}	$E_t x_{t+1}$-$E_{t-1} x_t$	$E_t x_{t+1}$	$E_t x_{t+1}$	$E_t x_{t+1}$-x_{t-1}	$E_t x_{t+1}$-$E_{t-1} x_t$
r_t^n	0.629**				0.474**			
	(0.16)				(0.09)			
r_t^n X EXP	0.098				0.232**			
	(0.11)				(0.05)			
x_{t-1}		0.511***				0.496**		
		(0.03)				(0.14)		
x_{t-1} X EXP		-0.046*				0.072		
		(0.02)				(0.08)		
$x_{t-1} - x_{t-2}$			-0.107*				-0.225	
			(0.04)				(0.12)	
$x_{t-1} - x_{t-2}$ X EXP			-0.026				0.065	
			(0.02)				(0.08)	
$E_{t-2}x_{t-1} - x_{t-1}$				-0.792***				-0.944***
				(0.04)				(0.05)
$E_{t-2}x_{t-1} - x_{t-1}$ X EXP				-0.013				0.151
				(0.03)				(0.07)
α	34.625***	29.830***	51.497***	41.439***	21.329***	21.070***	26.300***	23.319***
	(2.57)	(0.68)	(0.07)	(2.35)	(0.58)	(0.49)	(0.15)	(1.99)
N	4511	4421	4331	4308	3430	3321	3214	3214
R^2	0.0976	0.321	0.0204	0.612	0.0410	0.118	0.0135	0.549
A.I.C	63109.6	60662.2	61133.2	60460.2	50485.8	48502.8	46967.5	46943.0
B.I.C	63122.4	60675.0	61145.9	60473.0	50498.1	48515.0	46979.6	46955.1

(I) Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Heterogeneity in Forecasts

Forecast error information may reduce the heterogeneity in subjects' forecasts by providing a common focal point. As a measure of heterogeneity in expectations, we calculate the standard deviation (in basis points) of forecasts each period at the session level. Histograms and kernel density functions depict the distribution of heterogeneity in Figure 3 by treatment and repetition.

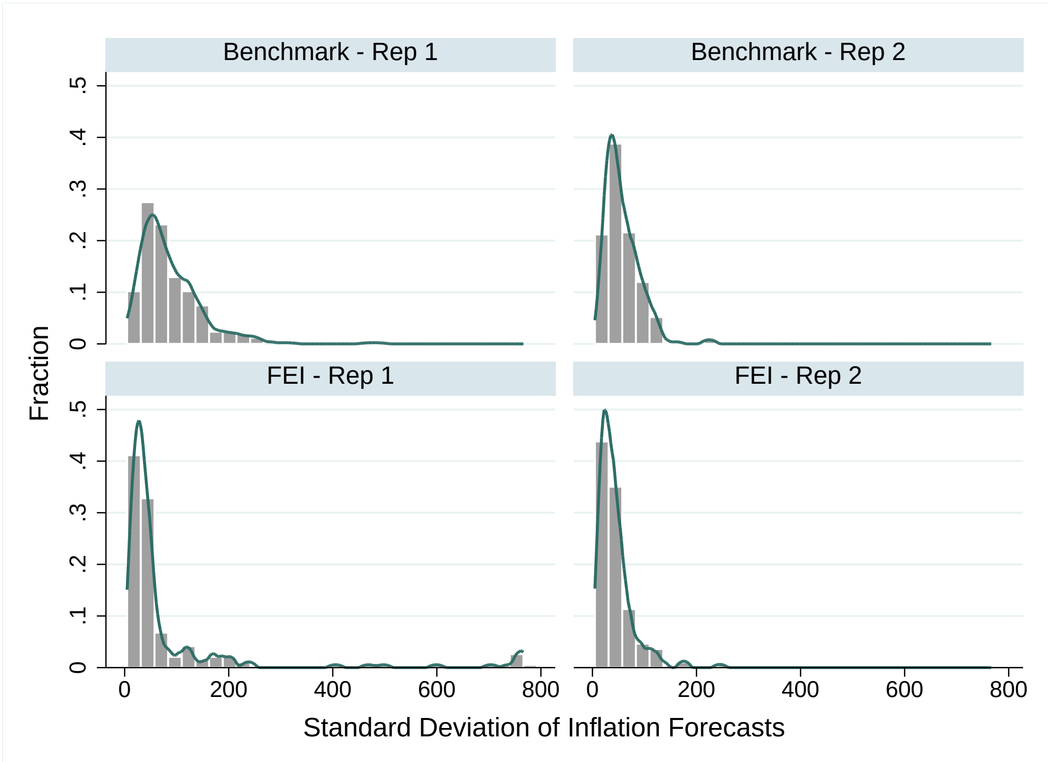
The distributions of forecast heterogeneity are relatively skewed toward zero when subjects are presented with sailent forecast error information. That is, there is considerably less disagreement in inflation and output forecasts when subjects have common information to coordinate on. Two-sample Kolmogorov-Smirnov rejects the null hypothesis that the distribution functions are identical across treatments for either of the repetitions ($p < 0.01$ for both inflation and output gap disagreements). The median inflation disagreement in Repetition 1 (Repetition 2) is 70 (50) bps in the Benchmark treatment and 38 (35) bps in the FEI treatment. Similarly, the median output disagreement in Repetition 1 (Repetition 2) is 159 (191) bps in the Benchmark treatment and 119 (131) bps in the FEI treatment. While inflation disagreements lessen over time, output disagreements worsen for both treatments. This is consistent with our finding from Table where we observe relatively large standard errors when we estimate the various models using output gap forecasts for experienced subjects.

Macroeconomic Stability

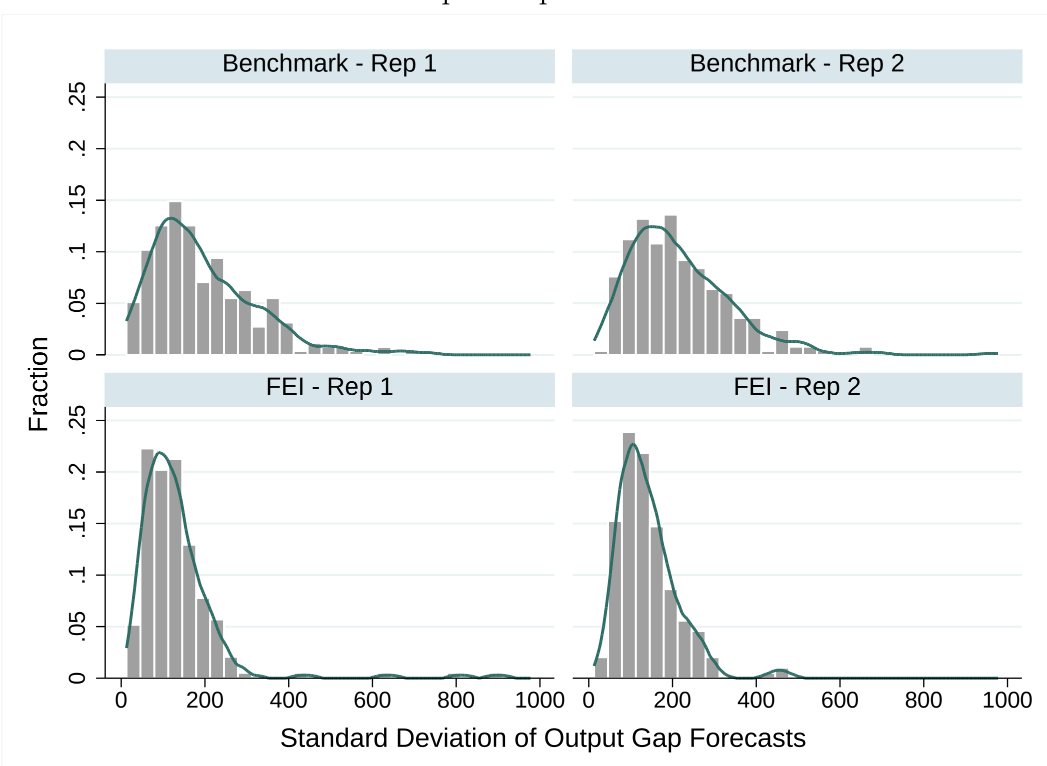
We now turn our attention to aggregate outcomes and compare the volatility of the output gap and inflation across treatments. Figures 6 and 9 presents time series

Figure 3: Heterogeneity in inflation and output gap forecasts

Inflation Forecasts



Output Gap Forecasts



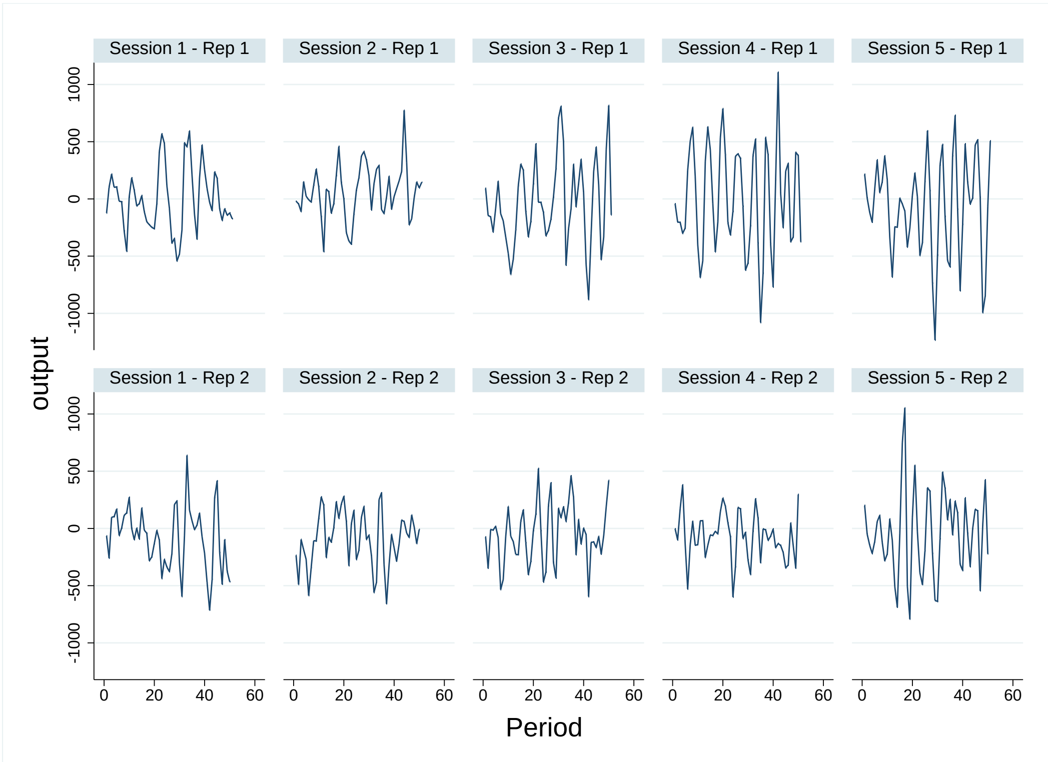
of the output gap and inflation across sessions and repetitions for each treatment while Table 6 provides the associated summary statistics. Consider behaviour of inexperienced subjects in Repetition 1. Visually, we can detect significant differences in both output and inflation across treatments. In the Benchmark economies, the aggregate variables appear more volatile and reach greater extremes than in the FEI economies. The mean standard deviation of the output gap (inflation) is 149.89 (58.43) basis points higher in Benchmark treatment. Wilcoxon rank sum tests reject the null hypothesis that the distributions of output gap and inflation variability across the two treatments are identical ($p = 0.014$ for both output gap and inflation). This coincides with our earlier finding that inexperienced subjects in the B treatment are relatively more responsive to their forecast errors and more adaptive than subjects in the FEI treatment. The average autocorrelation of output in the first repetitions of B and FEI are 0.46 and 0.26, respectively.

This highly reactive behaviour in the Benchmark treatment dampens on average with learning in the second repetition. The mean standard deviation of output (inflation) falls by 84.7 (32.12) basis points, and a signed-rank test weakly rejects the null hypothesis that there are small differences across repetitions ($p = 0.138$ for output gap and $p = 0.08$ for inflation). This is consistent with the findings in the previous section that, with learning, there are minimal differences across repetitions for any of the learning models. Subjects somewhat decrease their reliance on lagged output in favour of lagged forecast errors and contrarian beilefs when forming their expectations, resulting in increased mean reversion.

The opposite occurs in the FEI treatment. In the second repetition, the mean standard deviation of output (inflation) significantly increases by 63.14 (30.77) basis points ($p = 0.068$ for both variables). This increase in volatility is generated by a more extreme reaction to forecast errors. Given the considerable changes across repetitions in both treatments, there are no significant differences between the B and FEI treatments in Repetition 2 ($p = 0.806$).

Figure 6: Time series of the output gap by session and repetition

Benchmark



FEI

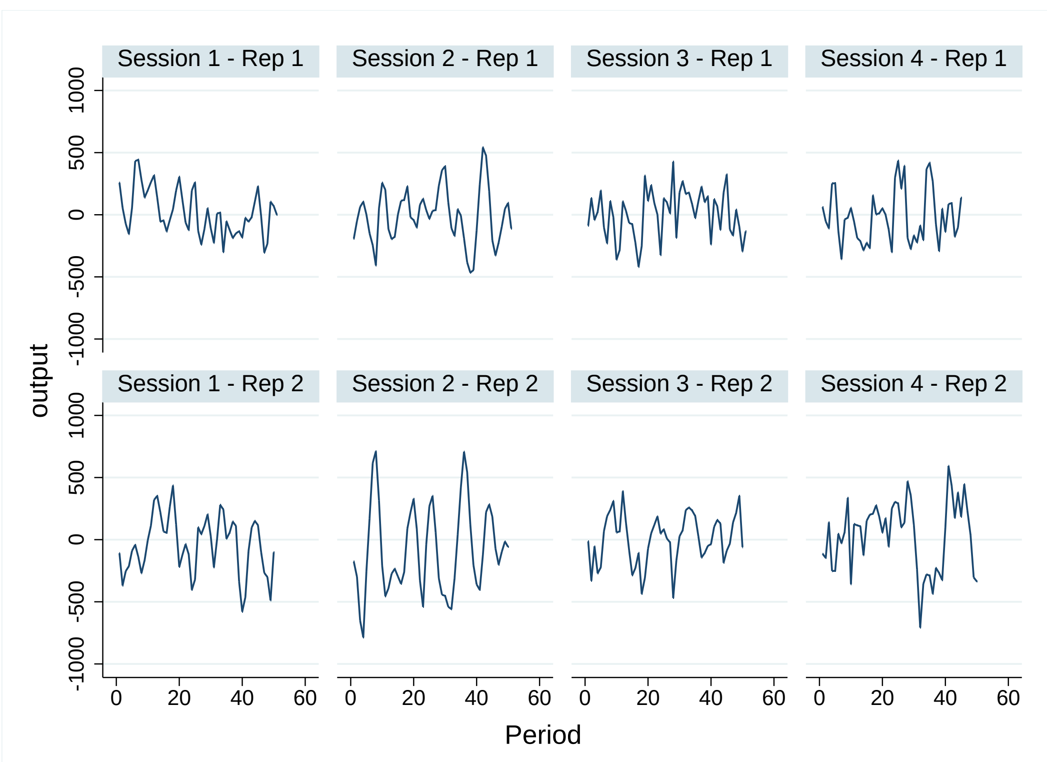
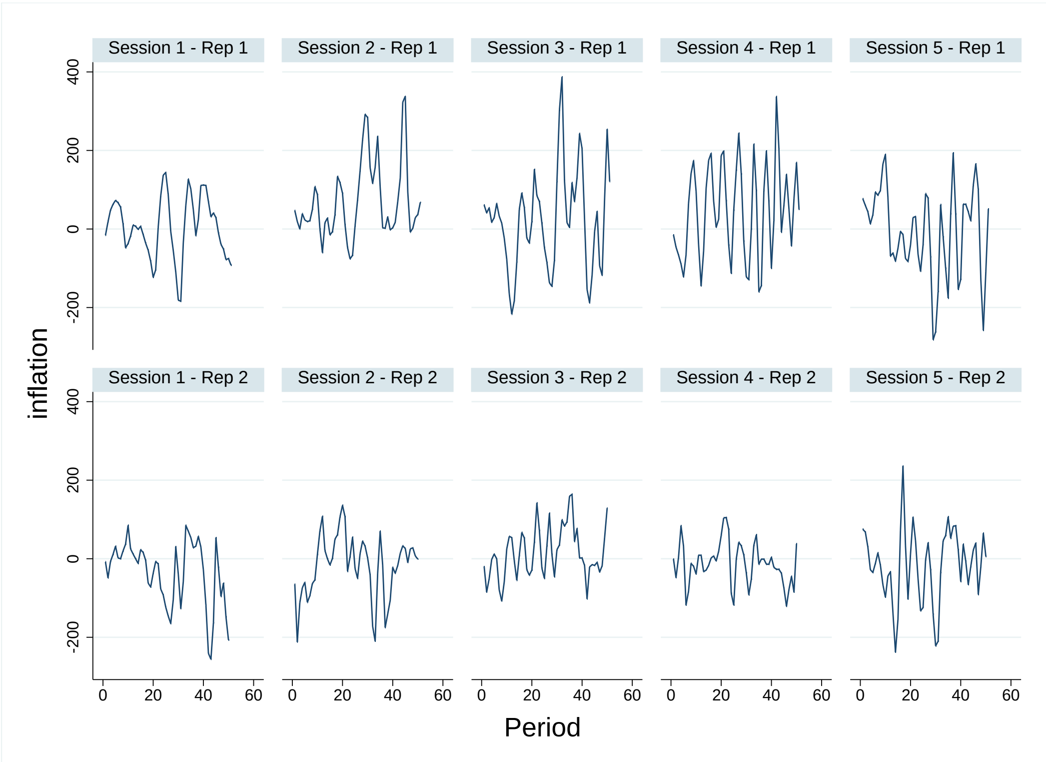


Figure 9: Time series of inflation by session and repetition

Benchmark



FEI

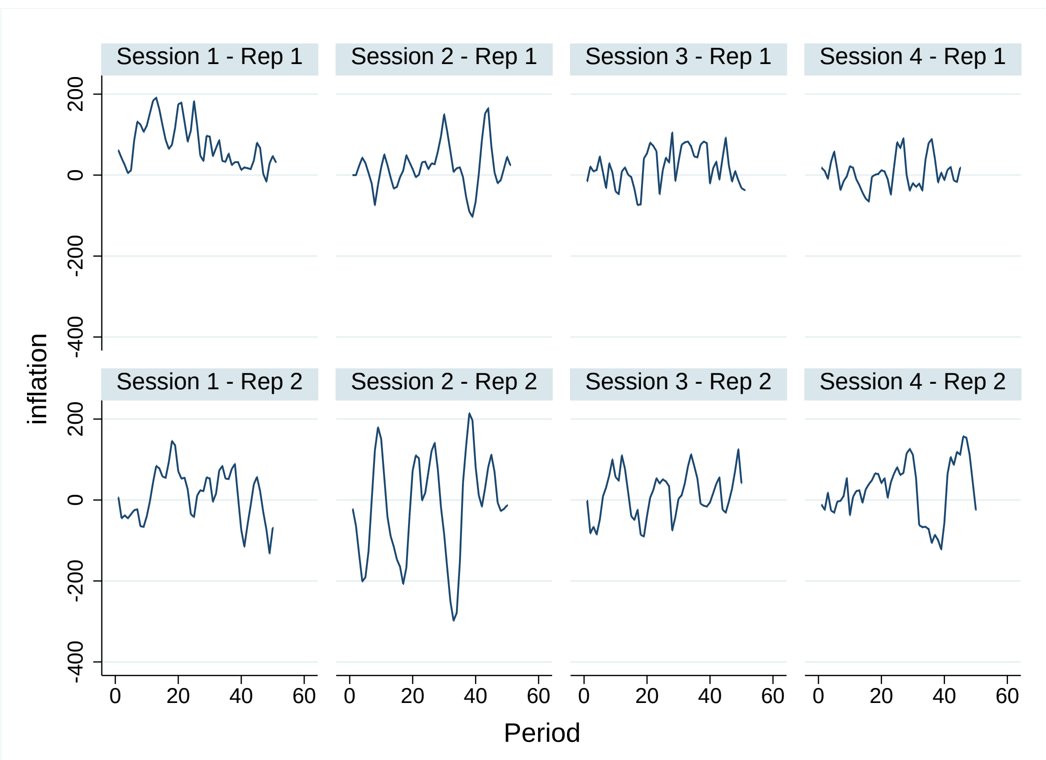


Table 6: Standard deviation of output and inflation[I]

Treatment		Output Gap		Inflation	
		Rep. 1	Rep. 2	Rep 1.	Rep. 2
B					
	mean	350.84	266.14	106.03	73.91
	min	224.41	204.04	76.91	53.38
	max	459.76	370.30	126.31	91.15
	p-value[a]		0.138		0.08
FEI					
	mean	200.95	264.09	47.60	78.37
	min	182.59	195.55	37.32	55.12
	max	220.33	355.73	54.89	127.78
	p-value[b]		0.068		0.068
	p-value[c]	0.014	0.806	0.014	0.806

(I) Summary statistics for the standard deviation of output gap and inflation calculated at the session-repetition level are presented.
(a,b) p-values associated with a signed-rank test across repetitions.
(c) p-values associated with a rank-sum test across treatments.

5 Discussion

This paper reports the findings from a laboratory experiment that explores the effects of experimental design features on forecasting behavior. The experimental environment is modeled as a reduced-form New Keynesian economy, where aggregate expectations formed by subjects are used to generate macroeconomic dynamics. This experiment specifically studies how forecast error information and learning influence expectation formation. In the benchmark environment, subjects must look up historical information and infer their forecast errors by comparing time series of their forecasts to realized values. The results of this treatment are compared to a second environment where subjects are provided salient information on their forecast errors in the previous period.

Four heuristics of expectation formation are compared to identify well-fitting models under different information structures: rational, adaptive, trend-chasing and constant-gain learning. While subjects do significantly utilize random shocks and past outcomes in their forecasts, forecasting behavior is best described by constant-gain learning under both the benchmark and the forecast error information environments.

Inexperienced subjects generally attempt to correct past forecast errors by significantly raising (lowering) their forecasts in response to past under- (over-) forecasting. However, when it comes to inflation forecasting, the reactions are significantly less extreme when subjects are provided with precise information about their forecast errors. In other words, inexperienced subjects with only visual information overreact to their forecast errors compared to those with additional numerical information. Presenting inexperienced subjects with accessible and salient forecast error information also draws their attention away from aggregate shocks when forming both forecasts. While these subjects are less 'rational' than what would be predicted by the rational expectations model, they incur smaller forecast errors because they receive immediate, more precise feedback and correct themselves. This results in a significantly lower forecast errors and volatility.

After extensive learning, experienced subjects continue to utilize forecast error information. Those with salient forecast error information significantly increase their usage of the aggregate shock in forming their forecasts, leading to more extreme forecasts, outcomes and forecast errors. As a result, they become increasing overreactive to their errors, perpetuating greater volatility.

The benchmark treatment can be viewed as an environment with informational frictions. Subjects must actively seek out and interpret relevant information about forecast accuracy on a second screen, leaving them prone to inattentiveness, extrapolative or over-reactive behavior that generate disagreements. Similar to KP (2014) and Roos and Luhan (2013), we find that most subjects will not utilize information if it comes at a cognitive cost. Instead, they overly rely on easy-to-interpret information such as historical information and trends. With limited time and capacity to interpret information, we observe that subjects rationally select a coarse subset of variables and heuristics to condition their expectations on - a finding consistent with the notion of rational inattention developed by Sims (2003). By providing a common and accessible forecasting heuristic to all subjects, heterogeneity in expectations is efficiently and effectively reduced.

The findings of this experiment suggest that the design of an experimental interface matters. Providing salient forecast error information will encourage subjects to utilize that information and will alter how a subject forms beliefs. Indeed, the focal information can potentially serve as an effective coordinating device. Over time, however, some subjects reduce their reliance on the supplementary information, leading other subjects to also find it less useful. Consistent with Assenza et al. (2013)

and Pfajfar and Zakelj (2014) , we find that providing subjects the opportunity to learn matters. Subjects in the Benchmark treatment are able to reduce their forecast errors substantially by altering their reliance on various pieces of information. However, with learning comes the opportunity for coordinating devices, such as the focal forecast error information, to fail.

It is worth emphasizing that had we only ran one repetition per session, it would have been easy to conclude that behaviour across the information treatments is significantly different. A second repetition shows that forecasting behaviour changes with learning and the relative benefits of focal information reduce. Switching between forecasting rules has been well-observed by Hommes (2011) and Pfajfar and Zakelj (2014) over long horizons, but these experiments are typically conducted as one long repetition. Given that this is a coordination game that rewards forecast accuracy, subjects will mimic the behaviour they believe is driving historical aggregate behaviour and can result in long stretches of non-rational forecasting. Stationary repetition allows subjects to more effectively 'learn away' suboptimal forecasting rules that may have emerged in the beginning of the session when they experimented with various strategies.

Our experiment demonstrates the ability to influence expectations and overall economic activity. Practically speaking, policy makers can encourage constant gain learning by making forecast errors more salient. This can be accomplished by encouraging both firms and households to update their expectations more frequently and communicate effectively current inflation and demand statistics in such a way that is retained by the general public. Financial planning and commercial bank websites can play an important role by providing an application that allows individuals to track their expectations and forecast accuracy over time.

More generally, central bank communication is an area where laboratory experiments have the potential to be particularly insightful. Further experiments can shed light on what information subjects are more likely to respond to and coordinate on. Filardo and Hofmann (2013) have recently observed in the United States that while qualitative and calendar-based forward guidance of monetary policy has been effective at influencing interest rate expectations, communication of more complex threshold-based policies beginning in December 2012 are associated with increased volatility and disagreement in financial markets. This is just one example where the clarity and ease of understanding of information can lead to better coordination of expectations. Finally, our treatment variation in information was conducted across different groups. Instead, one could consider an experiment where focal information is presented unexpectedly. How and whether subjects would respond to new information after learning to coordinate their beliefs with others is an open question that is particularly relevant in a world where policy makers are increasingly communicating to the public.

6 References

Adam, K. (2007). Experimental evidence on the persistence of output and inflation*. The Economic Journal, 117(520), 603-636.

Assenza, T., Heemeijer, P., Hommes, C., Massaro, D. (2013). Individual expectations and aggregate macro behavior. Working Paper No. 13-016/II. Tinbergen Institute Discussion Paper.

Blume, A., Gneezy, U. (2000). An experimental investigation of optimal learning in coordination games. Journal of Economic Theory, 90(1), 161-172.

Demertzisa, M. Viegib, N. (2008). Inflation targets as focal points. International Journal of Central Banking, 4(1), 55-87.

Demertzis, M. Viegi, N. (2009). Inflation targeting: a framework for communication. The BE Journal of Macroeconomics, 9(1).

Duffy, J. (2012). Macroeconomics: A Survey of Laboratory Research. Working Paper.

Figlewski, S., Wachtel, P. (1981). The formation of inflationary expectations. The Review of Economics and Statistics, 1-10.

Filardo, A. Hofmann, B. (2014). Forward guidance at the zero lower bound1. International banking and financial market developments, 3, 37-53.

Keane, M. P., Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. The American Economic Review, 714-735.

Kryvtsov, O. Petersen, L. (2014). Expectations and Monetary Policy: Experimental Evidence. Working Paper.

Mehta, J., Starmer, C., Sugden, R. (1994). Focal points in pure coordination games: An experimental investigation. Theory and Decision, 36(2), 163-185.

Milani, F. (2007). Expectations, learning and macroeconomic persistence. Journal of monetary Economics, 54(7), 2065-2082.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. American Economic Review, 85(5), 1313-1326.

Orphanides, A., Williams, J. C. (2005). The decline of activist stabilization policy: Natural rate misperceptions, learning, and expectations. Journal of Economic dynamics and control, 29(11), 1927-1950.

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

Pfajfar, D. Santoro, E. (2013). News on Inflation and the Epidemiology of Inflation Expectations. Journal of Money, Credit and Banking, 45(6), 1045-1067.

Pfajfar, D. Zakelj, B. (2012). Uncertainty and Disagreement in Forecasting Inflation: Evidence from the Laboratory. Working Paper.

Pfajfar, D. Zakelj, B. (2013). Inflation Expectations and Monetary Policy Design: Evidence from the Laboratory. Working Paper.

Pfajfar, D., Zakelj, B. (2014). Experimental evidence on inflation expectation formation. Journal of Economic Dynamics and Control, 44, 147-168.

Roos, M. W. Luhan, W. J. (2013). Information, Learning and Expectations in an Experimental Model Economy. Economica, 80(319), 513-531.

Sims, C. (2003). Implications of rational inattention. Journal of Monetary Economics, 50(3), 665-690.

Tversky, A., Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive psychology, 5(2), 207-232.

Woodford, M. (2003). Interest and Prices: Foundations of a theory of monetary policy. Princeton University Press.