Learning Integrated Inflation Forecasts in a Simple Multi-agent Macroeconomic Model

Blake LeBaron[1]*, Karen Smith[2]‡

[1]Brandeis University
[2] Drew University
May 2024

Abstract

This paper implements a model with a population of heterogeneous macro forecasters. Their objectives are to forecast output and inflation, both inputs in standard New Keynesian macro models. The model is implemented by first calibrating the agents to professional forecasters at the micro level. Model runs then try to replicate both the dynamics, bias and cross sectional heterogeneity of forecasts and the economy. These are done both in a model with static forecasters, and one where the forecasters are learning from each other in a social fashion. We find that expectations about the inflation process which conjecture near random walk behavior can be self-fulfilling, yielding inflation volatility and persistence on the order of magnitude of U.S. macro data. However, our forecasting populations often fall short of the heterogeneity of predictions from survey data. In some cases, monetary policy can be used to shift the model from its volatile/persistent equilibrium over to a more stable, strongly mean reverting inflation rate.

Keywords: forecasting, heterogeneity, agent-based macroeconomics, behavioral macroeconomics

*Corresponding author: Blake LeBaron, Brandeis University, 415 South Street, Mailstop 32, Waltham, MA 02453 - 2728, blebaron@brandeis.edu.

†Drew University.

‡Acknowledgements: We are grateful to Herbert Dawid (the editor) and two anomymous referees for useful suggestions. Also, Alan Isaac and seminar participants at the Eastern Economic Association meetings in Feb 2024 provided useful feedback.

1 Introduction

Heterogeneity has become a standard for modeling in macroeconomics and finance. Agent-based models have played a role as a core tool in trying to understand the endogenous dynamics between agents and their expectations, in either macroeconomic or financial settings. This paper analyzes the dynamics of a multiple agent system designed to reproduce features from a standard New Keynesian macro model built from a set of agents learning about their own endogenous dynamics. It is close in spirit to the evolutionary learning mechanisms used in Arifovic, Bullard & Kostyshyna (2012) (ABK), and more recently in Arifovic, Grimaud, Salle & Vermandel (2020). They refer to their learning as "social learning." This concept was most clearly defined in Vriend (2000) with origins going back to Arifovic (1994). It refers to agents learning from each other as opposed to learning individually. It is also related to the concept of "epidemiological expectations" which has recently been surveyed in Carroll & Wang (2022). We also follow much of the same structure used to study learning in a macro setting from Bullard & Mitra (2002). The distinction in our approach is to represent heterogeneity with a large and rich population of simulated forecasts making predictions on both inflation, and future output. The rule structure is heavily influenced by standard time series models that either line up with theory or data in important ways.

We extend the work of ABK in several directions. First, we simplify the learning structures by abandoning the genetic algorithm (GA) learning mechanism. GA's are often powerful tools for generating novelty in a learning environment, but their complexity can often hide important features. We move to a very transparent system with a rich but finite set of forecasting rules. Rule selection is still similar to ABK, but we have complete control over the set of rules used which gives us more power to scrutinize the interactions between expectations and the time series they endogenously generate. In the basic New Keynesian framework we show this interaction is very tight. ABK condition forecasts on a stochastic real interest rate which is the underlying state variable in the economy, and its only shock. We shift our expectations to inflation and output themselves since there is a much stronger case that these are observable.

Our choices for forecasting agents are driven by model structure, simplicity, and comparisons to professional forecaster information. Our hope is that a successful forecaster identified here will yield insights about individual and professional forecasts collected on macro economic variables. We show that simple autoregressive (AR) forecasters lead to data consistent with their expectations and keep the simulated model close to some theoretical predictions. Another interesting distinction between our simplified agent-based system and ABK is that our results are often much tighter in terms of simulation outcomes. Our cross sections of very long time series runs show almost no variability in the estimated long run moments, a strong defense of ergodicity in most of our runs. Furthermore, we have very few parameter combinations generating mixed results in terms of stability. Models either converge or explode, but we rarely see both for the same parameter sets. This is probably a result of our very simple models, and very long simulation runs.

Our agent inflation predictions are extended to include integrated moving average processes (IMA). They capture the extreme persistence in U.S. inflation well, (Stock & Watson 2007) and (Pivetta & Reis 2007). It is generally a useful time series representation for many features that are common in macro series. In a recent discussion Franses (2020) shows how it is universally applicable, and tied to simple exponential filters. The connection between the IMA(1,1) and filtering methods is strong. This intellectual connection goes back to Muth (1960), but is emphasized in Hyndman & Billah (2003). The fact that the IMA does well is also related to trying to model trends and cycles in macro time series.[1] It also connects to results yielding an ARIMA(1,1,1) representation where an integrated series faces a slowly changing growth rate.

Using results from a common panel of macro predictors for the U.S. from the Survey of Professional Forecasters, we calibrate approximate time series predictors that line up with the professional forecasters in terms of both known forecast bias, and cross sectional variability.[2] We then put this population of forecasters together, and first see how they do as a group. Their predictive models are combined with a standard benchmark new Keynesian model to generate dynamics in output and inflation subject to monetary policy from a Taylor rule. We test whether these simple models can behave reasonably, and generate data that is at least consistent with their cross sections of predictors. In other words, is the model and the set of heterogeneous forecasters self-consistent? In our second stage we release the predictive models to allow them to learn over time using a combination of evolution and social learning where accurate forecasting rules are rewarded and replicated. We are again interested in their ability to generate macro features, and whether they continue to appear similar to actual populations of forecasters in the data in terms of dispersion and bias.

Our results agree with ABK in several areas. First, we replicate the learnability of the rational expectations equilibrium in the model for certain parameters and restrictions in our learning models. Agents generate a world which is self-consistent with their forecasting models. We also replicate the possibility of sunspot equilibria that are not efficient in terms of inflation volatility. In particular, we show that inflation

[1] A recent success of this approach is Bianchi, Nicolo & Song (2023).

[2] See (Smith 2022) for many references and examples, and Clements, Rich & Tracy (2023) for a recent survey. Also, see Branch (2004) who shows that a panel of agents can be aligned with some of the forecaster data.

self-confirms a near integrated process which is not far from what we see in the actual macro series.[3] We are able to control whether the agents move to this equilibrium through both policy, and the time ordering of our learning process. Similar to ABK, we explore the stability properties of the model and its connections to E-stability. In some specifications we get much stronger convergence than E-stability would predict which is similar to ABK. However, in other model specifications our convergence is exactly on target with E-stability predictions, a strong divergence from ABK.

One of our results speaks very directly to policy. We show that monetary policy can be used to select the equilibrium dynamics, and can pull the model out of the integrated learning equilibrium. In doing this, a central bank would greatly reduce the long term volatility of observed inflation rates, as it drives these forecasting rules out of the population. This learning population dynamic aligns with commentary about making sure that long term expectations for inflation do not become entrenched in consumer/worker beliefs.

At the aggregate level our forecasters generate reasonable inflation time series. When we turn to the micro level, and compare with the data on professionals some interesting distinctions emerge. First, we observe much less heterogeneity in our forecasters than in the set of professionals. We can get closer to the professionals by adding noise to our agents, but the noise level needed seems very large, and it destroys some of the coherence in the aggregate forecast behavior. We also show that our forecasters individual errors are mostly uncorrelated, a departure from most studies of professionals. This difference shines a very sharp light on professional forecaster behavior, since our mechanical forecasters, who optimize mean squared error from inflation targets, will have a difficult time generating this suboptimal aspect of human behavior.

We have strong connections to the heterogeneous agent modeling world. In macroeconomics this literature is quite extensive, and Dawid & Delli Gatti (2018) and Branch & McGough (2018) are two excellent surveys. Our use of a finite set of models as opposed to an open ended learning mechanism is closely related the heterogeneous agent model style (HAM) pioneered in Brock & Hommes (1997), but extended to thinking about many strategies theoretically in Brock, Hommes & Wagener (2005) and computationally in LeBaron (2021). In macroeconomics the modeling of heterogeneity using small sets of predictive models, and discrete choice learning dynamics, has several early examples in Branch & McGough (2009), Branch

[3]This result has a strong connection to Tortorice (2018) that considers a single agent learning setup where the agent is unsure whether the observed data is integrated or not. In his structure learning about stationarity is able to generate more realistic outputs from a standard DSGE model. Another related paper is Evans, Honkapohja, Sargent & Williams (2013) who examine a two predictor learning experiment in a cobweb model. They find that for certain parameter combinations there is a strong appeal to a model with a changing nonstationary parameter.

& Evans (2011), De Grauwe (2011), and Anufriev, Assenza, Hommes & Massaro (2013).[4] We diverge from these important early papers by having a larger set of predictors, based on observable series, and tied to time series model predictions. Also, we stay closer to ABK by using a tournament selection mechanism as opposed to a discrete choice system.

Recent research has worked to calibrate the predictors to results from human experiments as in Assenza, Heemeijer, Hommes & Massaro (2021) and surveyed in Hommes (2021). Two important connections from the social learning world are Hommes, Makarewicz, Massaro & Smits (2017) and Anufriev, Hommes & Makarewicz (2019). The first lines up with experimental data in a New Keynesian setting, and the second aligns the GA based learners with data from learning to forecast style experiments. This line of research of fitting computational agent-based models to results from experiments has its fundamental origins in Arifovic (1996), and the general field has recently been surveyed in Arifovic & Duffy (2018).

Section 2 introduces the macro time series sample along with the forecaster data. Several stylized facts are presented along with time series estimates for the models that will eventually be used in the agent-based simulation. Section 3 presents the New Keynesian macro model, and also introduces the structure for the agent-based expecations. Section 4 presents the results from the various agent-based computer simulations, and section 5 concludes.

2 Forecaster calibration

2.1 Long series

We start by looking at the properties of traditional long macro series for the U.S. downloaded from the Federal Reserve Bank of St. Louis (FRED). Inflation uses both the CPI measure, and the GDP deflator which are available back to 1947. Output is measured with the output gap, which is GDP divided by potential GDP. This is available at quarterly frequency back to 1949. All these series end with the fourth quarter 2022. All series are quarterly, seasonally adjusted, annualized, and measured in percentage units. Figure 1 plots the two series. Both show strong persistent patterns. The presence of regular business cycles is clear on the output panel. Inflation is less regular, but takes several long deviations from its mean.

Figure 2 displays the autocorrelation patterns for these two series, and their first differences. Also, the theoretical autocorrelations for an autoregressive model with one lag (AR(1)) are matched to the first

[4] The latter of these considers a heterogenous population of forecasters with up to 12 different types, and appeals to theoretically extending this to an infinite population of forecasters.

order autocorrelation and presented for comparison. For inflation we see a large amount of persistence as indicated by positive autocorrelations that decay more slowly than the AR(1) model. The first difference of the series displays a negative autocorrelation at the first lag, dropping to near zero after this. This graphical pattern is consistent with an integrated process whose first difference follows a moving average with one lag, or an IMA(1,1) process which we will be using extensively. Output appears slightly more complicated with autocorrelations below the simple AR values, and small, but persistently negative first difference autocorrelations which could indicate over differencing.

Table 1 gives summary statistics and results from our benchmark forecasting methods for the three long FRED series. For the two inflation series the means are both near 3. The standard deviation is much higher for the CPI series at 3.28 versus 2.57 for the GDP deflator. We will generally stay with the deflator as being a more consistent reference series over our long range data.

Three simple models are compared in terms of forecast performance, an AR(1), IMA(1,1), and a naive benchmark random walk (RW). The last model uses the value at t as a forecast for $t+1$. All of the estimates are done in sample, and are an optimistic estimate of forecast performance. We are only interested in showing the reasonableness of these simple time series models, not their expected values in practice. In all cases we compare the properties of forecast errors as defined by

$$e_{t+1} = x_{t+1} - f_{t+1,t} \tag{1}$$

where x_{t+1} is the forecast target, and $f_{t+1,t}$ is the forecast for $t+1$ given time t information. e_{t+1} is our forecast error. The rows labeled RMSE show the root mean squared forecast error for the forecasting models. All forecasts are one quarter ahead. For inflation the performance of the two time series models is relatively similar, and they beat the sample standard deviation by a large amount. In all cases the RW performed worst. However, it is not too far from the other models.

We also estimate two standard measures of forecaster performance, bias and error autocorrelations. Bias is measured as the sample mean for e_{t+1}, and autocorrelation estimates the first order autocorrelation for e_{t+1} These are both standard performance measures, and for a good predictor, both should be near zero. Our mechanical forecaster errors are generally unbiased. The autocorrelations for the forecast errors are small for the IMA model, but relatively large and negative for both the AR and the RW forecasts for inflation, indicating model misspecification. Figure 3 shows the autocorrelation patterns for the three forecasts for inflation (GDP deflator). The IMA model leaves very little indication of error correlation, but the other two show strong negative correlations at the first lag. This indicates some model misspecification for the latter two models, further strengthening the case for the IMA being a reasonable time series model for inflation.

Values for output show a similar level of variability in the standard deviation as the GDP deflator series. Also, it is a little biased as evident from the mean of -1.51. The set of three forecasts are again used, and this time the performance of all three is similar. The AR(1) performs best, but the random walk, and the IMA are the same. Figure 4 displays forecast error autocorrelations for output. All three models show similar patterns of first positive, and then consistent negative autocorrelations. This distinct shape suggests that something is wrong with all three models here. There is probably value in a longer lagged AR, or a more complex model, but we will not undertake this search. Our objective is to stay with very simple and parsimonious candidates for our agent-based simulations.

Table 2 presents the estimated parameters for the two time series models. The AR model picks up strong persistence in all the series, but it should be noted that the parameter stays a good distance away from 1. The MA parameter, θ, is estimated as less than zero for the two inflation series. Section 3 will show the direct connection between the IMA model and an exponential filter with a gain of $1 + \theta$. Our inflation estimates correspond to filtering gain parameters in the range of $0.5 - 0.6$. Finally, output shows greater persistence, but the MA parameter is near zero. This suggests that when confronted with the possibility of a true random walk the output series would take this as opposed to a noisy random walk.

2.2 Forecasters

We have two objectives in this section. First, to provide some supporting evidence for reasonable mechanical time series predictors that will stand in for professional forecasters in our heterogenous predictor simulations. Second, we will also document a small set of stylized facts about near term professional forecasts that we will be trying to replicate in our agent-based simulations. For this we add time series drawn from the Survey of Professional Forecasters (SPF) available from the Federal Rerserve Bank of Philadelphia. This is a common data set used to evaluate the expectations generated by professional forecasters. We will examine their inflation forecasts for eventual comparisons with our simulated agent-based expectations. These series are available quarterly from 1983, and we end with the fourth quarter in 2022. The series record several different horizons, but we concentrate only on the one quarter ahead forecasts.

The first two rows of table 3 present summary statistics for inflation in the subperiod where we have professional forecaster data. The underlying series here come from the FRED database. The mean of about 2.8 for both series corresponds to this period being known as a time of low and relatively stable inflation. The standard deviation is also significantly lower than the longer samples shown in table 1. It is also reduced significantly by using the trimmed series.

Next we turn to forecaster comparisons. We continue to use only a one quarter ahead forecast. Longer range forecasts are interesting, but our modeling calibration probably lines up best with short term forecasts. Also, our time series forecasts have their best chance at competing in this realm. In forecasting, timing is critical, and it is important to note that we are using a forecast which is generated in the middle of the current quarter (t), and directed at one quarter ahead (t+1). This forecast survey is done almost one month into a quarter, so forecasters have already seen 1/3 of the current quarter's data. This gives it a slight unfair disadvantage over the time series models used for comparison since they use the entire quarter t realisation.

Returning to table 3 we see the RMSE for the same benchmark forecasting models estimated in sample along with the mean forecast from the professionals, labeled SPF. All models are relatively close to the professionals in terms of RMSE performance. The naive RW is always the worst, and for the raw CPI series RMSE is even larger than the standard deviation of the series. Turning to bias we see little evidence for strong bias in any of the forecasts. The forecast error autocorrelations tell a slightly different story. Both the naive RW model, and the professionals show a large amount of autocorrelation in their forecast errors. It is interesting to note that while similar in magnitude these autocorrelations have opposite signs. The autocorrelations patterns for the IMA forecast errors countinue to be small. There is a slight blip for the CPI series with a positive 0.15 value, but this is not large relative the the correlations we see from the professionals. Figure 5 plots the autocorrelations for the professional forecast errors for both the CPI and trimmed CPI series. In both cases there is a strong positive correlation in the forecast errors at one quarter, and persisting out 2-3 quarters depending on the series.

The last row of table 3 looks at the cross sectional dispersion of the professional forecasts. The Survey of Professional Forecasters reports several numbers for these. We will use the simple cross sectional standard deviation. We want to normalize this to line up with our model simulations. A first thought might be to normalize relative to the mean inflation rate, but since this is so close to zero in the later parts of the data, it seems to make more sense to use something else. We will normalize relative to the sample standard deviation. This value is reported in the last line. For the raw series it is about 0.34, and for the trimmed series it is 0.61. The numerator in these estimates is exactly the same. The difference is driven by the much lower standard deviation in the trimmed series. For our results we will take the 0.34 value as our target, since this is using the series that is the forecasters' target.

This section presented evidence from both observed inflation series, and professional forecasters. It shows that the AR(1) and IMA(1,1) should be considered reasonable models when compared with professional forecasters in the world of short term inflation forecasts. They are on the same order of magnitude in terms of RMSE, and their error autocorrelations are better (smaller) than those of the professionals.[5] We also presented two other facts on the professional forecasters. Even for short horizon forecasts, their errors show large autocorrelations. These seem strange since it is a common indicator of inefficiency in a forecast. Finally, they also generate a large amount of cross sectional dispersion. We will eventually compare this with simulated agents. One of our goals is to see how difficult it will be to replicate this dispersion.

3 Forecasters in a simple New Keynesian model

3.1 Base model

The model we use for the dynamics of output and inflation comes from Woodford (2003) and is studied extensively in a learning environment in Bullard & Mitra (2002). This is the basic structure that is used in the social learning model of Arifovic et al. (2012) (ABK), and we follow much of their notation. It is a standard 3 equation New Keynesian model with a single stochastic state variable in the form of a changing real interest rate. The familiar three equation system is given by,

$$z_t = z^e_{t+1} - \sigma^{-1}(r_t - \pi^e_{t+1}) + \sigma^{-1} r^n_t \tag{2}$$

$$\pi_t = \kappa z_t + \beta \pi^e_{t+1} \tag{3}$$

$$r_t = \phi_\pi \pi_t + \phi_z z_t, \tag{4}$$

with the real interest rate following an AR(1) process as in,

$$r^n_t = \rho r^n_{t-1} + \epsilon_t, \tag{5}$$

[5]Since the time series models are not out of sample, and use full quarter t information, they probably should fall slightly behind the professionals in a fair race. It is not our objective here to make exact comparisons, but to show they are reasonable stand in models for our agent-based forecasters.

and ϵ_t is drawn from a normal, $N(0, \sigma_\epsilon^2)$. Output is z_t in percentage deviation from potential output, the output gap, and π_t is inflation as a deviation from a steady state inflation rate. Variables shown as x_{t+1}^e are subjective expectations of time $t+1$ variables using only information known up to time t, and sometimes only $t-1$. To simplify notation we will follow this convention,

$$x_t^e = E_{t-1}^*(x_t) \tag{6}$$

$$x_{t+1}^e = E_{t-1}^*(x_{t+1}) \tag{7}$$

where $E^*()$ represents the subjective expectation of our agents using mechanical time series forecasts. Equation 4, the Taylor rule, is the standard feedback from the observed outcomes to the policy maker's nominal interest rate response.

This is the timing considered by ABK, and we will stay with this format to keep our results closely comparable to theirs. It does bring up a few simultaneity issues which need to be addressed in an agent based model. Most of these are related to the timing in the Taylor rule used for interest rate policy. As written here interest rates are set given contemporaneous information on inflation and output. This has been criticized as not being realistic, and other formulations consider lagged information, contemporaneous expectations, or other expectation formations as Taylor rule inputs. These may have varying impacts on the dynamics of the model and the ease of tranfering it into a computational agent-based form.[6]

In setting up the simulation we will need to break down some of the sequencing and simultaneity of the basic model. We can write the linear system as follows with $y_t = (z_t, \pi_t)$:

$$y_t = A_0 y_t + A_1, \tag{8}$$

$$A_0 = \begin{bmatrix} \sigma^{-1}\phi_z & \sigma^{-1}\phi_\pi \\ \kappa & 0 \end{bmatrix}, \tag{9}$$

$$A_1 = \begin{bmatrix} z_{t+1}^e + \sigma^{-1}(r_t^n + \pi_{t+1}^e) \\ \beta\pi_{t+1}^e \end{bmatrix}. \tag{10}$$

[6] One of the most intuitive and easiest to implement from an agent-based perspective is to use contemporaneous expectations π_t^e and z_t^e in the Taylor rule. This and many other possibilities are covered in detail in Bullard & Mitra (2002).

This can be solved for the current output and inflation levels from

$$y_t = (I - A_0)^{-1} A_1. \tag{11}$$

Setting A_1 requires knowledge of r_t^n the exogenous shock, and also the one period ahead expectations, which in our model are determined from $t-1$ information. This is all can done easily in a computational agent-based world with $t-1$ values of all variables [7] The exact modeling sequence is given by:

1. Starting at time t determine the natural rate, r_t from its past value and the current shock.
2. Now use this along with $t-1$ information to generate $t+1$ expectations for the matrix A_1.
3. Use equation 11 to get time t values for the state variables.

The model can also be lined up to a rational expectations steady state where the perceived law of motion (PLM) lines up with the actual law of motion (ALM). One can follow Bullard & Mitra (2002) and write the system as,

$$y_t = B y_{t+1}^e + \chi r_t^n \tag{12}$$

$$B = \frac{1}{\sigma + \phi_z + \kappa\phi_\pi} \begin{bmatrix} \sigma & 1 - \beta\phi_\pi \\ \kappa\sigma & \kappa + \beta(\sigma + \phi_z) \end{bmatrix} \tag{13}$$

$$\chi = \frac{1}{\sigma + \phi_z + \kappa\phi_\pi} \begin{bmatrix} 1 \\ \kappa \end{bmatrix} \tag{14}$$

Assume that all agents follow the following PLM

$$y_t = \alpha + c r_t^n, \tag{15}$$

and expectations given by,

$$y_{t+1}^e = a + c\rho r_t^n. \tag{16}$$

[7] If the modeler were willing commit to $t+1$ expectations as linear in current output and inflation, then the model could be further compacted, but we cannot make this assumption.

The actual law of motion then follows,

$$y_t = Ba + (Bc\rho + \chi)r_t^n. \tag{17}$$

The MSV solution aligning the actual and perceived laws of motion yields

$$y_{t+1}^e = \bar{a} + \bar{c}r_t^n, \tag{18}$$

with

$$\bar{a} = 0, \quad \bar{c} = (I - \rho B)^{-1}\chi. \tag{19}$$

This formulation gives a simple analytic answer to a learning equilibrium in the model. It also shows that since (z_t, π_t) are linear functions of r_t^n they will also follow the same AR(1) followed by r_t^n.

3.2 Stability

Bullard & Mitra (2002) show that a necessary and sufficient condition for a rational expectations equilibrium to be unique and stable on certain learning dynamics (E-stability) the following condition must hold,[8]

$$\kappa(\phi_\pi - 1) + (1 - \beta)\phi_z > 0. \tag{20}$$

In the social learning environment used here (or in ABK), it is not clear if this condition will guarantee stability. An interesting result from ABK is that for some parameter values where this condition is violated, they still see some convergence. Also, it should be noted that for the case where $\phi_z = 0$ the condition corresponds to the "Taylor principle" in which stability requires that central banks respond more than 1-1 to changes in actual or perceived inflation.

3.3 Expectations

Our objective is to replace the genetic algorithm (GA) learning system of ABK with a simpler, more tractable, learning approach. Our methodology will implement some expectational models inspired by our earlier empirical observations. Our mechanical forecasters will only be approximations to true market dynamics, but we will test these on their generated data, and show that often their forecasting mistakes are small, and

[8] They actually show that this condition applies in many different Taylor rule formulations.

uncorrelated.[9]

Our first and simplest set of expectations uses an AR(1). For inflation this would look like,

$$\pi_t^e = g_j \pi_{t-1}, \quad \pi_{t+1}^e = g_j^2 \pi_{t-1}. \tag{21}$$

In the learning process, agents are not assumed to know the full dynamics of π_t, but in our first base case they attempt to learn the autoregressive parameter, g_j. The different forecasting models in a given family are indexed by j. The population of agents is assumed to sit on different values of g_j, and learning involves population weight shifting over time to better predictor rules. The set of values on this coefficient are allowed to vary on the interval, $g_j \in [0,1]$. Both 0 and 1 are allowed, corresponding to the limits of white noise, and a random walk. Similar logic applies to predicting output yielding,

$$z_t^e = g_j z_{t-1}, \quad z_{t+1}^e = g_j^2 z_{t-1}. \tag{22}$$

The AR(1) provides a useful test model to check our results. Also, equation 18 suggests this can be a good approximation in the theoretical ALM for the model.

In our second set of experiments we continue to hold z_t to the AR(1) format with learning. However, we acknowledge the empirical evidence that an IMA(1,1) may be a good approximation for inflation dynamics as it performs reasonably well in comparison with professional forecasters in near term forecasting experiments. Agents make predictions as if inflation follows,

$$\pi_t - \pi_{t-1} = \theta \epsilon_{t-1} + \epsilon_t. \tag{23}$$

ϵ_t is a standard white noise process with $E(\epsilon_t) = 0$ and variance σ_ϵ^2. The above process can be written in terms of forecasts as,

$$E_{t-1}(\pi_t) = \pi_{t-1} + \theta \epsilon_{t-1},$$

$$\pi_{t-1} = \pi_{t-2} + \theta \epsilon_{t-2} + \epsilon_{t-1},$$

$$\pi_{t-1} = E_{t-2}(\pi_{t-1}) + \epsilon_{t-1},$$

[9] The aggregation of expectations as performed here, and in many other papers should always be viewed cautiously as their application in a standard New Keynesian macro model involves a large number of assumptions (see Branch & McGough (2009). Brock & Haslag (2015) is an example of detailed micro level analysis of inflation with heterogeneous forecasters.

and

$$\epsilon_{t-1} = \pi_{t-1} - E_{t-2}(\pi_{t-1}).$$

Substituting into the previous equation gives,

$$E_{t-1}(\pi_t) = \pi_{t-1} + \theta(\pi_{t-1} - E_{t-2}(\pi_{t-1}))$$

finally, with our notation,

$$\pi_t^e = -\theta\pi_{t-1}^e + (\theta+1)\pi_{t-1}, \tag{24}$$

which corresponds to a standard exponential filter with gain parameter equal to $(\theta + 1)$ or,

$$\pi_t^e = (1-h_j)\pi_{t-1}^e + h_j\pi_{t-1}. \tag{25}$$

Under the assumption of an integrated process, the expectation at $t+1$ given $t-1$ information is

$$\pi_{t+1}^e = \pi_t^e. \tag{26}$$

Again, h_j is the model parameter that agents will be trying to learn.

As in a typical social learning model it is assumed that there is a population of forecasting agents following models with parameters g_j and/or h_j depending on the simulation. A given fraction of the population, or weight, sits at each parameter, and final forecasts are a weighted average of each of the individual predictors. These become the aggregate expectations used in the model dynamics. For example,

$$\pi_t^e = \sum_{j=1}^{J} w_j \pi_{t,j}^e \tag{27}$$

represents the aggregate one period ahead inflation forecast. The variable w_j represents the weight on predictor j, and J is the total number of predictive models. The individual predictors, represented with $\pi_{t,j}^e$ are not updated every period. They acknowledge that real forecasts are "sticky", and they are updated each period with a fixed update probability, p_{update}. This both reflects results from actual forecasters, and individual surveys, as well as making sure the learning algorithms don't try to synchronize with agent forecasting behavior.

3.4 Learning dynamics

The dynamic learning model is computational, but designed to be simple and tractable. The learning framework is a simplified version of other social learning methods. It is a bridge between the more complex, open ended genetic algorithm (GA) models as used in ABK, and simpler heuristic switching models (HSM) in which agents chose the best model from a small finite set of possibilities.[10] The model attempts to construct a large, albeit finite, space of models that span most reasonable forecasting possibilities.[11]

Rule selction is closely aligned with the GA space which uses a pairwise tournament selection as opposed to the discrete choice methods common in HSM modeling. In each period each rule is paired with another rule. Their performance is compared, and a fraction of the weight from the weaker rule is transferred to the stronger rule. This can often be a small set of models, on the order of 2-4. We follow this approach here, enumerating models over the single forecast parameter indexed by j, g_j and h_j.

As in most learning models, model performance, or fitness (f), is measured as mean squared error (MSE) off the forecast target,

$$f_j = \frac{1}{T}\sum_{t=1}^{T}(\pi_t - \pi_t^e)^2. \tag{28}$$

This flat sample average is not feasible, so it is estimated in real time with a constant gain recursive algorithm that would map directly into more complicated forms from recursive least squares learning,

$$f_{t,j} = (1 - g_f)f_{t-1,j} + g_f(\pi_t - \pi_t^e)^2, \tag{29}$$

for all models j, at each period t.

In each period the learning algorithm sweeps through all rules j, and for each one chooses a random partner, k, from the set of rules. If $f_{t,k} < f_{t,j}$, then a fraction of the weight on rule j is moved to rule k. This fraction will be fixed to f_{switch} for all runs. For example, if 0.20 fraction of the population is on rule 1, and rule 2 is estimated to be better, than $f_{switch} * 0.20$ weight is moved from rule 1 to rule 2. This learning dynamic stays close to the tournament selection methods of ABK. It also is related to the discrete choice dynamics used in many agent-based models. We find this method to be a good combination of simplicity and effectiveness for our highly stylized models.

[10] The origins of this methodology go back to Brock & Hommes (1997) and Lux (1998).
[11] An example of this strategy in finance is LeBaron (2021).

4 Results

4.1 Parameters, benchmarks, and calibration

The model is now ready to be implemented. For all cases we will be using the three equation model described in section 3. We follow the expectation timing, and sequential nature described in that section. Table 4 summarizes the parameter values used. The first set, $[\beta, \sigma, \kappa]$, represent the core NK parameters and are drawn directly from ABK, who get these from an early benchmark in Woodford (2003). The most critical parameters in the model are the Taylor rule parameters, ϕ_π and ϕ_z. We will be using many different values for these, as we are interested in how the learning model interacts with changes in monetary policy. We will concentrate on the pair $(\phi_\pi, \phi_z) = (1.0, 0.05)$ for many runs since they are in a reasonable range, and generate some interesting results. We set the autoregressive parameter for the shock, ρ to 0.35 which is again drawn from these earlier papers. We chose to calibrate the size of the shocks in the process to bring the variability of inflation and output close to that in the U.S. time series. This sets $\sigma_{r^n} = \sigma_\epsilon / \sqrt{1-\rho^2} = 0.70$. Our calibration experiment for this parameter will be presented shortly. Our simulations are run for $T = 20,000$ quarters, and we find this gives the models plenty of time to settle into a steady state. When we present cross sectional averages and standard deviations over several runs we use a cross section of $N = 25$.

The parameter range for our AR(1) forecasters is set to $[0, 1]$, and there are 101 rules spread uniformly across this range with a step size of 0.01. It is important to note, that we include values of 0 and 1, which correspond to a white noise, and random walk predictors, respectively. We are not considering either explosive, or negative parameters in the AR model. For the IMA model the gain parameter is allowed to vary between $[0.05, 1]$ again with a step size of 0.01. At the largest extreme, it is also a pure random walk. On the low side, it is not allowed to go to zero, since this implies a constant predictor.

The fraction of weight that is switched when an improved rule is detected is $f_{switch} = 0.05$. This implies a relatively small amount of shifting agents. It keeps the model from thrashing around too much, but still is large enough to get us reasonable convergence in most cases.[12] The value p_{update} defines the probability that agents update their actual forecasts each period. It will be held at 0.5 for all runs. Finally, g_f represents the gain value for the rule fitness updates. It also is fixed for all runs, and was determined to give a relatively smooth level of rule selection as models settle into steady states.

[12] This parameter might be considered an equivalent to the "intensity of choice" parameter in many agent-based models. It determines how strongly changes in fitness get reflected in changes in populations.

We also will be using our multiple forecasting models in different situations. We concentrate our work most heavily on inflation forecasts. For all runs the output forecast models will stay with the simple AR(1) mechanism. For inflation we will start with the AR(1) as our baseline, but will move to the IMA(1,1) model to see if it can generate more realistic series. Finally, we will explore situations where both types of predictors are active for inflation modeling. In these cases we want to know when and where the strategies might drive each other out of existence.

Our first set of simulations tests the model structure, and the learning components of the model to make sure the setup is functioning in the rational expectations benchmark. The model will be run using AR(1) forecasts for both inflation and output, as they are in the benchmark simulations of ABK. Agent expectations are set using equation 18 where all future expectations are linear functions of the natural rate, r_t^n, and the PLM and ALM for inflation and output both follow an AR(1). Table 5 gives the results for a set of runs using these criteria. Keep in mind that here, the expectations are not learned, and are fixed to their theoretical values. The table fixes the noise level $\sigma_r = 0.7$. We will show later that this is a reasonable level for the shock noise.

Since this table structure will be used in several future runs, it is important to describe it here. The first two columns correspond to the parameters in the Taylor rule, ϕ_π and ϕ_z. The third column, "explode", reports the fraction of runs which are exploding (unstable).[13] Since in this set of runs, the expectations are locked down, the fraction of explosive runs is zero by design. The fourth column is the "E-stability" condition described in equation 20. The sign of this condition will often be critical in determining the theoretical model conditions for stability and learnability. The next three columns report the first order autocorrelation $ACF(\pi_t)$, the first order autocorrelation of the first difference $ACF(\Delta\pi_t)$, and the mean forecasting parameter for the AR(1) inflation forecasts $\bar{g}(\pi)$. The AR(1) structure of the natural rate determines the theoretical value for the correlations and forecasting parameter for both inflation and output which should be $\rho = 0.35$. It is clear in these columns that across all runs, the inflation correlations, and the predictive models line up with this value. It is important to remember here that the predictive models are only watching the data go by, and lining up with it. They are not "live" in the sense of being used by agents. A similar pattern occurs for output in the columns labeled $ACF(z)$ and $\bar{g}(z)$ which show output autocorrelations, and the optimal rule forecast parameter. All estimates are again consistent with the AR(1) model, and for both output and inflation remain consistent across all Taylor rule parameters.

[13]We define explosive runs as either π_t yields a value of true in the python $isnan()$ function, or if the maximum absolute value of inflation is larger than 10. Most explosive runs take inflation into the nan range in python.

The final columns in table 5 report the standard deviations for the inflation and output series respectively. It should be noted here that the output standard deviation is a little high as compared to values from table 1, but most importantly the inflation variability is very low. Future experiments will be looking at these values very carefully.

The next two tables, 6 and 7, present the basic core results of the paper. In particular, table 7 presents the key direction for the paper in terms of how learning impacts inflation dynamics. In both cases the model is now made live in that agents predictive models are now used as inputs into the NK model. All runs use the calibrated value of $\sigma_r = 0.7$ for the exogenous noise level. Predictions feed into dynamics, which then feed back to model accuracy and agent behavioral adjustments. This is the typical feedback cycle in any forecasting agent-based model.

Table 6 begins with the AR model specification for both the inflation and output values. Turning first to output, z_t, there is a slight increase in the autocorrelations beyond the $\rho = 0.35$ to values closer to 0.40. This is true both in the generated time series and the optimal models. This increase in autocorrelation is even larger for the inflation series which pushes it closer to values ranging between 0.5 and 0.7 depending on the Taylor rule parameters. The time series values are self-consistent with the AR(1) forecasting parameters used by the agents.[14] The standard deviation for the inflation series is shown in the column labeled, σ_π. It is well below actual values for U.S. inflation variability as presented in table 1. Often by more than an order of magnitude.

Figure 6 shows the generated model time series for both inflation and output from a representative run for $\phi_\pi = 1.0$, $\phi_z = 0.05$, and $\sigma_r = 0.7$. Both series are generally uniform, and display few strong patterns. Their possible AR(1) persistence is not obvious in these very long time series, but it is present. Their autocorrelation patterns are presented in figure 7. The figure shows the ACF for output and inflation along with comparisons with a simple AR(1) lined up with the first order autocorrelation. We see in both cases the strong pattern of positive autocorrelations which is consistent with an AR(1) specification in both cases.

Summarizing, what we see in this first set of simulations is a model generating time series which follow the model specifications, and are self-consistent with the predictor models in use by the agents. The magnitude of the autocorrelations is slightly too high in both cases from where theory predicts they should be. Finally, the level of variability in the inflation time series is well below the actual series, and its model

[14] This increase in persistence for both inflation and output has an interesting connection to learning models developed in Hommes & Zhu (2016). They see a similar pattern, but in a very different learning framework.

specification of an AR(1) does not line up well with the actual U.S. time series. This leads us to the second experiment where we will change the inflation prediction model family to an integrated moving average.

In table 7 we present the simulation results for the model with agents now using an IMA(1,1) model for predicting future inflation. Output still uses the AR(1) forecasting framework. As mentioned in section 3.3 we implement this model with the mathematically equivalent, but more intuitive, form of an exponential moving average. The results in this table are at the core of this paper's message. We can see right away that they are dramatically different from learning in the AR(1) world.

The first key difference is that several of the model specifications explode. These all correspond to Taylor rules with the weakest inflation response, $\phi_\pi = 0.90$. In the stable cases we see something very different in the generated inflation time series. First, the inflation standard deviation as increased by almost a factor of 10 in most cases, often yielding a level of variability consistent with the actual series. For example, at $\phi_\pi = 1.0, \phi_z = 0.05$ the inflation standard deviation is 2.35 which is close to the actual data. The second key features which have changed are the autocorrelation patterns in the inflation series. In all cases the first order autocorrelation is almost 1, and the autocorrelation for the first differences are all negative, and in the range of -0.1 to -0.2 which are close to what was displayed in figure 2. They correspond to a filter gain parameter near 0.8 as reported in the column labeled $\bar{g}(\pi)$. The autocorrelations and parameters for output series are relatively unchanged from the AR(1) set of experiments which should be expected since the agents are using the same forecasting model.

Figure 8 displays the simulated time series for inflation and output. The output series looks similar to previous case, but inflation has changed dramatically. It now shows long persistent deviations from its mean that are much closer to what we see in actual inflation data, and are visually consistent with the integrated model the agents are now using for their forecasts. The increase in the level of inflation variability shows directly in the figure, and is consistent with the values from table 7.

Figure 9 turns to the time series autocorrelations to see how well they line up with the perceived model specification used by agents. Output patterns have not changed as they are consistent with the simple AR(1) process used for agent forecasting. As expected the features of inflation have changed. The level of inflation itself is extremely persistent with a value near 1 which decays slowly. The first difference of inflation shows a small negative autocorrelation of about -0.2 at the first lag, and then drops to zero after this. This is perfectly aligned with the IMA(1,1) used by agents, and shown to be broadly consistent with the data.

In terms of time series specification, we have inflation lining up with our previous results. However, they are not quite the same. For example, for $(\phi_\pi, \phi_z) = (1.0, 0.05)$ we see a first order autocorrelation in the first difference of inflation from table 7 of -0.173. Although not reported, the cross-sectional standard deviation of this estimate is extremely small with a value of (0.001). This comes from ergodicity, and our very long time series. This would correspond to an IMA model parameter of $\theta = -0.1785$ which compares to estimated values from CPI and the GDP deflator of -0.45 and -0.39 respectively. Their estimated standard errors are also small at (0.03), and (0.02) respectively. Although, not a formal statistical test, our model exhibits the correct qualitative properties of an IMA(1,1), but quantitatively it is far from the data. Also, remember that the corresponding gain level is $1 + \theta$, so as θ approaches zero from below our forecasting model becomes arbitrarily close to a random walk. These parameters generate much less autocorrelation in the real U.S. inflation series which can be seen by comparing figures 2 and 9. From these figures it is clear that model inflation is much more persistent than actual inflation. Another quick check is revealed by looking at figure 8. Inflation takes long swings, but some of the swings seem last for extremely long periods of time that seem to be on the order of nearly 400 quarters, or 100 years. So a kind of rough quick check shows quantitatively that the model generates too much persistence inflation relative to the data.[15]

All the early runs presented have used a fixed noise level for the exogenous natural rate shocks, $\sigma_r = 0.7$. This level was determined to line up the variability of inflation and output to the U.S. data presented in table 1. Table 8 shows how this noise level was determined. The previous run cross sections are simulated in this case using the IMA forecast specification from table 7. The Taylor rule parameters are fixed at one of the central values of $(\phi_\pi, \phi_z) = (1.0, 0.05)$. The noise level, σ_r, is then swept through values from 0.1 to 1.0. We then look at the variability as measured by the standard deviations for inflation and output in the final two columns. From table 1 we see both the inflation standard deviation (GPD deflator) and output estimated near 2.5. We look for a noise level that tries to get us reasonably close on the inflation target, since we are more interested in inflation dynamics. However, we do not want to move output to something unreasonable. This motivates our choice of $\sigma_r = 0.7$. It generates an inflation standard deviation of 2.3, and an output standard deviation of 3.1. We are pretty much on target with inflation, but slightly high on the output side. It is clear from the table that there is no single noise level that will line up both data targets. All future runs in the paper will stay with this level of shock variability.

It is important to summarize the results we have at this point since they are at the core of the paper. The basic model is simulated with agents using an AR(1) model for forecasting output. Their forecasts

[15] One needs to be careful with this comparison, since our benchmark for comparison is a nonstationary model, and our simulation generates stationary time series by design. Of course, in the actual data we do not know what the true process is for inflation.

for inflation use both an AR(1) model which is motivated by theory, but also an IMA(1,1) model which is motivated by the data. In both cases the models generate time series which are self-consistent with the forecasting model, either AR(1) or IMA(1,1). In the case of the AR(1) inflation forecast, inflation variability is well below the data, and the time series dynamics follow a simple AR(1) which is not consistent with the data. This changes with the IMA(1,1) specification, yielding time series which are much more realistic. In the remainder of the paper will explore many of the details about what is happening in the learning dynamics around this model.

4.2 Dual model learning

The results in the previous section suggest a mechanism for a self-consistent learning model generating realistic inflation dynamics. The model is still slightly contrived in that it forces the integrated model on the learning agents. They have no choice but to use something in its very specific class. In the extreme they could perceive a random walk for inflation (gain level of 1), or converge to a very long range mean (gain level near 0). There is no option for even considering other models. This section allows for agents to choose between the AR(1) and the IMA(1,1) forecast specifications using the ones that best fit their generated data.

Table 9 takes the previous simulation runs, but now agents are able to chose from a family of AR(1) forecasters or the IMA(1,1) model. It is a combination of the two previous forecasting experiments in one place.[16] The table presents many similar results to table 6 with output close to that from the basic AR(1) learning world. The table incorporates several extensions from the previous table. The column labeled "integrated" indicates the mean model weight put on the integrated forecasting rules. The values here are often close to zero which indicates agents have converged to the AR(1) forecast set, and its more stable inflation output. There is one exception. For the parameters of $(\phi_\pi, \phi_z) = (1.0, 0.0)$ the agents prefer the integrated forecast. Results in this case show a highly persistent inflation series, and a small, but negative autocorrelation for the first difference which is consistent with results from previous IMA learning experiments. The table prints out one parameter for the inflation forecast rules in $\bar{g}(\pi)$. This number varies depending on the dominant forecast rule in use. When this is an AR rule, it represents the AR forecasting parameter, and when it is the integrated rule, it is the gain from the integrated forecast. This makes these numbers directly comparable to the previous tables.

Some insight into what is going on can be seen in figure 10. This figure takes a single model run for

[16] One small technical difference here is that the range of AR models is now held below 1.0 in parameter space. This is because that parameter corresponds to a pure random walk, and that is already covered when the gain parameter is 1.0 in the set of filtering rules.

$(\phi_\pi, \phi_z) = (1.0, 0.05)$ and displays the final mean squared errors (MSE) for the two different forecasting models across their parameter ranges. The AR model is dominant here, so it is clear why most of the runs lock down on this option. The minimum MSE value is roughly consistent with the run averages presented in table 9 showing $\bar{g}(\pi)$ of 0.64.

If agents are started with a uniform prior across all the learning models, it looks likely they will be drawn to the basic AR forecasters. Is there any possible route to a convergence to the IMA model, and is there any possibility for some kind of path dependence in learning? This is explored in the next experiments which "burns in" the initial time series patterns using IMA learning only during the first 5000 quarters. At this point learning is opened up to either model and agents are free to drop back to the AR(1) model if it is shown to do better. It is a simple experiment to understand if the learning dynamics is globally attracted to one single model, or if there is only local attraction where agents can get stuck with the IMA model.

In table 10 we see the output of these experiments. The results change dramatically. Agents lock down on the IMA model and stay there as shown by the "integrated" column. The estimated and model parameters are very close to the original IMA forecasting parameters from table 7. Inflation is highly persistent, looks like an IMA(1,1) process, and is often close to the variability of the data as seen from the column labeled σ_π. The presence of the AR(1) forecast possibility has had little or no impact on the results.

Figure 11 repeats the previous comparison for the relative forecast performance. This is a much trickier plot to perceive. It is hard to see, but the integrated model is doing the best at a gain level near 0.6. The set of AR models does present a near competitor model at value of 0.99 which represents a near random walk for inflation, so it is not exactly choosing a stationary model here. The random walk and the IMA(1,1) are very close both in terms of performance since they are not too far from each other in model space.

These runs of the competing forecasting models along with the IMA burn in form a core benchmark model for our paper. They allow for all the interesting dynamics that integrated inflation forecasting gives, but its existence is there for interesting reasons. The agents have locked down on a potentially suboptimal, and volatility magnifying, forecasting system in the basic NK model. Instability is an endogenous feature of the interaction between forecasts and model dynamics. It is essentially a sunspot equilibrium in the usual macroeconomic sense. For these reasons we will examine some other details of this model run.

In figure 12 we plot histograms of where the forecast model weight sits for the different forecasts. Output reports the distribution around the AR parameter, and the inflation does the same for the gain parameter. This is done for one run with $(\phi_\pi, \phi_z) = (1.0, 0.05)$. Technically, the model also includes AR forecasters for inflation, but in this run the weights on all of these were zero, so plotting them was unnecessary. The graph shows relatively little forecast dispersion in both cases. The AR parameter is distributed tightly around a value slightly larger than 0.40, and the inflation gain is slightly more disperse, and distributed around a value slightly larger than 0.8. The picture is consistent with the 25 run summary in table 10 which gives values for these forecast parameters of 0.41, and 0.81 respectively. Figure 13 presents the time series of the weighted forecast parameters for inflation and output forecasts. The time series converge to their respective values after a relatively short period of time, and remain stable after this point.[17] There are no obvious large swings, so the dynamics of the market can be viewed as coming from fixed stable versions of the forecasting models, and not complicated dynamics across different models over time.

Given the importance of this set of simulation runs it is important to delve a little deeper into what we are seeing in terms of forecast rule performance. Table 11 takes the same set of runs from table 10 and checks the forecast rules using tests similar to those we applied to the professional forecaster data. The table looks at the inflation forecasts only. For clarity the table eliminates the three sets of parameters which generated explosive runs. We look at several critical forecast performance measures which include forecast bias, forecast error autocorrelation, and root mean squared prediction errors (RMSE). The sample moments are means estimated from the 25 runs, each using the last half of the 20,000 length time series. The standard deviation across the 25 simulations is presented in parenthesis for each value. The first two of these values measure the efficiency of our forecasts. Bias is represented by our model estimate of $E(e_{t,\pi})$ in the table. It shows no evidence for bias at all in the weighted forecast rules. The next column, labeled $ACF(e_{t,\pi})$ displays the autocorrelation of the model forecast errors. Correlations in these errors would also indicate some form of irrationality or inefficiency in our predictors. For all our Taylor rule specification these are also essentially zero. To an outside researcher testing these models they would see no evidence of irrationality. This is interesting since the optimal models are estimated evolutionarily, and not through an online recursive least squares process. This simple reinforcement learning process in the model yields very efficient looking predictors. Similar patterns hold for the output predictors in the columns $E(e_{t,z})$ and $ACF(e_{t,z})$.

The RMSE errors for the two predictors are presented in the RMSE columns, alongside the target standard deviation for comparison. Obviously, the standard deviation is essentially the RMSE using the sample mean. The models are doing generally very well in that the RMSE is much lower than the standard devi-

[17] Convergence may appear impressive in this figure, but it should be noted that this model is run for $20,000$ quarters. A quick eyeball estimate suggests the model takes almost 200 quarters, or nearly 50 years, to get the gain parameter above 0.7. In actual data 50 years of time series is available, but this suggests that model convergence would only start to appear in the late 20th century. The speed of learning in social learning models can be slow, and this issue remains an important research question.

ations. This should be as expected since the agents appear to have reasonable models and very long time series to use them with.

Comparisons with the survey of professional forecasters (SPF) from table 3 are interesting. CPI and trimmed CPI are both slightly less volatile than the simulations, and their time series dynamics are much more complex (if not nonstationary). Their bias is also low as in the simulations, but they generate a lot of forecast error autocorrelation. This is well known about forecasters, but its comparison to the simulated agents is quite stark here. Finally, they do produce larger RMSE forecast errors. This may occur because the nature of their forecasting problem is much more difficult than in a basic simulated NK model facing only one exogenous shock. However, the fact that their errors are so persistent, even at the short horizon, suggests this feature may be special to human professional forecasters.

In all cases our cross sectional standard deviations that are displayed in parenthesis in table 11 show very low dispersion across runs. This confirms the precision with which we are estimating the sample moments, and it also backs up assumptions about ergodicity in our runs. For example, the estimated forecast error autocorrelations for $(\phi_\pi, \phi_z) = (1.0, 0.05)$ is 0.009, and in this case the standard deviation of (0.008) shows that the runs are comfortably centered around zero in the cross sectional distribution. Also, our inflation volatility estimates are very reliable. For the same policy parameters we show a volatility level of $\sigma_\pi = 2.37$, with a cross sectional standard deviation of (0.182), showing that the inflation volatility levels are very consistent across all 25 runs. As a practical matter, the cross sectional simulations are probably unnecessary given the strong ergodic properties of these models.

4.3 Model stability

We have generally been dealing with issues of model stability in this paper. In this section we look at these in more detail. We have seen two parts to this problem. First, we have seen that some of our simulations generate significantly more volatility by shifting agent predictive models to the integrated and nonstationary predictors. Second, we also have seen some models exploding for different Taylor rules. In ABK they showed that the connection of explosive models to the e-stability conditions was not that sharp in a social learning setting. We turn first to the question of forcing our models away from the integrated world through monetary policy.

Table 12 takes our runs with the dual learning approach from table 10. This is the case with the dual models operating, but with the initial IMA burn in period. The question is whether changing the Taylor policy parameters can coax the learners back over to the more stable AR model equlibrium. The table looks at two different output parameters, $\phi_z = (0, 0.05)$, and steadily increases the inflation responsiveness, ϕ_π, from 1.0 to 4.0. For the case of $\phi_z = 0.0$ the fraction of integrated runs, "integrated", quickly drops from 1.0 to 0.0 as ϕ_π reaches 1.75. At this point the results mimic those of basic AR learning along with relatively low inflation volatility. For the case of $\phi_z = 0.05$ there is a similar reduction, but it is slower. The fraction of integrated runs doesn't reach 0.0 until $\phi_\pi = 2.5$. It is important note for feasibility this is a very large value for the inflation reaction implying a more than 2 for 1 response in the nominal interest rate to inflation. It would probably be out of the range of what policy makers would ever implement. Also, in this transition the level of inflation volatility, as measured by standard deviation, falls by nearly a factor of 10. This is a much larger reduction than some of the fine tuning of inflation volatility that appears to be happening for larger values of ϕ_π. The large impact is driven by the fact that this policy change does not just marginally change the feedback from inflation to future nominal interest rates, but it radically moves the entire forecasting equilibrium from one model to another.

In our next two tables we look directly into the e-stability conditions and how well they predict the behavior of our learning systems. These experiments are important for policy makers since they suggest the boundary for which monetary policy might fail explosively. Our model is quite different from the simpler learning systems that sit behind e-stability conditions, and in particular those considered in Bullard & Mitra (2002). In some of their experiments ABK show that with social learning parameter combinations that are unstable under e-stability are actually stable in their simulations. We are interested what the stability situation looks like for our models,

In table 13 we look at the dual model learning case where the model is started with equal weighting across all rules. Previously, in table 9 we showed that for many of our parameter choices this case converged to agents using AR inflation forecasts. The table also showed that in some cases the model was unstable and exploding. This seemed related, but not perfectly, to the signs on the e-stability conditions. Now we turn to a more extensive test. In table 13 we connect the analytic stability condition to our runs. The e-stability condition (equation 20) is clearly connected to the probabilities of model instability, but it is not exact. All e-stable cases (E-stable > 0) give us stability, but we have some analytically unstable cases with (E-stable < 0) where we have stable model runs. There are several cases with $\phi_z = 0.05$ and $\phi_\pi \geq 0.6$ that are all stable from a simulation perspective, but unstable according to the e-stable condition. This is an interesting result in that it lines up with some cases in ABK where social learning seemed to have stronger stability implications. Here, it can be thought of as coming from the fact that a set of expectations that is well settled into the stationary AR parameter draws the model into a stable outcome from which it cannot escape.

In table 14 we turn to the e-stable condition in cases where the model converges to the IMA predictors. The table again looks at runs for the dual model learning situation with the IMA startup period. We present the e-stable value from equation 20, along with the fraction of our 25 simulation runs which explode. As can be seen in the table the e-stable condition aligns perfectly with model stability. A negative value of e-stable implies an explosive model. It is also interesting that there are no cases where we have fractions of explosive runs. Models are either unstable or stable which implies a very tight connection between this type of social learning model, and some of the underlying learning analytics. This is not the case for ABK where there can occasionally be fractions of models with stability under certain parameter combinations.

4.4 Agent heterogeneity

Our model naturally generates a set of heterogeneous model predictors. We have seen a first glimpse of this in figure 12 which shows a relatively compact distribution of our forecasters. We explore this issue of agent heterogeneity more completely in table 15. This table concentrates on our benchmark set of runs with $(\phi_\pi, \phi_z) = (1.0, 0.05)$ as we've used before. The model structure is our dual learning model with the initial IMA(1,1) burn in period of 5000. Remember in table 10 we saw that this converged reliably to the IMA/integrated model. The first line in table 15 is an exact replication of our previous runs. It adds one extra column labeled $\sigma(E^*(\pi))/\sigma_\pi$. This represents the cross sectional heterogeneity of the inflation forecasts as measured by the standard deviation normalized by the inflation time series standard deviation. These estimates should be compared with a similar number at the bottom of table 3. This value is 0.34 for the CPI forecasts and 0.61 using the trimmed CPI forecasts. The comparable value from the simulation, displayed in the right most column, is significantly lower with a value of 0.015. The standard model generates agent heterogeneity levels which are well off that from the professional forecasters.

One possible solution is to add noise to the forecasters. This is done by adding a random amount of noise to the MSE of the forecast each period. The level of noise is normalized by the mean MSE across the inflation model as in

$$\sigma_{MSE} = f_{noise} E(MSE), \tag{30}$$

where f is the parameter which will be swept from 0 to 0.20 in the table. Each rule then has noise added to its performance which follows a Gaussian with zero mean, and standard deviation given by σ_{MSE}. Results for different values of f_{noise} (labeled noise in table 15) are presented. It is clear that as the noise level increases forecast dispersion increases as shown in the last column. The general structure of the model appears not to change much, but there is a strong drop in inflation volatility from 2.49 to 0.82 as the model noise level hits its maximum. Increases in forecast dispersion come with a price of smoothing out inflation forecasts, and stabilizing the time series output of the model.

Figure 14 shows a time series of the rule dispersion measure for the no noise case, and the model using a $f_{noise} = 0.05$ fraction noise level. The no noise case shows that the low rule dispersion is consistent across the entire run and remains well below the value of 0.34 from the data. It is interesting that there is initially substantial dispersion in rules, but this drops off quickly as agent learning takes place. The results are more complicated in the noise case which ends up with a value near 0.15 which is consistent with the cross sectional summary from table 15. This model is initially trained with inflation forecasts using only the IMA model until period 5000. After that the agents are allowed to chose between the IMA model and the AR model. In the case of no noise this period gives a barely visible impact on the time series. In the noise case the change is large with a significant jump in forecast heterogeneity as agents spread out across rules. We can see further evidence of this in figure 15 which shows the cross sectional dispersion for the inflation forecasts at the end of the run. Most forecasts remain with the IMA model, but they are now spread across many gain parameters. Also, the model allows for some leakage into a few very persistent (near random walk) AR models as well. This rule dispersion is consistent with the MSE comparisons from figure 11 that showed a very flat range across many gain parameters, and a near equivalence between the large parameter AR models and many of the IMA models. This lack of model specification clarity means that when noise is added agents are not very confident in which models to use, and forecast dispersion is possible.

5 Conclusions

We have analyzed a simple agent-based approach in a New Keynesian model with learning agents. We build off models with social learning along with heterogeneous agent approaches with a form of tournament selection. Our boundedly rational agents form forecasts by choosing rules from a reasonable set of tractable forecasting models. Our goal is to replace and reinvigorate Arifovic et al. (2012) with a slightly different technology. Our new set of agents shows a strong tendency for self-confirming their beliefs about time series dynamics even though we are operating in a space where we cannot be completely sure if perceived laws of motion in the model are lining up with the simulation actual laws of motion. We perform many tests showing that forecasting agent perceptions would not be disturbed by the time series they generate.

Our most important results are concerned with modifying our inflation forecasts to predictive rules which align with well known observed dynamics in U.S. macro time series. The use of an IMA model for forecasting, as implemented with an exponential filter, radically changes our model time series to reflect these expectations. Inflation volatility increases to reasonable levels, and also increases persistence. Analysis of the time series reveals structure that would by characterized as an IMA process by standard identification and forecast error analysis. We have generated a sunspot equilibrium which diverges from the minimum state variable representation, and generates very realistic dynamics. All this is done with a stripped down New Keynesian model operating with only one shock.

Agent-based modeling is often concerned with coevolution, or how different strategies evolve against each other. We play the AR and IMA models against each other in several time series races. Generally, the AR dominates, but if the agents are started using IMA forecasts initially, then the AR types cannot successfully invade. The system remains stuck in the higher volatility IMA steady state. We show that for certain policy rules the dynamics can be shifted back to the AR steady state, but this might require a very large response of the nominal interest rate to inflation.

One of the more interesting results in ABK is that convergence is possible in some cases where the e-stability conditions do not hold. We explore this for our models, and get mixed, but interesting results. For our AR models we see strong convergence for many unstable policy values. We conjecture that this might be related to the "learning support" that does not allow for explosive parameters in the model. When we look at the more empirically valid IMA models, we find the the e-stability conditions align exactly with our simulations in that they perfectly forecast when models are unstable and do not converge. Therefore, our results are somewhat more cautious than those of ABK in that for some types of learning models the classic e-stability conditions are relevant.

ABK also look at possible sunspot equilibria in their learning models. They find that certain common factor sunspots are possible generating prediction models relying on additional state variables. We view our IMA equilibrium as a close relative to a sunspot, but technically it is different. We do generate inflation dynamics showing a large increase in variability, but it does this by conditioning on a particular model, and not by adding an extra information variable as is usually done in a classic sunspot model.

Finally, we look into the issue of forecast heterogeneity. For professional forecasters we have good information on the dispersion of their forecasts. We estimate this value and compare with equivalent measures in our models. For our basic simulations this value may be off by almost a factor of 10, with the professionals generating much more heterogeneity. We were able to reduce this discrepancy by adding noise to our agents, but even for large noise levels the computational agents were still less heterogeneous than the professionals. This result asks some deep questions about the dynamics of social learning, tournament selection, and the fitness landscape of our models. However, the easy answer might be that in the real world agents have much more data to look at than in a stylized NK model. We have made the forecasting problem too easy.

Forecast dispersion is one of two stylized facts from the professional forecasters that we find difficult to replicate. The other is the autocorrelation of prediction errors. In all our simulated equilibrium, the socially learned forecasts show less autocorrelation in their errors, diverging from our empirical results with the SPF. This element of forecast irrationality was already puzzling, but it is interesting that a simplified model with a very crude learning mechanism does not come close to replicating this aspect of the data. We may need to explore the forecasts at a more detailed individual level, or think more carefully about less rational agents for our simulated forecasters.

There are still many things left to understand about this model. We have left output out of the picture at the moment, by assuming it simply is best modeled with the rational expectations perceived law of motion even though this may be empirically unrealistic. Also, the timing of the model is not correct. Waves of inflation periods look realistic, but their periodicity is far longer than in real data. This needs to be addressed. One solution is to allow ourselves more shocks. We purposefully tied our hands to a single shock to see how we could do. Allowing for the types of shocks which are now common in most modern macro models would probably change our data replication ability.

Agent-based modeling is fundamental to understanding the endogenous heterogeneity in models where predicting the future is critical. The application to macroeconomic forecasting is important because it is a policy relevant problem that is often perceived from the perspective of a homogeneous agent. Also, it is an area where we possess relatively good data on the magnitude of forecast heterogeneity. Social learning models, as pioneered in Arifovic (1994) remain a critical tool in understanding the varied behaviors of many standard models when populated by diverse sets of adaptive learning agents.

References

Anufriev, M., Assenza, T., Hommes, C. & Massaro, D. (2013), 'Interest rate rules and macroeconomic stability under heterogeneous expectations', *Journal of Economic Dynamics and Control* **17**, 1574–1604.

Anufriev, M., Hommes, C. & Makarewicz, T. (2019), 'Simple forecasting heuristics that make us smart: Evidence from different market experiments', *Journal of the European Economic Association* **17**(5), 1538–1584.

Arifovic, J. (1994), 'Genetic algorithm learning and the cobweb model', *Journal of Economic Dynamics and Control* **18**, 3–28.

Arifovic, J. (1996), 'The behavior of the exchange rate in the genetic algorithm and experimental economies', *Journal of Political Economy* **104**, 510–541.

Arifovic, J., Bullard, J. & Kostyshyna, O. (2012), 'Social learning and monetary policy rules', *The Economic Journal* **123**, 38–76.

Arifovic, J. & Duffy, J. (2018), Heterogeneous agent modeling: Experimental evidence, *in* C. Hommes & B. LeBaron, eds, 'Handbook of Computational Economics: Heterogeneous Agent Modeling', Vol. 4, North Holland.

Arifovic, J., Grimaud, A., Salle, I. & Vermandel, G. (2020), Social learning and monetary policy at the effective lower bound, Technical report, Bank of Canada.

Assenza, T., Heemeijer, P., Hommes, C. H. & Massaro, D. (2021), 'Managing self-organization of expecations through monetary policy: A macro experiment', *Journal of Moneary Economics* **117**, 170–186.

Bianchi, F., Nicolo, G. & Song, D. (2023), Inflation and real activity over the business cycle, Technical Report 31075, NBER.

Branch, W. (2004), 'The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations', *Economic Journal* **114**, 592–621.

Branch, W. A. & Evans, G. W. (2011), 'Monetary policy and heterogeneous expectations', *Economic Theory* **47**, 365–393.

Branch, W. A. & McGough, B. (2009), 'A new Keynesian model with heterogeneous expectations', *Journal of Economic Dynamics and Control* **33**, 1036–1051.

Branch, W. A. & McGough, B. (2018), Heterogeneous expectations and micro-foundations in macroeconomics, *in* C. Hommes & B. LeBaron, eds, 'Handbook of Computational Economics: Heterogeneous Agent Modeling', Vol. 4, North Holland.

Brock, W. A. & Haslag, J. H. (2015), Inflation dynamics and inflation uncertainty in a model with heterogeneous forecasts, Technical report, University of Missouri-Columbia.

Brock, W. A. & Hommes, C. H. (1997), 'A rational route to randomness', *Econometrica* **65**, 1059–1097.

Brock, W. A., Hommes, C. H. & Wagener, F. O. O. (2005), 'Evolutionary dynamics with many trader types', *Journal of Mathematical Economics* **41**, 7–42.

Bullard, J. & Mitra, K. (2002), 'Learning about monetary policy rules', *Journal of Monetary Economics* **49**, 1105–1129.

Carroll, C. D. & Wang, T. (2022), Epidemiological expecations, Technical Report 30605, NBER, Cambridge, MA.

Clements, M. P., Rich, R. W. & Tracy, J. (2023), Surveys of professionals, *in* R. Bachmann, G. Topa & W. van der Klaauw, eds, 'Handbook of Economic Expectations', Vol. 1, Academic Press, chapter 3, pp. 71–106.

Dawid, H. & Delli Gatti, D. (2018), Agent-based macroeconomics, *in* C. Hommes & B. LeBaron, eds, 'Handbook of Computational Economics: Heterogeneous Agent Modeling', Vol. 4, North-Holland.

De Grauwe, P. (2011), 'Animal spirits and montetary policy', *Economic Theory* **47**, 423–457.

Evans, G. W., Honkapohja, S., Sargent, T. J. & Williams, N. (2013), Bayesian model averaging, learning and model selection, *in* T. J. Sargent & J. Vilmunen, eds, 'Macroeconomics at The Service of Public Policy', Oxford University Press, chapter 7.

Franses, P. H. (2020), 'IMA(1,1) as a new benchmark for forecast evaluation', *Applied Economics Letters* **27**(17), 1419–1423.

Hommes, C. (2021), 'Behavioral and experimental macroeconomics and policy analysis: A complex systems approach', *Journal of Economic Literature* **59**(1), 149–219.

Hommes, C., Makarewicz, T., Massaro, D. & Smits, T. (2017), 'Genetic algorithm learning in a New Keynesian macroeconomic setup', *Journal of Evolutionary Economics* **27**, 1133–1155.

Hommes, C. & Zhu, M. (2016), Behavioral learning equilibria, persistence amplification, and monetary policy, Technical report, University of Amsterdam.

Hyndman, R. J. & Billah, B. (2003), 'Unmasking the Theta method', *International Journal of Forecasting* **19**, 287–290.

LeBaron, B. (2021), 'Microconsistency in simple empirical agent-based financial models', *Computational Economics* **58**(1), 83–101.

Lux, T. (1998), 'The socio-economic dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distributions', *Journal of Economic Behavior and Organization* **33**, 143–165.

Muth, J. (1960), 'Optimal properties of exponentially weighted forecasts', *Journal of the American Statistical Association* **55**, 299–306.

Pivetta, F. & Reis, R. (2007), 'The persistence of inflation in the U.S.', *Journal of Economic Dynamics and Control* **31**(4), 1326–1358.

Smith, K. (2022), Cross-country inflation expecations evidence of heterogeneous and synchronized markets, Technical report, Brandeis International Business School.

Stock, J. & Watson, M. (2007), 'Why has U.S. inflation become harder to forecast?', *Journal of Money, Credit and Banking* **39**(1), 3–33.

Tortorice, D. L. (2018), 'The business cycle implications of fluctuating long run expectations', *Journal of Macroeconomics* **58**, 266–291.

Vriend, N. J. (2000), 'An illustration of the essential difference between individual and social learning, and its consequences for computational analysis', *Journal of Economic Dynamics and Control* **24**, 1–19.

Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton, NJ.

Table 1: U.S. inflation and output: Long series

	π(cpi)	π (deflator)	Output gap
Mean	3.53	3.20	-1.51
Std	3.28	2.57	2.34
Forecast errors			
RMSE(AR)	2.23	1.62	1.03
RMSE(IMA)	2.23	1.57	1.06
RMSE(RW)	2.40	1.68	1.06
Bias(AR)	0.00	0.02	-0.00
Bias(IMA)	0.04	0.03	-0.01
Bias(RW)	0.015	0.01	-0.01
ACF(1)-AR	-0.11	-0.22	0.11
ACF(1)-MA	0.06	-0.01	0.01
ACF(1)-RW	-0.27	-0.34	0.07

Summary statistics and forecasting errors: U.S. 1949:1 - 2022:4. Inflation is measured using CPI and GDP deflator. Output gap is estimated as GDP/potential GDP. All series are quarterly, seasonally adjusted. Source: FRED. AR = AR(1), IMA = IMA(1,1), and RW = Random Walk (Naive).

Table 2: Model parameter estimates

	π(cpi)	π (deflator)	Output growth
AR(1)	0.73	0.79	0.89
	(0.03)	(0.02)	(0.02)
IMA(1,1)	-0.45	-0.39	0.05
	(0.03)	(0.02)	(0.02)
Gain	0.55	0.61	1.05

Parameter estimates: Quarterly, 1949:1 through 2022:4. Numbers in parenthesis are standard errors. Gain refers to the corresponding Kalman gain for the estimated IMA(1,1) which is $(1+\theta)$.

Table 3: Recent data: Forecaster/model comparisons

	π(cpi)	π (trimmed cpi)
Mean	2.85	2.78
Std	2.21	1.24
Forecast errors		
RMSE(AR)	1.98	0.79
RMSE(IMA)	2.07	0.81
RMSE(RW)	2.36	0.88
RMSE(SPF)	2.04	0.78
Bias(AR)	-0.00	-0.00
Bias(IMA)	0.07	0.06
Mean(RW)	0.03	0.05
Mean(SPF)	0.12	0.05
ACF(1)-AR	-0.02	-0.27
ACF(1)-IMA	0.14	-0.05
ACF(1)-RW	-0.31	-0.37
ACF(1)-SPF	0.29	0.39
$\sigma(SPF)/\sigma(\pi)$	0.34	0.61

Professional forecaster comparisons: Quarterly, 1983:1 through 2022:4. All forecasts are 1 quarter ahead. CPI and trimmed CPI. Source: FRED. AR = AR(1), IMA = IMA(1,1), and RW = Random Walk (Naive). SPF = mean 1 quarter ahead forecast from the Survey of Professional Forecasters. $\sigma(SPF)/\sigma(\pi)$ refers to the cross sectional standard deviation over the forecaster set normalized by the time series standard deviation.

Table 4: Parameter values

Parameter	Value(s)
β	0.99
σ	0.157
κ	0.024
ϕ_π	Multiple values (1.0)
ϕ_z	Multiple values (0.05)
ρ	0.35
σ_{r^n}	0.70
T (time periods)	20000 quarters
N (cross section)	25
AR Parameter	$[0.00, \ldots, 1.0]$
Gain parameter	$[0.05, \ldots, 1.0]$
Rules (J)	100
f_{switch}	0.05
p_{update}	0.5
g_f (MSE gain)	0.001

Table 5: AR(1) equilibrium expectations benchmark

ϕ_π	ϕ_z	explode	E-stable	$ACF(\pi)$	$ACF(\Delta\pi)$	$\bar{g}(\pi)$	$ACF(z)$	$\bar{g}(z)$	σ_π	σ_z
0.900	0.000	0.000	-0.002	0.351	-0.325	0.350	0.351	0.350	0.840	5.719
1.000	0.000	0.000	0.000	0.346	-0.327	0.346	0.346	0.346	0.815	5.550
1.100	0.000	0.000	0.002	0.350	-0.324	0.350	0.350	0.350	0.794	5.402
0.900	0.050	0.000	-0.002	0.345	-0.328	0.345	0.345	0.345	0.596	4.054
1.000	0.050	0.000	0.001	0.349	-0.326	0.351	0.349	0.351	0.586	3.991
1.100	0.050	0.000	0.003	0.352	-0.320	0.352	0.352	0.352	0.573	3.901
0.900	0.100	0.000	-0.001	0.352	-0.323	0.350	0.352	0.350	0.463	3.151
1.000	0.100	0.000	0.001	0.349	-0.325	0.349	0.349	0.348	0.454	3.092
1.100	0.100	0.000	0.003	0.350	-0.322	0.350	0.350	0.351	0.448	3.051

Equilibrium expectations: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. "explode" reports the fraction of runs exploding in an unstable fashion. "E-stable" is the learning stability condition. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. These are the fitted AR(1) values from the learning algorithm The final two columns present the standard deviations for $[\pi_t, z_t]$.

Table 6: AR(1) learning

ϕ_π	ϕ_z	explode	E-stable	$ACF(\pi)$	$ACF(\Delta\pi)$	$\bar{g}(\pi)$	$ACF(z)$	$\bar{g}(z)$	σ_π	σ_z
0.900	0.000	0.000	-0.002	0.777	-0.150	0.771	0.456	0.457	0.591	4.136
1.000	0.000	0.000	0.000	0.740	-0.160	0.737	0.448	0.447	0.530	4.059
1.100	0.000	0.000	0.002	0.707	-0.178	0.706	0.438	0.437	0.491	3.973
0.900	0.050	0.000	-0.002	0.657	-0.205	0.660	0.418	0.419	0.369	3.160
1.000	0.050	0.000	0.001	0.629	-0.216	0.630	0.412	0.412	0.354	3.119
1.100	0.050	0.000	0.003	0.606	-0.222	0.609	0.404	0.405	0.342	3.078
0.900	0.100	0.000	-0.001	0.600	-0.230	0.604	0.399	0.399	0.284	2.565
1.000	0.100	0.000	0.001	0.581	-0.236	0.579	0.398	0.394	0.277	2.543
1.100	0.100	0.000	0.003	0.579	-0.234	0.579	0.396	0.394	0.274	2.520

AR(1) learning: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. "explode" reports the fraction of runs exploding in an unstable fashion. "E-stable" is the learning stability condition. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. These are the fitted AR(1) values from the learning algorithm The final two columns present the standard deviations for $[\pi_t, z_t]$.

Table 7: IMA(1,1) learning

ϕ_π	ϕ_z	explode	E-stable	$ACF(\pi)$	$ACF(\Delta\pi)$	$\bar{g}(\pi)$	$ACF(z)$	$\bar{g}(z)$	σ_π	σ_z
0.900	0.000	1.000	-0.002	-	-	0.525	-	0.520	-	-
1.000	0.000	0.000	0.000	0.994	-0.119	0.871	0.448	0.446	3.381	4.065
1.100	0.000	0.000	0.002	0.983	-0.136	0.854	0.434	0.433	1.966	3.968
0.900	0.050	1.000	-0.002	-	-	0.544	-	0.540	-	-
1.000	0.050	0.000	0.001	0.992	-0.173	0.808	0.413	0.413	2.346	3.119
1.100	0.050	0.000	0.003	0.984	-0.178	0.802	0.408	0.408	1.566	3.082
0.900	0.100	1.000	-0.001	-	-	0.584	-	0.579	-	-
1.000	0.100	0.000	0.001	0.991	-0.200	0.775	0.396	0.396	1.846	2.540
1.100	0.100	0.000	0.003	0.984	-0.200	0.770	0.392	0.392	1.334	2.519

IMA(1,1) learning: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. "explode" reports the fraction of runs exploding in an unstable fashion. "E-stable" is the learning stability condition. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. For output these are the AR(1) forecast parameter, and for inflation is the filter gain parameter. The final two columns present the standard deviations for $[\pi_t, z_t]$.

Table 8: IMA(1,1) noise calibration

ϕ_π	ϕ_z	σ_r	explode	E-stable	$ACF(\pi)$	$ACF(z)$	σ_π	σ_z
1.000	0.050	0.100	0.000	0.001	0.993	0.414	0.342	0.447
1.000	0.050	0.200	0.000	0.001	0.992	0.412	0.654	0.890
1.000	0.050	0.300	0.000	0.001	0.993	0.413	1.025	1.338
1.000	0.050	0.400	0.000	0.001	0.993	0.418	1.359	1.789
1.000	0.050	0.500	0.000	0.001	0.993	0.413	1.701	2.224
1.000	0.050	0.600	0.000	0.001	0.992	0.415	2.022	2.678
1.000	0.050	0.700	0.000	0.001	0.992	0.414	2.375	3.120
1.000	0.050	0.800	0.000	0.001	0.993	0.409	2.760	3.557
1.000	0.050	0.900	0.000	0.001	0.993	0.414	3.063	4.016
1.000	0.050	1.000	0.000	0.001	0.993	0.413	3.388	4.455

IMA(1,1) noise calibration: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. σ_r is the noise level represented by the standard deviation of the real interest rate. “explode” reports the fraction of runs exploding in an unstable fashion. “E-stable” is the learning stability condition. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. These are the fitted AR(1) values from the learning algorithm The final two columns present the standard deviations for $[\pi_t, z_t]$.

Table 9: Dual model learning

ϕ_π	ϕ_z	explode	integrated	E-stable	$ACF(\pi)$	$ACF(\Delta\pi)$	$\bar{g}(\pi)$	$ACF(z)$	$\bar{g}(z)$	σ_π	σ_z
0.900	0.000	1.000	1.000	-0.002	-	-	0.989	-	0.520	-	-
1.000	0.000	0.000	1.000	0.000	0.994	-0.116	0.880	0.448	0.447	3.376	4.063
1.100	0.000	0.000	0.080	0.002	0.741	-0.170	0.730	0.438	0.436	0.615	3.983
0.900	0.050	0.000	0.000	-0.002	0.658	-0.202	0.659	0.421	0.421	0.369	3.161
1.000	0.050	0.000	0.000	0.001	0.636	-0.212	0.639	0.413	0.413	0.357	3.125
1.100	0.050	0.000	0.000	0.003	0.625	-0.218	0.627	0.411	0.411	0.349	3.084
0.900	0.100	1.000	1.000	-0.001	-	-	0.992	-	0.583	-	-
1.000	0.100	0.000	0.000	0.001	0.593	-0.231	0.595	0.399	0.400	0.280	2.550
1.100	0.100	0.000	0.000	0.003	0.580	-0.235	0.581	0.395	0.394	0.275	2.528

Dual model learning: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. "explode" reports the fraction of runs exploding in an unstable fashion. "integrated" reports the fraction of runs where agents perceive the series as following the integrated process. "E-stable" is the learning stability condition. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. These are the fitted AR(1) values from the learning algorithm The final two columns present the standard deviations for $[\pi_t, z_t]$.

Table 10: Dual model learning, IMA start period

ϕ_π	ϕ_z	explode	integrated	E-stable	$ACF(\pi)$	$ACF(\Delta\pi)$	$\bar{g}(\pi)$	$ACF(z)$	$\bar{g}(z)$	σ_π	σ_z
0.900	0.000	1.000	1.000	-0.002	-	-	0.989	-	0.520	-	-
1.000	0.000	0.000	1.000	0.000	0.994	-0.128	0.866	0.445	0.444	3.247	4.053
1.100	0.000	0.000	1.000	0.002	0.983	-0.136	0.857	0.431	0.432	1.949	3.970
0.900	0.050	1.000	1.000	-0.002	-	-	0.997	-	0.540	-	-
1.000	0.050	0.000	1.000	0.001	0.993	-0.169	0.810	0.417	0.416	2.370	3.128
1.100	0.050	0.000	1.000	0.003	0.983	-0.180	0.798	0.408	0.408	1.537	3.080
0.900	0.100	1.000	1.000	-0.001	-	-	0.999	-	0.579	-	-
1.000	0.100	0.000	1.000	0.001	0.992	-0.202	0.774	0.395	0.397	1.905	2.541
1.100	0.100	0.000	1.000	0.003	0.985	-0.203	0.766	0.393	0.393	1.333	2.521

Dual model learning, IMA start period: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. "explode" reports the fraction of runs exploding in an unstable fashion. "integrated" reports the fraction of runs where agents perceive the series as following the integrated process. "E-stable" is the learning stability condition. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. These are the fitted AR(1) values from the learning algorithm The final two columns present the standard deviations for $[\pi_t, z_t]$.

Table 11: IMA(1,1) forecast performance

ϕ_π	ϕ_z	$E(e_{t,\pi})$	$ACF(e_{t,\pi})$	$RMSE(\pi)$	σ_π	$E(e_{t,z})$	$ACF(e_{t,z})$	$RMSE(z)$	σ_z
1.000	0.000	0.000	0.003	0.356	3.247	-0.006	-0.008	3.630	4.053
		(0.001)	(0.010)	(0.003)	(0.352)	(0.059)	(0.008)	(0.024)	(0.045)
1.100	0.000	0.000	-0.001	0.351	1.949	0.002	-0.005	3.583	3.970
		(0.001)	(0.017)	(0.003)	(0.205)	(0.018)	(0.010)	(0.028)	(0.043)
1.000	0.050	-0.000	0.009	0.278	2.370	-0.008	-0.003	2.844	3.128
		(0.000)	(0.008)	(0.001)	(0.182)	(0.035)	(0.005)	(0.011)	(0.022)
1.100	0.050	-0.000	0.009	0.275	1.537	-0.007	-0.003	2.811	3.080
		(0.000)	(0.008)	(0.002)	(0.104)	(0.016)	(0.006)	(0.020)	(0.031)
1.000	0.100	0.000	0.006	0.228	1.905	0.005	-0.005	2.335	2.541
		(0.000)	(0.010)	(0.002)	(0.223)	(0.024)	(0.007)	(0.012)	(0.018)
1.100	0.100	0.000	0.011	0.227	1.333	-0.004	-0.001	2.318	2.521
		(0.000)	(0.010)	(0.002)	(0.092)	(0.011)	(0.007)	(0.014)	(0.022)

IMA(1,1) forecast performance: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. $e_{t,x}$ refer to the forecast errors for the two predictions. $E(e_{t,x})$ is the mean prediction error, or bias. $ACF(e_{t,x})$ is the autocorrelation of the forecast errors. $RMSE(x)$ is the root mean squared prediction error on the given target x. σ_π and σ_z are again the standard deviations for $[\pi_t, z_t]$. All numbers in parenthesis are the standard deviations from the cross section of runs.

Table 12: Policy and stability: Dual model learning, IMA start

ϕ_π	ϕ_z	explode	integrated	E-stable	$ACF(\pi)$	$ACF(\Delta\pi)$	$\bar{g}(\pi)$	$ACF(z)$	$\bar{g}(z)$	σ_π	σ_z
1.000	0.000	0.000	1.000	0.000	0.994	-0.121	0.872	0.447	0.445	3.292	4.054
1.250	0.000	0.000	0.520	0.006	0.941	-0.137	0.865	0.418	0.419	1.148	3.876
1.500	0.000	0.000	0.160	0.012	0.813	-0.171	0.798	0.394	0.394	0.638	3.693
1.750	0.000	0.000	0.000	0.018	0.697	-0.196	0.731	0.382	0.380	0.453	3.572
2.000	0.000	0.000	0.000	0.024	0.639	-0.215	0.671	0.377	0.373	0.406	3.455
2.500	0.000	0.000	0.000	0.036	0.566	-0.241	0.588	0.369	0.363	0.355	3.248
3.000	0.000	0.000	0.000	0.048	0.519	-0.255	0.540	0.361	0.354	0.323	3.077
3.500	0.000	0.000	0.000	0.060	0.499	-0.264	0.513	0.357	0.351	0.301	2.902
4.000	0.000	0.000	0.000	0.072	0.482	-0.273	0.496	0.348	0.343	0.285	2.772
1.000	0.050	0.000	1.000	0.001	0.992	-0.166	0.817	0.417	0.419	2.371	3.124
1.250	0.050	0.000	1.000	0.006	0.971	-0.185	0.792	0.399	0.399	1.150	3.028
1.500	0.050	0.000	0.840	0.013	0.927	-0.193	0.777	0.391	0.389	0.791	2.943
1.750	0.050	0.000	0.640	0.018	0.888	-0.213	0.745	0.378	0.377	0.637	2.868
2.000	0.050	0.000	0.280	0.025	0.771	-0.218	0.710	0.367	0.366	0.451	2.782
2.500	0.050	0.000	0.000	0.036	0.586	-0.240	0.614	0.358	0.353	0.296	2.627
3.000	0.050	0.000	0.000	0.049	0.526	-0.255	0.557	0.354	0.350	0.266	2.514
3.500	0.050	0.000	0.000	0.060	0.505	-0.263	0.525	0.354	0.346	0.251	2.409
4.000	0.050	0.000	0.000	0.072	0.484	-0.272	0.501	0.347	0.341	0.237	2.306

Policy and stability: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. "explode" reports the fraction of runs exploding in an unstable fashion. "integrated" reports the fraction of runs where agents perceive the series as following the integrated process. "E-stable" is the learning stability condition. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. These are the fitted AR(1) values from the learning algorithm The final two columns present the standard deviations for $[\pi_t, z_t]$.

Table 13: Stability check: Dual model learning

ϕ_π	ϕ_z	E-stable	explode
0.500	0.000	-0.012	1.000
0.600	0.000	-0.010	1.000
0.700	0.000	-0.007	0.920
0.800	0.000	-0.005	0.960
0.900	0.000	-0.002	0.000
1.000	0.000	0.000	0.000
1.100	0.000	0.002	0.000
1.200	0.000	0.005	0.000
1.300	0.000	0.007	0.000
1.400	0.000	0.010	0.000
1.500	0.000	0.012	0.000
0.500	0.050	-0.011	0.560
0.600	0.050	-0.009	0.000
0.700	0.050	-0.007	0.000
0.800	0.050	-0.004	0.000
0.900	0.050	-0.002	0.000
1.000	0.050	0.001	0.000
1.100	0.050	0.003	0.000
1.200	0.050	0.005	0.000
1.300	0.050	0.008	0.000
1.400	0.050	0.010	0.000
1.500	0.050	0.013	0.000

Policy and stability: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. “E-stable” is the learning stability condition. “explode” reports the fraction of runs exploding in an unstable fashion.

Table 14: Stability check: Dual model learning, IMA start

ϕ_π	ϕ_z	E-stable	explode
0.500	0.000	-0.012	1.000
0.600	0.000	-0.010	1.000
0.700	0.000	-0.007	1.000
0.800	0.000	-0.005	1.000
0.900	0.000	-0.002	1.000
1.000	0.000	0.000	0.000
1.100	0.000	0.002	0.000
1.200	0.000	0.005	0.000
1.300	0.000	0.007	0.000
1.400	0.000	0.010	0.000
1.500	0.000	0.012	0.000
0.500	0.050	-0.011	1.000
0.600	0.050	-0.009	1.000
0.700	0.050	-0.007	1.000
0.800	0.050	-0.004	1.000
0.900	0.050	-0.002	1.000
1.000	0.050	0.001	0.000
1.100	0.050	0.003	0.000
1.200	0.050	0.005	0.000
1.300	0.050	0.008	0.000
1.400	0.050	0.010	0.000
1.500	0.050	0.013	0.000

Stability check: Dual model learning, IMA start Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. "E-stable" is the learning stability condition "explode" reports the fraction of runs exploding in an unstable fashion.

Table 15: Increasing model noise: Dual model learning, IMA start

ϕ_π	ϕ_z	noise	explode	integrated	$ACF(\pi)$	$ACF(\Delta\pi)$	$\bar{g}(\pi)$	$ACF(z)$	$\bar{g}(z)$	σ_π	$\sigma(E^*(\pi))/\sigma_\pi$
1.000	0.050	0.000	0.000	1.000	0.993	-0.171	0.814	0.416	0.417	2.493	0.015
1.000	0.050	0.025	0.000	1.000	0.985	-0.198	0.735	0.419	0.418	1.678	0.107
1.000	0.050	0.050	0.000	1.000	0.979	-0.201	0.720	0.424	0.425	1.381	0.139
1.000	0.050	0.075	0.000	1.000	0.970	-0.197	0.716	0.426	0.428	1.169	0.173
1.000	0.050	0.100	0.000	1.000	0.964	-0.189	0.719	0.438	0.441	1.057	0.189
1.000	0.050	0.150	0.000	1.000	0.953	-0.189	0.707	0.443	0.452	0.927	0.221
1.000	0.050	0.200	0.000	1.000	0.941	-0.185	0.702	0.452	0.466	0.826	0.240

Increasing model noise: Dual model learning, IMA start: Parameters ϕ_π and ϕ_t are the policy terms from the interest rate rule. “noise” is the noise level added to the fitness objective as a fraction of the mean(MSE). “explode” reports the fraction of runs exploding in an unstable fashion. “integrated” reports the fraction of runs where agents perceive the series as following the integrated process. The ACF columns report the first order autocorrelation for π (level and first difference) and z. The two values for $\bar{g}$ are the weighted forecast parameters across all agents for π and z. These are the fitted AR(1) values from the learning algorithm The column labeled $\sigma(E^*(\pi))/\sigma_\pi$ measures the cross sectional standard deviation in inflation forecasts divided by the time standard deviation of realized inflation, π_t.

Figure 1: U.S. Inflation and Output Gap

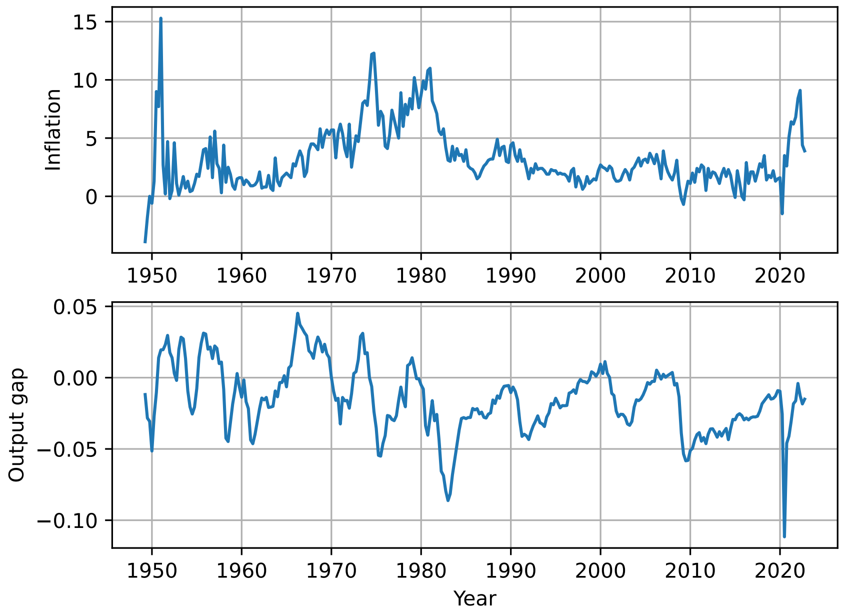

Figure 2: U.S. Inflation and Output Gap: Autocorrelations

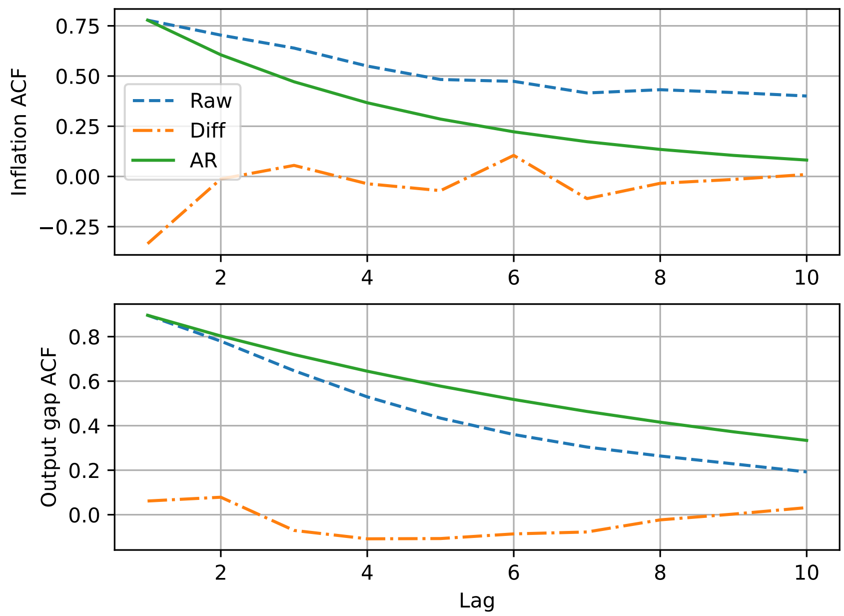

Figure 3: Inflation model forecast error autocorrelations

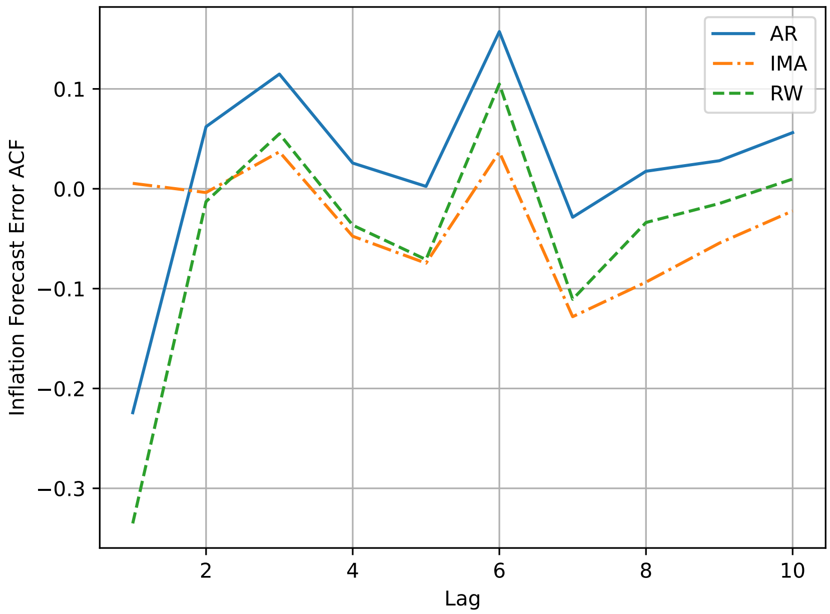

Figure 4: Output model forecast error autocorrelations

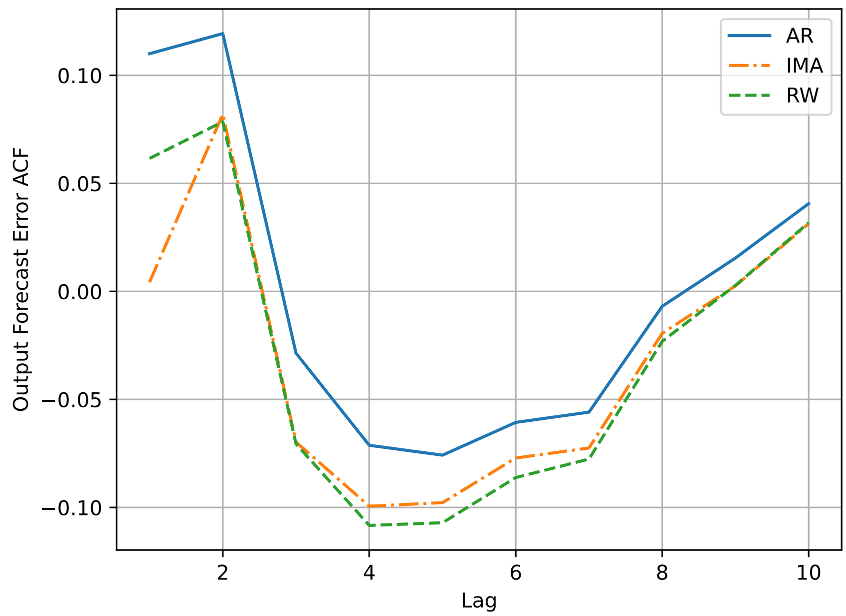

Figure 5: Professional forecast error autocorrelations

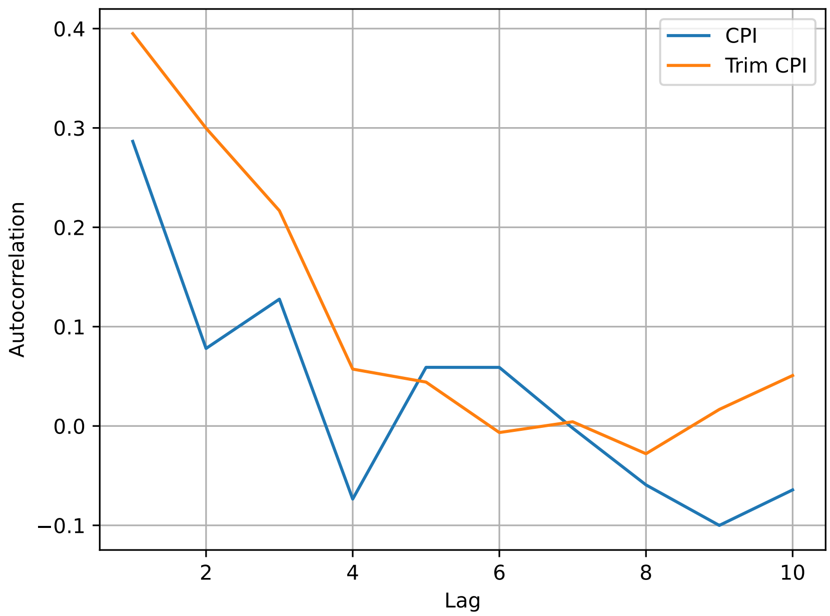

Figure 6: Model time series: AR predictors

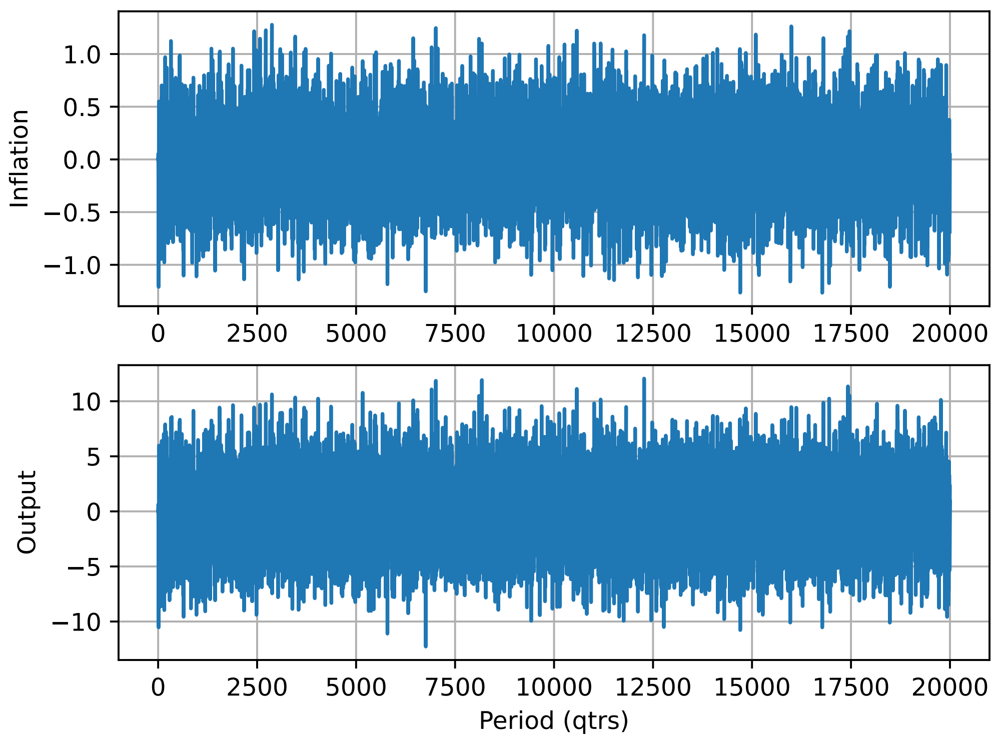

Figure 7: Inflation and output autocorrelations: AR predictors

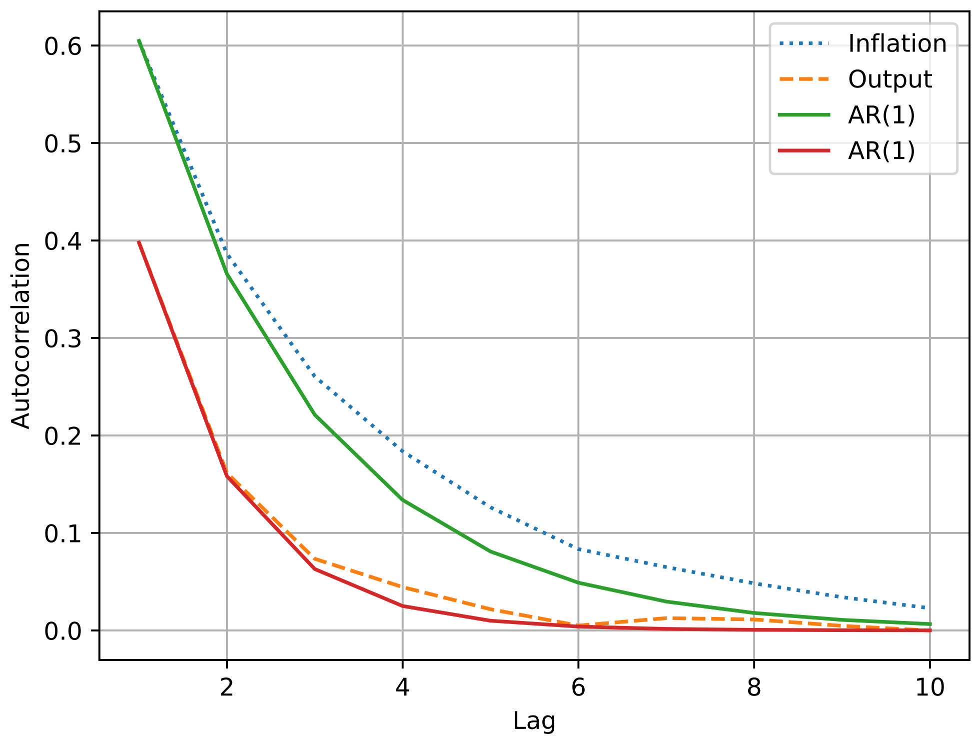

Figure 8: Model time series: IMA predictors

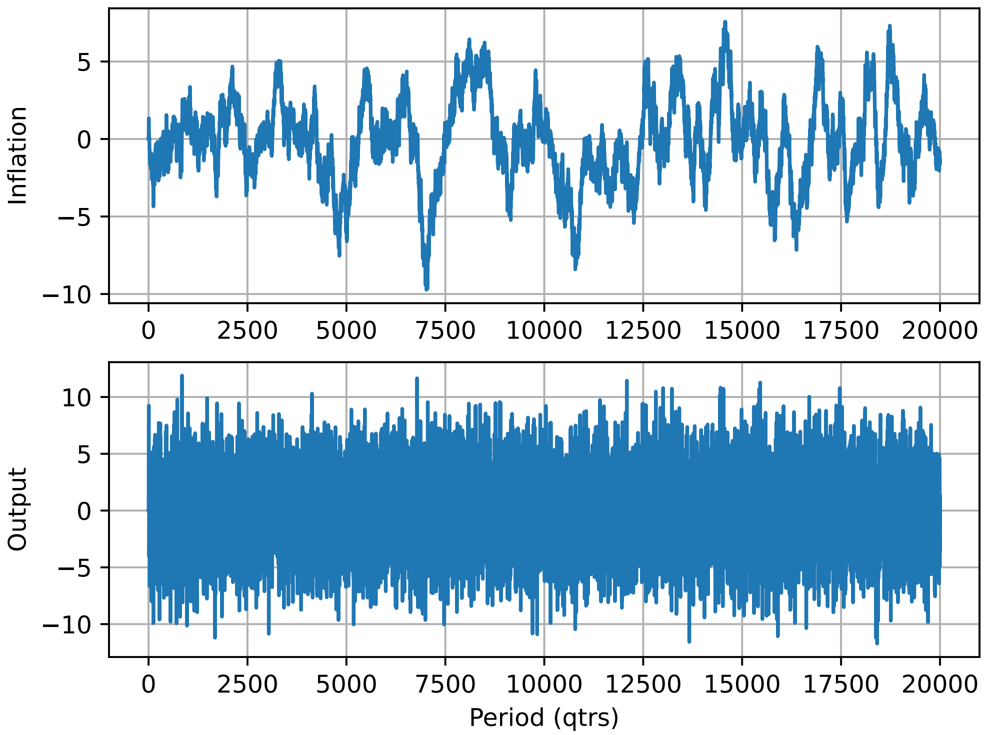

Figure 9: Inflation and output autocorrelations: IMA predictors

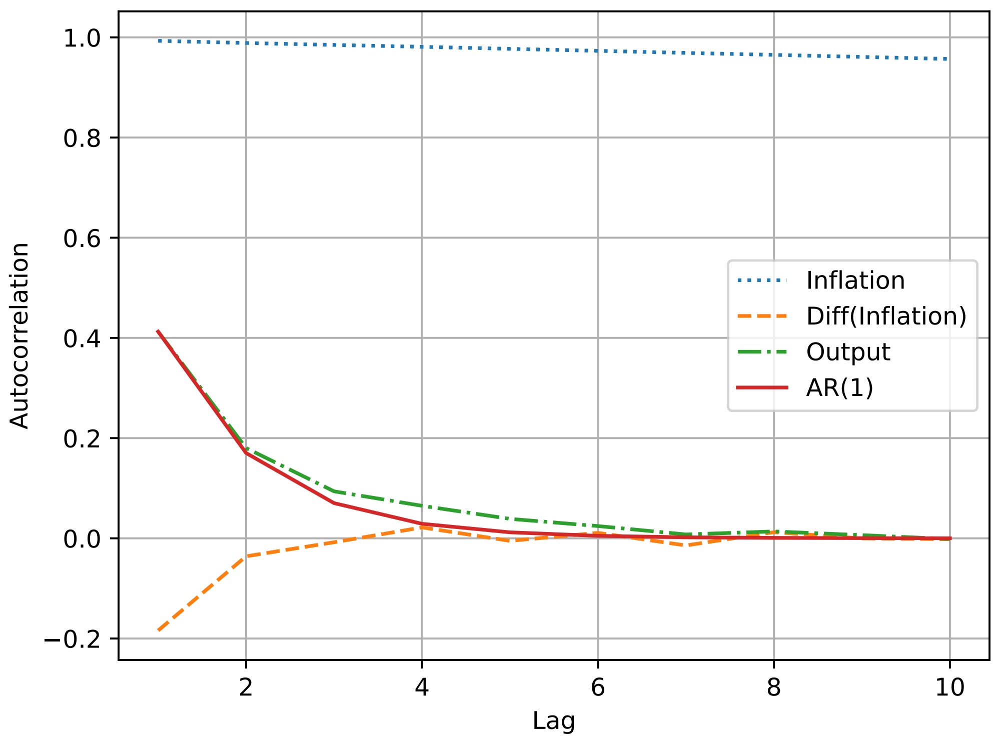

Figure 10: Forecast mean squared error

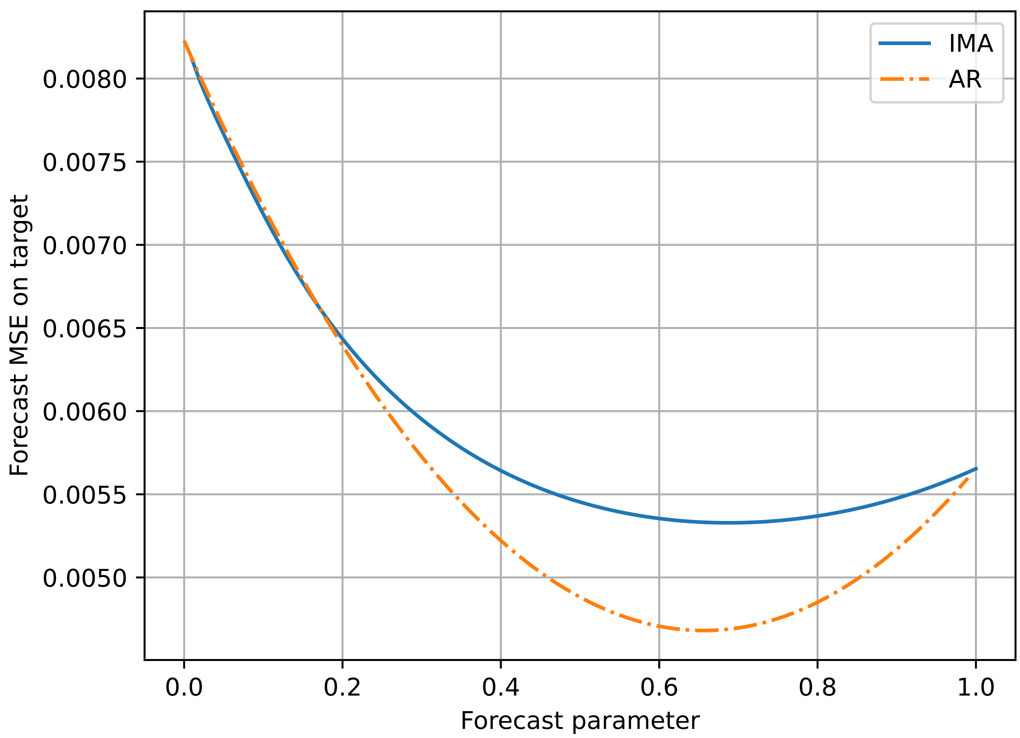

Figure 11: Forecast mean squared error: IMA start

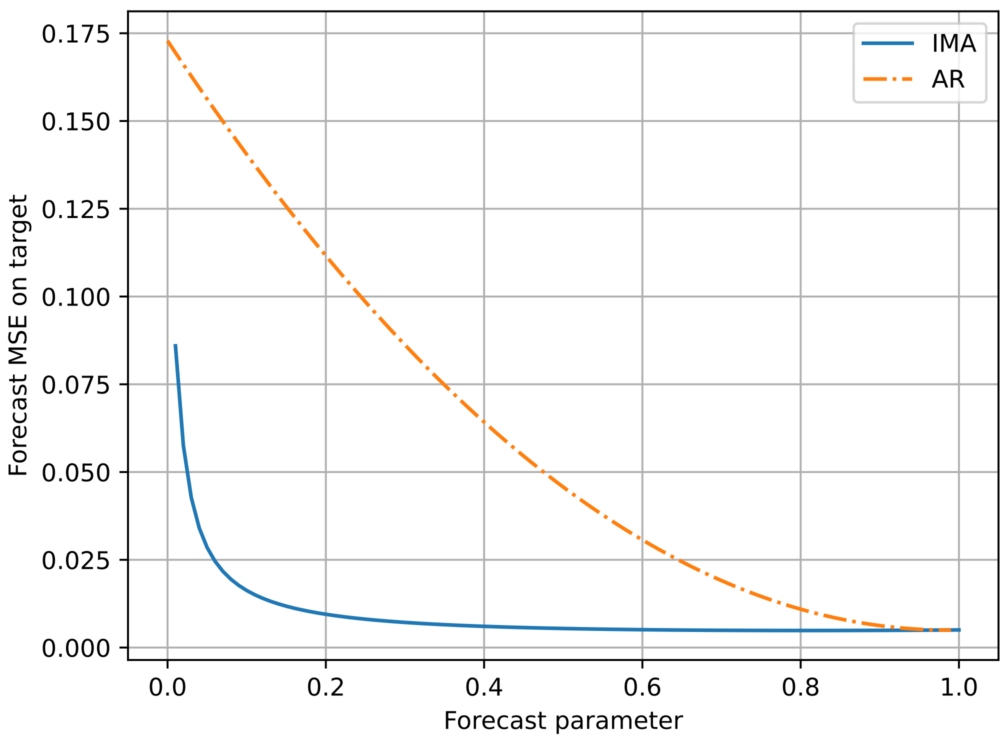

Figure 12: Forecast weight cross section: IMA/AR predictors

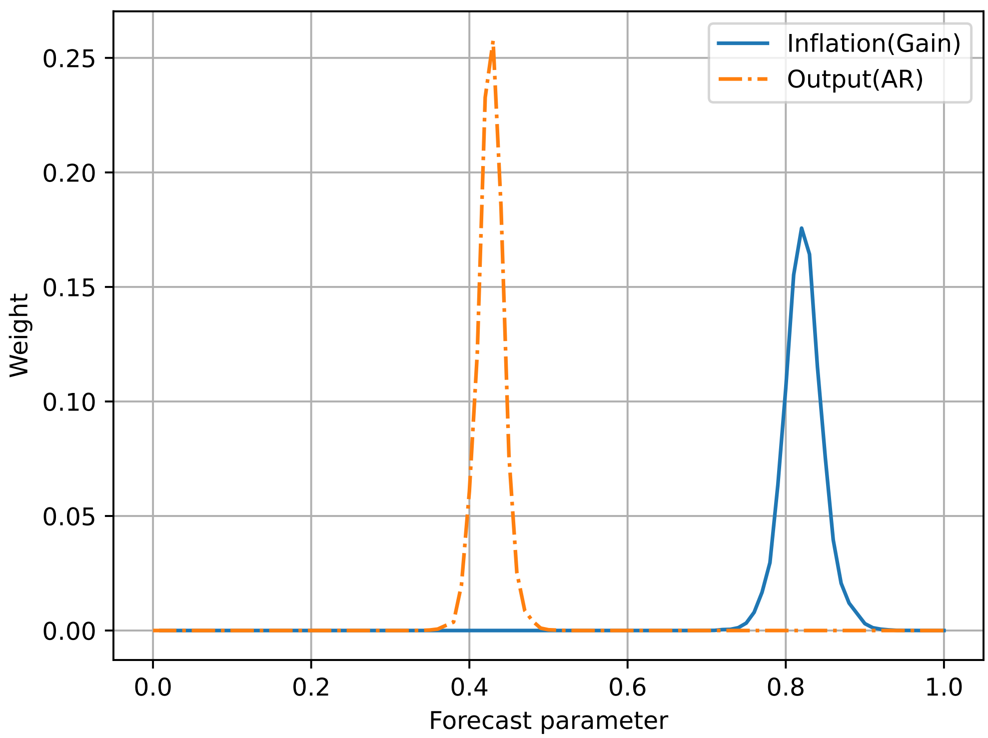

Figure 13: Forecast parameter time series: IMA/AR predictors

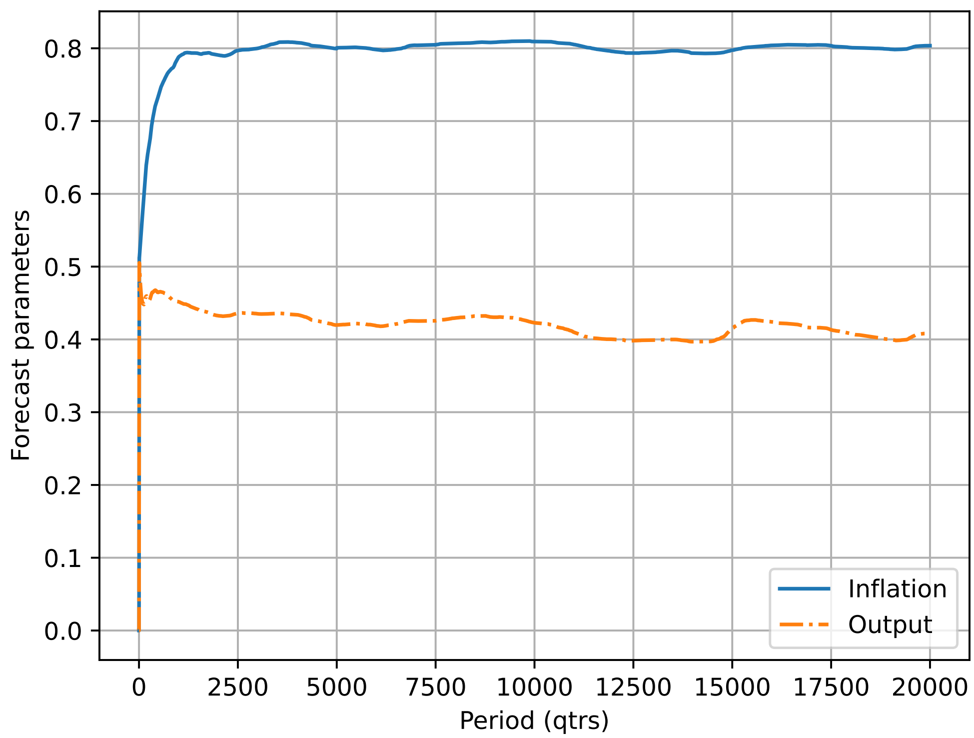

Figure 14: Forecast dispersion

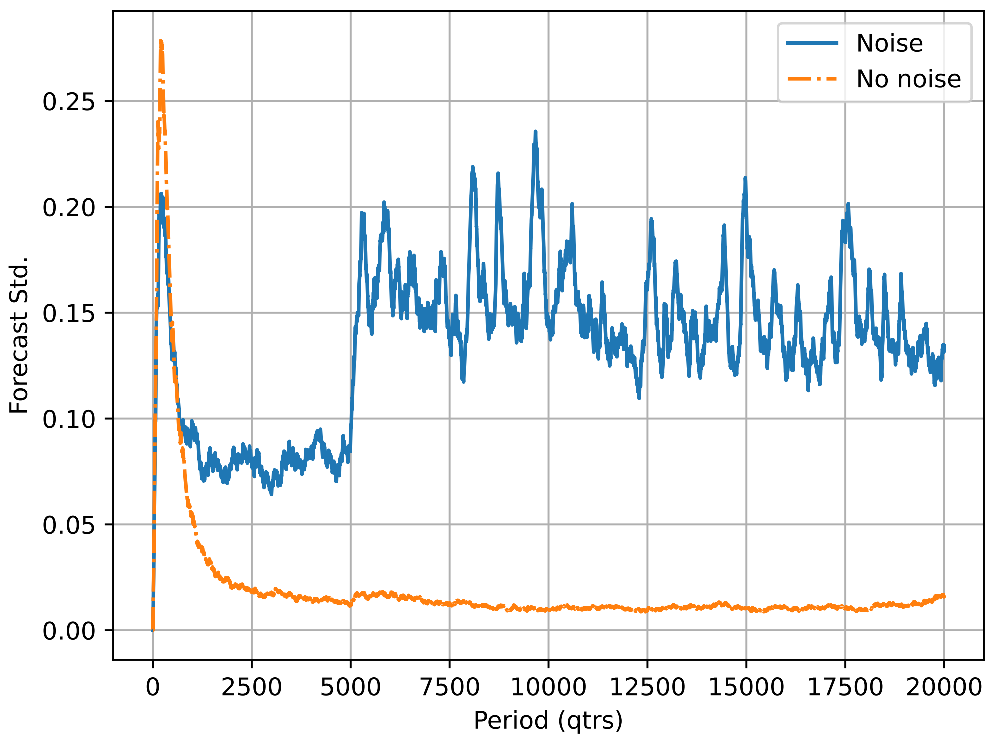

Figure 15: Forecast weight cross section with added noise