Misspecified Expectations Among Professional Forecasters*

Julio L. Ortiz [†] Federal Reserve Board

July 2024

Abstract

Analyzing professional forecasts, I find that a model of expectation formation in which respondents misperceive the true law of motion of the data generating process, which in turn causes them to form an erroneous view of its underlying persistence, tends to outperform alternative models when fit to forecast errors and revisions. Misspecified expectations outperforms the alternatives for a variety of macroeconomic aggregates both in and out of sample. Misspecified expectations is successful in fitting the data in part because it allows forecast errors to be longer lived. I conclude that misspecified expectations can serve as a suitable approach to model expectation formation among professional forecasters.

Keywords: Non-rational expectations. Expectation formation. Noisy information. Overreactions. Misspecification. **JEL Codes**: D83, D84, E70

^{*}I would like to thank editor Ken West, and three anonymous referees for their helpful comments which greatly improved the paper. I would also like to thank Francesco Bianchi, Carola Binder, Thiago Ferreira, Jaime Marquez, Stephen Terry, and Fabian Winkler for their valuable feedback and helpful conversations. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

[†]Email: julio.l.ortiz@frb.gov.

Address: 20th and C Street NW, Washington, DC 20551.

1 Introduction

There is ample survey-based evidence that forecast errors are predictable on the basis of information available to respondents in real time. This finding has animated a rich literature studying violations of full information rational expectations (FIRE) (Bordalo et al., 2020; Broer and Kohlhas, 2022; Coibion and Gorodnichenko, 2012, 2015; Farmer et al., 2021; Fuhrer, 2018; Kohlhas and Walther, 2021; Kucinskas and Peters, 2022; Ryngaert, 2023). While several theoretical departures from FIRE can explain forecast inefficiency, the literature has not yet settled on a benchmark non-FIRE model (Reis, 2020). This paper undertakes a formal analysis of competing theories of expectation formation in an effort to move toward establishing such a benchmark in the context of professional forecasting.

In this paper I study a noisy information rational expectations model along with three leading non-FIRE models which are: overconfident expectations (Daniel et al., 1998), diagnostic expectations (Bordalo et al., 2020), and misspecified expectations (Fuster et al., 2010), the latter of which assumes that forecasters misperceive the objective law of motion, causing them to formulate an erroneous view of the underlying persistence of the data generating process. My baseline application of misspecified expectations assumes that the variable of interest follows an AR(2) process, but that forecasters perceive it to follow an AR(1).¹

Using panel data from the U.S. Survey of Professional Forecasters, I estimate these models of expectation formation and find that misspecified expectations best fits observed forecast errors and revisions across a range of macroeconomic variables. Among the subset variables that exhibit overreactions, misspecified expectations appears to outperform the alternative models for variables that exhibit less objective persistence.

In general, misspecified expectations outperforms the alternatives because it is able to produce overreaction to new information but also overextrapolation across horizons. When perceiving the persistence of the process to be larger than it is, forecasters under misspecified expectations place

¹In online Appendix C.3, I consider an environment in which the data follows an AR(1) process, but that forecasters misperceive the AR(1) coefficient.

a larger Bayesian weight on new information than is called for. At the same time, when projecting their forecasts forward, forecasters apply the relatively larger misperceived persistence, causing their forecast errors to be longer lived. Hence, under misspecified expectations, these Bayesian forecasters exhibit a bias in their updating equation and in their prediction equation. On the other hand, forecasters in the alternative models project their forecasts according to the objective persistence of the process.

Taken together, my results suggest that a framework in which decision-makers adopt parsimonious models of richer underlying processes, causing them to misperceive the true persistence of the process, can serve as a useful benchmark to describe expectation formation among professional forecasters for a variety of macroeconomic aggregates.

I begin by illustrating a simple forecasting problem in a noisy information environment. The macroeconomic variable of interest follows an AR(2) process and is unobservable to forecasters. Forecasters issue predictions of this aggregate based on information gleaned from a contemporaneous private signal. In this linear Gaussian setting, forecasters employ the Kalman filter to obtain the optimal forecast which is consistent with the conditional expectation.

I next examine the aforementioned three models of expectation formation and assess their ability to fit the data relative to the rational baseline and to each other. In this paper, I focus on straightforward biases which can be flexibly embedded into more complex macroeconomic models. I estimate these models via maximum likelihood estimation (MLE) and rank the models based on encompassing weights. My results show that misspecified expectations outperforms the alternatives both in and out of sample for real GDP growth, which I focus on for my baseline results, as well as a range of other macroeconomic variables. While I abstract away from learning in my set up, I analyze the relative performance of the models over rolling windows of the data as well as by forecaster "age" in the survey, and argue that learning about the data generating process, which could favor misspecified expectations, does not drive my results.

Misspecified expectations provides a successful fit to a variety of moments. I start by verifying that all three estimated biased models generate forecaster-level overreaction based on the Bordalo

et al. (2020) errors-on-revisions regression. I also examine the consensus-level analog to this regression (Coibion and Gorodnichenko, 2015) and find that the estimated non-rational models are able to qualitatively match this moment. Next, I show that the estimated misspecified expectations model is quantitatively able to produce overshooting (Angeletos et al., 2020; Bianchi et al., 2022) in aggregate expectations. In addition, misspecified expectations offers a superior fit to certain moments such as persistent disagreement across horizons and updating behavior based on the estimated weight placed on priors and news. The fact that misspecified expectations is better able to match cross-horizon features of the data, such as persistent disagreement, and updating weights is consistent with the intuition that misspecified expectations outperforms the other models because it operates through both the predict and update equations.

While misspecified expectations successfully fits the data for a wide range of macroeconomic forecasts, it is important to note that my results are specific to the context of professional forecasting. The relative rankings of the candidate models that I examine may be different when analyzing household or firm expectations, or when studying micro-level expectations rather than forecasts about aggregate outcomes. Nonetheless, given that professional forecasters are arguably the most well-informed agents in the economy, the literature on survey expectations has found these predictions attractive in part to infer a lower bound on economy-wide information frictions and biases (Bianchi et al., 2022; Carroll, 2003; Coibion and Gorodnichenko, 2015; Cornand and Hubert, 2022).

My paper relates to a number of both longstanding and more recent contributions to the literature on survey expectations and violations of FIRE. One strand of the literature studies non-rational biases in survey forecasts. Bordalo et al. (2020) find that forecast errors and revisions are negatively correlated at the individual level. I use this correlation as the relevant measure of overreactions in Section 5. To explain this particular violation of FIRE, Bordalo et al. (2020) propose a theory of diagnostic expectations. In earlier contributions, Daniel et al. (1998) and Moore and Healy (2008) propose theories of overconfident expectations in which forecasters believe their private information to be more precise than it truly is. Other notable contributions can be found in the finance literature (Kent and Hirshleifer, 2015; Kent et al., 1998, 2001). Inspecting the term structure of forecast uncertainty, Binder et al. (2022) also documents evidence consistent with overconfidence. Other behavioral biases have also been proposed in the literature such as intrinsic expectations (Fuhrer, 2018), relative overconfidence (Broer and Kohlhas, 2022), and overpersistence bias (Rozsypal and Schlafmann, 2023). The misspecified expectations model that I estimate implies that forecasters tend to overextrapolate, in line with overpersistence bias.

Another strand of the literature studies rational deviations from FIRE. For instance, Azeredo da Silveira et al. (2020) proposes a theory of noisy memory in which forecasters optimize over their history of past signals.² Moreover, Farmer et al. (2021) argues that many of the forecasting anomalies in the data can be rationalized in a learning environment. More closely related to the notion of misspecification that I study, Branch and Evans (2006) and Pfajfar (2013) develop models featuring intrinsic heterogeneity in which agents optimally adopt different predictor functions. Therefore, while overconfident expectations and diagnostic expectations correspond to behavioral biases, misspecified expectations can arise due to a behavioral bias or because adopting parsimonious forecasting models is optimal.

The rest of the paper is organized as follows. Section 2 details the noisy information setting and the rational baseline model. Section 3 outlines the menu of biased models. Section 4 estimates these models via MLE and assesses their fit to the data. Section 5 further discusses why misspecified expectations outperforms the alternatives. Section 6 concludes.

2 A Baseline Rational Expectations Model

I begin by outlining a noisy information rational expectations model which will serve as the baseline non-FIRE theory against which I will compare the other non-FIRE models. The model outlined here is in the spirit of Coibion and Gorodnichenko (2015) and Bordalo et al. (2020), though under the assumption of a richer driving process which later allows me to model misspecified expectations

²Afrouzi et al. (2023) similarly propose a model of working memory to explain biases in expectations. The costly retrieval of past information in their model can be interpreted as emanating from memory constraints or availability biases.

as a world in which forecasters assume simpler dynamics.

Suppose that a forecaster wishes to predict some aggregate variable, x_t , which follows an AR(2) process:

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + w_t, \quad w_t \sim N(0, \sigma_w^2)$$

At a given point in time, t, the forecaster, indexed by i, has access to a noisy private signal,

$$y_t^i = x_t + v_t^i, \quad v_t^i \sim N(0, \sigma_v^2),$$

The forecaster's objective is to minimize her mean squared errors. Hence, the optimal forecast of x_t is the conditional expectation $\mathbb{E}(x_t | \mathcal{I}_t^i)$, where \mathcal{I}_t^i denotes the forecaster's information set at time t. In this linear Gaussian setting, the forecaster can employ the Kalman filter to obtain an optimal forecast. From the Kalman filter, we obtain the following predict and update equations for x_t ,

Predict:
$$x_{t|t-1}^i = \rho_1 x_{t-1|t-1}^i + \rho_2 x_{t-2|t-1}^i$$
 (1)

Update:
$$x_{t|t}^{i} = x_{t|t-1}^{i} + \kappa_1 (y_t^{i} - x_{t|t-1}^{i})$$
 (2)

$$x_{t-1|t}^{i} = x_{t-1|t-1}^{i} + \kappa_2 (y_t^{i} - x_{t|t-1}^{i}),$$
(3)

where κ_1 and κ_2 denote the Kalman gains, which are the optimal weights placed on the private signal when updating the current prediction $x_{t|t}^i$ and the previous prediction $x_{t-1|t}^i$, respectively.

This baseline model yields a number testable predictions. For instance, the forecast error, $x_t - x_{t|t}^i$, should be uncorrelated with anything residing in the forecaster's information set at time t. In addition, the forecast revision, $x_{t|t}^i - x_{t|t-1}^i$, is uncorrelated with anything in the forecaster's information set through t - 1. These orthogonality conditions, however, tend to be violated in the data. I next consider a menu of alternative theories of expectation formation which deviate from this benchmark model.

3 Alternative Models of Beliefs

In this section, I outline three models of expectation formation which can each, in principle, match well-known features of survey expectations.

3.1 Overconfident Expectations

Overconfidence is a theory in which individuals believe their private signals to be more informative than they truly are (Daniel et al., 1998). As a result, forecasters trust their signals more than is warranted and consequently place excessive weight on new private information.

Rather than accurately processing their respective private signals, forecasters misperceive the amount of noise associated with private information:

$$y_t^i = x_t + v_t^i \qquad v_t^i \stackrel{\text{i.i.d.}}{\sim} N(0, \check{\sigma}_v^2),$$

where $\check{\sigma}_v = \alpha_v \sigma_v$ and $\alpha_v \in [0, 1]$. The predict and updating rules are as in the rational case outlined in the previous section, but with excessively large weights, $\hat{\kappa}_1$ and $\hat{\kappa}_2$, placed on new information.

3.2 Diagnostic Expectations

Diagnostic expectations is a theory that is rooted in the Kahneman and Tversky (1972) representativeness heuristic. According to diagnostic expectations, agents form their beliefs subject to a cognitive friction in which they conflate the objective likelihood of a type in a group with its representativeness (i.e., the frequency of the type within the group *relative* to a reference group). This is formalized in Gennaioli and Shleifer (2010) who define the representativeness of a type τ for group *G* as:

$$R(\tau, G) = \frac{h(T = \tau | G)}{h(T = \tau | - G)}$$

where $h(T = \tau | G)$ denotes the distribution of variable T in group G. According to diagnostic expectations, subjective probabilities are formed based on a distorted density,

$$h^{\varphi}(T=\tau|G) \propto h(T=\tau|G)R(\tau,G)^{\varphi}.$$

Under rational expectations, $\varphi = 0$, while under diagnostic beliefs, $\varphi > 0$.

In this context, forecasters form expectations about the future state of the world, and they judge future outcomes that are more representative to be more likely. Following Bordalo et al. (2019, 2020), since the latent state described in the previous section follows a Markov process with an objective conditional distribution of $f(x_t|x_{t-1})$ and a signal history, $Y_t^i = \{y_t^i\}_{t=0}$, we can characterize a diagnostic forecaster's perceived distribution as,

$$f^{\varphi}(x_t|Y_t^i) \propto f(x_t|Y_t^i) \left[\frac{f(x_t|Y_t^i)}{f(x_t|Y_{t-1}^i \cup x_{t|t-1}^i)} \right]^{\varphi}.$$

In this case, the reference group reflects a "no news" scenario in which today's signal is uninformative, $y_t^i = x_{t|t-1}^i$. Therefore, when $\varphi > 0$, the forecaster exhibits representativeness with respect to recent news. As a result, diagnostic expectations as devised here is a theory of overreaction to new information, and we can interpret this bias as reflecting imperfect memory and associativeness.

A diagnostic forecaster's current-period forecast of x_t is,

$$\hat{x}_{t|t}^{i} = x_{t|t}^{i} + \varphi \big[x_{t|t}^{i} - x_{t|t-1}^{i} \big],$$

and based on the assumed AR(2) dynamics, her one-step ahead forecast is

$$\hat{x}_{t+1|t}^{i} = \rho_1 \hat{x}_{t|t}^{i} + \rho_2 \hat{x}_{t-1|t}^{i}$$

3.3 Misspecified Expectations

I next consider a bias in which forecasters adopt simple models of the world in the spirit of Fuster et al. (2010) and Molavi (2022), causing them to misperceive the true persistence of the data generating process. Applications of misspecified expectations that result in misspecified persistence include Gabaix (2019) and Rozsypal and Schlafmann (2023).³

In this paper, I focus on a version of misspecification which is closest to natural expectations as modeled in Fuster et al. (2010). As with natural expectations, I assume that forecasters neglect longer lags in the data generating process. In its simplest form, the underlying state follows an AR(2) process:

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + w_t, \quad w_t \sim N(0, \sigma_w^2)$$

but forecasters treat x_t as an AR(1) process when devising their predictions

$$x_t = \hat{\rho} x_{t-1} + u_t$$

where $u_t = (\rho_1 - \hat{\rho})x_{t-1} + \rho_2 x_{t-2} + w_t$.⁴ Importantly, forecasters still understand the information structure. If the perceived persistence loads excessively onto the first lag, then forecasters will overextrapolate.

The predict and update equations in this case are,

Predict:
$$\hat{x}_{t|t-1}^i = \hat{\rho} \hat{x}_{t-1|t-1}^i$$
 (4)

Update:
$$\hat{x}_{t|t}^{i} = \hat{x}_{t|t-1}^{i} + \hat{\kappa}(y_{t}^{i} - \hat{x}_{t|t-1}^{i}).$$
 (5)

³In principle, it could optimal for forecasters to use more parsimonious models rather than estimate the richer dynamics governing the true data generating process (Branch and Evans, 2006; Brock and Hommes, 1997; Pfajfar, 2013). On the other hand, forecasters could be predisposed to use simpler models due to cognitive frictions.

⁴This form of misspecification is technically different from Fuster et al. (2010). First, I do not explicitly model an AR(2) in levels and assume that agents forecast an AR(1) in growth rates. Second, here, the perceived persistence is estimated from the data rather than defined to be a function of the true autocorrelation parameters. Third, Fuster et al. (2010) define natural expectations as a weighted average of rational expectations and the naive AR(1) expectations. I do not make such an assumption under misspecified expectations.

3.4 Other Models

In general, existing alternatives to FIRE can be categorized into one of two groups: models that generate underreactions and models that generate overreactions. Theories of underreactions include revision smoothing (Scotese, 1994), sticky information (Mankiw and Reis, 2002), noisy information (Lucas, 1972), rational inattention (Woodford, 2001; Sims, 2003), and adaptive expectations (Cagan, 1956; Nerlove, 1958), among others. In this paper, I model different theories of overreaction within a noisy information environment. As a result, the models that I consider feature some scope for underreaction. I abstract away from theories that can only generate underreactions, however, mainly because they are unable to speak to the robust evidence of overreaction among individual forecasters.

Aside from the models considered in this paper, other theories that produce overreaction in expectations include imperfect memory (Afrouzi et al., 2023; Azeredo da Silveira et al., 2020), multi-frequency forecasting under aggregation constraints (Bürgi and Ortiz, 2023), asymmetric attention (Kohlhas and Walther, 2021), and learning (Farmer et al., 2021). I abstract away from these models because they cannot be flexibly nested into the current setting.⁵ These rich theories require additional structure, different environments, and, crucially, additional parameters which could pose an identification challenge.

While I limit the set of models considered here to a rational benchmark, overconfident expectations, diagnostic expectations, and misspecified expectations, I supplement my results with various robustness checks. Overall, one might view these models as the most tractable among a broader set of theories of overreaction. Such tractability is desirable, particularly when embedding non-FIRE expectations into medium- and large-scale quantitative models. Moreover, extending beyond the set of models considered here, my results imply that a successful non-FIRE theory should be able to match particular features of the data, as I will show in Section 5.

⁵I revisit learning in the subsequent section and show that it does not drive my results.

4 Model Estimation

I next estimate the four models of expectation formation outlined above. I begin by reviewing the data and estimation approach, and then report the results.

4.1 Data: U.S. Survey of Professional Forecasters

To fit the parameters governing information frictions and biases in each model, I use professional forecasts from the U.S. Survey of Professional Forecasters (SPF). The SPF is a quarterly survey managed by the Federal Reserve Bank of Philadelphia. The survey began in the fourth quarter of 1968, and provides forecasts from several forecasters across a range of horizons. The SPF has been used in many studies in the survey expectations literature (Capistran and Timmermann, 2009; Clements, 2015; Coibion and Gorodnichenko, 2012).

For my baseline results, I estimate the models using data on real GDP growth, however, I also estimate and analyze the relative performance of the models for other macroeconomic variables. I consider forecasts made from 1992Q1 to 2019Q4. I choose this later sample in order to avoid estimating the model over a longer period that potentially encompasses different regimes. In addition, the survey itself has undergone several changes since 1968 including a redefining of output in 1992 from GNP to GDP.⁶ Moreover, in Section 4.5 I utilize the full sample to estimate the models over rolling windows and examine their relative performance over time.

The estimation procedure that I detail below requires unbroken sequences of observations, so I keep only forecasters' longest spell of reported forecasts. Furthermore, because entry and exit from the SPF may potentially be non-random (Engelberg et al., 2011), I keep only forecasters who have a minimum spell length of eight quarters. These choices yield a sample of 77 unique forecasters from 1992Q1 to 2019Q4. In online Appendix C.4 I re-estimate the models for different macroeconomic variables by specifying minimum spell lengths of four and 12 quarters, and find that my results are qualitatively unchanged.

⁶I estimate the models over the full sample in online Appendix C.2.

Table 1: S	PF Summary	Statistics
------------	------------	------------

	Mean	Median	Std. Dev.	25%	75%
One-quarter ahead forecast error	-0.620	-0.243	1.859	-1.353	0.798
One-quarter ahead forecast revision	-0.131	-0.051	0.882	-0.483	0.250
Real-time real GDP growth	2.550	2.585	1.853	1.580	3.541

Note: The table reports summary statistics for annualized real GDP growth forecast errors and revisions as well as realtime real GDP growth. The sample of forecast errors and revisions is constructed using the SPF and spans 1992Q1-2019Q4, with 77 unique forecasters and 1,520 forecaster-quarter observations.

Table 1 reports summary statistics for real-time real GDP growth as well as the associated forecast errors and revisions in the SPF. I follow the literature in specifying real-time forecast errors rather than forecast errors based on revised figures (Bordalo et al., 2020; Coibion and Gorodnichenko, 2015; Kohlhas and Walther, 2021). In online Appendix A, I provide additional details regarding the baseline sample.

4.2 Maximum Likelihood Estimation

I estimate the models following a three-step MLE procedure. First, I collect the parameters of the fundamental variable in a vector, $\theta = (\rho_1 \ \rho_2 \ \sigma_w)$, and estimate θ via MLE using x_t as the observation which corresponds to the real-time macroeconomic time series.⁷ I fix the parameters obtained in the first step according to their point estimates so that they are the same across all models. Second, I estimate the signal noise dispersion parameter, σ_v , in the rational expectations model using the panel of forecast errors, forecast revisions, and lagged forecast errors from the SPF, and calibrate it to this point estimate across the remaining three models.⁸ Finally, for each of the remaining three models, I estimate the bias parameters, α_v , φ , and $\hat{\rho}$ via MLE once again using the SPF data.

I take this approach for two reasons. First, by setting the fundamental and information parame-

⁷I collect these real-time variables from the U.S. SPF Error Statistics dataset.

⁸I choose to fit the models to these three observations as they jointly capture over- and underreaction in macroeconomic expectations. Whereas contemporaneous errors and revisions exhibit overreaction, error persistence reflects underreaction (Ryngaert, 2023).

Table 2: Real GDP Growth Estimates

Panel A: Fundamental Parameters					
First order autocorrelation	ρ_1	0.434			
		(0.009)			
Second order autocorrelation	ρ_2	-0.006			
		(0.002)			
Persistent innovation dispersion	σ_w	1.663			
		(0.012)			
Panel B: Information/Bias Parameters					
		Rational	Overconfident	Diagnostic	Misspecified
		Expectations	Expectations	Expectations	Expectations
Private noise dispersion	σ_v	1.530	1.530	1.530	1.530
-		(0.058)	-	-	-
Overconfidence	α_v		0.722		
			(0.041)		
Diagnosticity	φ			0.233	
				(0.026)	
Perceived persistence	$\widehat{ ho}$				0.564
					(0.018)
Panel C: Model Selection					
Log likelihood		-8280.3	-8089.2	-8084.7	-8051.9
AIC		16571	16190	16181	16116
Encompassing weight		0.000	0.000	0.462	0.539

Note: Panel A reports estimates for the first stage MLE which estimates the parameters governing the fundamental process. Panel B reports the parameter estimates based on the second and third steps of the estimation procedure. Standard errors reported in parentheses. For each model, panel C reports the maximized log likelihood, AIC, and encompassing weight (defined in equation 6).

ters to be consistent across all three biased models, I am able to evaluate overconfident expectations, diagnostic expectations, and misspecified expectations solely on the basis of the biases that they generate. Second, this approach makes the identification of the parameters transparent and feasible. The fundamental parameters are identified from the macroeconomic time series while the information friction and the bias parameters are identified from survey expectations. More importantly, one would not otherwise be able to jointly identify σ_v and α_v in the overconfident expectations model.⁹ Online Appendix B provides additional details on the construction of the likelihood functions and the state space formulation of the different models.

⁹Though σ_v and α_v cannot be jointly estimated, in online Appendix C.6 I jointly estimate $\{\sigma_v, \varphi\}$ and $\{\sigma_v, \hat{\rho}\}$ and show that it allows the biased models to better match over- and underreactions observed in the data.

Table 2 reports my baseline results. The estimates of the parameters governing the evolution of real GDP growth are reported in panel A and imply a relatively low persistence for real GDP growth. Panel B reports the rest of the parameters. The estimate of private noise dispersion, obtained from the estimated rational expectations model, implies a signal-to-noise ratio of $\frac{\sigma_w}{\sigma_v} \approx 1.09$.

Turning to the bias parameter estimates, the estimated degree of overconfidence is 0.72, indicating that forecasters believe their private signals to be more precise than they truly are. Moreover, the estimated degree of diagnosticity is about 0.23, implying that forecasters overreact to new information. Finally, the perceived persistence in the misspecified expectations model is about 0.56. Given that the first order autocorrelation of real GDP growth is estimated to be 0.43, the misspecified expectations model implies that forecasters overestimate the first order autocorrelation of GDP growth and neglect the partial reversal which characterizes its second lag, thereby generating overreactions.

Finally, panel C reports model selection statistics. The maximized log likelihood for each model is reported in the first row. Comparing across columns, it is evident that the estimated misspecified expectations model produces the highest log likelihood across all models. The second row of panel 2 reports the corresponding Akaike information criterion (AIC) for each model. The AIC provides an estimate of the expected, relative distance between a given fitted model and the unknown true model that generated the data (Akaike, 1973, 1974). In this case, comparing the AIC across models, does not change the relative rankings that one would obtain by simply comparing the maximized likelihoods.¹⁰

I formally compare the relative fit of each model by employing encompassing weights (West, 2001; Harvey et al., 1998; West, 2006). Encompassing weights are obtained by relating realizations to their model-based counterparts through a constrained linear regression. In this particular setting, I estimate the weights, w_k ,

$$y_{it}^{\text{data}} = \sum_{k=1}^{4} w_k y_{it}^{\text{model } k} + \varepsilon_{it}, \quad \text{subject to } \sum_{k=1}^{4} w_k = 1.$$
(6)

where w_k is the encompassing weight for model k. I specify y to be the one-quarter ahead forecast.

¹⁰The results remain unchanged when selecting models on the basis of the Bayesian information criterion.

	Rational	Overconfident	Diagnostic	Misspecified
	Expectations	Expectations	Expectations	Expectations
СРІ	0.000	0.000	0.142	0.859
GDP Deflator	0.000	0.026	0.771	0.203
Housing starts	0.239	0.000	0.000	0.761
Industrial production	0.000	0.000	1.000	0.000
Payroll employment	0.000	0.000	0.000	1.000
Real consumption expenditures	0.000	0.077	0.146	0.777
Real federal government spending	0.000	0.000	0.000	1.000
Real GDP	0.000	0.000	0.462	0.539
Real nonresidential investment	0.000	0.000	0.000	1.000
Real residential investment	0.000	0.000	0.624	0.376
Real state and local government spending	0.000	0.000	0.000	1.000
Unemployment rate	0.745	0.000	0.000	0.255
3-month Treasury bill	0.000	0.287	0.000	0.713
10-year government bond	0.000	0.254	0.000	0.746

Table 3: Model Fit Across Macroeconomic Variables

Note: The table reports encompassing weights (see equation 6) for each for 14 macroeconomic variables covered in the SPF.

The estimated encompassing weights are reported in the final row of Table 2 and indicate that the misspecified expectations model outperforms the others since it has the largest weight, 0.539, when compared to the other models.

4.3 Other Macroeconomic Variables

I estimate the model for 13 other macroeconomic variables in the SPF. Table 3 reports encompassing weights for each variable. For most of these aggregates, misspecified expectations provides the best fit to the data as indicated by the relatively larger estimated weights reported in the rightmost column. Diagnostic expectations, however, registers larger encompassing weights for certain macroeconomics series such as the GDP deflator, industrial production, and real residential investment. Overconfident expectations does not yield particularly large encompassing weights for any variable. Finally, based on the encompassing weights, rational expectations provides the best fit to unemployment rate forecasts.

Misspecified expectations outperforms the other biased models for most variables, in part be-



Figure 1: Estimated Persistence Bias

Note: The bars depict the persistence bias, defined as $\hat{\rho} - \rho_1$. CPI denotes consumer price index. PGDP denotes GDP deflator. HOUS denotes housing starts. IP denotes industrial production. EMP denotes payroll employment. RCONS denotes real consumption expenditures. RFED denotes real federal government expenditures. RGDP denotes real GDP. RNRES denotes real non-residential investment. RRES denotes real residential investment. RSL denotes real state & local government expenditures. UNEMP denotes the unemployment rate. TB3M denotes the 3-month Treasury bill. GB10Y denotes the 10-year government bond.

cause a subset of these variables exhibit underreactions. Figure 1 plots the bias in the misspecified expectations model, $\hat{\rho} - \rho_1$ for each macroeconomic variable. Note that this bias takes on positive and negative values, allowing it to account for variables that exhibit overreactions and underreactions. The alternative models, on the other hand, cannot generate forecaster-level underreaction.

Moreover, among the macroeconomic variables that exhibit overreactions, I find that misspecified expectations tends to offer a better fit for less persistent series. Figure 2, visualizes this pattern by plotting the encompassing weight against the sum of the autoregressive coefficients for each variable. With some exceptions, the figure suggests that less persistent variables, based on the sum of the autoregressive coefficients, are variables for which the encompassing weight on misspecified expectations is higher.¹¹ This is because the scope for misspecified expectations to generate overreactions, $\hat{\rho} - \rho_1$, is greater when ρ_1 is low.

To supplement the conclusions drawn from Table 3, I also conduct a likelihood ratio test for non-nested models (Vuong, 1989). I compare the misspecified expectations model to each of the

¹¹A similar pattern emerges when reproducing this figure for the out-of-sample estimates described in Section 4.4. This figure is plotted in online Appendix C.7.

Figure 2: Misspecified Expectations Fits Less Persistent Variables



Note: The figure plots the encompassing weight on the misspecified expectations models against the sum of the autoregressive coefficients for each variable. The figure displays only variables for which forecasters overextrapolate. CPI denotes consumer price index. PGDP denotes GDP deflator. IP denotes industrial production. RCONS denotes real consumption expenditures. RFED denotes real federal government expenditures. RGDP denotes real GDP. RNRES denotes real non-residential investment. RRES denotes real residential investment. RSL denotes real state & local government expenditures.

other models separately. Under the null hypothesis, misspecified expectations is observationally equivalent to one of the other models. When the test statistic exceeds the critical value, we reject the null and conclude that there is evidence in favor of the misspecified expectations model relative to the model against which is being compared. The results of this exercise, reported in online Appendix C.1, generally accord with those from the model comparisons based on encompassing weights.

4.4 Out-of-Sample Performance Across Models

In addition to providing a superior in-sample fit for a wider range of macroeconomic variables, the misspecified expectations model also provides a better out-of-sample fit relative to the other models. I assess the out-of-sample fit by estimating each model based on data from the first half of the sample, 1992Q1-2005Q4, and then using these estimates along with data from the latter half of the sample to construct and compare encompassing weights.

	Rational Expectations	Overconfident Expectations	Diagnostic Expectations	Misspecified Expectations
СРІ	0.000	0.257	0.000	0.743
GDP Deflator	0.000	0.000	0.987	0.013
Housing starts	0.502	0.000	0.047	0.451
Industrial production	0.000	0.041	0.959	0.000
Payroll employment	0.000	0.000	0.813	0.188
Real consumption expenditures	0.000	0.000	0.000	1.000
Real federal government spending	0.000	0.000	0.591	0.409
Real GDP	0.000	0.000	0.000	1.000
Real nonresidential investment	0.000	0.111	0.416	0.472
Real residential investment	0.007	0.000	0.000	0.994
Real state and local government spending	0.029	0.000	0.151	0.820
Unemployment rate	1.000	0.000	0.000	0.000
3-month Treasury bill	0.387	0.003	0.000	0.610
10-year government bond	0.000	0.653	0.106	0.240

Table 4: Out-of-Sample Model Fit

Note: The table reports encompassing weights (see equation 6) for each for 14 macroeconomic variables covered in the SPF.

Table 4 reports out-of-sample encompassing weights for each model across the same set of macroeconomic variables. Here, again, misspecified expectations provides the best fit to more of the variables. However, diagnostic expectations now outperforms all other models for the GDP de-flator, industrial production, payroll employment, and real federal government expenditures. Over-confident expectations outperforms the other models for the 10-year government bond. Finally, the rational expectations model outperforms the others for housing starts and the unemployment rate.

These out-of-sample results might lead one to wonder whether forecasts could be improved in real time. There is a longstanding literature which finds that statistical models are often unable to outperform survey forecasts out of sample (Ang et al., 2007; Faust and Wright, 2013). Based on my results, one cannot necessarily conclude that forecasters could improve their forecasts out of sample because I include the contemporaneous individual and consensus forecast errors in my set of MLE observations, which are not known to forecasters at the time in which they issue their forecasts.

For a similar reason, my results are not inconsistent with Eva and Winkler (2023) which finds that error predictability regressions, such as the Coibion and Gorodnichenko (2015) regression,

perform relatively poorly out of sample. The objective of this paper is to determine which non-FIRE model provides the best fit to the data. To answer this question, I estimate the different models using information available to the econometrician, which is not the same as the information available to forecasters in real time.

4.5 Learning About the Data Generating Process

Learning, which is a compelling and realistic feature of real-time forecasting, is another theory which can explain predictable forecast errors (Evans and Honkapohja, 2001; Farmer et al., 2021). Though Farmer et al. (2021) emphasize that initial beliefs rather than model misspecification are ultimately responsible for generating overreactions in their learning environment, in this section I explore whether learning may nonetheless bias my results in favor of misspecified expectations. Focusing on real GDP growth, I first examine the evolution of the encompassing weights for each model over rolling windows. Second, because the rolling windows approach does not account for composition, I split my sample by forecaster "age," estimate the models for each sub-sample, and then compare the encompassing weights across models.

Rolling Windows Approach

If learning were to bias my results in favor of misspecified expectations, then it would likely do so by providing a better fit to the data in the early part of the sample, during which time forecasters in the survey presumably employ simple forecasting models, and then providing a worse fit as forecasters learn about the underlying process.

To examine whether this is the case, I construct 25-year-long sub-samples with a three-year rolling window using the full pre-COVID-19 sample of the data (1968Q4-2019Q4). I repeat the MLE procedure and estimate the encompassing weights for each model from each sub-sample. Figure 3 plots the encompassing weight, w_k for each model over time. Values closer to one indicate greater evidence for a given model while values closer to zero imply less evidence for the model.

Panel A plots w_k for the rational expectations model over time. Here, we see that there is no





Note: The figure plots the encompassing weight, w_k , for each model over rolling windows based on equation (6). The dates on the horizontal axis correspond to the end date of each 25-year window.

evidence for the rational model as its encompassing weight is lower than the other models across all sub-samples. Panel B plots w_k for the overconfident expectations model. In the earlier sub-samples, I find that the encompassing weight for this model was higher than the other non-rational models, however, the weight then declines and remains close to zero for all remaining sub-samples.

Panel C plots w_i for the diagnostic expectations model. Diagnostic expectations outperforms misspecified expectations in four of the ten sub-periods. Meanwhile, the misspecified expectations model outperforms diagnostic expectations in six of the ten sub-samples, and outperforms all models in five of the ten sub-samples. Overall, the patterns observed in Figure 3 do not indicate that learning over time favors misspecified expectations.

	Rational	Overconfident	Diagnostic	Misspecitied
	Expectations	Expectations	Expectations	Expectations
Below median age Above median age	$0.000 \\ 0.000$	0.000 0.234	$0.000 \\ 0.000$	1.000 0.766

Table 5: Encompassing Weights by Forecaster Experience

Note: The first row reports encompassing weights based on a subsample of the data that reflects forecasters below the median age in the sample. The second row reports encompassing weights based on a subsample of the data that reflects forecasters above the median age in the sample.

Experience as a Proxy for Learning

The conclusions drawn from the exercise in the previous section rely on the assumption that we observe the same forecasters over time. In reality, this is not the case. Because entry and exit in the SPF can affect when a forecaster begins the learning process, I next explore whether learning affects the relative rankings of the candidate models by comparing "experienced" and "inexperienced" forecasters, where I proxy experience with forecaster age (tenure) in the sample.¹² I drop the eight-quarter spell length sample restriction and define inexperienced forecasters as those whose respective ages fall below the unconditional sample median while experienced forecasters are those whose ages reside above the median.¹³

Table 5 reports encompassing weights for each model. The results imply that misspecified expectations outperforms the other models among experienced forecasters. Overall, while learning is likely reflected in professional forecasts, it does not appear drive the relative model rankings to favor misspecified expectations.

¹²Because the SPF is anonymized, I am unable to determine the "experience" of a given forecaster if she has a history of participating in other surveys as well. However, in this case my approach would label a seasoned forecaster who enters the SPF as an "inexperienced" forecaster, which would work against misspecified expectations if more seasoned forecasters are less likely to use simple forecasting models. Nonetheless, I find that misspecified expectations outperforms the other models across both groups.

¹³I have also completed this exercise by identifying experienced and inexperienced forecasters based on their *maximum* age attained in the sample rather than their current age in the sample. The results remain unchanged when applying this alternative definition.

5 Misspecified Expectations Fits Important Moments

Of the different models considered, misspecified expectations most often outperforms the others. To understand why, I focus on my baseline real GDP growth results and examine five features of the data that have been studied in the literature: overreaction, underreaction, overshooting (i.e., over- and underreaction), persistent disagreement, and updating behavior. I summarize my findings below and report the results in online Appendix D.

I begin by noting that the estimated misspecified expectations model performs as well as the others when examining common measures of over- and underreaction. I measure overreactions by estimating the Bordalo et al. (2020) forecast errors-on-revisions regression using simulated data. To measure underreactions, I aggregate the simulated forecasts to the consensus level and estimate the Coibion and Gorodnichenko (2015) errors-on-revisions regression. Overall, each of the three non-rational models is able to generate individual overreactions and aggregate underreactions, as in the data.

I next examine each model's ability to generate simultaneous over- and underreaction in the form of delayed overshooting. Delayed overshooting is a phenomenon documented by Angeletos et al. (2020) which finds that forecasters initially underreact to a shock and later overreact. In principle, each of the models considered here can generate delayed overshooting since each features scope for overreaction and underreaction. However, when simulating impulse responses of forecast errors for each estimated model, I find that only the misspecified expectations model generates a sign switch in the forecast error, which indicates overshooting.

Finally, I study two more features of the data, where I find stronger evidence favoring misspecified expectations. First, misspecified expectations does a better job of generating persistent disagreement among forecasters across horizons. Second, I find that misspecified expectations is better able to match the relative weights that forecasters place on priors and news. Misspecified expectations offers an improvement relative to other models when matching these moments because the misspecified expectations bias also enters into forecasters' predict equation. As a result, forecasters in the misspecified expectations model exhibit longer-lived errors based on their private signals, generating persistent disagreement. In addition, the misspecified expectations bias need not load excessively onto the update equation, which allows the model to better match relative weights placed on priors and news.

6 Conclusion

At present, a host of non-FIRE theories exist in the literature. As mentioned in Reis (2020), however, there is little agreement on a suitable non-FIRE benchmark. This paper offers a partial answer to this question by showing that misspecified expectations, a model of expectation formation in which forecasters' perceived law of motion differs from the objective law of motion and causes misperceptions in the underlying persistence, outperforms other non-FIRE models for a variety of macroeconomic variables. Misspecified expectations can reproduce patterns of overreaction, underreaction, overshooting, persistent disagreement, and it delivers empirically similar updating weights. Embedding the misspecified expectations described here into a quantitative model only requires introducing two parameters into an otherwise standard model. A promising avenue for future research could be to examine whether there is evidence favoring misspecified expectations in other settings.

References

- Afrouzi, Hassan, Spencer Kwon, Augustin Landier, Yueran Ma, and David Thesmar (2023), "Overreaction in Expectations: Evidence and Theory." *The Quarterly Journal of Economics*, 138.
- Akaike, Hirotugu (1973), "Information Theory and an Extension of the Maximum Likelihood
 Principle–In: Second International Symposium on Information Theory (Eds) BN Petrov, F."
 Csaki. BNPBF Csaki Budapest: Academiai Kiado, 267–281.
- Akaike, Hirotugu (1974), "A New Look at the Statistical Model Identification." *IEEE Transactions* on Automatic Control, 716–723.
- Ang, Andrew, Geert Bekaert, and Min Wei (2007), "Do macro variables, asset markets, or surveys forecast inflation better?" *Journal of Monetary Economics*, 4.
- Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry (2020), "Imperfect Macroeconomic Expectations; Evidence and Theory." *NBER Macroeconomics Annual 2020*, 35.
- Azeredo da Silveira, Rava, Yeji Sung, and Michael Woodford (2020), "Optimally imprecise memory and biased forecasts." Working Paper 28075, National Bureau of Economic Research.
- Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma (2022), "Belief Distortions and Macroeconomic Fluctuations." *American Economic Review*, 112.
- Binder, Carola, Tucker S. McElroy, and Xuguang S. Sheng (2022), "The Term Structure of Uncertainty: New Evidence From Survey Expectations." *Journal of Money, Credit, and Banking*, 54, 39–71.
- Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020), "Overreaction in Macroeconomic Expectations." *American Economic Review*, 110, 2748–2782.
- Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2019), "Diagnostic Expectations and Stock Returns." *Journal of Finance*, 74, 2839–2874.

- Branch, William A. and George W. Evans (2006), "Intrinsic heterogeneity in expectation formation." *Journal of Economic Theory*, 127, 264–295.
- Brock, William A. and Cars H. Hommes (1997), "A Rational Route to Randomness." *Econometrica*, 65, 1059–1095.
- Broer, Tobias and Alexandre Kohlhas (2022), "Forecaster (Mis-)Behavior." *The Review of Economics and Statistics*, 1–45.
- Bürgi, Constantin and Julio L. Ortiz (2023), "Overreaction Through Anchoring." Working Paper.
- Cagan, Phillip (1956), "The Monetary Dynamics of Hyperinflation." Studies in the Quantity Theory of Money, 25–117, The University of Chicago Press.
- Capistran, Carlos and Allan Timmermann (2009), "Disagreement and Biases in Inflation Expectations." *Journal of Money, Credit and Banking*, 41, 365–396.
- Carroll, Christopher D (2003), "Macroeconomic expectations of households and professional forecasters." *The Quarterly Journal of Economics*, 118, 269–298.
- Clements, Michael P. (2015), "Are Professional Macroeconomic Forecasters Able to Do Better Than Forecasting Trends?" *Journal of Money, Credit and Banking*, 47, 349–381.
- Coibion, Olivier and Yuriy Gorodnichenko (2012), "What Can Survey Forecasts Tell Us About Information Rigidities?" *Journal of Political Economy*, 120, 116–159.
- Coibion, Olivier and Yuriy Gorodnichenko (2015), "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review*, 109, 465– 490.
- Cornand, Camille and Paul Hubert (2022), "Information frictions across various types of inflation expectations." *European Economic Review*, 146.

- Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998), "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance*, 53, 1839–1885.
- Engelberg, Joseph, Charles F. Manski, and Jared Williams (2011), "Assessing the Temporal Varioation of Macroeconomic Forecasts by a Panel of Changing Composition." *Journal of Applied Econometrics*, 26, 1059–1078.
- Eva, Kenneth and Fabian Winkler (2023), "A Comprehensive Empirical Evaluation of Biases in Expectation Formation." Finance and Economics Discussion Series 2023-042.
- Evans, G.W. and S. Honkapohja (2001), *Learning and Expectations in Macroeconomics*. Princeton University Press.
- Farmer, Leland, Emi Nakamura, and Jón Steinsson (2021), "Learning About the Long Run." Working Paper 29495, National Bureau of Economic Research.
- Faust, Jon and Jonathan H. Wright (2013), "Chapter 1 Forecasting Inflation." *Handbook of Economic Forecasting*, 2, 2–56.
- Fuhrer, Jeff (2018), "Intrinsic Expectations Persistence: Evidence from Professional and Household Survey Expectations." Working Paper.
- Fuster, Andreas, David Laibson, and Brock Mendel (2010), "Natural Expectations, Macroeocnomic Fluctuations." *Journal of Economic Perspectives*, 24, 67–84.
- Gabaix, Xavier (2019), "Behavioral Inattention." *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 261–343.
- Gennaioli, Nicola and Andrei Shleifer (2010), "What Comes to Mind?" *The Quarterly Journal of Economics*, 125, 1399–1433.
- Harvey, David I., Stephen J. Leybourne, and Paul Newbold (1998), "Tests for forecast encompassing." *Journal of Business & Economic Statistics*, 16, 254–259.

- Kahneman, Daniel and Amos Tversky (1972), "Subjective probability: a judgment of representativeness." *Cognitive Psychology*, 3, 430–454.
- Kent, Daniel and David Hirshleifer (2015), "Overconfident Investors, Predictbable Returns, and Excessive trading." *Journal of Economic Perspectives*, 29, 61–88.
- Kent, Daniel, David Hirshleifer, and Avanidhar Subrahmanyam (1998), "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance*, 53, 1839–1885.
- Kent, Daniel, David Hirshleifer, and Avanidhar Subrahmanyam (2001), "Overconfidence, Arbitrage, and Equilibrium Asset Pricing." *Journal of Finance*, 56, 921–965.
- Kohlhas, Alexandre N. and Answer Walther (2021), "Asymmetric Attention." *American Economic Review*, 111, 2879–2925.
- Kucinskas, Simas and Florian Peters (2022), "Measuring Under- and Overreaction in Expectation Formation." *Review of Economics and Statistics*.
- Lucas, Jr., Robert E. (1972), "Expectations and the Neutrality of Money." *Journal of Economic Theory*, 4, 103–124.
- Mankiw, Gregory and Ricardo Reis (2002), "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics*, 117, 1295–1328.
- Molavi, Pooya (2022), "Simple Models and Biased Forecasts." Working Paper.
- Moore, Don and Paul J. Healy (2008), "The Trouble with Overconfidence." *Psychological Review*, 115, 502–517.
- Nerlove, Marc (1958), "Adaptive Expectations and Cobweb Phenomena." *Quarterly Journal of Economics*, 72, 227–240.

- Pfajfar, Damjan (2013), "Formation of rationally heterogeneous expectations." *Journal of Economic Dynamics and Control*, 37, 1434–1452.
- Reis, Ricardo (2020), "Comment on 'Imperfect Expectations: Theory and Evidence'." *NBER Macroeconomics Annual 2020*, 35, 99–111.
- Rozsypal, Filip and Kathrin Schlafmann (2023), "Overpersistence Bias in Individual Income Expectations and its Aggregate Implications." *American Economic Journal: Macroeconomics*, 15, 331–71.
- Ryngaert, Jane (2023), "What Do (and Don't) Forecasters Know About U.S. Inflation?" *Journal of Money, Credit and Banking.*
- Scotese, Carol A. (1994), "Forecast Smoothing and the Optimal Under-utilization of Information at the Federal Reserve." *Journal of Macroeconomics*, 16, 653–670.
- Sims, Christopher (2003), "The Implications of Rational Inattention." *Journal of Monetary Economics*, 50, 665–690.
- Vuong, Quang H. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, 57, 307–333.
- West, Kenneth D. (2001), "Tests for forecast encompassing when forecasts depend on estimated regression parameters." *Journal of Business & Economic Statistics*, 19, 29–33.
- West, Kenneth D. (2006), "Chapter 3 forecast evaluation." volume 1 of *Handbook of Economic Forecasting*, 99–134, Elsevier.
- Woodford, Michael (2001), "Imperfect Common Knowledge and the Effects of Monetary Policy." *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund Phelps.*