Behavioral Impulse Responses

Bryan Kelly, Semyon Malamud, Emil Siriwardane, and Hongyu Wu*

This version: March 21, 2025

Abstract

We develop the concept of a Behavioral Impulse Response (BIR), which uses the dynamics of forecast errors to trace out how deviations from full-information rational expectations (FIRE) are corrected over time. BIRs based on professional forecasts of macroeconomics outcomes and corporate earnings imply that violations of FIRE occur much more frequently than suggested by existing tests. These deviations tend to correct gradually, often over several quarters, with sizable variation in correction speeds across different forecast targets and forecasters. Our theoretical analysis highlights why BIRs provide a simple yet powerful set of moments that can be used to discipline models of belief formation.

JEL: C52, C53, D83, D84, E7, E17, G17, G14

*Bryan Kelly is at AQR Capital Management, Yale School of Management, and NBER. Semyon Malamud is at Swiss Finance Institute, EPFL, and CEPR, and is a consultant to AQR. Emil Siriwardane is at Harvard Business School and NBER. Hongyu Wu is at Yale School of Management. We are grateful to Nick Barberis, John Y. Campbell, Robert Engle, Sam Hanson, Max Miller, Marco Sammon, Josh Schwartzstein, Andrei Shleifer, Jeremy Stein, Adi Sunderam, Luis Viceira, and seminar participants at Yale SOM for their helpful comments. AQR Capital Management is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR.

1 Introduction

There is a growing recognition within macroeconomics and finance that households, investors, and corporate managers do not always adhere to full-information rational expectations (FIRE) when making predictions.[1] According to this paradigm, individuals should form statistically efficient forecasts that optimally utilize all available information. Yet, in reality, forecasts often deviate from FIRE due to behavioral biases (Bordalo et al., 2020b; Afrouzi et al., 2023), imperfect information (Coibion and Gorodnichenko, 2012, 2015), and model uncertainty (Timmermann, 1993; Schwartzstein and Sunderam, 2021; Farmer et al., 2023). The prevalence of these deviations raises a natural next question: do people "wake up" rapidly when their forecasts depart from FIRE or do corrections occur more gradually?

To appreciate the practical importance of this question, consider a world in which corporate managers are fully rational, but investors overreact to positive news (Stein, 1996). If investors correct themselves quickly after an overreaction, then managers may ignore short-lived stock price inflation when making hiring, investment, and capital structure decisions. Conversely, managers may react more strongly when markets overreact if such errors are long-lived, thereby exerting tangible effects on the economy.

In this paper, we develop a new tool called a behavioral impulse response (BIR) that characterizes the process by which deviations from FIRE are corrected over time. As their name suggests, BIRs draw inspiration from a long tradition in macroeconomics that uses vector autoregressions (VARs) to study how shocks propagate dynamically through a system (Sims, 1980). In a similar spirit, BIRs capture how forecast errors evolve following a forecast revision.[2] BIRs are extremely portable and easy to estimate, relying only on OLS regressions

[1] For instance, departures from rationality have been used in models of the macroeconomy (Sims, 2003; Woodford, 2003; Mankiw and Reis, 2002; Gabaix, 2020; Pflueger et al., 2020), corporate practice (Baker et al., 2007; Malmendier, 2018), and capital markets (Barberis et al., 2015a, 2021).

[2] Properly speaking, impulse responses are defined as a response to "shocks" or other residuals in an economic model. BIRs are responses to innovations in forecasts, which are always catalyzed by an impulse of some sort (information arrival, behavioral error, etc.), but not necessarily structural shocks in the traditional sense (e.g, in a structural VAR).

and an observable term structure of forecasts. Furthermore, as we show more precisely below, they deliver a simple yet powerful set of moments that can help differentiate between models of belief formation and thus guide the development of theory.

To understand how BIRs work, consider a sequence of forecasts $F_t, F_{t+1}, ..., F_{t+n-1}$ for a fixed target y_{t+n}, such as 2025Q4 GDP growth, and define $e_t = y_{t+n} - F_t$ as the realized error associated with the time-t forecast, F_t. In an influential study, Coibion and Gorodnichenko (2015) develop a test for whether F_t adheres to FIRE by studying the conditional expectation of e_t following a forecast revision, $R_t = F_t - F_{t-1}$, at time t. BIRs build on this test by jointly studying the conditional mean of *all* future forecast errors e_t, e_{t+1}, ..., e_{t+n-1} following a revision R_t. More formally, we define the BIR, $B(j)$, at horizon j as follows:

$$B(j) = \mathbb{E}_t[e_{t+j}|R_t = 1] - \mathbb{E}_t[e_{t+j}|R_t = 0], \quad 0 \leq j < n.$$

The first point on the BIR, $B(0)$, is the object of interest in Coibion and Gorodnichenko (2015). Under FIRE, it—and the rest of the BIR curve—must equal zero because forecast errors should not be predictable. Negative values of $B(0)$ are often interpreted as evidence of overreaction, since upward revisions are reliably followed by more disappointing realizations of y_{t+n}. Conversely, positive values of $B(0)$ are generally viewed as underreaction because upward revisions are consistently followed by more positive forecast errors.[3]

Each subsequent point along the BIR curve then measures how future forecast errors are expected to evolve following a unit revision at time t. Equivalently, because the difference between two consecutive forecast errors reveals the revision between the two periods, the shape of the BIR captures how long the deviation from FIRE at time t remains in future

[3]There is some debate in the literature about the precise interpretation of $B(0)$ (Afrouzi et al., 2023; Kučinskas and Peters, 2022). We follow Coibion and Gorodnichenko (2015) and Bordalo et al. (2020b) in referring to $B(0) < 0$ as overreaction. Regardless, $B(0) \neq 0$ is a clear violation of FIRE.

forecasts. To see this more explicitly, notice that:

$$B(j-1) - B(j) = \mathbb{E}_t[R_{t+j}|R_t = 1] - \mathbb{E}_t[R_{t+j}|R_t = 0],$$

where $R_{t+j} = F_{t+j} - F_{t+j-1}$ is the one-period revision made at $t+j$. In the extreme, if the deviation from FIRE at t is permanent, then all future revisions will be zero on average so that $B(0) = ... = B(n-1) \neq 0$. This type of behavior will manifest as a perfectly flat BIR, whose level is determined by the nature of the reaction at t.

The BIR will take on a different shape when deviations from FIRE are short-lived. For illustration, assume that forecasts overreact at time t, meaning $B(0) < 0$. If forecasters correct this deviation immediately, all future forecast errors should average to zero, resulting in $B(1) = ... = B(n-1) = 0$. Additionally, the fact that forecasters correct their time-t overreaction after one period implies that their revision at $t+1$ is predictably negative. This correction process is directly embedded in the shape of the BIR (see Figure 1), since $\mathbb{E}_t[R_{t+1}|R_t = 1] - \mathbb{E}_t[R_{t+1}|R_t = 0] = B(0) - B(1) = B(0) < 0$.

The estimation of BIRs is straightforward and parallels the large literature on local projection methods in macroeconomics (see Jordà (2023) for a recent survey). In our empirical work, we estimate them using two datasets that are commonly used in the belief-formation literature. The first is the Survey of Professional Forecasters (SPF), a survey that contains forecasts by multiple individuals of various macroeconomic and financial series (e.g., Mankiw et al. (2003a), Coibion and Gorodnichenko (2012)). The second is the Institutional Brokers' Estimate System (IBES), a panel of individual equity analyst forecasts of earnings for publicly traded U.S. stocks (e.g., Bouchaud et al. (2019), Van Binsbergen et al. (2023)). Both datasets are sampled quarterly. Consistent with prior research, aggregate BIRs estimated by pooling forecasts of all series made by all forecasters show that individuals in both datasets tend to overreact (i.e., $B(0) < 0$), especially at long horizons (Giglio and Kelly, 2017; Bordalo et al., 2020b; Afrouzi et al., 2023). This overreaction is corrected gradually

over time. For example, when forecasters in SPF predict macroeconomic outcomes that will be realized in three quarters, about 40% of the overreaction from their time-t forecasts persists in forecasts made two quarters later.

Across forecasters, statistical tests based on BIRs show that deviations from FIRE are common in both the SPF and IBES, occurring far more frequently than suggested by existing tests in the literature. For the longest available forecast horizon in SPF (three quarters), almost 60% of forecasters reject the null of FIRE using BIRs, compared to less than 20% using the Coibion and Gorodnichenko (2015) test (henceforth CG-test). Importantly, these rejection rates account for the multiple hypotheses testing problem following Benjamini and Hochberg (1995). The difference between the two tests of FIRE is even more pronounced in IBES, where forecaster-level rejection rates via BIRs are well over 70% for horizons exceeding six quarters and are about ten times higher than those implied by CG tests. At the series level (i.e. forecast targets), BIRs also demonstrate a substantially enhanced ability to detect statistical deviations from FIRE.

For most series in SPF and IBES, these deviations tend to be in the form of overreaction, at least at longer horizons (Bordalo et al., 2020b, 2024). The speed with which they are corrected, however, varies widely across series. This is true in both the SPF and IBES. To illustrate, consider the BIR for six-quarter ahead forecasts in IBES, the longest horizon for which a reasonably sized panel of forecasts is available. For some series, correction occurs almost immediately, within one quarter, while for others, little to no correction is evident for nearly a year. Some of these differences can be attributed to series persistence, with less persistent series exhibiting more overreaction but relatively faster correction.[4] At the same time, correction speeds still vary meaningfully even within series that have comparable degrees of under or overreaction at time t, suggesting that the size of deviations from FIRE

[4]Overreaction also tends to be slightly higher for series with more volatility (e.g., based on stock returns), though correction speeds are somewhat invariant to series volatility.

and the correction process are likely governed by distinct properties of the data-generating process.

We also document economically meaningful differences in the level and shape of BIRs across forecasters, again focusing on forecast horizons of three and six quarters in SPF and IBES, respectively. Roughly 70% of forecasters in both datasets exhibit overreaction to news at time t. Moreover, there is sizable variation in correction speed across forecasters in SPF and IBES, much more so than across series. For instance, about one-fourth of forecasters in SPF overreact to news at time t yet show no correction in forecasts made two quarters later. Conversely, overreaction for an equal number of SPF forecasters is fully corrected within two quarters. Furthermore, corrections vary strongly within forecasters who react similarly to news at time t. This fact implies that deviations from FIRE and the process by which they are corrected are likely driven by distinct cognitive factors. Interestingly, BIRs for both IBES and SPF forecasters converge to FIRE over time, consistent with some form of learning. These broad patterns are also present in a subsample of SPF forecasters who cover the majority of possible series, alleviating concerns that selection of forecasters into certain series mechanically drives cross-sectional variation in BIRs.

In the latter part of the paper, we use the estimated BIRs to assess the strengths and weaknesses of different models of belief formation.[5] Our analysis is organized around a general class of models—what we call the "linear autonomous class"—that encompasses a wide range of belief formation processes that have been studied in prior work, including adaptive expectations (Cagan, 1956; Nerlove, 1958), diagnostic expectations (Bordalo et al., 2020a; Bianchi et al., 2024a), noisy information and sticky information (Giannoni and Woodford, 2003; Coibion and Gorodnichenko, 2012), and many others. Our primary theoretical

[5] A natural concern with such an exercise is that the BIRs we estimate may not purely reflect forecaster beliefs and could at least partially reflect strategic considerations. For example, equity analysts in IBES may alter their forecasts to enhance their relationship with corporate managers. For this reason, we focus on properties of BIRs that are robust across both IBES and SPF, as the latter contains only macroeconomic outcomes. See Section 4 for additional discussion.

contribution is to derive a closed-form expression for the BIR of the entire linear autonomous class.

To illustrate the paper's main proposition, we focus on three popular models of belief formation, starting with the model of diagnostic expectations used by Bordalo et al. (2020a). A calibrated version of the model using estimates from the literature shows that diagnostic expectations succeed in matching the level of BIRs observed empirically, both in the SPF and IBES datasets. Our theoretical analysis further reveals that models of diagnostic expectations imply an abrupt correction process: in standard applications, any overreaction at time t is fully corrected within one period. While this appears to be a reasonable approximation for some forecasters in SPF and IBES, most have a correction process that is fairly slow and smooth.

Next, we analyze the model from Afrouzi et al. (2023), who study belief formation when forecasters face information processing costs. By design, their model is quantitatively successful at matching the level of BIRs and the dependence of overreaction on the persistence of the forecast target. Costly information processing also implies a correction process that is smooth and decelerating, with most of the gains occurring in the first few periods, especially for series with low persistence. However, in the data, there are a meaningful number of forecasters for whom the correction process is extremely slow at first, only to accelerate after many periods. Moreover, the correction process does not depend as heavily on series persistence in the data as it does in the model.

The third model we study features extrapolative expectations, as in Hirshleifer et al. (2015). Under common calibrations used in the literature (e.g., Nagel and Xu (2022)), forecasters with extrapolative expectations overreact to a degree that resembles what is observed in SPF and IBES. Nevertheless, these models imply a correction process that is much too slow relative to the data.

In our last theoretical exercise, we move outside the class of linear autonomous models and

demonstrate how BIRs can be paired with simulations to study virtually any model of belief formation. We specifically analyze the model from Farmer et al. (2023), in which Bayesian agents must form forecasts of a process consisting of a unit root and a transitory component, both of which are unobserved. Farmer et al. (2023) show how model uncertainty of this kind severely slows Bayesian learning, leading to persistent deviations from FIRE, like those documented by Coibion and Gorodnichenko (2015). Using their simulation code, we ask whether model uncertainty can also generate realistic BIRs. We find that model uncertainty shows some promise in matching empirical BIRs, even those with a slow correction process, but the success depends critically on the initial priors of the forecasters. Whether the initial priors we consider are plausible is an open and interesting question for future research.

Literature Review As mentioned above, BIRs extend the influential tests of FIRE designed by Coibion and Gorodnichenko (2015). This extension yields two primary benefits. First, tests of FIRE via BIRs offer more statistical power because they are based on joint restrictions across the entire BIR curve, as opposed to tests of a single restriction like the one proposed by Coibion and Gorodnichenko (2015). Second, and more importantly, BIRs shed light on the dynamics of correction by examining the temporal evolution of forecasts for a fixed target. Our theoretical analysis shows that these dynamics are useful for disciplining models of belief formation. In this respect, our approach is closely related to recent work by Augenblick and Rabin (2021), who develop a test of Bayesian updating based on how forecasts of a fixed target evolve through time.[6]

Our paper also complement previous research that has estimated the impulse response of forecast errors to identified macroeconomic shocks, mainly to characterize information frictions facing forecasters (Coibion and Gorodnichenko, 2012), explain discrepancies between individual and consensus forecasts (Angeletos et al., 2021), and examine forecasters' reactions

[6] A large literature tests FIRE based on forecasts, including Mincer and Zarnowitz (1969), Lovell (1986), Thomas (1999), Mankiw et al. (2003b), Kohlhas and Walther (2021), Bianchi et al. (2022), and many others.

to specific structural shocks (Kučinskas and Peters, 2022).[7] These studies use forecast errors of a moving target (e.g., the time-series of one-period ahead forecast errors), whereas BIRs are based on forecast errors of a fixed target. As mentioned above, this subtle distinction is precisely what allows BIRs to characterize the correction process. Another nice feature of BIRs is they do not require one to take a stance on whether forecasters—or econometricians for that matter—can observe and identify structural shocks, which is especially challenging in macroeconomic settings (Bauer and Swanson, 2023). Like CG tests, BIRs rely on the much weaker assumption that forecasters are aware of their own history of revisions.

The remainder of the paper is organized as follows. In Section 2, we formally define BIRs and describe a simple way to estimate them. Section 3 then estimates BIRs in both the SPF and IBES datasets. Section 4 contains our theoretical analysis, which focuses primarily on how to use BIRs to assess models of belief formation. Section 5 highlights how BIRs can be paired with simulations to evaluate more complex models of belief formation, like the one studied by Farmer et al. (2023). Section 6 concludes. Additional details and results are also contained in a separate online appendix.

2 Behavioral Impulse Responses: Definition and Estimation

In this section, we formally define BIRs and highlight some of their basic properties. We then discuss why BIRs are useful for testing theories of belief formation. Finally, we describe how to estimate BIRs on any dataset in which forecasts of a fixed target (e.g., 2024Q4 GDP) are tracked over time.

2.1 Definition and Basic Properties

The notation employed in this paper closely follows the conventions used in fixed-income asset pricing research (e.g., Cochrane and Piazzesi (2005)). Let y_t denote the realization

[7] In recent work, Bianchi et al. (2024b) embed forecasts from a machine learning model in a structural asset pricing model to study how the stock market reacts to different types of news.

of a stochastic time series, such as GDP or corporate earnings growth. $F_t^{(n)}$ is the forecast of y_{t+n} made at time t, or the n-period ahead forecast of y. For now, we consider only forecasts of a single time series made by a single individual. It is straightforward to extend the notation to accommodate multiple time series and forecasters, as we do in the empirical analysis contained in Section 3. $e_t^{(n)} = y_{t+n} - F_t^{(n)}$ is the realized forecast error associated with the forecast $F_t^{(n)}$. For a fixed target y_{t+n}, the sequence of forecasts from t onwards is therefore given by $F_t^{(n)}, F_{t+1}^{(n-1)}, ..., F_{t+n-1}^{(1)}$. Similarly, $e_t^{(n)}, e_{t+1}^{(n-1)}, ..., e_{t+n-1}^{(1)}$ is the sequence of forecast errors. Finally, $R_t^{(n)} = F_t^{(n)} - F_{t-1}^{(n+1)}$ is defined as the one-period forecast revision of y_{t+n} made at t.

Next, we formally define the behavioral impulse response function.

Definition 1 *The behavioral impulse response function (BIR) for forecast horizon n is given by*

$$B(n,j) = \mathbb{E}\left[e_{t+j}^{(n-j)} | R_t^{(n)} = 1\right] - \mathbb{E}\left[e_{t+j}^{(n-j)} | R_t^{(n)} = 0\right], \quad j = 0, 1, ..., n-1 \tag{1}$$

Given a unit revision made at t, $\{B(n,j)\}_{j=0}^{n-1}$ traces out the expected sequence of forecast errors, holding fixed the forecast target y_{t+n}. As their name suggests, BIRs are rooted in the vast macroeconomics literature on impulse response functions, dating back at least to Sims (1980). In the usual macroeconomic setting, an impulse response function traces how an outcome of interest, like inflation, responds dynamically to a shock, like a surprise change in the stance of monetary policy. Analogously, in our setting, a behavioral impulse response traces out how forecast errors evolve following a revision R_t at time t.

Under full-information rational expectations (FIRE), BIRs take a very simple form that is useful for hypothesis testing: they must be zero for all forecast horizons, n, and periods j after the time-t revision. This follows from the fact that forecast errors are unpredictable under FIRE.

Remark 1 Under FIRE, $B(n,j) = 0, \quad \forall n, j$

The initial point along the BIR curve, $B(n,0) = \mathbb{E}\left[e_t^{(n)} | R_t^{(n)} = 1\right] - \mathbb{E}\left[e_t^{(n)} | R_t^{(n)} = 0\right]$, is familiar to behavioral economics research on belief formation (Bordalo et al. (2020b); Coibion and Gorodnichenko (2015); Farmer et al. (2023)). It measures the expected forecast error associated with a revision made at time t. When $B(n,0) < 0$, an upward revision made today predicts a more negative future forecast error. As mentioned in the introduction, this is typically interpreted as an overreaction to the news because the information that prompted the upward revision results in a disappointing realization of y_{t+n}. By the same token, $B(n,0) > 0$ can be interpreted as underreaction to news relative to FIRE. The literature on belief formation has consistently found $B(n,0) < 0$ at the individual level and has built theories trying to explain this phenomenon (Bordalo et al. (2020b); Afrouzi et al. (2023)).

While the extant literature focuses primarily on the link between *today's* error and today's revision, the main innovation in this paper is to also study the link between *future* errors and today's revision. For example, if a forecaster tends to overreact to the news today, how quickly, if ever, do they realize their mistake? Do they revise their forecast immediately, or does it take time? The remaining points ($j > 0$) along the BIR curve answer these questions.

To see this more formally, first note that the difference between the forecast error at times $t+1$ and t reveals the revision at $t+1$:

$$\begin{aligned} e_{t+1}^{(n-1)} - e_t^{(n)} &= \left[y_{t+n} - F_{t+1}^{(n-1)}\right] - \left[y_{t+n} - F_t^{(n)}\right] \\ &= -\left[F_{t+1}^{(n-1)} - F_t^{(n)}\right] \\ &= -R_{t+1}^{(n-1)}\,. \end{aligned}$$

Or, more generally, $e_{t+1}^{(n-j)} - e_t^{(n-j+1)} = -R_{t+1}^{(n-j)}$. From this identity, it follows that:

$$B(n, j-1) - B(n, j) = \mathbb{E}\left[R_{t+j}^{(n-j)} | R_t^{(n)} = 1\right] - \mathbb{E}\left[R_{t+j}^{(n-j)} | R_t^{(n)} = 0\right], \quad j > 0.$$

This equation says that the difference between two successive points on the BIR curve reveals the expected revision between those dates, conditional upon a unit revision made at time t. Thus, BIRs embed the trajectory of future revisions following a revision made at time t. Under FIRE, these revisions should also not be predictable by $R_t^{(n)}$, implying a perfectly flat BIR.

Stylized Example Returning to the question posed above, consider a data-generating process where the rational forecast of y_{t+n} is always zero, like, for example, an i.i.d process with zero means. Next, suppose that at time t the forecaster irrationally revises its forecast up from $F_{t-1}^{(n+1)} = 0$ to $F_t^{(n)} = 1$, implying $R_t^{(n)} = 1$. Further, suppose that the forecaster comes to its senses in the next period, revising the forecast back down and keeping it at the rational forecast after that.

Figure 1 plots the implied BIR in this stylized example for $n = 10$. The first point on the BIR curve is $B(n, 0) = -1$ because the revision made at time t leads to a forecast that overshoots by one unit. The remaining points along the BIR curve are all zero because the forecaster realizes the mistake and makes no more errors. The mistake correction also implies that the revision at that time $t + 1$ is $R_{t+1}^{(n-1)} = -1$. The magnitude of this correction is naturally embedded in the BIR because $B(n, 0) - B(n, 1) = \mathbb{E}\left[R_{t+1}^{(n-1)} | R_t^{(n)} = 1\right] - \mathbb{E}\left[R_{t+1}^{(n-1)} | R_t^{(n)} = 0\right] = -1$

As this example highlights, BIRs measure the initial degree of over- or under-reaction to news at time t and how subsequent revisions then respond. This is the sense in which BIRs summarize the dynamics of belief formation. Consequently, any theory of belief formation must match the level and shape of BIRs that are estimated empirically. BIRs are also useful

because they offer a natural way to study how belief formation varies across forecasting environments. For instance, they can be used to characterize how belief formation varies across individuals, stocks, or the nature of news. We highlight these applications later in Section 3, after describing how to estimate BIRs.

2.2 Estimation and Inference

Estimation BIRs can be easily estimated using any data that tracks forecasts of a fixed target through time. Given these datasets often contain multiple series and forecasters, we now extend the baseline BIR model as follows. Let $y_{s,t+n}$ denote the realization of series s at time $t+n$, where s could be GDP growth or the earnings per share of a publicly listed firm. $F^{(n)}_{i,s,t}$ is the forecast made by individual i at time t of the realization of series s in n periods and $e^{(n)}_{i,s,t} = y_{s,t+n} - F^{(n)}_{i,s,t}$ is the associated forecast error. $R^{(n)}_{i,s,t} = F^{(n)}_{i,s,t} - F^{(n+1)}_{i,s,t-1}$ is the revision made by forecaster i at time t for series s's realization at $t+n$.

The estimation process for BIRs closely aligns with the methodologies used in the macroeconomics literature. Specifically, assuming a linear conditional mean function, BIRs can be estimated by running n different regressions of forecast errors on past revisions:

$$e^{(n-j)}_{i,s,t+j} = c^{(n)}_j + \psi^{(n)}_j R^{(n)}_{i,s,t} + v^{(n-j)}_{i,s,t+j}, \quad \forall j = 0,,,,.n-1, \tag{2}$$

in which case the $B(n,j) = \psi^{(n)}_j$. This estimation strategy is akin to the local projection methods described originally by Jordà (2005) and summarized recently by Jordà (2023). Non-linear conditional mean functions are straightforward to accommodate as well. The alternative to estimating impulse responses with local projections is through a vector autoregression (VAR). As shown by Plagborg-Møller and Wolf (2021), the VAR-based and local projection approaches deliver the same impulse response in the population, though their finite-sample properties may differ. In general, VARs provide more efficient estimates but at the cost of higher bias, partly because they depend on appropriately defining the

lag length of the model (Li et al. (2022)). With these considerations in mind, we proceed with local projections mainly because they are easier to adapt to panel settings of the kind considered in our empirical work and because our applications use large panels in which precision is less of a concern.

The estimation strategy in regression (2) imposes homogeneity of BIRs across individuals, time, and targets (series). The resulting estimates should, therefore, be viewed as roughly the average BIR within a given dataset. The assumption of homogeneity can be easily relaxed by estimating BIRs on different subsamples of the data, like for each individual i or series s. We explore BIR heterogeneity later in our empirical work.

Inference of Point Estimates Jordà (2023) discusses the main inference issues when estimating local projections using macroeconomic data. As he points out, the residuals of a local projection regression generally have a moving average structure that can affect the construction of standard errors. Consequently, corrections targeting serial correlation in errors, such as the popular Newey-West HAC estimators, are preferable. In our panel context, clustering standard errors within each forecaster-series cell, (i, s), is a simple way to account for these regression error structures, though more efficient solutions are possible. In our empirical work, we also cluster standard errors within each revision date and series cell, (t, s), to account for common information shocks that may induce revisions for a series to be cross-sectionally correlated at a given time.

Hypothesis Testing For many applications, it is desirable to test the coefficient restrictions of an entire BIR curve jointly. For instance, testing whether an estimated BIR can reject the null hypothesis of FIRE amounts to a joint test of whether $B(n, 0) = B(n, 1) = ... = B(n, n-1) = 0$. An important advantage of this joint test is that it offers more power to reject FIRE than testing the single restriction of $B(n, 0) = 0$, as is commonly done in the literature (e.g., Coibion and Gorodnichenko (2015)).

To implement such a test, notice that equation (2) can be estimated for all j in a single panel regression that appropriately stacks the data. Specifically, suppose there are $i = 1, ..., I$ forecasters, $s = 1, ..., S$ series, and $t = 1, ..., T$ periods. This means the last n-period forecast is made at $T - n$. The entire BIR curve can then be estimated in one pass by running the following stacked regression:

$$\begin{aligned}
e_{i,s,t}^{(n)} &= c_0^{(n)} + \psi_0^{(n)} R_{i,s,t}^{(n)} + v_{i,s,t}^{(n)} \\
e_{i,s,t+1}^{(n-1)} &= c_1^{(n)} + \psi_1^{(n)} R_{i,s,t}^{(n)} + v_{i,s,t+1}^{(n-1)} \\
e_{i,s,t+2}^{(n-2)} &= c_2^{(n)} + \psi_2^{(n)} R_{i,s,t}^{(n)} + v_{i,s,t+1}^{(n-2)} \\
&\vdots \\
e_{i,s,t+n-2}^{(2)} &= c_{n-2}^{(n)} + \psi_{n-2}^{(n)} R_{i,s,t}^{(n)} + v_{i,s,t+n-2}^{(2)} \\
e_{i,s,t+n-1}^{(1)} &= c_{n-1}^{(n)} + \psi_{n-1}^{(n)} R_{i,s,t}^{(n)} + v_{i,s,t+n-1}^{(1)},
\end{aligned}$$

for all i, s, t. This one-shot regression approach simply requires stacking the data used to run each individual regression in (2) and then allowing the intercept and slope to vary with j. Here, it is critical to cluster standard errors within each revision date and series cell, (t, s), since revisions will clearly be mechanically correlated along this dimension. Hypothesis testing on the vector $\boldsymbol{\psi} = (\psi_0^{(n)}, \psi_1^{(n)}, ..., \psi_{n-1}^{(n)})'$ of BIR coefficients then proceeds in the standard fashion. Testing whether two BIR curves from subsamples of the same overall dataset are equal can be achieved using the same stacking procedure, then allowing the intercepts and slopes to vary with j and the subsamples of interest (e.g., a particular series).

3 BIRs in Practice

In this section, we establish some basic properties of BIRs using forecasts from two standard datasets in macroeconomics and finance: the SPF and IBES. The SPF is a quarterly survey, administered by the Federal Reserve Bank of Philadelphia, that contains predictions from

professional forecasters on several macroeconomic and financial outcomes (e.g., nominal GDP). We cover the fifteen series listed in Table A1 of the Internet Appendix, which largely draws on Bordalo et al. (2020b). IBES contains analyst-by-firm level forecasts of quarterly EPS, semi-annual EPS, annual EPS, and long-term EPS growth over a number of horizons. We focus on quarterly EPS forecasts because they are the most populated for longer forecast horizons. We largely follow standard practice from the literature when cleaning and processing both datasets. These steps are detailed in Internet Appendix IA.1, where we also define forcasted growth rates, forecast errors, and discuss how we handle outliers. For a given series s, forecast horizon n, and date t, a forecaster must have a complete set of forecasts from $t-1$ to $t+n$ to be included in the analysis.

3.1 Aggregated BIRs

We begin our empirical analysis by estimating aggregate BIRs based on forecasts pooled across all series and forecasters.

3.1.1 SPF

Figure 2a contains the estimated BIR from the SPF dataset for $n = 3$, the longest available horizon in the SPF, along with 95% confidence bands based on standard errors that are clustered by forecaster-series, (i, s), and series-date, (s, t). Table 1 presents the full set of point estimates contained in the plot, their standard errors, and the number of observations used to estimate the BIR. Plots for shorter-horizon BIRs ar presented in Figure A2a of the Internet Appendix. The first point on the BIR in Figure 2a, $B(n, 0)$, is negative and statistically different from zero. This finding reflects the well-known fact that individual forecasters in the SPF typically overreact to new information (Bordalo et al., 2020b).[8]

[8]Our baseline estimate of $B(3, 0)$ is not directly comparable to any single Coibion and Gorodnichenko (2015) regression coefficient reported in Bordalo et al. (2020b) because we use slightly different definition of growth rates and pool forecasts across all series (see Internet Appendix IA.1.1).

The novelty of BIRs lies in the subsequent parts of the curve, like for instance $B(3,1)$. Its magnitude implies that following a unit revision to a forecast made at t of y_{t+3}, the magnitude of the forecast error next quarter is roughly two-fifth of the forecast error made at time t. In other words, on average, 60% of the initial overreaction in the time-t forecast is undone within one period. The next point on the BIR curve, $B(3,2)$, is roughly equal to $B(3,1)$, suggesting no further correction from $t+1$ and $t+2$. By $t+3$, forecasters have more or less corrected their initial overreaction. This finding is not surprising given that $B(3,3)$ is reflects the so-called nowcast of y_{t+3} that is made midway through quarter $t+3$

The bottom two rows of Table 1 show p-values for two different joint tests of each BIR curve. The first, labeled $p(FIRE)$, tests the null hypothesis that the estimated BIR is generated under FIRE. As mentioned above, this is a joint test of whether $B(n,j) = 0$ for all j. At all horizons, we reject the null of FIRE with p-values below 0.001. The last row of Table 1 presents p-values from a weaker test of FIRE, namely that all points along the BIR curve are equal, even if they differ from zero. Recall from the discussion in Section 2.1 that this is equivalent to testing whether future revisions can be predicted by current ones, which should not be the case under FIRE. One might also label this as a test of forecaster "stubbornness" since it uncovers whether forecasters tend to stick with their forecast after a revision. Unsurprisingly, this null is easily rejected for all n.

3.1.2 IBES

Figure 2b plots the average BIR in the IBES data for $n = 6$, the longest horizon for which we have a reasonably sized panel of forecasts. Internet Appendix IA.1.2 provides complete details on how we clean and process IBES and Figure A2b presents BIRs for all possible horizons. The confidence bands in Figure 2b are based on standard errors that are clustered by analyst-stock and stock-date, and the full set of point estimates are also reported in Table 2. The first point on the BIR, $B(6,0)$, indicates that the average IBES forecaster

overreacts to news received at t, much like the average SPF forecaster.[9] The remaining points of the curve show that this deviation from FIRE is corrected gradually and at a relatively slower pace compared to the SPF. By $t+1$, only about 20% of the overreaction in the time-t forecasts has been corrected, and the initial deviation's half-life extends beyond a year. Unsurprisingly, as shown in the bottom two rows of Table 2, the BIR easily rejects the null of FIRE.

The BIR in Figure 2b supports a crucial aspect of the hypothesis proposed by Bordalo et al. (2024), henceforth BGLS, which centers on the established observation that high stock valuation ratios at time t are predictive of lower future returns (Campbell and Shiller, 1988). Contrary to theories positing time-varying discount rates as the explanation, BLGS argue that return predictability stems from overreaction in forecasts of future cash flows. The basic idea is that the market irrationally exaggerates its response to the news at time t; positive news leads to overly optimistic expectations of future cash flows and unwarranted inflation of stock prices. This overvaluation is corrected over time as earnings growth fails to meet expectations, resulting in a reversion in prices and predictability low returns. At the heart of this process is the market's eventual acknowledgment of its initial error. Figure 2b offers direct evidence of this correction, illustrating that IBES forecasters do adjust their forecasts, albeit slowly, after an initial overreaction to the news.[10]

3.2 How pervasive are deviations from FIRE?

Relative to existing methods, statistical tests of FIRE based on BIRs are more powerful because they are based on joint restrictions on the conditional dynamics of current and

[9] Bouchaud et al. (2019) document underreaction in consensus forecasts for 1-year forecasts of annual, not quarterly, EPS. We show in Internet Appendix IA.2.5.1 that their focus on consensus forecasts and annual EPS drives the differences between our results.

[10] Bordalo et al. (2024) focus primarily on forecasts of long-term growth (LTG) for S&P 500 firms. BIRs are harder to estimate for LTG forecasts because they have an ambiguous horizon, ranging from 3-5 years. Similar to BLGS, however, we find overreaction in individual forecasts at this horizon, i.e., $B(6,0) < 0$. In Internet Appendix IA.2.5.3, we also confirm similar BIRs for large stocks, like those studied by BLGS.

future forecast errors. In this subsection, we showcase this feature of BIRs by testing the null hypothesis of FIRE in the cross-section of series and forecasters. We then benchmark the resulting rejection rates against those from the popular test proposed by Coibion and Gorodnichenko (2015), henceforth the CG-test.

Forecast Targets (Series) Table 3a summarizes the fraction of series in the SPF and IBES that reject one of two tests of FIRE. The first is CG-test of whether the initial point $B(n, 0)$ on the BIR equals zero. The second is based on whether the entire BIR curve equals zero. As discussed in Section 2, this is equivalent to jointly testing whether forecast revisions at time t predict contemporaneous forecast errors and subsequent revisions. To construct the table, we estimate BIRs for every horizon n and series s in either the SPF or IBES, pooling data across all forecasters covering each series. A series-horizon cell must have a least 10 observations to be included in the analysis. For each horizon n, the table reports the fraction of series for which the null of FIRE is rejected under both tests at a 5% confidence level. For both tests, we adjust individual p-values to account for the multiple testing problem following the procedure in Benjamini and Hochberg (1995).

Consistent with Bordalo et al. (2020b), the first row of Table 3a shows that the large majority of series in the SPF reject the CG-test of FIRE, regardless of horizon. The second row contains the fraction of series that reject the null of FIRE based on the entire BIR curve. Rejection rates are moderately larger for all forecast horizons when using this more stringent test.

The third and fourth rows of Table 3a present the results for series in IBES. Rejection rates using the CG-test are fairly low for almost all horizons, typically around 5%. Rejection rates are substantially higher when using the entire BIR, especially for longer horizon forecasts. At shorter horizons, forecasts for about 30% of series reject the null of FIRE using BIR-based tests, roughly eight times the amount that do with CG-tests. The discrepancy between CG and BIR-based tests becomes even more pronounced for longer-horizon forecasts.

For forecast horizons over six quarters, rejection rates are between seven and sixty times higher using BIRs and approach 60%. This is at least partly driven by the fact that the number of restrictions increases with horizon.

Forecasters Table 3b summarizes tests of FIRE in the cross section of forecasters. Unlike Table 3a, which is reflects BIRs estimated for each series, this table is based on BIRs that are estimated for each forecaster i. Under CG-tests, slightly less than 20% of forecasters in the SPF reject the null hypothesis of FIRE at both $n = 2$ and $n = 3$. These rejection rates increase threefold when using the entire BIR. The contrast between the two tests is even more striking for IBES forecasters: across all forecasting horizons, about 5% of equity analysts in IBES reject CG-tests. The fourth row of Table 3b shows that rejection rates rise by a factor of five to nearly 30% based on BIRs for shorter-horizon forecasts ($n \leq 3$). For longer-horizon forecasts, over 70% of IBES forecasters reject the null of FIRE, which is nearly ten times the rate observed under CG-tests.

To summarize, the results in Table 3 show that deviations from FIRE are common in both the SPF and IBES datasets, occurring much more frequently than suggested by existing tests in the literature. Table A5 in the Internet Appendix highlights the importance of correcting for the multiple hypothesis testing problem in this exercise. For example, without the Benjamini and Hochberg (1995) correction, almost 30% of forecasters in SPF reject FIRE based on CG-tests, compared to about 20% after the correction. The Benjamini and Hochberg (1995) correction impacts rejection rates even more for IBES because the number of forecasters and series is much larger. Under CG-tests, rejection rates fall by a factor of about four for both series and forecasters after the correction. Rejection rates are much higher overall—and the attenuation from the correction is much weaker—when using BIRs to test for deviations from FIRE. BIRs offer an improved ability to statistically detect these deviations by imposing joint restrictions on the dynamics of forecast errors and revisions.

This observation parallels Jordà (2023), who argues that hypothesis testing using traditional impulse responses offer more statistical power than tests based on individual point estimates.

3.3 Series-Level Analysis

While the preceding analysis points to frequent deviations from FIRE in the cross section of series and forecasters, it is silent on the magnitude of these deviations and the speed with which they are corrected. In this subsection, we quantify variation in the level and shape of BIRs varies across series and explore possible correlates of correction speed.

3.3.1 How much do BIRs vary across series?

Table 4a summarizes how the initial level, $B(3,0)$, of BIRs varies across all fifteen series in the SPF. The table focuses on $n = 3$, the longest possible forecast horizon for BIRs in the SPF. The median series in SPF exhibits overreaction in the CG-sense, such that forecasts at time t overshoot their target by -0.35 units for every unit revision at t. All but three series in SPF overreact, again consistent with the findings of Bordalo et al. (2020a).[11]

Table 4b examines how much correction speeds vary across series. $W_j = B(3,j)/B(3,0)$ is a natural measure of correction speed, as it captures how much deviations from FIRE at time t remain in forecasts made j periods later. High values of W_j indicate a slow correction process. For example, suppose $B(3,0) < 0$ and $W_1 = 1$. In this case, all of the overreaction in the forecast made at time t remains in the forecast made one quarter later. On the other hand, a value of $W_1 = 0$ means deviations from FIRE are on average corrected immediately. To construct Table 4b, we split series s into two groups based on the sign of $B(3,0)$. Series that initially overreact ($B(3,0) < 0$) are further sorted into terciles based on W_1. For all series in a given tercile, we estimate a pooled BIR and report estimates of W_j, with standard errors based on the delta method. The bottom two rows of Table 4b report the pooled

[11] The three series that exhibit initial underreaction are the T-Bill yield, growth in GDP Price Index (PGDP), and growth in housing starts (HOUSING)

$B(3,0)$ estimate and the number of series in each group. The last column of the table also contains estimates from a pooled BIR of the three series that underreact in the SPF.

Correction speeds vary meaningfully across series in the SPF. In the overreacting and low-W_1 group, the small and statistically insigificant W_i indicates that overreaction at time t is fully undone within one quarter. In contrast, for overreacting and high-W_1 series, forecasts made one and even two periods later still embed about half of the overreaction from the time-t forecast. Interestingly, the initial degree of overreaction is similar across the two groups, a fact that we return to in more detail below. Figure 4a visualizes the results in Table 4b by plotting the BIRs for the low and high-W_i overreacting series.

Next, we characterize variation in the level and shape of series in IBES. We focus on forecast horizons of six quarters ($n = 6$), again in order to study the properties of longer-horizon forecasts while still maintaining a reasonably sized panel of forecasts. The second row of Table 4a summarizes the level of BIRs in the cross-section of IBES. Forecasts made at time t for the median IBES series overreact, with a CG-coefficient of $B(6,0) = -0.35$. Moreover, the majority of series in IBES exhibit overreaction, though the proportion is slightly lower than in the SPF. The distribution of $B(6,0)$ in IBES also has much wider tails compared to the SPF.

Table 4c shows how correction speeds vary across series in IBES. The table is assembled similarly to Table 4b, with the exception that over and underreacting series are both split into quartiles based on W_1. This more granular sorting is possible due to the larger number of stocks in IBES relative to the series in SPF. The table reveals a striking degree of heterogeneity across series in IBES in terms of correction speed. For example, for overreacting and high-W_1 series, all of the overreaction from time t is still present in forecasts made one year later, whereas virtually none remains for the overreacting and low-W_1 group. Similar amounts of heterogeneity are present in series that underreact.

Figure 4b provides a visual depiction of BIR heterogeneity across series in IBES. For

readability, the plot shows the pooled BIR for series in the top and bottom quartiles of W_1, both for under and overreacting series. Nearly 20% of stocks in IBES overreact and correct so slowly that their BIRs are convex, whereas a similar fraction of series overreact but correct quickly enough to generate concave BIRs. The orange and blue lines show that there is comparable heterogeneity in correction speeds among stocks that underreact.

3.3.2 What factors determine correction speed?

We now study some potential sources of variation in correction speed across series, focusing on the following candidates: (i) the size of the initial deviation from FIRE, (ii) series volatility, and (iii) series persistence.

Initial Reaction Magnitude We first examine the relationship between the degree of over or underreaction at time t and the speed of subsequent correction. This link is important from the perspective of behavioral theories. To see why, suppose that the speed of correction is fully determined by the size of the initial deviation from FIRE at time t. For instance, one might expect series with large overreactions also to have slower correction speeds. Under this scenario, a realistic behavioral model could use a single aspect of the data-generating process to modulate both the degree of overreaction and the speed of correction. Conversely, if the speed of correction varies independently from the degree of overreaction, a model would need the dynamics of belief formation to depend on multiple aspects of the data-generating process.

The results in Table 4 provide an early indication that correction speeds are not completely pinned down by the degree of initial reaction. Consider, for example, the analysis of the SPF series in Table 4b. Across the first three columns, the level of overreaction is virtually identical. Yet, correction speeds W_1 vary meaningfully—in other words, some series overreact and correct quickly, whereas others overreact equally as much but correct slowly. More formally, we can reject the null hypothesis that W_1 is equal for the three groups at

conventional confidence levels.[12] In Table 4c, which contains results for IBES, there appears to be a positive relationship between $B(6,0)$ and W_1. When moving from the first to the fourth column, the speed of correction declines by construction, and the magnitude of $B(6,0)$ also increases, though not monotonically, suggesting an imperfect link between the two.

Figure 4c explores this issue in more detail using a series of double sorts, again focusing on IBES due to the much larger number of series. Specifically, series that overreact are first split into groups based on the magnitude of $B(6,0)$, then further split based on W_1. Within each resulting group, we then estimate a pooled BIR. The process is then repeated for series that underreact, resulting in a total of eight estimated BIRs. By design, this sorting process creates variation in correction speed W_1 while roughly holding fixed the degree of initial reaction $B(6,0)$. In the plot, series with similar levels of reaction are grouped by color, and those with slow (fast) corrections are depicted by a dashed (solid) line.

The figure reveals significant variation in correction speeds, even after accounting for the size of the initial deviation from FIRE. In the most overreacting group, depicted in blue, the dashed blue line represents the pooled BIR for the slow-correcting series, while the solid line represents the fast-correcting series. While we cannot reject the null hypothesis that the initial degree of overreaction, $B(6,0)$, differs between the two groups, the plot shows that their correction speeds are quite different. In the fast-correcting group, forecasts made one year later retain 48% of the initial overreaction from time t. In contrast, 75% of the initial overreaction persists in the slow-correcting group a year later, and these differences are statistically significant at conventional confidence levels ($p = 0.005$). Similar patterns emerge when looking at the other series in the plot. Another way to see the distinction between correction speeds and the initial reaction is to compare slow-correcting series that differ in their initial reaction, like those depicted by the dashed orange and blue lines. Overreaction at time t is over ten times larger for the series embedded by the dashed blue line, yet the

[12] The null of equal W_1 for the first and third groups is also rejected, as is the null of equal W_2 across all three groups.

correction process for the two groups appears very similar. We further confirm the relatively weak link between $B(n,0)$ and W_1 at the series level for both SPF and IBES in Internet Appendix Figure A6.

Volatility Next, we explore how correction speed varies with the properties of the underlying data-generating process for each series, starting with volatility. Figure 5a displays two pooled BIRs, one for series with high return volatility in IBES and another for series with low volatility. Volatility is defined as the average monthly stock return volatility in quarter t when the revision used for the BIR is made. Stock return volatilities are taken from Jensen et al. (2023). Low and high-volatility series are, respectively, those in the bottom and top terciles of stock return volatility.

The first thing to note from the figure is that overreaction appears more pronounced in series with high volatility, though this difference is not statistically significant. The correction processes for high-volatility stocks are comparable, albeit a bit slower, to low-volatility stocks during the first year. For high-volatility stocks, about 70% of the overreaction from time t remains in forecasts made one year later compared to about 50% for low-volatility stocks. Correction then accelerates after t+4 for high-volatility stocks. Figure 5b shows that a similar pattern holds in SPF, at least in the sense that series with high volatility exhibit larger initial overreaction. Volatility, in this case, is defined using the unconditional standard deviation of the residuals from fitting an AR(p) model, where p is chosen using the AIC criteria.

Persistence In recent work, Afrouzi et al. (2023) present experimental and observational evidence that overreaction is larger for series with low persistence. Motivated by their analysis, we investigate the dependence of correction speed on series persistence. Figure 5c shows pooled BIRs estimated for series that have low and high cash-flow persistence. Following Bouchaud et al. (2019), cash-flow persistence is defined based on an AR(1) model of each series' annual operating cash flows scaled by assets. Series with low persistence

are in the bottom tercile of AR(1) coefficient and vice versa. Consistent with Afrouzi et al. (2023), Figure 5c shows that series with low persistence have more overreaction. The correction process for this group is also slightly faster as well, though it is hard to draw strong conclusions given the size of the confidence bands. Figure 5d shows a similar story holds in the SPF, at least in the sense that low persistence is associated with more overreaction.

3.4 Forecaster-Level Analysis

In this subsection, we mirror the analysis from above to analyze how BIRs vary across forecasters rather than series.

3.4.1 How much do BIRs vary across forecasters?

The first row in Table 5a presents summary statistics for $B(0)$ across forecaster-level BIRs in SPF. As in Bordalo et al. (2020a), the median forecaster in SPF overreacts, with a CG-coefficient of -0.35. Moreover, nearly three-quarters of forecasters in SPF overreact. The table also shows that the initial degree of overreaction varies quite strongly across forecasters.

Table 5b summarizes differences in correction speeds across forecasters in SPF. Each column in the table contains estimates of W_j from a BIR that is pooled across forecasters that either overreact or underreact. Within the set of overreactors, groups are formed by splitting forecasters into terciles based on their individual W_1. For underreactors, we only split forecasters into two groups due to their more limited size. The table reveals a striking amount of variation in correction speed across forecasters. Consider, for instance, the set of overreacting forecasters in the first three columns. For those that correct the fastest, overreaction in time-t forecasts more or less reversed within one quarter and is fully undone within two. For those in the bottom tercile, 117% ($t = 8.74$) of the overreaction from time t remains in forecasts made one quarter later, and 95% ($t = 8.15$) remains two quarters later.

Figure 6a visualizes the results in Table 5b by plotting the BIRs for the slowest and fastest correcting forecasters.

For completeness, we repeat the preceding analysis for equity analysts in IBES. The second row in Table 5a shows how their individual $B(6,0)$ varies in the cross-section. Similar to SPF forecasters, most equity analysts in IBES overreact for $n = 6$. The median CG-coefficient is also comparable across the two datasets, though the tails of the distribution are somewhat wider in IBES.

Table 5c summarizes correction speeds across IBES analysts. Much like Table 5b, the table is constructed by sorting analysts based on the sign of $B(6,0)$, splitting them further based on W_1, then estimating pooled BIRs within each group. Because there are more analysts in IBES compared to SPF, analysts are sorted into quartiles based on W_1, as opposed to halves. Correction speeds vary strongly across equity analysts. For analysts who overreact and correct slowly, on average 128% ($t = 4.92$) of the overreaction at time t persists in forecasts made a year later. In contrast, for fast-correcting analysts, only 22% ($t = 3.37$) of the initial overreaction remains after one year. A similar degree of heterogeneity is present within the set of analysts that underreact. As Figure 6b shows, these differences in correction speeds give rise to both concave and convex BIRs.

Given the variation in BIRs across series documented in Section 3.3.2, one concern is that the forecaster-level variation in Table 5 and Figure 6 is driven by selection. To see why, suppose the entire BIR is fully determined by the properties of the data-generating process (e.g., its persistence) and that all forecasters are identical. For instance, all forecasters may be equally diagnostic in the model of Bordalo et al. (2019) or have the same priors in Farmer et al. (2023). Further, suppose that each forecaster covers only a subset of possible series. In this scenario, BIRs will appear to vary across forecasters, but only because of series are governed by different data-generating processes. There are a few reasons, however, to think that these selection issues are not driving our results. For one, compared to Figure 4a, Figure

6a shows there is substantially more variation across forecasters than across series in SPF. Moreover, in Internet Appendix Section IA.2.4, we document similar levels of heterogeneity across forecasters in SPF that cover all possible series. Thus, the variation in Table 5 and Figure 6 likely reflects properties of forecasters, not series.

3.4.2 Why do some forecasters correct faster than others?

Initial Reaction Forecasters may differ in correction speed simply because some over or underreact more than others. It is straightforward to see from Table 5 that this is not the case—some forecasters overreact and correct quickly on average, whereas others overreact by the same degree but correct slowly. For example, in Table 5b, $B(0)$ for overreacting SPF forecasters in the middle and top terciles of W_1 are roughly the same in magnitude, and we cannot reject the null that they are equal. However, for forecasters in the middle tercile, the forecast made at $t+2$ contains over half of the overreaction from the forecast made at time t. On the other hand, for forecasters in the top tercile, nearly all of the overreaction remains after two quarters. Similarly, in Table 5c, overreacting IBES forecasters in the bottom and top quartiles of W_1 have similar levels of initial overreaction but strongly differ in correction speed. These two groups are not small, representing 32% of all forecasters in IBES. The disconnect between initial reaction and correction speed in both SPF and IBES is also clear in Figures 6a and 6b. These patterns suggest that models of belief formation need at least two distinct belief-related parameters to jointly explain over or underreaction and the subsequent process of correction.

Experience Given the large and heterogenous deviations from FIRE documented in Figure 6, a natural question is whether BIRs are a static feature of IBES analysts or whether behavioral responses to information change over time. BIRs offer a simple way to study this question because they can be estimated over different phases of an analyst's career. Figure 7 implements this idea for analysts with at least one decade for forecasts in either SPF (Figure

7a) or IBES (Figure 7b). The threshold of ten years is roughly the median experience in both datasets. For a given forecaster i, we compute BIRs using the first five years of forecasts, requiring at least ten observations to be included. B_i^{early} is the average level of forecaster i's BIR during this "inexperienced" phase of their career. Forecasters are then sorted into three terciles based on B_i^{early}. This means that the most overreacting forecasters are in the bottom tercile and the underreacting are in the top. Within each tercile, we then estimate two pooled BIRs: one using data during the first five years of each forecaster's career and another based on forecasts for all subsequent years when they are more "experienced." The figure shows the resulting BIRs from both career phases for the group of forecasters who overreacted early in their careers. In the Internet Appendix Section IA.2.5.4 , we document similar results for the smaller set of underreactors.

Figure 7 reveals a robust relationship between BIRs and experience in both the SPF and IBES datasets: forecasters appear to become more rational as their careers progress. In SPF, overreaction is fairly large and relatively slow correcting for inexperienced forecasters, with nearly two-thirds of their overreaction at time t still present in forecasts made one quarter later. For experienced forecasters in SPF, overreaction in time-t forecasts is economically small and statistically insignificant, and any overreaction is fully corrected within one period. Figure 7b paints a similar picture for equity analysts in IBES, confirming that the findings from the SPF are not driven by any issues with how individual contributors are tracked in that dataset.[13]

3.5 State Dependence

Our final empirical analysis asks whether the level and shape of BIRs varies with the nature of news received by forecasters. A simple way to measure the sentiment of news received by each forecaster is based on the sign and size of their revision. For example, a large and positive revision is an indication that a forecaster received very positive news. Figure 8a

[13]For a complete discussion of these issues, see the SPF documentation.

implements this idea in IBES by estimating BIRs for $n = 6$ based on the sign of $R_{i,s,t}$ and its absolute magnitude. Specifically, large positive revisions are defined based on whether $|R_{i,s,t}|$ is in the top tercile of all positive revisions. All remaining positive revisions are considered small. Small and large negative revisions are defined analogously.

The most striking aspect of Figure 8a is that equity analysts *underreact* to negative news but *overreact* to positive news. Moreover, for negative news, underreaction is relatively larger for smaller negative news. At the same time, the correction speed for large negative news is much faster than for small negative news. On the other hand, overreaction to positive news is similar regardless of whether the news is large or small, as is the process of correction. In Internet Appendix IA.2.5.5, we show this result holds for different forecast horizons as well. Interestingly, Figure 8b shows that this asymmetry is not present in the SPF, where overreaction is present for both positive and negative news. Overreaction is larger, and the speed of correction is faster for positive news, though these differences are not estimated with great precision.[14]

One potential explanation for this result is that analysts in IBES may downplay negative news and overstate positive news in an effort to enhance their relationship with the management of the stocks they cover. However, the fact that underreaction is smaller for very bad news is somewhat challenging for this incentive-based view. Presumably, analysts have the strongest incentive to downplay the very worst news. We view pinning down the microfoundations of this result as an interesting avenue for future theory. Given it may reflect incentives, we do not consider this particular pattern when discussing the implications for theories of belief formation in later sections.

Figure 8c asks a related question: do forecasters react differently to news during recessions? Recession dates are taken directly from the Business Cycle Dating Committee of

[14]This type of pattern is harder to rationalize under the view that deviations from FIRE arise due to information frictions (Mankiw and Reis, 2002; Coibion and Gorodnichenko, 2012), given the ability of professional forecasters to observe the realization of y should presumably not depend on the sentiment of the news.

the National Bureau of Economic Research (NBER). The figure shows that BIRs for IBES forecasters depend heavily on the state of the economy. Outside of recessions (orange line), equity analysts exhibit a modest amount of overreaction that is then corrected almost entirely in one quarter. In contrast, they are almost seven times more overreactive during recessions (blue line). The speed of correction during recessions is somewhat linear. For $n = 6$, roughly half of the overreaction at time t is still embedded in forecasts one year later. We show in Internet Appendix Figure A9 that some of this overreaction can be attributed to the Covid-19 pandemic. Relative to IBES forecasters, Figure 8d shows that the belief formation process varies much less with the state of the economy for SPF forecasters.[15] While overreaction does appear marginally larger during recessions, we cannot reject the null hypothesis that the two BIRs are different ($p = 0.302$).

4 Using BIRs to Assess Behavioral Models of Belief Formation

In this section, we derive BIRs for a general class of behavioral models that allows for fairly flexible departures from rationality. This class encompasses a wide range of the most popular models in the literature, including extrapolative expectations (Hirshleifer et al., 2015; Nagel and Xu, 2022), diagnostic expectations (Bordalo et al., 2020b), and costly information acquisition (Afrouzi et al., 2023). We then assess the strengths and weaknesses of different models by comparing their implied BIRs with those estimated from Section 3, focusing on properties that are robust across SPF and IBES. This approach should alleviate concerns that forecasts in IBES do not fully reflect the true beliefs of equity analysts, given possible career incentives to cater EPS forecasts to firm management. Such incentives are presumably weaker or non-existent for SPF forecasters because their targets are macroeconomic and broad financial market outcomes. In other words, the properties of BIRs that exist in both

[15] For the SPF sample, the recession periods are 1969Q4-1970Q4, 1973Q4-1975Q1, 1980Q1-1980Q3, 1981Q3-1982Q4, 1990Q3-1991Q1, 2001Q1-2001Q4, 2007Q4-2009Q2, and 2020Q1-2020Q2.

datasets are more likely to reflect the true beliefs of forecasters and are, therefore, appropriate for calibrating models of belief formation.

4.1 General Setup

Throughout this section, we study the forecasting problem for a stochastic process $y_t,\ t \geq 0$. The information set of the forecaster is generated by a set of signals $S_t \in \mathbb{R}^P$, where some of the S_t signals can be lagged versions of y_t. Formally, this information set is described by the filtration $\mathcal{F}_t = \sigma(\{(y_\tau, S_\tau)\}_{\tau \leq t})$ generated by $(y_t, S_t) \in \mathbb{R}^{P+1}$. The stochastic process $Z_t,\ t \geq 0$, is adapted to $\mathcal{F} = \{\mathcal{F}_t\}$ if Z_t is $\mathcal{F}_t$-measurable for each $t \geq 0$. We use $E_t[y_{t+n}] = E_t[y_{t+n}|\mathcal{F}_t]$ to denote forecast of y_{t+n} under FIRE and $F_t[y_{t+n}]$ to denote the forecast based on potentially incomplete information or behavioral biases. When there is no ambiguity, we denote multi-period forecasts by $F_t^{(n)} = F_t[y_{t+n}]$ for $n \geq 1$.

To derive BIRs, we focus on a particular class of models in which forecasts depend linearly on the history of rational forecasts. This class, which we label as linear autonomous forecasts, is formally defined as follows.

Definition 2 *Forecasts F_t are defined as linear autonomous if they take the following form:*

$$F_t[y_{t+n}] \ = \ \sum_{\tau=0}^{\infty} \gamma(n,\tau) E_{t-\tau}[y_{t+n}] \ + \ \nu_t^{(n)} \tag{3}$$

for some coefficients $\gamma(n,\tau)$ and an independent noise process $\nu_t^{(n)}$ with $\mathrm{Cov}(\nu_t^{(n)}, \nu_{t-1}^{(n)}) = 0$ and $\mathrm{Var}[\nu_t^{(n)}] = \sigma_\varepsilon^2(n)$.

The coefficients $\gamma(n,\tau)$ define the loading of the forecast on current and past rational forecasts. Though it is of course possible to define a broader class of processes in which $F_t[y_{t+n}]$ is an arbitrary function of rational forecasts, we think it is natural to focus on linear functions because they preserve the linearity of forecasts across targets, meaning $F_t[a_1 y_{t+n} + a_2 \tilde{y}_{t+n}] \ = \ a_1 F_t[y_{t+n}] \ + \ a_2 F_t[\tilde{y}_{t+n}]$.

A key property of linear autonomous forecasts is that they encompass many popular behavioral models of belief formation, including adaptive expectations (Cagan, 1956; Nerlove, 1958), extrapolative expectations (Barberis et al., 2015b; Hirshleifer et al., 2015), noisy information and sticky expectations (Giannoni and Woodford, 2003; Coibion and Gorodnichenko, 2012), diagnostic expectations (Bordalo et al., 2020a), smooth diagnostic expectations (Bianchi et al., 2024a), misestimation of persistence (Gabaix, 2019), and costly information processing (Afrouzi et al., 2023). Appendix A shows how to cast each of these models in the linear autonomous class. As one example, consider the model of diagnostic expectations from Bordalo et al. (2020a). In this model, forecasts are given by:

$$F_t[y_{t+n}] = E_t[y_{t+n}] + \theta(E_t[y_{t+n}] - E_{t-J}[y_{t+n}]), \tag{4}$$

where the diagnosticity parameter θ modulates the size of the behavioral bias in the forecast, and J controls the number of revisions that drive the bias. $\theta = 0$ corresponds to FIRE and positive values of θ imply forecasts overreact. This model of beliefs clearly fits into the linear autonomous class in Definition 2. Specifically, for $J = 1$, $\gamma(n, 0) = 1 + \theta$, $\gamma(n, 1) = -\theta$, and $\gamma(n, \tau) = 0$ for $\tau > 1$.

4.2 BIRs for Linear Autonomous Forecasts

We now derive BIRs within the class of linear autonomous forecasts. To do so, we make the following assumptions about the data-generating process for y_t.

Assumption 1 *y_t evolves according to:*

$$y_{t+1} \;=\; \beta' S_t \;+\; \varepsilon_{t+1}\,, \tag{5}$$

for some vector $\beta \in \mathbb{R}^P$ and an i.i.d. stochastic process ε_t with zero mean and variance σ^2_ε.

Assumption 2 *The signals S_t evolve according to*

$$S_{t+1} \;= \rho S_t \;+\; \xi_{t+1}, \tag{6}$$

where $|\rho| < 1$ is a scalar that controls the (equal) persistence of signals, and ξ_{t+1} is an i.i.d. stochastic process with zero mean.

Assumption 3 *The process $q_t = \rho^{-1} E_t[y_{t+1}]$ is stationary with a long-run variance σ_q^2.*

Assumption 1 says that the true y_{t+1} is a linear function of lagged signals plus an error term. This assumption is without loss of generality because linearity can always be achieved by expanding the set of signals to include a non-linear transformation of the "true" set of signals (Kelly et al., forthcoming). Moreover, because S_t can include lags of y, Assumption 1 nests cases in which y_t is an arbitrary AR(q) process. Assumption 2 is stronger and puts a simple dynamic structure on the signals that inform each forecast. Combined with Assumption 1, it implies that the FIRE-forecast can be written as:[16]

$$E_t[y_{t+n}] \;=\; \rho^{n-1} E_t[y_{t+1}] \;=\; \rho^n q_t. \tag{7}$$

Assumption 3 then imposes that this rational forecast is stationary. Finally, we make one more additional assumption on the weight function $\gamma(n, \tau)$ to make the analysis of BIRs more tractable.

Assumption 4 (Separability) *The weight function $\gamma(n, \tau)$ that defines the belief formation process can be written as:*

$$\gamma(n, \tau) \;=\; k_0(n)\, \mathbf{1}_{\tau=0} \;+\; \delta(n)\kappa(\tau), \tag{8}$$

[16] Under a more general structure in which $S_{t+1} = \Psi S_t + \xi_{t+1}$, it is straightforward to show that forecasts under FIRE are given by $E_t[y_{t+n}] = \beta'(\Psi)^{n-1} S_t$.

where, without loss of generality, we impose the normalization $\sum_{\tau=0}^{\infty}\kappa(\tau)\rho^{2\tau} = 1$.

The notion of separability in Assumption 4 is satisfied in all of the behavioral models discussed above. For example, continuing with the example of diagnostic expectations from above and setting $J = 1$, $\kappa(0) = (1-\rho^2)^{-1}$, $\kappa(1) = -(1-\rho^2)^{-1}$, and $\kappa(\tau) = 0$ for $\tau > 1$. Moreover, $\delta(n) = \theta(1-\rho^2)$ and $k_0(n) = 1$. Under Assumptions 1-4, we can now derive the paper's main proposition regarding the shape of BIRs in the linear autonomous class of belief formation models.

Proposition 2 (BIRs in the Linear Autonomous Class) *Under Assumptions 1-4, BIRs are given by:*

$$B(n,j) = \frac{-\sigma_\varepsilon^2(n)\mathbf{1}_{j=0} + A(n,j) - \rho^2 A(n+1,j+1)}{D(n)}, \tag{9}$$

where

$$\begin{aligned}
A(n,j) &= (1-k_0(n-j))(k_0(n)+\delta(n)) - \delta(n-j)\Big[k_0(n)\Gamma(j) + \delta(n)\Xi(j)\Big] \\
D(n) &= D_{MZ}(n) + \rho^2 D_{MZ}(n+1) \\
&\quad - 2\rho^2(k_0(n+1)k_0(n) + k_0(n)\delta(n+1) + k_0(n+1)\delta(n)\Gamma(1) + \delta(n)\delta(n+1)\Xi(1)) \\
D_{MZ}(n) &= k_0(n)^2 + 2k_0(n)\delta(n) + \delta(n)^2\Xi(0) \\
\Xi(j) &= \sum_{\tau=0}^{j}\kappa(\tau) + \sum_{\tau_1=1}^{\infty}\sum_{\tau_2=0}^{\infty}\kappa(\tau_1+j)\kappa(\tau_2)\rho^{2\max(\tau_1,\tau_2)} \\
\Gamma(j) &= \sum_{\tau=0}^{j}\kappa(\tau) + \sum_{\tau=1}^{\infty}\kappa(\tau+j)\rho^{2\tau}
\end{aligned} \tag{10}$$

Proposition 2 shows that BIRs are a complex function of $\kappa(\tau)$ and $\delta(n)$, the two objects that define the linear autonomous class of forecasts.

4.3 BIRs in Leading Behavioral Models

We now apply Proposition 2 to solve for the BIR in several leading behavioral models.

4.3.1 Diagnostic Expectations

To start, we continue with the example of diagnostic expectations from Bordalo et al. (2020a). In this case, Proposition 2 implies that the BIR is as follows.

Corollary 3 *In the model of diagnostic expectations (Bordalo et al., 2020a) given by Equation (4), the BIR for $J = 1$ and diagnosticity parameter θ equals:*

$$B(n, j) = \frac{A}{D}\mathbf{1}_{j=0} \tag{11}$$

where $\delta = \theta(1 - \rho^2)$ and:

$$\begin{aligned} A &= -\delta[1 + \delta\Xi] \\ D &= D_{MZ} + \rho^2 D_{MZ} - 2\rho^2(1 + \delta) \\ D_{MZ} &= 1 + 2\delta + \delta^2\Xi \\ \Xi &= (1 - \rho^2)^{-1} \end{aligned} \tag{12}$$

To visually illustrate the implications of Corollary 3, Figure 9a displays BIRs for varying diagnosticity and signal persistence parameters. Using diagnosticity estimates from Bordalo et al. (2020a) and Bordalo et al. (2019), we set $\theta \in \{0.5, 0.9\}$. For signal persistence, we explore two extremes, setting $\rho \in \{0.1, 0.9\}$.

Figure 9a reveals several strengths of the diagnostic expectations framework. First and foremost, the initial level $B(n, 0)$ of the model-implied BIR lines up well with those found in the SPF and IBES datasets (Figures 2a and 2b). Moreover, holding θ fixed, a lower signal persistence ρ results in a more pronounced overreaction, which is directionally consistent

with Bordalo et al. (2020a), Afrouzi et al. (2023), and the analysis of persistence that we present in Section 3.3.2.

While Figure 9a demonstrates how diagnostic expectations can match the level of observed BIRs, it also highlights a potential shortcoming of the model. To understand the core issue, note that for $J = 1$, the n-period forecast at time t depends on the rational forecast and the rational revision: $F_t^{(n)} = \mathbb{E}_t[y_{t+n}] + \theta(\mathbb{E}_t[y_{t+n}] - \mathbb{E}_{t-1}[y_{t+n}])$. In turn, the forecast error next period depends on the rational forecast error and the rational revision: $e_{t+1}^{(n-1)} = (y_{t+n} - \mathbb{E}_{t+1}[y_{t+n}]) + \theta(\mathbb{E}_{t+1}[y_{t+n}] - \mathbb{E}_t[y_{t+n}])$. By definition, both components are unpredictable by *any* variable at time t, including the diagnostic time-t revision $R_t^{(n)}$, hence why the BIR goes to zero immediately. The abrupt discontinuity in the model-implied BIR turns out to be a more general feature of theories in which forecasts depend on finite lags of rational revisions, including the model of smooth diagnostic expectations in Bianchi et al. (2024a). We demonstrate this point formally via the following proposition.

Proposition 4 *Suppose that $F_t[y_{t+n}]$ depends only on revisions in rational forecasts between t and $t-J$ (i.e., $\kappa(\tau) = 0 \quad \forall \tau > J$ and $\sum_{\tau=0}^{J} \kappa(\tau) = 0$). Then, for any $j \geq J$, $B(n, j) = 0$.*

Proposition 4 demonstrates that models like diagnostic and smooth diagnostic expectations rely on an abrupt correction mechanism. Specifically, these models imply that forecasters, given a lag structure J, fully rectify any overreaction at t by time $t+J$, resulting in a discontinuity in the model-implied BIR at $B(n, J)$. The purple line in Figure 9a illustrates this result by plotting the BIR for $J = 3$, $\theta = 0.9$, and $\rho = 0.9$. The flat and negative BIR for $j = 0, 1, 2$ indicates an initial overreaction to news followed by no correction for two periods. At $j = 3$, diagnostic forecasters then "wake up" and fully reverse their overreaction.

For both SPF and IBES forecasters, this pattern of abrupt correction appears inconsistent with the data. For instance, for $n = 3$, Figure 2a shows that SPF forecasts correct about 60% of their overreaction at time t within one period. Figure 2b demonstrates an even slower correction process for IBES forecasters. Moreover, 6a shows that correction is much slower

for a subset of SPF forecasters. Similarly, IBES forecasters exhibit a relatively slow and smooth correction process (Figure 2b), some much more so than others (Figure 6b).

Heterogeneous speeds of correction, like those seen in Figure 6b, also pose a challenge for models featuring diagnostic expectations. The reason why is that the level and shape of the diagnostic-BIR are largely pinned down by the diagnosticity and persistence parameters, θ and ρ. Consequently, it is difficult for such models to produce two BIRs that have identical initial levels of overreaction but differ in their speeds of correction.

4.3.2 Costly Information Processing

Next, we consider the belief formation process proposed by Afrouzi et al. (2023), who develop a model to accommodate the fact that overreaction to contemporaneous news is larger for longer-horizon forecasts and series with less persistence, both in experimental and observational data. Afrouzi et al. (2023) focus on the special case in which y_t is an AR(1) process with mean μ and persistence parameter ρ. We show in Appendix A that this simple version of the model is a part of the linear autonomous class. Applying Proposition 2 to this model yields the following BIR.

Corollary 5 *Suppose $y_{t+1} = (1-\rho)\mu + \rho y_t + \varepsilon_t$ and that forecasts follow the model of costly information processing in Afrouzi et al. (2023), whereby:*

$$F_t[y_{t+n}] \;=\; const(n) \;+\; E_t[y_{t+n}] \;+\; \delta(n)E_t[y_{t+n}], \tag{13}$$

with

$$\delta(n) \;=\; (1-\rho^n)\rho^{-n} \min\left\{1,\; \left(\frac{\omega\underline{\tau}}{(1-\rho^n)^2}\right)^{1/(1+\gamma)}\right\} \tag{14}$$

for some $\omega, \underline{\tau}, \gamma \geq 0$. *Then, the BIR is given by:*

$$B(n,j) \;=\; A(n,j)\frac{1-\rho^2}{D(n)} \tag{15}$$

where

$$\begin{aligned} A(n,j) \;&=\; -\delta(n-j)\Big[1+\delta(n)\Big] \\ D_{MZ}(n) \;&=\; (1+\delta(n))^2 \\ D(n) \;&=\; D_{MZ}(n)+\rho^2 D_{MZ}(n+1) \;-\; 2\rho^2(1+\delta(n+1))(1+\delta(n))\,. \end{aligned} \tag{16}$$

Moreover, $B(n,j)$ *is monotone increasing and concave in* j.

Figure 9b uses Corollary 5 to plot BIRs for two different parameterizations from Afrouzi et al. (2023), one for when $\rho = 0.2$ and another for $\rho = 0.95$. These values roughly correspond to the average AR(1) coefficients we respectively estimate for the low- and high-persistence series in the SPF in Section 3.3.2. The variance σ_ε^2 of ε_t is fixed at 0.2 across both parameterizations. Additionally, the scale and convexity of information processing costs are set to $\omega = 1$ and $\gamma = 2$, respectively. Following Section 6.2 of Afrouzi et al. (2023), the baseline precision $\underline{\tau}$ that agents have about the long-run mean of y equals $(1-\rho^2)/\sigma_\varepsilon^2$.

The Figure 9b depicts the model-implied BIRs when y has either high (orange line) or low (blue line) persistence. The model was designed to match the fact that overreaction is larger for less persistent series and it does so well, though the degree of overreaction for the low-persistence series is a bit higher in the model than in the data. What is more interesting in the speed of correction within the model. For series with low persistence, corrections occur smoothly yet fairly quickly compared to series with high persistence. In other words, correction speed in the model appears tightly tied to $B(0)$. In Section 3.3.2, we found this link to be much weaker in the data.

Corollary 5 also highlights another shortcoming of the model, namely that the model-

implied BIR is concave in j: following an initial overreaction, forecasters in the model gradually correct their mistakes but at a decelerating rate. This means that the model cannot match the BIRs for the slowest-correcting forecasters in the SPF and IBES datasets. For instance, in Figure 6a, we find that for some forecasters in SPF, the BIR is convex in j and thus implies an extremely slow correction process.

4.3.3 Extrapolative Expectations

Lastly, we consider a model in which forecasters extrapolate from past realizations of y. The specific setup we consider largely draws on Hirshleifer et al. (2015) while also bearing similarities to Nagel and Xu (2022). Unlike Barberis et al. (2015b), who develop an asset pricing model centered on price extrapolation, both Hirshleifer et al. (2015) and Nagel and Xu (2022) focus on the extrapolation of past fundamentals. We follow both papers in assuming that the true data generating process has a constant mean, so that $y_{t+1} = \bar{y} + \varepsilon_{t+1}$, where ε_{t+1} is i.i.d.[17] In addition, forecasts are assumed to evolve according to:

$$\begin{aligned} F_t[y_{t+1}] &= \mu_t \\ \mu_{t+1} &= (1-\rho-\tilde{\rho})\bar{\mu} + \rho\mu_t + \tilde{\rho}y_{t+1}, \end{aligned} \tag{17}$$

where $\tilde{\rho}$ modulates the degree of extrapolation from the most recent realization of y, ρ controls the impact of extrapolation made from past realizations of y, and $\bar{\mu}$ is the forecasters perceived long-run mean. Hirshleifer et al. (2015) provide several possible microfoundations for this particular belief formation process, like, for instance, if the forecaster perceives a time-varying growth rate for y. Applying Proposition 2 to this model then yields the following BIR.

Corollary 6 (BIRs for extrapolative expectations) *In the model of extrapolative ex-*

[17] In Appendix C, we allow for an autoregressive structure in y_t, so that $E_t[y_{t+1}] = \bar{y} + \rho_y(y_t - \bar{y})$. The expressions for BIRs are more involved in this case.

pectations given by (58)*, the BIR is given by:*

$$B(n,j) \;=\; -\frac{\rho^j}{(\rho+\tilde{\rho})^j}\frac{1-\rho(\rho+\tilde{\rho})}{(1+(\rho+\tilde{\rho})^2-2\rho(\rho+\tilde{\rho}))} \;<\; 0\,. \tag{18}$$

In particular, $B(n,j)$ *is independent of* n*, and is increasing and concave in* j*.*

Consistent with Hirshleifer et al. (2015) and Nagel and Xu (2022), Corollary 6 shows that extrapolative expectations can generate both overreaction ($B(n,0) < 0$)) and the pattern of smooth decay observed in empirically estimated BIRs. However, the corollary also reveals that the extrapolative expectations model in Equation (58) fails to accurately reflect how the level of BIRs varies with the forecasting horizon n (see, for instance, Table 2). This follows from the fact that n appears nowhere in the expression for the BIR.

To visualize the BIR implied by extrapolative expectations, Figure 9c plots the BIR using the following parameters: $n = 6$, $\bar{\mu} = 0$, $\tilde{\rho} = 0.018$, and $\rho = 1 - \tilde{\rho}$. Under this parameterization, the model of extrapolative expectations in (58) reduces to the one considered by Nagel and Xu (2022), who choose the value of $\tilde{\rho}$ based on the analysis of inflation expectations in Malmendier and Nagel (2015). The initial level of the BIR, in this case, equals -0.5 and is comparable to what is observed empirically. However, in addition to its lack of dependence on forecast horizon n, the BIR is essentially flat under this parameterization of extrapolative expectations. The implication is that following overreaction to the news at time t, extrapolative agents counterfactually correct their mistakes at a very slow pace — essentially, all of their initial overreaction remains after five quarters.

4.4 Discussion

The analysis in this section leverages BIRs to assess the strengths and weaknesses of prominent behavioral models of belief formation. A key finding is that many models can match the initial level of BIRs, though unsurprisingly the model in Afrouzi et al. (2023) most accurately captures the dependence on series persistence. However, all models have a harder

time matching the slow pace of correction observed for many forecasters in SPF and IBES. Moreover, BIRs often display heterogeneity that is incompatible with these models, most notably the fact that correction speeds vary even for forecasters that appear to react equally to news at time t.

More broadly, the results in this section also highlight the benefits that BIRs offer over traditional impulse responses of forecast errors to identified structural shocks. For one, the discontinuous correction process that is inherent to models of diagnostic expectations is not readily apparent in the impulse responses studied by Angeletos et al. (2021) and **?**. In addition, our analysis of how BIRs vary across forecasters and stocks would be considerably more difficult to implement with traditional impulse response methods because of the latter's dependence on well-identified shocks. For instance, with traditional impulse responses, analyzing whether equity analysts overreact or underreact to specific stocks would require identifying exogenous shocks to those stocks. Furthermore, BIRs capture how forecast errors dynamically respond to all news received at time t, not just a particular shock, which is often what matters for economic outcomes (e.g., prices).

5 BIRs and Model Uncertainty

In this section, we demonstrate how BIRs can be paired with simulations to explore models of belief formation that lie outside of the linear autonomous class studied in Section 4. We specifically consider belief formation for Bayesian forecasters in the presence of model uncertainty (e.g., Timmermann (1993)). Our specific setup mirrors that of Farmer et al. (2023), henceforth FNS. Using their formulation of model uncertainty, which operates through an unobserved components model, we characterize the conditions under which a Bayesian forecaster can generate BIRs that are observed in the SPF and IBES datasets.

5.1 Setup

Farmer et al. (2023) study belief formation about the following data generating process (DGP) for y_t:

$$\begin{aligned} y_t &= \mu_t + x_t, \\ \mu_t &= \mu_{t-1} + \sqrt{\gamma}\sigma\eta_t, \quad \eta_t \sim N(0,1) \\ x_t &= \rho x_{t-1} + \sqrt{1-\gamma}\sigma\omega_t, \quad \omega_t \sim N(0,1) \end{aligned} \tag{18}$$

In this model, the target series y_t is the sum of a random walk, μ_t, and an AR(1) process, x_t, both of which are unobserved. The parameters governing y_t are ρ, σ, and γ. They respectively control the persistence of x_t, the conditional variance of y_t, and the share of this variance attributable to μ_t shocks. Like the states, x_t and μ_t, these parameters are also unknown to forecasters, who only observe quarterly realizations of y_t.

FNS endow forecasters with beliefs about the initial states, μ_1 and x_1, which we specify below. In addition, forecasters have the following initial priors about the parameters governing y_t:

$$\rho \sim N(\mu_p, \sigma_p^2), \quad \gamma \sim \mathcal{B}(\alpha_\gamma, \beta_\gamma), \quad \sigma^2 \sim \mathcal{IG}(1.25, 0.5625), \tag{19}$$

where $\mathcal{B}$ is the beta distribution and $\mathcal{IG}$ is the inverse-gamma distribution. We define $\Theta = (\mu_p, \sigma_p^2, \alpha_\gamma, \beta_\gamma)$ as the parameter vector that defines the forecasters' initial priors about ρ and γ. As in FNS, the priors for σ^2 are held fixed throughout our analysis.

In the latter part of their paper, FNS simulate y_t by assuming parameter values for ρ, γ, and σ and then study the time-series of forecasts for different initial priors Θ. Using their replication kit, we conduct a similar exercise to explore the conditions under which Bayesian forecasters in this setting will generate BIRs that resemble the ones found in the SPF and

IBES datasets. Our simulation exercise specifically proceeds as follows. First, we simulate five different paths of y_t according to the DGP in (18), each of length $T = 260$ (for the IBES, we use $T = 190$ to match the time frame of the real data). When matching SPF forecasts, we set the true $\rho = 0.02$ and $\gamma = 0.02$. For IBES forecasts, we set $\rho = 0.5$ and $\gamma = 0.25$. For both SPF and IBES analysis, we set $\sigma = 0.5$. As we discuss below, these parameters were chosen so that the simulated y_t roughly resembles the average realized series in the SPF and IBES datasets, at least in terms of their observed persistence.

Next, we set forecasters' beliefs about the initial states to be $\mu_1 \sim N(y_1, 1)$ and $x_1 \sim N(0, 1)$, such that the two are perfectly negatively correlated given the DGP in Equation (18). Armed with their initial beliefs about the states (μ_1, x_1) and parameters of the DGP contained in Θ, forecasters observe simulated realizations of y_t and update their beliefs according to Bayes Law. After ten years of realizations, they then form forecasts of y_{t+n} in each quarter t based on their beliefs at that time. This updating sequence exactly follows FNS and generates a panel of forecasts for each simulated sample of y_t. For each sample of simulated forecasts, we estimate a BIR, denoted by $\tilde{B}(n, j; \Theta)$. We use this notation to reiterate the dependence of the forecasts—and hence the BIR—on the initial beliefs Θ.

We then search over the initial belief vector Θ to best match the average individual BIR observed in the SPF or IBES datasets, as plotted in Figures 2a and 2b, respectively. That is, we select Θ to minimize $\sum_j [B(n, j) - \tilde{B}(n, j; \Theta)]^2$. As in the data, we use $n = 3$ for the SPF and $n = 6$ for the IBES BIRs. For each optimal Θ^*, we then simulate $n = 100$ different samples of y and estimate the BIRs as described above for the given initial beliefs Θ^*. Finally, we compute the average BIR across the n samples, along with confidence bands based on the standard deviation of each point on the BIR divided by $\sqrt{n}$.

5.2 Main Findings

Figure 10a plots the resulting simulated BIR along with the true one from the SPF. The figure shows that the model fits the data fairly well. The initial levels of the BIRs are close, though there is slightly more overreaction in the model. The speed of correction in the model also aligns well with the data. In the model, over half of the overreaction in the time-t forecast is eliminated by $t+1$, slightly lower than the 60% observed empirically. Figure 10c shows the distribution of initial beliefs, governed by Θ^*, and the true parameter values (vertical dotted line). As in FNS, overreaction occurs when beliefs about ρ and γ are biased upward. The true ρ and γ equal 0.02, whereas agents in the model have priors for ρ and γ that are centered around 0.18 and 0.3, respectively.

Figure 10b presents the simulated BIR (in blue) when targeting the average individual BIR from IBES (in orange). Once again, the model is able to generate a relatively realistic BIR. The initial level of overreaction in the model, as captured by $B(n,0)$, is very close to the data. In addition, the slow speed of correction embedded in the true BIR is also present in the model. In the data, roughly 80% of the overreaction in the time-t forecast is still present in the forecast made one quarter later, compared to about 90% in the model. There is virtually no additional correction between $t+1$ and $t+2$ in the data, whereas an additional 25 pp of correction occurs in the model. The fact that the model can generate relatively slow corrections is interesting given the behavioral models we explored in 4 had a harder time doing so. With that said, the speed of correction does accelerate too quickly in the model for $j > 2$ relative to the data.

Figure 10d displays the initial belief distributions that underlie Figure 10b. Unlike the SPF, overreaction in this case occurs through a mixture of downward biased beliefs about ρ and upward biased beliefs about γ: beliefs about ρ are centered at 0.05, well below the true $\rho = 0.5$, and beliefs about γ are centered around 0.99, well above the true $\gamma = 0.25$. To develop some intuition for how beliefs about how γ and ρ modulate the speed of correction,

the green line in Figure 10b shows a simulated BIR for initial beliefs that are centered around $\rho = 0.95$ and $\gamma = 0.90$. These beliefs were chosen to match the initial level of overreaction in the data while also ensuring a rapid correction process. The results of this exercise suggest upward-biased beliefs about γ may be the driving force behind overreaction at time t, whereas beliefs about ρ may play a bigger role in the correction process.

It may be tempting to view Figure 10 as evidence that model uncertainty can generate realistic BIRs at the individual level, even when corrections to overreactions at time t are slow. This interpretation hinges on two key questions. First, does our simulated DGP realistically approximate the series forecasted by SPF forecasters and IBES analysts? While a definitive answer to this question is outside of the scope of this paper, Figure 11 offers a preliminary answer by comparing sample paths of the simulated y with the average realized y in both the SPF and IBES datasets. For instance, Figures 11a and 11b show the cross-sectional average realized SPF series alongside three sample paths of the simulated y, and Figure 11c and 11d do the same for the IBES series. Visually, the true and simulated series appear similar, at least in terms of their persistence. The AR(1) coefficients for the average and simulated realized SPF series equal 0.33 and 0.37, respectively. Similarly, the AR(1) coefficients for the average and simulated IBES series are 0.95 and 0.88, respectively.[18]

The second question is whether forecasters' initial beliefs Θ^* about the parameters governing y in our simulation are reasonable, as they are crucial for driving both the initial degree of overreaction and the speed of correction. This question is challenging to answer definitively and presents a valuable direction for future research. For the time being, we take Figure 10 as evidence of the potential for model uncertainty to account for deviations from FIRE observed in survey data. As FNS argue, model uncertainty is especially appealing for explaining deviations from FIRE in professional forecasts, as there are strong career incentives to generate forecasts that are free of behavioral biases. More broadly, the results

[18]For both the SPF and IBES simulations, the AR(1) coefficient is based on the average AR coefficient across the $n = 100$ simulations.

in this section highlight how BIRs can be paired with simulations to explore belief formation in situations where a closed-form BIR is not readily available.

6 Conclusion

In this paper, we develop the concept of a behavioral impulse response function, which captures how forecast errors and revisions of a fixed target dynamically respond following a forecast revision made at time t. Intuitively, BIRs capture how much and in which direction forecasts deviate from FIRE at time t, as well as the subsequent correction process. BIRs offer more powerful tests of FIRE relative to existing methods like the one proposed by Coibion and Gorodnichenko (2015) because they impose joint restrictions on how all future forecast errors and revisions respond to a revision at time t. They can be easily estimated using OLS regressions on any dataset that features a term structure of forecasts.

We estimate BIRs using individual forecasts of macroeconomic and financial variables from the SPF and earnings per share from the IBES. Our analysis reveals that BIRs from both datasets strongly reject the null hypothesis of FIRE, showing a clear tendency for individual forecasters to overreact when forecasting distant targets. On average, deviations from FIRE are corrected gradually in both datasets, though correction speeds vary meaningfully across forecast targets and forecasters. Although popular behavioral models of belief formation can match the level of BIRs, they often have a harder time matching their shape. The slow correction that is observed for many forecasters (and series) poses a particular challenge for existing models, as does the fact that correction speed varies meaningfully even holding fixed the size of the initial deviation from FIRE.

We also highlight how BIRs can be paired with simulations to explore belief formation in more complicated settings, such as when forecasters are uncertain about the true model of the world. Here, we uncover suggestive evidence that model uncertainty may explain the dynamic deviations from FIRE observed in BIRs. However, further research is needed

to assess the plausibility of the initial beliefs underpinning these sorts of models. Moving forward, we recommend that researchers studying belief formation incorporate BIRs to refine and validate their models, comparing derived or simulated BIRs against empirical estimates like those presented in this study.

References

Afrouzi, Hassan, Spencer Y Kwon, Augustin Landier, Yueran Ma, and David Thesmar, "Overreaction in expectations: Evidence and theory," *The Quarterly Journal of Economics*, 2023, p. qjad009.

Angeletos, George-Marios, Zhen Huo, and Karthik A Sastry, "Imperfect macroeconomic expectations: Evidence and theory," *NBER Macroeconomics Annual*, 2021, *35* (1), 1–86.

Augenblick, Ned and Matthew Rabin, "Belief Movement, Uncertainty Reduction, and Rational Updating," *The Quarterly Journal of Economics*, 02 2021, *136* (2), 933–985.

Baker, Malcolm, Richard S Ruback, and Jeffrey Wurgler, "Behavioral corporate finance," in "Handbook of empirical corporate finance," Elsevier, 2007, pp. 145–186.

Barberis, Nicholas, Lawrence J. Jin, and Baolian Wang, "Prospect Theory and Stock Market Anomalies," *The Journal of Finance*, 2021, *76* (5), 2639–2687.

___ , Robin Greenwood, Lawrence Jin, and Andrei Shleifer, "X-CAPM: An extrapolative capital asset pricing model," *Journal of Financial Economics*, 2015, *115* (1), 1–24.

___ , ___ , ___ , and ___ , "X-CAPM: An extrapolative capital asset pricing model," *Journal of financial economics*, 2015, *115* (1), 1–24.

Bauer, Michael D. and Eric T. Swanson, "An Alternative Explanation for the 'Fed Information Effect'," *American Economic Review*, March 2023, *113* (3), 664–700.

Benjamini, Yoav and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, *57* (1), 289–300.

Bianchi, Francesco, Cosmin L Ilut, and Hikaru Saijo, "Smooth Diagnostic Expectations," Technical Report, National Bureau of Economic Research 2024.

___ , **Sydney C. Ludvigson, and Sai Ma**, "Belief Distortions and Macroeconomic Fluctuations," *American Economic Review*, July 2022, *112* (7), 2269–2315.

___ , ___ , **and** ___ , "What Hundreds of Economic News Events Say About Belief Overreaction in the Stock Market," Working paper 2024.

Binsbergen, Jules H Van, Xiao Han, and Alejandro Lopez-Lira, "Man versus machine learning: The term structure of earnings expectations and conditional biases," *The Review of Financial Studies*, 2023, *36* (6), 2361–2396.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, "Diagnostic expectations and stock returns," *The Journal of Finance*, 2019, *74* (6), 2839–2874.

___ , ___ , ___ , **and** ___ , "Belief overreaction and stock market puzzles," 2020.

___ , ___ , **Rafael LaPorta, and Andrei Shleifer**, "Belief Overreaction and Stock Market Puzzles," *Journal of Political Economy*, 2024.

___ , ___ , **Yueran Ma, and Andrei Shleifer**, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–2782.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar, "Sticky expectations and the profitability anomaly," *The Journal of Finance*, 2019, *74* (2), 639–674.

Cagan, Phillip, "The monetary dynamics of hyperinflation," *Studies in the Quantity Theory if Money*, 1956.

Campbell, John Y. and Robert J. Shiller, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies*, 1988, *1* (3), 195–228.

Cochrane, John H. and Monika Piazzesi, "Bond Risk Premia," *American Economic Review*, March 2005, *95* (1), 138–160.

Coibion, Olivier and Yuriy Gorodnichenko, "What Can Survey Forecasts Tell Us about Information Rigidities?," *Journal of Political Economy*, 2012, *120* (1), 116–159.

___ **and** ___ , "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 2015, *105* (8), 2644–2678.

Farmer, Leland E., Emi Nakamura, and Jón Steinsson, "Learning about the Long Run," Working paper 2023.

Gabaix, Xavier, "Behavioral inattention," in "Handbook of behavioral economics: Applications and foundations 1," Vol. 2, Elsevier, 2019, pp. 261–343.

___ , "A Behavioral New Keynesian Model," *American Economic Review*, August 2020, *110* (8), 2271–2327.

Giannoni, Marc P and Michael Woodford, "How forward-looking is optimal monetary policy?," *Journal of Money, Credit and Banking*, 2003, pp. 1425–1469.

Giglio, Stefano and Bryan Kelly, "Excess Volatility: Beyond Discount Rates*," *The Quarterly Journal of Economics*, 08 2017, *133* (1), 71–127.

Hirshleifer, David, Jun Li, and Jianfeng Yu, "Asset pricing in production economies with extrapolative expectations," *Journal of Monetary Economics*, 2015, *76*, 87–106.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, "Is There a Replication Crisis in Finance?," *The Journal of Finance*, 2023, *78* (5), 246–2518.

Kelly, Bryan T, Semyon Malamud, and Kangying Zhou, "The virtue of complexity in return prediction," *Journal of Finance*, forthcoming.

Kohlhas, Alexandre N. and Ansgar Walther, "Asymmetric Attention," *American Economic Review*, September 2021, *111* (9), 2879–2925.

Kučinskas, Simas and Florian S. Peters, "Measuring Under- and Overreaction in Expectation Formation," *The Review of Economics and Statistics*, 11 2022, pp. 1–45.

Lazarus, Eben, Daniel J Lewis, James H Stock, and Mark W Watson, "HAR inference: Recommendations for practice," *Journal of Business & Economic Statistics*, 2018, *36* (4), 541–559.

Li, Dake, Mikkel Plagborg-Møller, and Christian K Wolf, "Local Projections vs. VARs: Lessons From Thousands of DGPs," Working Paper 30207, National Bureau of Economic Research July 2022.

Lovell, Michael C., "Tests of the Rational Expectations Hypothesis," *The American Economic Review*, 1986, *76* (1), 110–124.

Malmendier, Ulrike, "Behavioral Corporate Finance," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics - Foundations and Applications 1*, Vol. 1 of Handbook of Behavioral Economics: Applications and Foundations 1, North-Holland, 2018, pp. 277–379.

___ **and Stefan Nagel**, " Learning from Inflation Experiences *," *The Quarterly Journal of Economics*, 10 2015, *131* (1), 53–87.

Mankiw, N. Gregory and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve*," *The Quarterly Journal of Economics*, 11 2002, *117* (4), 1295–1328.

___ **,** ___ **, and Justin Wolfers**, "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual*, 2003, *18*, 209–248.

___ **,** ___ **, and** ___ , "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual*, 2003, *18*, 209–248.

Mincer, Jacob A and Victor Zarnowitz, "The evaluation of economic forecasts," in "Economic forecasts and expectations: Analysis of forecasting behavior and performance," NBER, 1969, pp. 3–46.

Nagel, Stefan and Zhengyang Xu, "Asset pricing with fading memory," *The Review of Financial Studies*, 2022, *35* (5), 2190–2245.

Nerlove, Marc, "Adaptive expectations and cobweb phenomena," *The Quarterly Journal of Economics*, 1958, *72* (2), 227–240.

Òscar Jordà, "Local Projections for Applied Economics," *Annual Review of Economics*, 2023, *15* (1), 607–631.

Pflueger, Carolin, Emil Siriwardane, and Adi Sunderam, "Financial Market Risk Perceptions and the Macroeconomy*," *The Quarterly Journal of Economics*, 03 2020, *135* (3), 1443–1491.

Plagborg-Møller, Mikkel and Christian K. Wolf, "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 2021, *89* (2), 955–980.

Schwartzstein, Joshua and Adi Sunderam, "Using Models to Persuade," *American Economic Review*, January 2021, *111* (1), 276–323.

Sims, Christopher A., "Macroeconomics and Reality," *Econometrica*, 1980, *48* (1), 1–48.

___ , "Implications of rational inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

Stein, Jeremy C., "Rational Capital Budgeting in an Irrational World," *Journal of Business*, 1996, *69* (4), 429–455.

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

Timmermann, Allan G., "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices," *The Quarterly Journal of Economics*, 1993, *108* (4), 1135–1145.

Woodford, Michael, *Imperfect Common Knowledge and the Effects of Monetary Policy*, Princeton University Press,

Òscar Jordà, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, March 2005, *95* (1), 161–182.

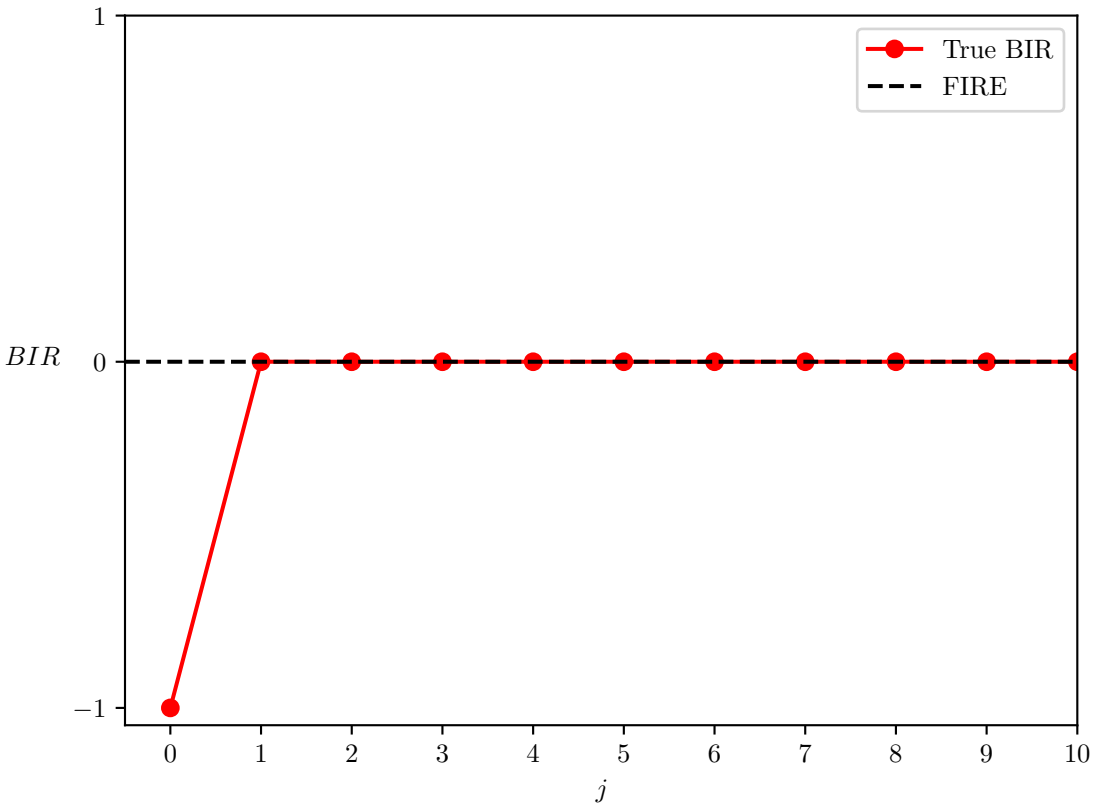
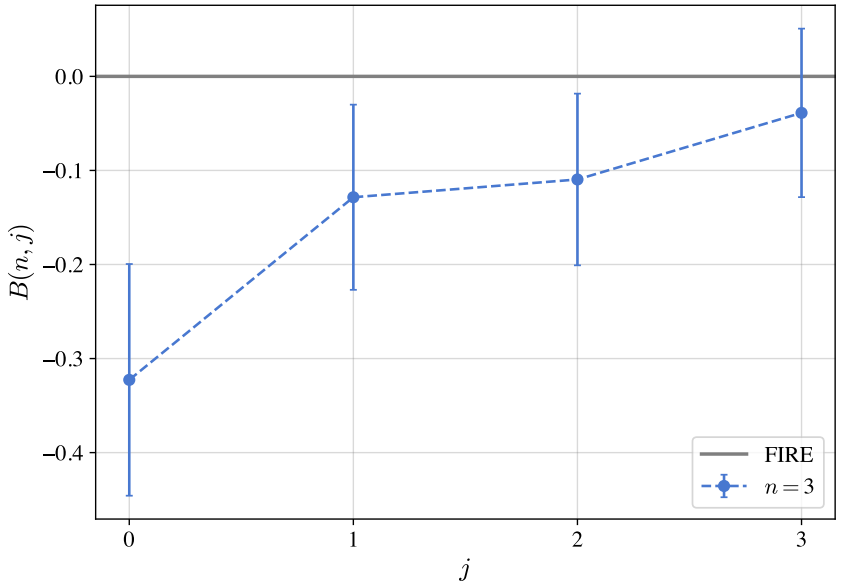


Figure 1: Stylized Example of a BIR. This figure shows a stylized BIR for an iid, mean zero forecast target. The forecaster generating the BIR is assumed to have a rational forecast up to time t, after which they irrationally revise their forecast up by one unit. They then realize their mistake at time $t+1$ and revise their forecast accordingly.

(a) Average BIR in SPF



(b) Average BIR in IBES

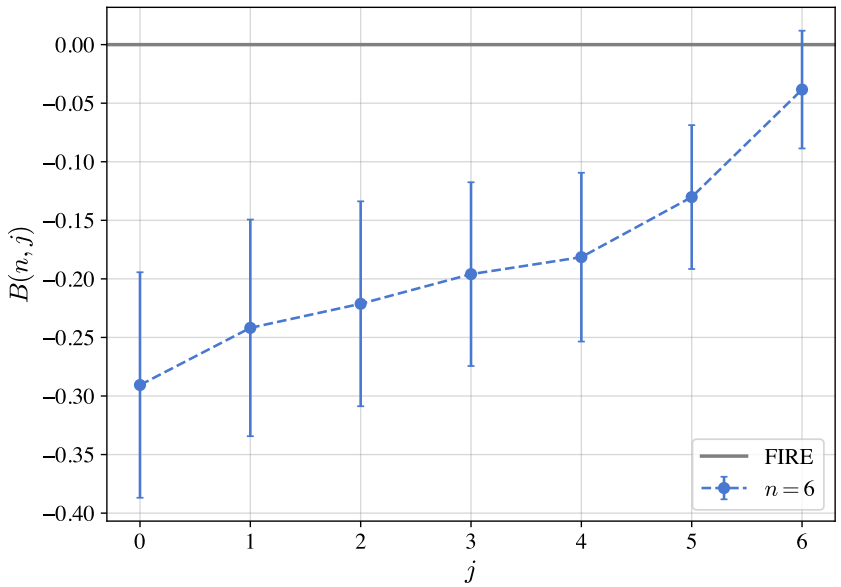


Figure 2: Average BIRs from SPF and IBES. Panels (a) and (b) show baseline BIR estimates for different ns in SPF and IBES, respectively. 95% confidence intervals in both panels are based on standard errors that are clustered by forecaster-series (i, s) and series-date (s, t), where t is the date of the first observed revision for a given forecast horizon n. The black line in both plots shows the BIR under FIRE. See Section 2.2 for complete estimation details.

(a) BIR in CBO

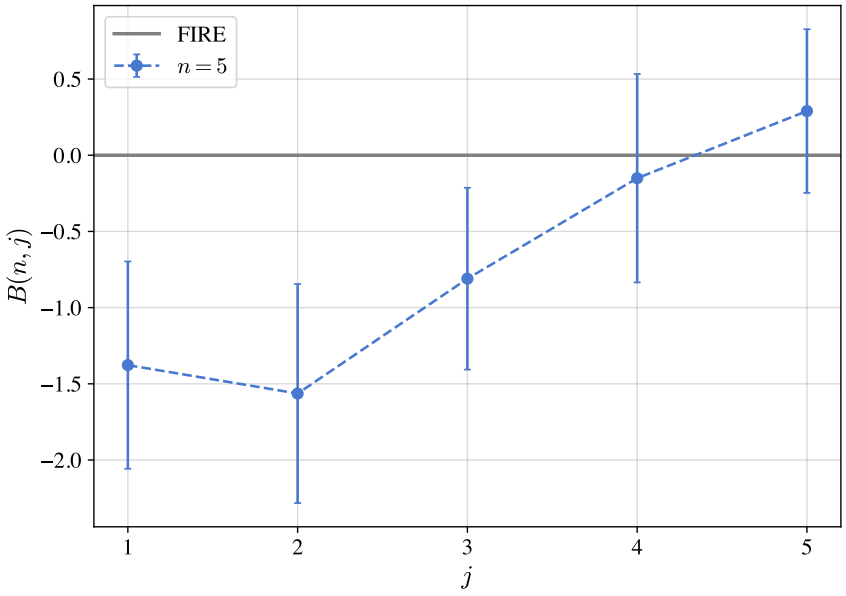


Figure 3: BIRs from CBO. This figure shows baseline BIR estimates for different ns in CBO real GDP growth forecasts. 95% confidence intervals are based on Newey-West standard errors with lags specified as $L = \lceil 1.3 \times T^{\frac{1}{2}} \rceil$, as proposed in Lazarus et al. (2018). The black line shows the BIR under FIRE. See Section 2.2 for complete estimation details.

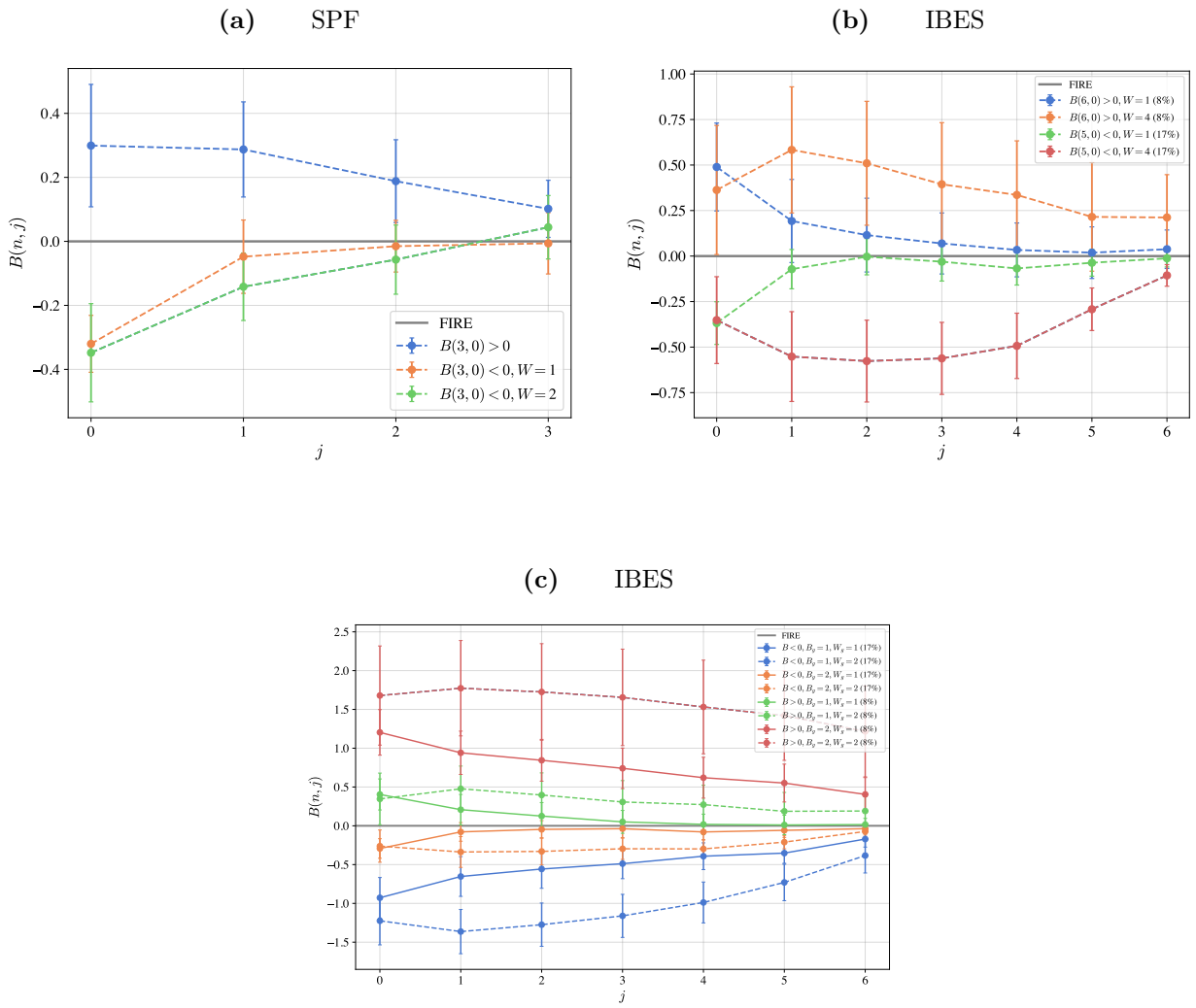


Figure 4: Variation in BIRs Across Series. Figure 4a visualizes variation in BIRs across series in SPF. To construct the plot, we estimate series-by-series BIRs and divide them based on the sign of $B(n,0)$. Within the set of series that exhibit underreaction ($B(n,0) > 0$), we estimate a pooled BIR and plot it in blue. Within the set of series that exhibit overreaction, we further split them into one of two groups based on $W(n,1) = B(n,1)/B(n,0)$, a measure of correction speed. Within each resulting group, we then estimate a pooled BIR and plot it. Figure 4b shows BIRs resulting from an analogous process in IBES, with the only difference being that we split series that exhibit underreaction into two groups based on $W(n,1)$. Figure 4c applies the same methodology, but splits series in IBES based on the magnitude of $B(n,0)$ and then $W(n,1)$. The legends in each plot show the percentage of series assigned to each group. 95% confidence intervals are based on standard errors that are clustered by forecaster-series, (i,s), and series-date, (s,t). The black line in each plot shows the BIR under FIRE. See Section 2.2 for complete estimation details.

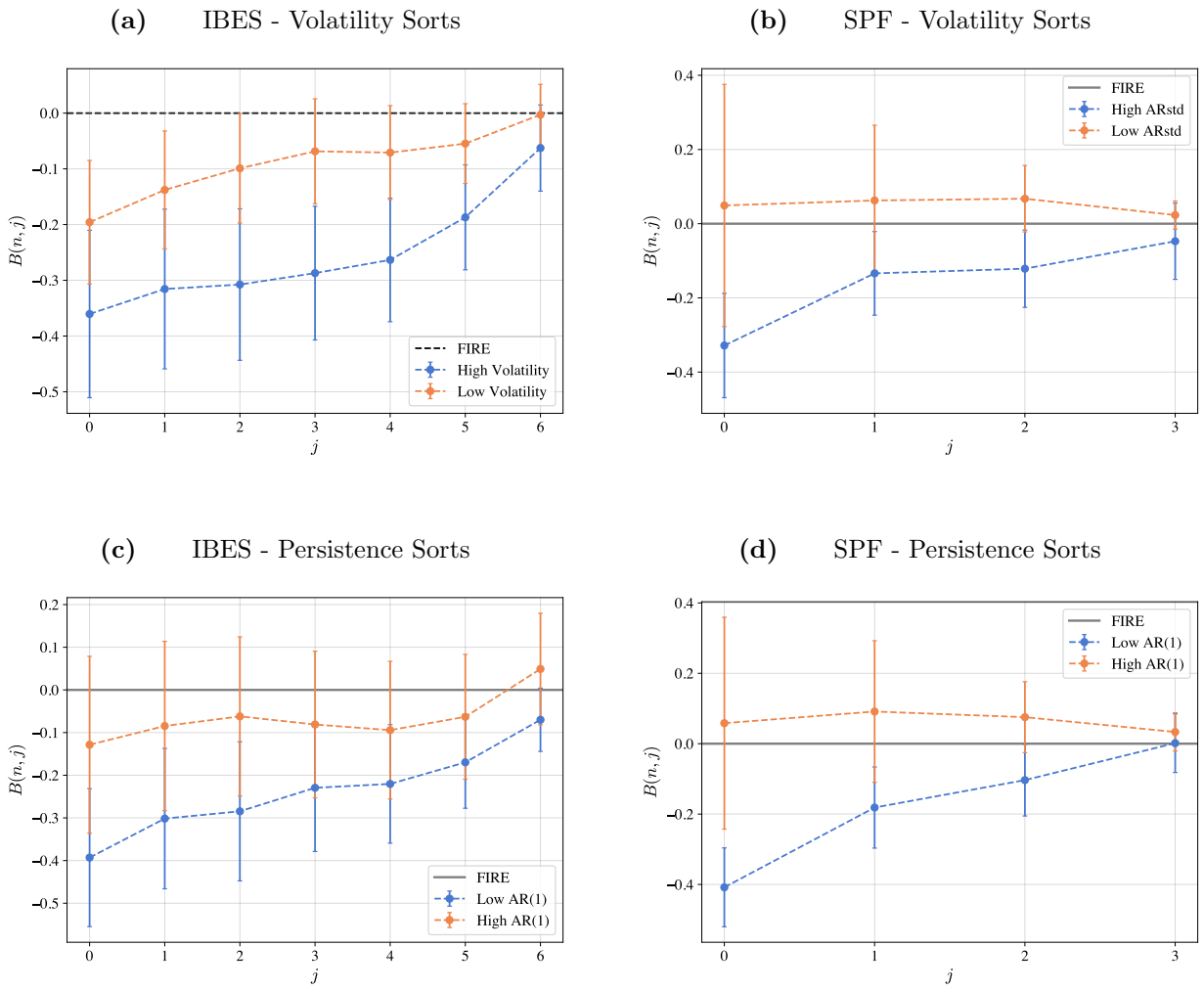
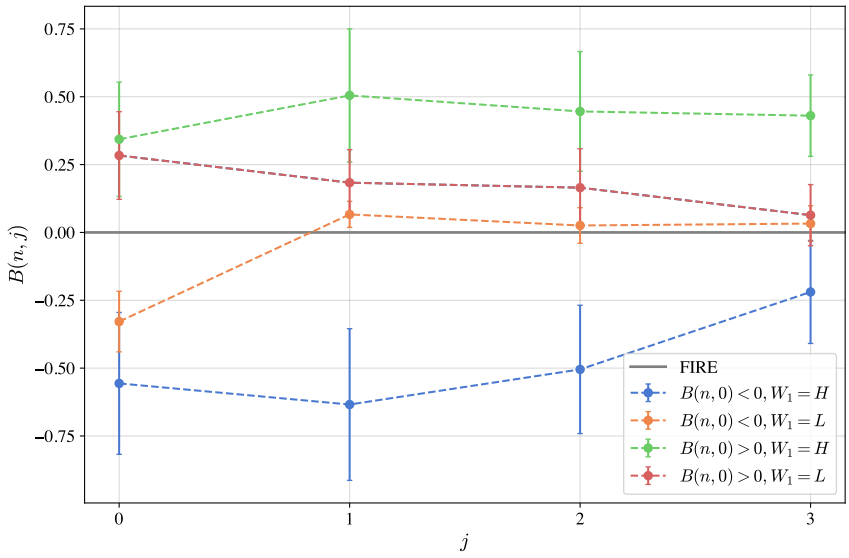


Figure 5: Correction Speeds and Series Characteristics. Figures 5a and 5b show pooled BIRs for series with high or low volatility in IBES and SPF, respectively. Volatility for each stock in IBES is defined as the average monthly stock return volatility in quarter t. In SPF, it is defined based on the residuals from an AR(p) model, where p is chosen according to the AIK . In both instances, low- and high-volatility series are defined based on terciles of volatility. Figures 5c and 5d show pooled BIRs based on sorts of persistence. In IBES, persistence is measured using an AR model for annual operating earnings scaled by assests, as in Bouchaud et al. (2019). In SPF, it is measured using an AR model of each series. We partition series into terciles by measure of persistence and estimate the BIRs on the top and bottom group. 95% confidence intervals are based on standard errors that are clustered by forecaster-series, (i, s), and series-date, (s, t). The black line in each plot shows the BIR under FIRE. See Section 2.2 for complete estimation details.

(a) SPF



(b) IBES

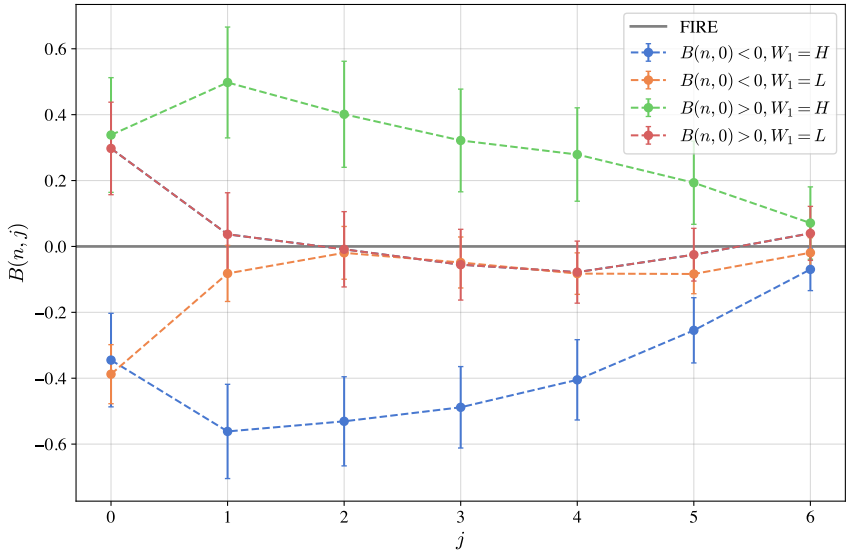


Figure 6: Variation in BIRs Across Forecasters. Figure 6b visualizes variation in BIRs across forecasters in SPF and IBES. For both datasets, BIRs are first estimated for each forecaster, then forecasters are split based on whether they initially over or underreact. Within the set of overreactors, forecasters are further split in to k bins based on $W_1 = B(n,1)/B(n,0)$. The same process is repeated for underreactors. A pooled BIR is then estimated for each bin and plotted in the figure. $k = 3$ in Figure 6a for SPF forecasters that overreact and $k = 2$ for those that underreact. $k = 4$ in Figure 6b for both types of IBES forecasters, and BIRs are only shown for the top and bottom quartiles of W_1. 95% confidence intervals are based on standard errors that are clustered by forecaster-series, (i, s), and series-date, (s, t). The black line in each plot shows the BIR under FIRE. See Section 2.2 for complete estimation details.

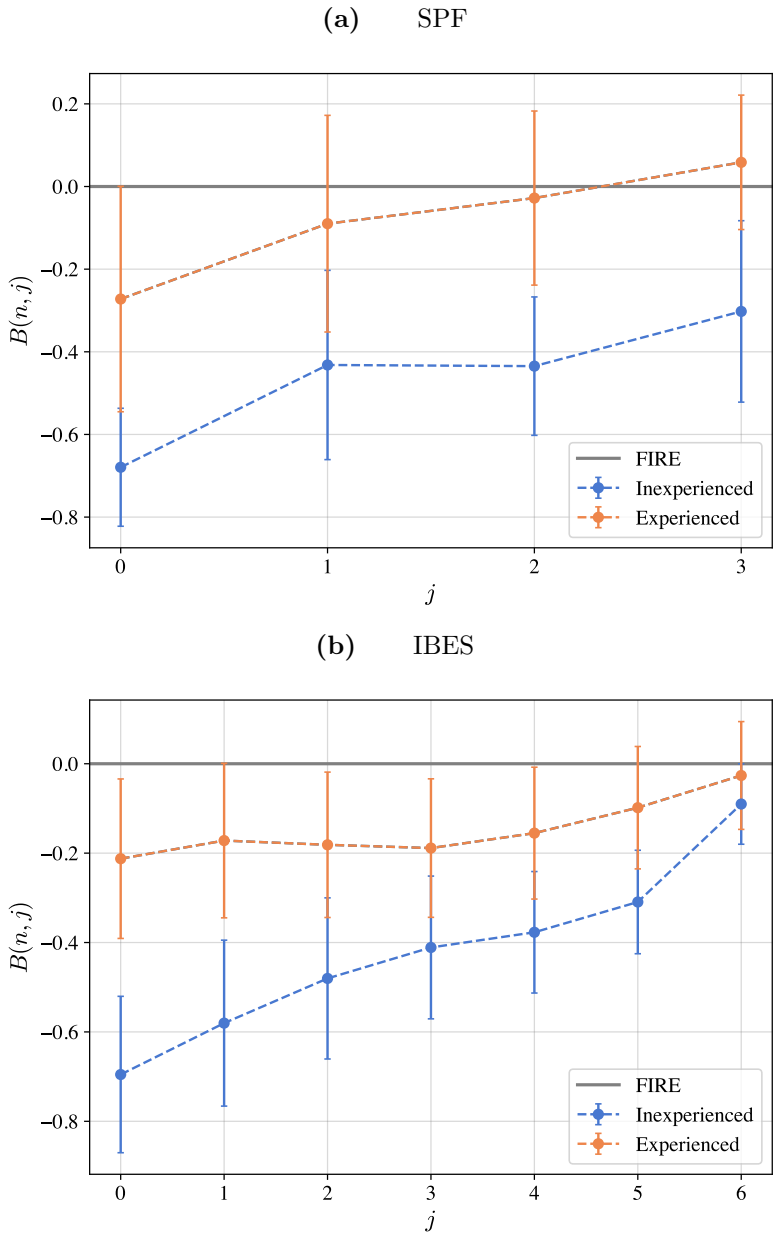


Figure 7: BIRs and Experience. This figure shows how BIRs vary with forecaster experience. We start by estimating BIRs for each forecaster using their first five years of data, requiring at least ten forecasts. We sort forecasters i by the average level of their BIR, denoted as B_i^{early}, and assign them to terciles. Within each tercile, we then estimate pooled BIRs using: (i) the first five-years of forecasts made by the forecaster (labeled as the "inexperienced" BIRs in blue); and (ii) all forecasts made after year five (labeled as the "experienced" BIR in orange). Figures 7a and 7b show the estimates for forecasters in SPF and IBES, respectively. For both plots, estimates are shown for forecasters that overreact early in their careers (i.e., the bottom tercile of B_i^{early}). Results for underreactors can be found in Internet Appendix Figure A7. The black line shows the full information rational expectations benchmark. 95% confidence intervals in the plot are based on standard errors clustered by analyst-stock, (i, s), and stock-date, (s, t)

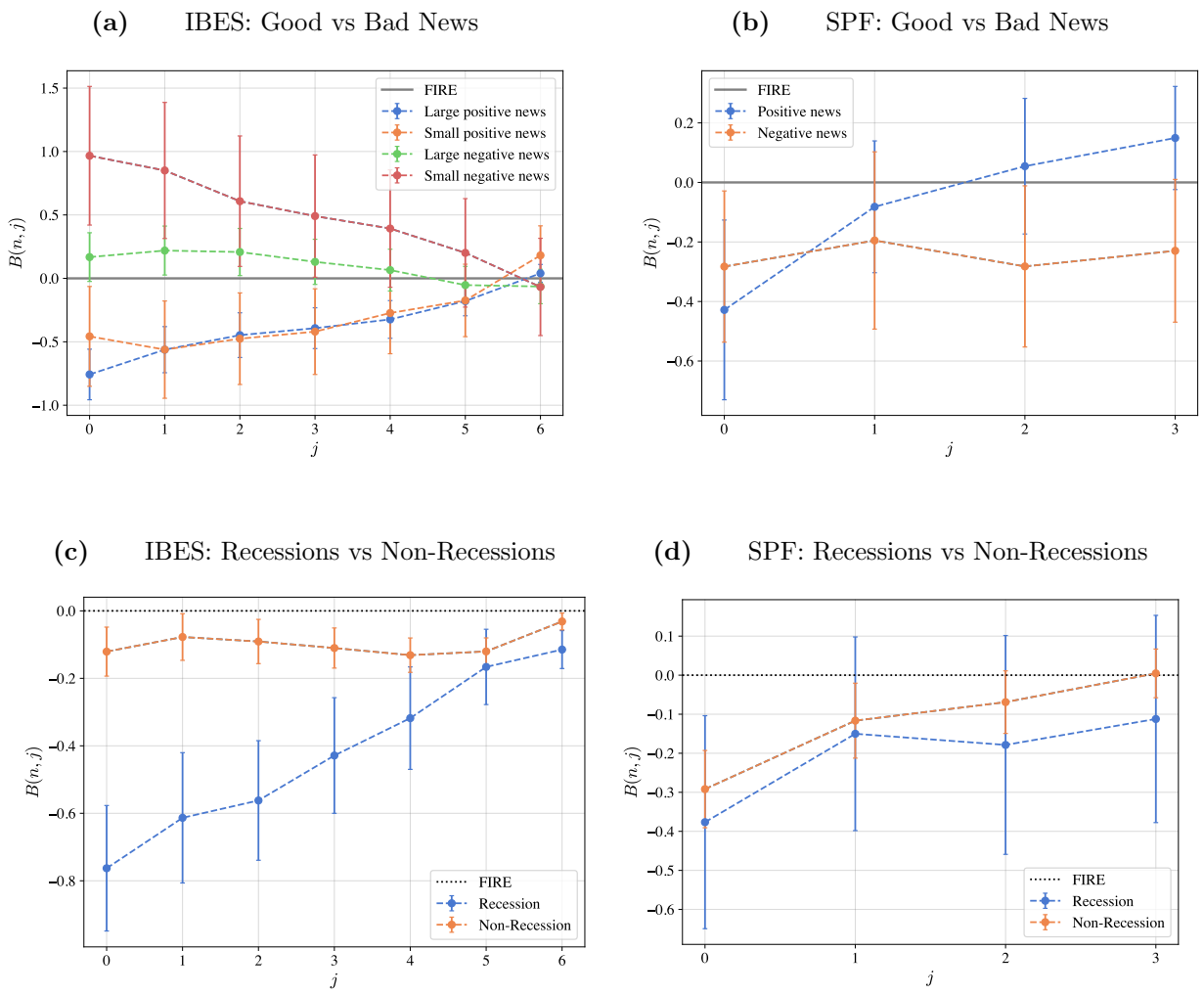


Figure 8: BIRs and the Nature of Shocks. This figure shows how BIRs vary with the nature of news received by forecasters. Panel (a) shows BIR that condition on the sign and size of the revision made by each forecaster. Large positive (negative) revisions are defined as the top tercile of positive (negative) revisions and small positive (negative) revisions are all others. Panel (b) shows a similar result for the SPF, but only conditions on the sign of revisions. Panel (c) conditions on whether revisions in IBES are made during NBER recessions. Panel (d) repeats panel (c) for SPF.

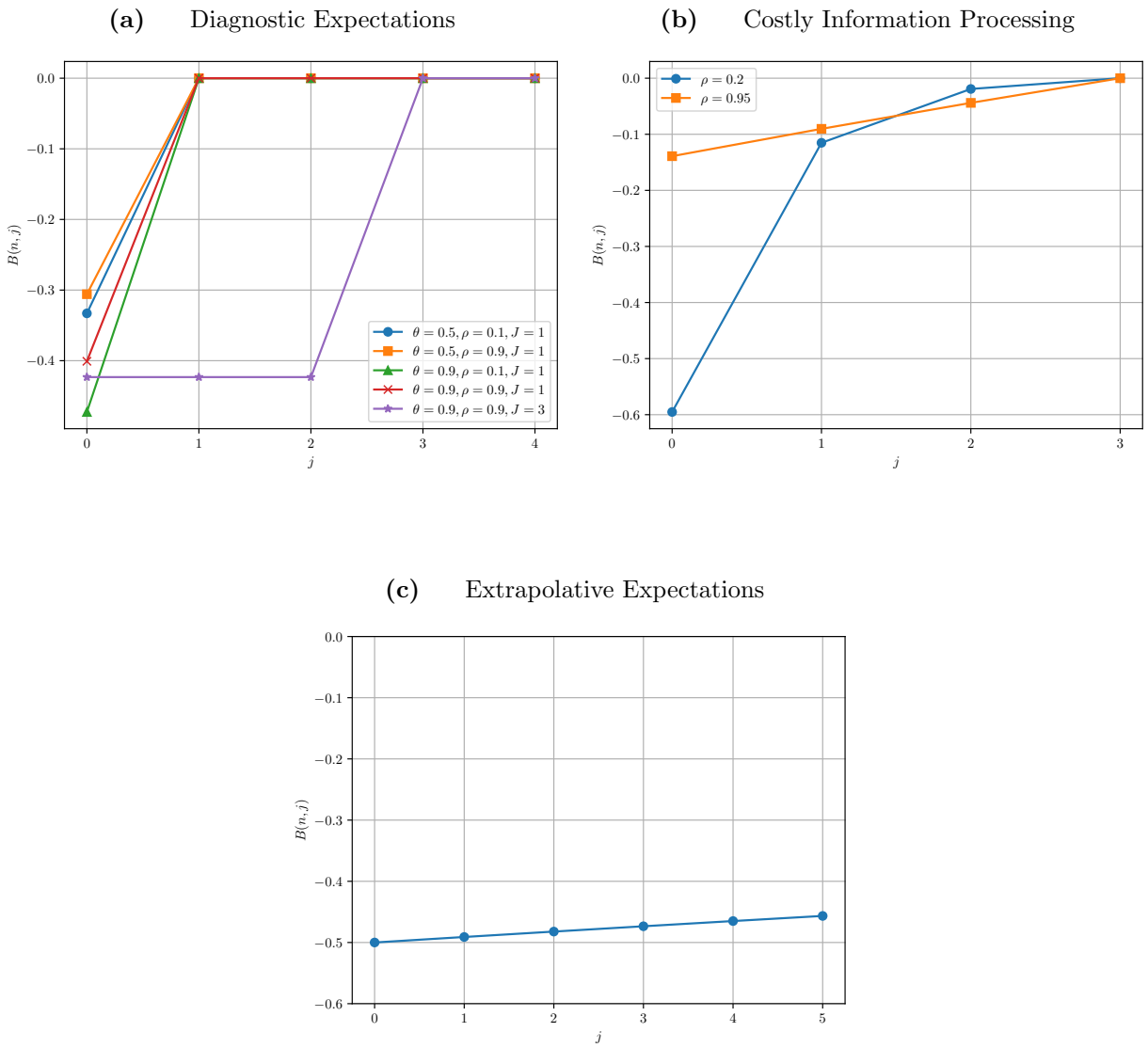


Figure 9: BIRs under Diagnostic and Extrapolative Expectations Panel (a) of this figure shows BIRs for different levels of diagnosticity (θ) and signal persistence (ρ) under the model of diagnostic expectations from Bordalo et al. (2020a). Panel (d) shows the BIR from Afrouzi et al. (2023), with the exact calibration detailed in Section 4.3.2. Panel (c) shows the BIR implied by the model of Nagel and Xu (2022), using the constant-gain learning parameter of 0.018 estimated in Malmendier and Nagel (2015).

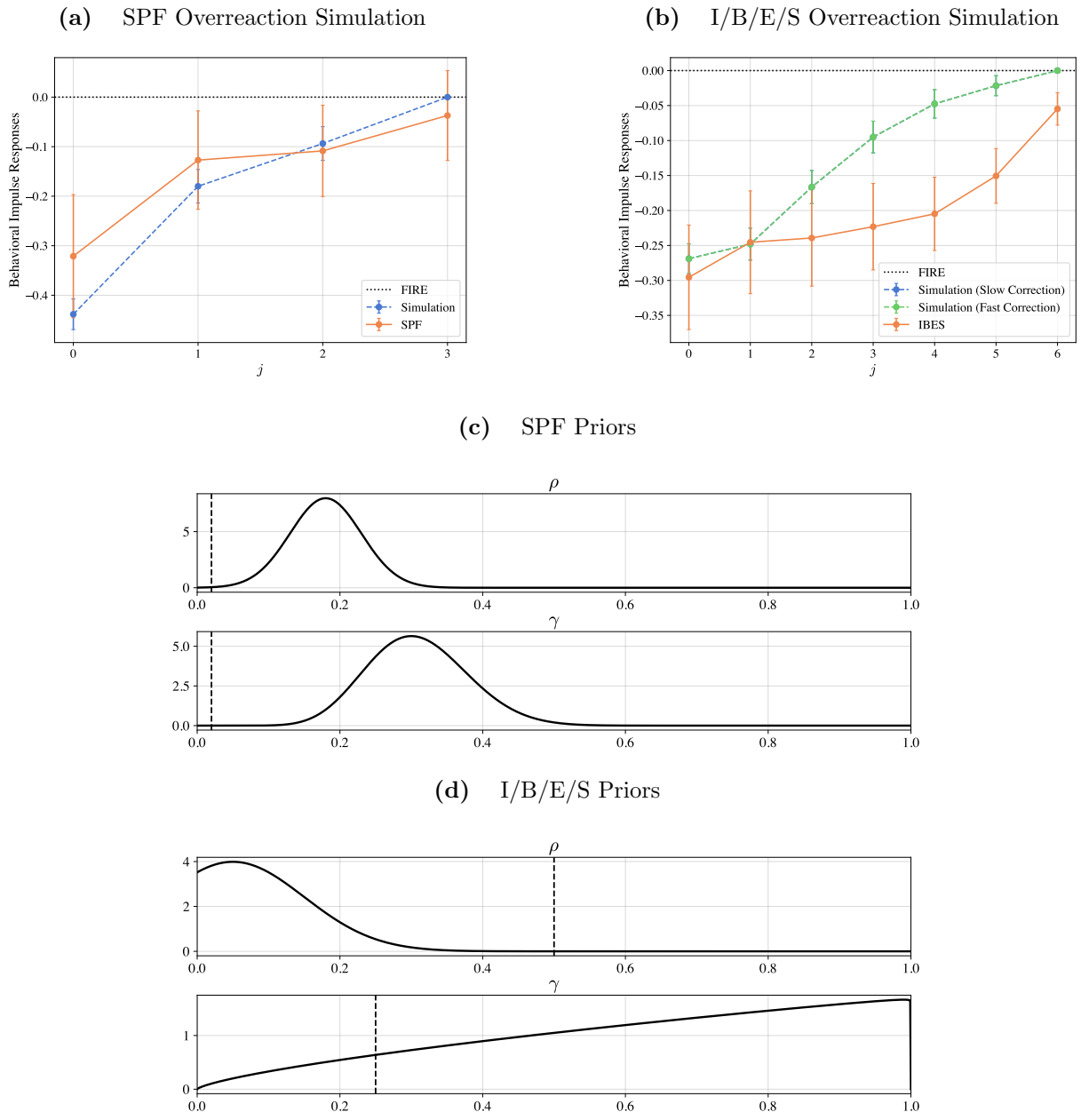
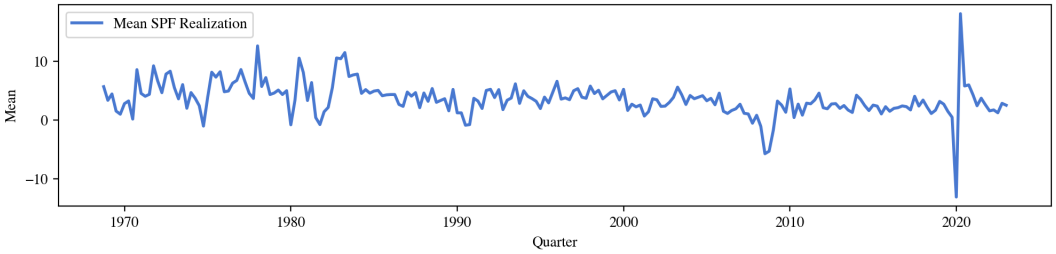
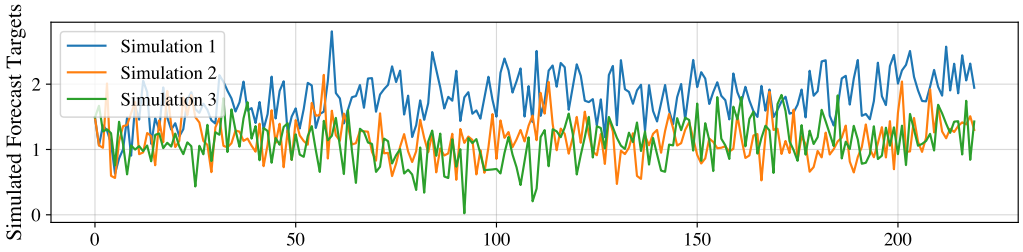


Figure 10: BIRs with Model Uncertainty Panel (a) of this figure shows BIRs estimated on the pooled SPF forecasts with $n = 3$. The plot also shows BIRs from the simulated model of Farmer et al. (2023) with the best-matched priors. Panel (b) shows the BIR for the I/B/E/S pooled forecasts with $n = 6$ and the simulated Bayesian forecasters in Farmer et al. (2023) with the best-matched priors. Panel (c) and (d) show the best-matched priors with respect to the SPF and I/B/E/S BIRs, respectively. See Section 5 for complete details.

(a) Quarterly Mean SPF Realization



(b) SPF Sample Simulated Series



(c) Quarterly Mean EPS



(d) EPS Sample Simulated Series

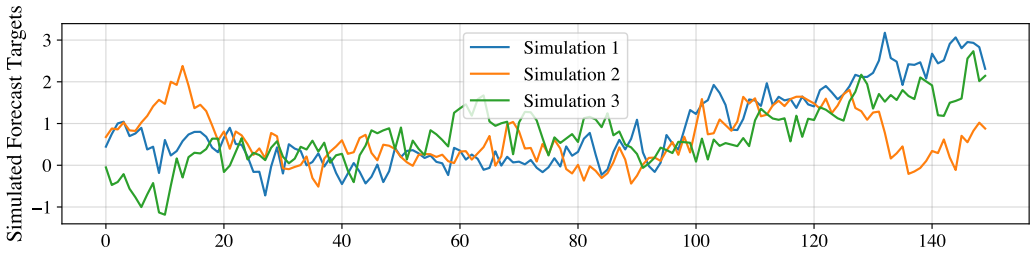


Figure 11: BIRs with Model Uncertainty Panel (a) of this figure shows the quarterly cross-sectional average realization of the SPF forecast targets. Panel (b) shows three randomly simulated paths according to Farmer et al. (2023) with $\rho = 0.18, \gamma = 0.30$. Panel (c) shows the quarterly cross-sectional average EPS in the I/B/E/S sample. Panel (d) shows three randomly simulated paths according to Farmer et al. (2023) with $\rho = 0.05, \gamma = 0.99$.

	(1)	(2)	(3)
$B(n,0)$	-0.323^{***}	-0.324^{***}	-0.261^{***}
	(0.063)	(0.061)	(0.067)
$B(n,1)$	-0.128^{**}	-0.173^{***}	-0.084^{**}
	(0.050)	(0.053)	(0.040)
$B(n,2)$	-0.110^{**}	-0.097^{**}	
	(0.047)	(0.047)	
$B(n,3)$	-0.039		
	(0.046)		
n	3	2	1
Adj. R^2 (%)	0.36	0.67	0.87
Observations	42,150	50,545	60,369
p(FIRE)	<0.001	<0.001	<0.001
p(Equal B)	<0.001	<0.001	<0.001

Table 1: BIRs Estimated from SPF Data. This table shows point estimates for BIRs based on the SPF forecast data. Columns (1), (2), and (3) show BIR estimates for $n = 3, 2, 1$, respectively. Each row in the table corresponds to a different j. Standard errors are listed in parentheses below each point estimate and are clustered by forecaster-series, (i, s), and series-date, (s, t). The row with $p(FIRE)$ reports the p-value from testing the null hypothesis of FIRE, namely that all points along the BIR curve equal zero, i.e. $B(n, j) = 0, \forall j$. The row with p(Equal B) reports the p-value from testing the null hypothesis that all points along the BIR curve are equal, which is a weaker test of FIRE. The adjusted-R^2 is based on averaging over the individual regressions used to estimate each point on the BIR curve (e.g., $B(n,0), ..., B(n,n)$). See Section 2.2 for complete estimation details. *, **, and *** denote p-values of less than 0.10, 0.05, and 0.01, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
$B(n,0)$	-0.291^{***}	-0.114^{***}	-0.007	0.013	0.028	0.056^{***}
	(0.049)	(0.036)	(0.029)	(0.025)	(0.021)	(0.016)
$B(n,1)$	-0.242^{***}	-0.119^{***}	-0.022	0.014	0.018	0.030^{**}
	(0.047)	(0.034)	(0.027)	(0.023)	(0.017)	(0.012)
$B(n,2)$	-0.221^{***}	-0.118^{***}	-0.046^{*}	0.007	0.022^{*}	
	(0.045)	(0.031)	(0.024)	(0.020)	(0.013)	
$B(n,3)$	-0.196^{***}	-0.108^{***}	-0.050^{**}	0.019		
	(0.040)	(0.026)	(0.020)	(0.017)		
$B(n,4)$	-0.181^{***}	-0.098^{***}	-0.034^{**}			
	(0.037)	(0.021)	(0.016)			
$B(n,5)$	-0.130^{***}	-0.019				
	(0.031)	(0.015)				
$B(n,6)$	-0.038					
	(0.026)					
n	6	5	4	3	2	1
Adj. R^2 (%)	0.30	0.09	0.02	0.00	0.01	0.03
Observations	49,081	94,863	165,708	273,368	473,827	716,490
p(FIRE)	<0.001	<0.001	<0.001	0.020	0.003	<0.001
p(Equal B)	<0.001	<0.001	<0.001	0.034	0.051	<0.001

Table 2: BIRs from IBES Data. This table shows point estimates for BIRs based on the I/B/E/S quarterly earnings forecast data. Each row in the table corresponds to a different j. Standard errors are listed in parentheses below each point estimate and are clustered by forecaster-series, (i,s), and series-date, (s,t). The row with $p(FIRE)$ reports the p-value from testing the null hypothesis of FIRE, namely that all points along the BIR curve equal zero, i.e. $B(n,j)=0, \forall j$. The row with p(Equal B) reports the p-value from testing the null hypothesis that all points along the BIR curve are equal, which is a weaker test of FIRE. The adjusted-R^2 is based on averaging over the individual regressions used to estimate each point on the BIR curve (e.g., $B(n,0),...,B(n,n)$). See Section 2.2 for complete estimation details. *, **, and *** denote p-values of less than 0.10, 0.05, and 0.01, respectively.

Data	Test	n				
		2	3	4	5	6
SPF	$B(n,0)=0$	0.73	0.80			
SPF	$B(n,j)=0,\forall j$	0.80	0.80			
SPF	Generalized BIR	0.80	0.80			
IBES	$B(n,0)=0$	0.07	0.04	0.03	0.04	0.06
IBES	$B(n,j)=0,\forall j$	0.28	0.31	0.35	0.43	0.48
IBES	Generalized BIR	0.60	0.60	0.57	0.57	0.48
# Series SPF		15	15			
# Series IBES		1299	1299	1299	1299	1299

(a) By Series

Data	Test	n				
		2	3	4	5	6
SPF	$B(n,0)=0$	0.14	0.17			
SPF	$B(n,j)=0,\forall j$	0.44	0.62			
SPF	Generalized BIR	0.65	0.62			
IBES	$B(n,0)=0$	0.09	0.07	0.07	0.03	0.07
IBES	$B(n,j)=0,\forall j$	0.32	0.40	0.53	0.62	0.72
IBES	Generalized BIR	0.90	0.91	0.91	0.90	0.72
# Forecasters SPF		190	190			
# Forecasters IBES		970	970	970	970	970

(b) By Forecaster

Table 3: How Pervasive are Deviations from FIRE? This table summarizes tests of FIRE across series and forecasters. Table 3a is based on BIRs estimated by for each series s and horizon n, whereas Table 3a is based on BIRs estimated for each forecaster i and horizon n. For each BIR, the first test of FIRE is based on whether $B(n,0)=0$, following Coibion and Gorodnichenko (2015). The second test of FIRE is based on whether $B(n,j)=0$ for all j. The third test of FIRE is based on whether the regression coefficients on all past revisions are jointly zero in the generalized BIR framework (**@Hongyu: refer to the equation in extension section**). For each test, we report the fraction of BIRs for which the null of FIRE is rejected with 95% confidence; also, we adjust individual p-values to account for the multiple testing problem following Benjamini and Hochberg (1995). The bottom rows of each table show the unique number of series and forecasters, respectively, in the cross-section for every n.

Data	N	mean	$p10$	$p25$	$p50$	$p75$	$p90$	$\mathbb{P}(B(n,0)<0)$
SPF	15	-0.22	-0.46	-0.40	-0.33	-0.14	0.18	80
IBES	1299	-0.32	-1.46	-0.88	-0.37	0.13	0.76	69

(a) Distribution of Initial Reaction in SPF and IBES

	W_1 bins for Overreactors			Underreactors
	1	2	3	
W_1	0.14	0.39***	0.53***	0.97***
W_2	0.03	0.11	0.51***	0.64***
W_3	0.01	-0.18	0.20	0.36**
$B(3,0)$	-0.32***	-0.36***	-0.34***	0.30***
# Series	4	4	4	3

(b) SPF: Correction Speed

	W_1 bins for Overreactors				W_1 bins for Underreactors			
	1	2	3	4	1	2	3	4
W_1	0.20	0.74***	1.00***	1.57***	0.39***	0.86***	1.03***	1.61***
W_2	0.01	0.67***	0.89***	1.64***	0.24	0.79***	0.96***	1.40***
W_3	0.08	0.57***	0.75***	1.60***	0.14	0.70***	0.92***	1.09***
W_4	0.18*	0.49***	0.68***	1.40***	0.07	0.62***	0.83***	0.92***
W_5	0.10	0.45***	0.56***	0.83***	0.04	0.57***	0.78***	0.59**
W_6	0.03	0.23***	0.30***	0.30**	0.08	0.47***	0.64***	0.58**
$B(6,0)$	-0.37***	-0.65***	-0.90***	-0.35***	0.49***	1.25***	1.16***	0.36**
% Forecasters	15	16	16	15	8	11	9	9

(c) IBES: Correction Speed

Table 4: Series-level Deviations from FIRE This table summarizes the size and correction of deviations from FIRE across series in the SPF and IBES. In Table 4a, we estimate series-level BIRs and sort series into one-hundred different groups based on the estimated $B(n,0)$. Within each group, we then estimate a pooled BIR and report summary statistics on the resulting $B(n,0)$. In Table 4b, we first sort series in SPF based on whether their $B(n,0)$ is negative (overreactors) or positive (underreactors). Within the set of overreactors, we further sort series into terciles on $W(n,1) = B(n,1)/B(n,0)$. Within each group, we then estimate a pooled BIR and report the resulting $W(n,j) = B(n,j)/B(n,0)$, a measure of correction speed. Because there is a limited number of series that exhibit underreaction in SPF ($B(n,0) > 0$), we do not sort further and instead report the estimated $W(n,j)$ when pooling across all series that underreact. Table 4c repeats this exercise for series in IBES, separately dividing series that overreact or underreact into quartiles based on $W(n,j)$. In all tables $n = 3$ for SPF and $n = 6$ for IBES.

Data	N	mean	$p10$	$p25$	$p50$	$p75$	$p90$	$\mathbb{P}(B(n,0)<0)$
SPF	190	-0.28	-1.14	-0.67	-0.36	0.05	0.47	72
IBES	970	-0.25	-1.58	-0.79	-0.27	0.27	1.01	63

(a) Distribution of Initial Reaction in SPF and IBES

	W_1 bins for Overreactors			W_1 bins for Underreactors	
	1	2	3	1	2
W_1	-0.20**	0.57***	1.14***	0.65***	1.47***
W_2	-0.08	0.48***	0.91***	0.58***	1.30***
W_3	-0.10	0.28**	0.39***	0.24	1.25***
$B(3,0)$	-0.33***	-0.61***	-0.56***	0.28***	0.34***
% Forecasters	24	24	24	14	14

(b) SPF: Correction Speed

	W_1 bins for Overreactors				W_1 bins for Underreactors			
	1	2	3	4	1	2	3	4
W_1	0.21**	0.81***	1.01***	1.63***	0.12	0.77***	0.98***	1.47***
W_2	0.05	0.67***	0.99***	1.54***	-0.03	0.66***	0.94***	1.19***
W_3	0.13	0.59***	0.86***	1.42***	-0.19	0.60***	0.95***	0.95***
W_4	0.21***	0.49***	0.77***	1.17***	-0.26	0.44***	0.89***	0.82***
W_5	0.22***	0.37***	0.66***	0.74***	-0.08	0.38***	0.89***	0.57***
W_6	0.05	0.21***	0.35***	0.20	0.13	0.29***	0.67***	0.21
$B(6,0)$	-0.39***	-0.78***	-0.99***	-0.34***	0.30***	1.03***	1.04***	0.34***
% Forecasters	16	16	16	16	9	9	9	9

(c) IBES: Correction Speed

Table 5: Forecaster-level Deviations from FIRE This table summarizes the size and correction of deviations from FIRE across forecasters in the SPF and IBES. In Table 4a, we estimate forecaster-level BIRs and sort forecasters into one-hundred different groups based on the estimated $B(n,0)$. Within each group, we then estimate a pooled BIR and report summary statistics on the resulting $B(n,0)$. In Table 4b, we first sort forecasters in SPF based on whether their $B(n,0)$ is negative (overreactors) or positive (underreactors). Within the set of overreactors, we further sort forecasters into terciles on $W(n,1) = B(n,1)/B(n,0)$. Within each group, we then estimate a pooled BIR and report the resulting $W(n,j) = B(n,j)/B(n,0)$, a measure of correction speed. Table 4c repeats this exercise for forecasters in IBES, separately dividing forecasters that overreact or underreact into quartiles based on $W(n,j)$. In all tables $n = 3$ for SPF and $n = 6$ for IBES.

A Linear Autonomous Forecasts and Leading Behavioral Models

We now show that many popular models of belief formation produce forecasts that a part of the linear autonomous class.

- **Adaptive Expectations** of Cagan (1956) and Nerlove (1958) model forecasts as:

$$F_t[y_{t+1}] = \delta y_t + (1-\delta) F_{t-1}[y_{t+1}]\,.$$

If y_t is an AR(1) process with the autocorrelation coefficient ψ, then $y_t = \psi^{-1} E_t[y_{t+1}]$. In this case, it is straightforward to show that:

$$F_t[y_{t+1}] = \sum_{\tau=0}^{\infty} (1-\delta)^\tau \delta \psi^{-1} E_{t-\tau}[y_{t+1}]\,,$$

which clearly satisfies the conditions of Definition 2.

- **Extrapolative Expectations** (Barberis et al., 2015b; Hirshleifer et al., 2015) take the form:

$$F_t[y_{t+1}] \;=\; y_t + \phi(y_t - y_{t-1}) \tag{20}$$

Assuming y_t follows an AR(1) process with coefficient ψ as above, we can rewrite extrapolative expectations as

$$F_t[y_{t+1}] \;=\; \psi^{-1} E_t[y_{t+1}] + \phi(\psi^{-1} E_t[y_{t+1}] - \psi^{-2} E_{t-1}[y_{t+1}]), \tag{21}$$

which again satisfies the conditions of Definition 2.

- **Noisy Information/Sticky Expectations** (e.g., Giannoni and Woodford (2003); Coibion and Gorodnichenko (2012)) postulate

$$F_t[y_{t+n}] \;=\; (1-\lambda)E_t[y_{t+n}] \;+\lambda F_{t-1}[y_{t+n}] + \;\varepsilon_t(n).$$

  Iterating this backward yields:

$$F_t[y_{t+n}] \;=\; \sum_{\tau=0}^{\infty}\lambda^{\tau}(1-\lambda)E_{t-\tau}[y_{t+n}] \;+\; noise_t\,. \tag{22}$$

- **Diagnostic Expectations** of Bordalo et al. (2020b) postulate that

$$F_t[y_{t+n}] = E_t[y_{t+n}] \;+\; \theta(E_t[y_{t+n}] - E_{t-J}[y_{t+n}]). \tag{23}$$

  for some lag $J \geq 1$, which is clearly a special case of Definition 2.

- **Smooth Diagnostic Expectations** introduced in (Bianchi et al., 2024a) postulate that

$$F_t[y_{t+n}] = E_t[y_{t+n}] \;+\; \delta(n)(E_t[y_{t+n}] - E_{t-J}[y_{t+n}]). \tag{24}$$

  for some lag $J \geq 1$, which is clearly a special case of Definition 2. (Bianchi et al., 2024a) provide a micro-foundation for the dependence of $\delta(n)$ on n, linking it to the volatility of the underlying stochastic process.

- **Mis-estimation of persistence.** As described in Gabaix (2019), the agent anchors the persistence parameter ψ to a default level of persistence, ψ_*,

$$\hat{\psi} = m\psi + (1-m)\psi_*, \tag{25}$$

where $m \in (0,1)$ controls the degree of irrationality. Assuming that

$$y_{t+1} \;=\; \psi y_t \;+\; \varepsilon_{t+1} \tag{26}$$

holds, we get

$$F_t[y_{t+n}] \;=\; (\hat{\psi}/\psi)^{n-1} E_t[y_{t+n}], \tag{27}$$

which is clearly a special case of Definition 2.

- **Limited Information Processing Capacity.** Afrouzi et al. (2023) develop a model of limited information capacity that generates deviations from FIRE with the strength of the deviations depending on the persistence ψ of the process as well as the forecasting horizon. They consider a simple AR(1) model,

$$y_t \;=\; (1-\rho)\mu \;+\; \rho y_{t-1} \;+\; \varepsilon_t, \tag{28}$$

so that

$$E_t[y_{t+n}] \;=\; \rho^n y_t \;+\; (1-\rho^n)\mu \tag{29}$$

and show that, with limited information capacity,

$$F_t[y_{t+n}] \;=\; \rho^j(\delta_1(n) y_t \;+\; \delta_2(n)\mu), \tag{30}$$

where the dependence of δ_1, δ_2 on n is subtle and non-linear. We discuss this model below. Here, we only note that it obviously fits Definition 2:

$$F_t[y_{t+n}] \;=\; \rho^n(\delta_1(n) y_t \;+\; \delta_2(n)\mu) \;=\; \delta_1(n) E_t[y_{t+n}] \;+\; \hat{\mu}(n) \tag{31}$$

for some constants $\hat{\mu}(n)$.

B Proofs for BIR Theory

Proposition 7 *Under the hypotheses of Proposition 2,*

$$\begin{aligned}
&\mathrm{Cov}(y_{t+n},\, F_t[y_{t+n}]) \;=\; \rho^{2n}\sigma_q^2(k_0(n)+\delta(n)\Gamma) \;>\; 0 \\
&\mathrm{Cov}(F_{t+j}[y_{t+n}],\, F_t[y_{t+n}]) \;=\; \sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \;+\; \rho^{2n}\sigma_q^2(k_0(n)k_0(n-j)+k_0(n-j)\delta(n)\Gamma+k_0(n)\delta(n-j)\Gamma(j)) \\
&+\; \rho^{2n}\delta(n-j)\delta(n)\sigma_q^2\Xi(j),
\end{aligned} \tag{32}$$

where

$$\Xi(j) \;=\; \sum_{\tau=0}^{j}\kappa(\tau) \;+\; \sum_{\tau_1=1}^{\infty}\sum_{\tau_2=0}^{\infty}\kappa(\tau_1+j)\kappa(\tau_2)\rho^{2\max(\tau_1,\tau_2)} \tag{33}$$

and

$$\Gamma(j) \;=\; \sum_{\tau=0}^{j}\kappa(\tau)+\sum_{\tau=1}^{\infty}\kappa(\tau+j)\rho^{2\tau} \tag{34}$$

In particular,

$$\begin{aligned}
&\mathrm{Var}[F_t[y_{t+n}]] \;=\; \sigma_\varepsilon^2(n) \;+\; \rho^{2n}\sigma_q^2(k_0(n)^2+k_0(n)\delta(n)\Gamma+k_0(n)\delta(n)\Gamma(0)) \\
&+\; \rho^{2n}\delta(n)^2\sigma_q^2\Xi(0)
\end{aligned} \tag{35}$$

and

$$
\begin{aligned}
&\mathrm{Var}[F_t[y_{t+n}] - F_{t-1}[y_{t+n}]] \\
&= \mathrm{Var}[F_t[y_{t+n}] \ + \ \mathrm{Var}[F_t[y_{t+n+1}] \ + \ 2\mathrm{Cov}(F_t[y_{t+n}], F_{t-1}[y_{t+n}]) \\
&= \sigma_\varepsilon^2(n) \ + \ \rho^{2n}\sigma_q^2(k_0(n)^2 + k_0(n)\delta(n)\Gamma + k_0(n)\delta(n)\Gamma(0)) \\
&+ \ \rho^{2n}\delta(n)^2\sigma_q^2\Xi(0) \\
&+ \ \sigma_\varepsilon^2(n+1) \ + \ \rho^{2(n+1)}\sigma_q^2(k_0(n+1)^2 + k_0(n+1)\delta(n+1)\Gamma + k_0(n+1)\delta(n+1)\Gamma(0)) \\
&+ \ \rho^{2(n+1)}\delta(n+1)^2\sigma_q^2\Xi(0) \\
&- 2\Big(\rho^{2(n+1)}\sigma_q^2(k_0(n+1)k_0(n) + k_0(n)\delta(n+1)\Gamma + k_0(n+1)\delta(n)\Gamma(1)) \\
&+ \ \rho^{2(n+1)}\delta(n)\delta(n+1)\sigma_q^2\Xi(1)\Big)
\end{aligned}
\tag{36}
$$

Proof of Propositions 7 and 2. We have

$$
\begin{aligned}
E_t[F_{t+j}[y_{t+n}]] \ &= \ E_t[\sum_{\tau=0}^{\infty}\gamma(n-j,\tau)\rho^{n+\tau-j}q_{t+j-\tau}] \\
&= \ \sum_{\tau=j+1}^{\infty}\gamma(n-j,\tau)\rho^{n+\tau-j}q_{t+j-\tau} \\
&+ \ q_t\sum_{\tau=0}^{j}\gamma(n-j,\tau)\rho^n \\
&= \ q_t\sum_{\tau=0}^{j}\gamma(n-j,\tau)\rho^n \ + \ \sum_{\tau=1}^{\infty}\gamma(n-j,\tau+j)q_{t-\tau}\rho^{n+\tau}
\end{aligned}
\tag{37}
$$

and

$$
\begin{aligned}
E_t[y_{t+n} - F_{t+j}[y_{t+n}]] \ &= \ E_t[y_{t+n}] - E_t[F_{t+j}[y_{t+n}]] \\
&= \ \rho^n q_t \ - \ q_t\sum_{\tau=0}^{j}\gamma(n-j,\tau)\rho^n \ - \ \sum_{\tau=1}^{\infty}\gamma(n-j,\tau+j)q_{t-\tau}\rho^{n+\tau} \\
&= \ \rho^n q_t\left(1 - \sum_{\tau=0}^{j}\gamma(n-j,\tau)\right) \ - \ \sum_{\tau=1}^{\infty}\gamma(n-j,\tau+j)q_{t-\tau}\rho^{n+\tau}
\end{aligned}
\tag{38}
$$

and hence

$$\begin{aligned}
&\mathrm{Cov}(E_t[y_{t+n} - F_{t+j}[y_{t+n}]], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \mathrm{Cov}(E_t[y_{t+n}], F_t[y_{t+n}]) - \mathrm{Cov}(E_t[y_{t+n}], F_{t-1}[y_{t+n}]) \\
&- \mathrm{Cov}(F_{t+j}[y_{t+n}], F_t[y_{t+n}]) + \mathrm{Cov}(F_{t+j}[y_{t+n}], F_{t-1}[y_{t+n}])\,.
\end{aligned} \tag{39}$$

We have

$$\begin{aligned}
&\mathrm{Cov}(y_{t+n}, F_t[y_{t+n}]) = \mathrm{Cov}(E_t[y_{t+n}], F_t[y_{t+n}]) = \rho^{2n}\mathrm{Cov}(q_t, k_0(n)q_t + \delta(n)\sum_{\tau=0}^{\infty}\kappa(\tau)\rho^{\tau}q_{t-\tau}) \\
&= \rho^{2n}\sigma_q^2(k_0(n) + \delta(n)\Gamma)
\end{aligned} \tag{40}$$

and, similarly, by stationarity,

$$\mathrm{Cov}(y_{t+n}, F_{t-1}[y_{t+n}]) = \mathrm{Cov}(y_{t+n+1}, F_t[y_{t+n+1}]) = \rho^{2(n+1)}(k_0(n+1) + \delta(n+1)\Gamma)\,. \tag{41}$$

Then,

$$
\begin{aligned}
\mathrm{Cov}(F_{t+j}[y_{t+n}], F_t[y_{t+n}]) \;&=\; \sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \\
&+\; \mathrm{Cov}(E_t[F_{t+j}[y_{t+n}]], F_t[y_{t+n}]) \\
&\underbrace{=}_{(37)}\; \sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \\
&+\; \mathrm{Cov}(q_t \sum_{\tau=0}^{j} \gamma(n-j,\tau)\rho^n \\
&+\; \sum_{\tau=1}^{\infty} \gamma(n-j,\tau+j) q_{t-\tau}\rho^{n+\tau}, \sum_{\tau=0}^{\infty} \gamma(n,\tau) q_{t-\tau}\rho^{n+\tau}) \\
&=\; \sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \\
&+\; \mathrm{Cov}(q_t k_0(n-j)\rho^n +\; q_t \delta(n-j) \sum_{\tau=0}^{j} \kappa(\tau)\rho^n \\
&+\; \delta(n-j) \sum_{\tau=1}^{\infty} \kappa(\tau+j) q_{t-\tau}\rho^{n+\tau}, \rho^n k_0(n) q_t + \delta(n) \sum_{\tau=0}^{\infty} \kappa(\tau) q_{t-\tau}\rho^{n+\tau}) \\
&=\; \sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \;+\; \rho^{2n}\sigma_q^2(k_0(n)k_0(n-j) + k_0(n-j)\delta(n)\Gamma + k_0(n)\delta(n-j)\Gamma(j)) \\
&+\; \rho^{2n}\delta(n-j)\delta(n)\sigma_q^2 \sum_{\tau=0}^{j} \kappa(\tau)\Gamma \\
&+\; \rho^{2n}\delta(n-j)\delta(n)\sigma_q^2 \sum_{\tau_1=1}^{\infty}\sum_{\tau_2=0}^{\infty} \kappa(\tau_1+j)\kappa(\tau_2)\rho^{2\max(\tau_1,\tau_2)} \\
&=\; \sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \;+\; \rho^{2n}\sigma_q^2(k_0(n)k_0(n-j) + k_0(n-j)\delta(n)\Gamma + k_0(n)\delta(n-j)\Gamma(j)) \\
&+\; \rho^{2n}\delta(n-j)\delta(n)\sigma_q^2\Xi(j)\,,
\end{aligned}
\tag{42}
$$

where we have defined

$$
\Gamma(j) \;=\; \sum_{\tau=0}^{j} \kappa(\tau) + \sum_{\tau=1}^{\infty} \kappa(\tau+j)\rho^{2\tau},\; \Gamma = \Gamma(0). \tag{43}
$$

Thus, by stationarity,

$$\begin{aligned}
&\mathrm{Cov}(F_{t+j}[y_{t+n}], F_{t-1}[y_{t+n}]) \;=\; \mathrm{Cov}(F_{t+j+1}[y_{t+n+1}], F_t[y_{t+n+1}]) \\
&= \;\rho^{2(n+1)}\sigma_q^2(k_0(n+1)k_0(n-j) + k_0(n-j)\delta(n+1)\Gamma + k_0(n+1)\delta(n-j)\Gamma(j+1)) \\
&+ \;\rho^{2(n+1)}\delta(n-j)\delta(n+1)\sigma_q^2\Xi(j+1)
\end{aligned} \tag{44}$$

Thus,

$$\begin{aligned}
&\mathrm{Cov}(y_{t+n} - F_{t+j}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \;\mathrm{Cov}(E_t[y_{t+n}], F_t[y_{t+n}]) \;-\; \mathrm{Cov}(E_t[y_{t+n}], F_{t-1}[y_{t+n}]) \\
&- \mathrm{Cov}(F_{t+j}[y_{t+n}], F_t[y_{t+n}]) \;+\; \mathrm{Cov}(F_{t+j}[y_{t+n}], F_{t-1}[y_{t+n}]) \\
&= \;\rho^{2n}\sigma_q^2(k_0(n) + \delta(n)\Gamma) - \rho^{2(n+1)}(k_0(n+1) + \delta(n+1)\Gamma) \\
&- \sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \;-\; \rho^{2n}\sigma_q^2(k_0(n)k_0(n-j) + k_0(n-j)\delta(n)\Gamma + k_0(n)\delta(n-j)\Gamma(j)) \\
&- \;\rho^{2n}\delta(n-j)\delta(n)\sigma_q^2\Xi(j) \\
&+ \;\rho^{2(n+1)}\sigma_q^2(k_0(n+1)k_0(n-j) + k_0(n-j)\delta(n+1)\Gamma + k_0(n+1)\delta(n-j)\Gamma(j+1)) \\
&+ \;\rho^{2(n+1)}\delta(n-j)\delta(n+1)\sigma_q^2\Xi(j+1) \\
&= \;-\sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \\
&+ \;\rho^{2n}\sigma_q^2(1 - k_0(n-j))((k_0(n) - \rho^2 k_0(n+1)) + \Gamma(\delta(n) - \rho^2\delta(n+1))) \\
&- \rho^{2n}\,\sigma_q^2((k_0(n)\delta(n-j)\Gamma(j) - \rho^2 k_0(n+1)\delta(n-j)\Gamma(j+1)))) \\
&- \;\rho^{2n}\,\sigma_q^2\,\delta(n-j)(\delta(n)\Xi(j) \;- \rho^2\delta(n+1)\Xi(j+1)) \\
&= \;-\sigma_\varepsilon^2(n)\mathbf{1}_{j=0} \\
&+ \;\rho^{2n}\sigma_q^2(1 - k_0(n-j))((k_0(n) - \rho^2 k_0(n+1)) + \Gamma(\delta(n) - \rho^2\delta(n+1))) \\
&- \rho^{2n}\,\sigma_q^2\delta(n-j)\Big[(k_0(n)\Gamma(j) - \rho^2 k_0(n+1)\Gamma(j+1)) \\
&+ (\delta(n)\Xi(j) \;- \rho^2\delta(n+1)\Xi(j+1))\Big]
\end{aligned} \tag{45}$$

Then, the regression coefficient is given by

$$\frac{\text{Cov}(y_{t+n} - F_{t+j}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}])}{\text{Var}[F_t[y_{t+n}] - F_{t-1}[y_{t+n}]]} \tag{46}$$

and the claim follows once we define

$$\begin{aligned}
D(n) &= D_{MZ}(n) + \rho^2 D_{MZ}(n+1) - 2\rho^2(k_0(n+1)k_0(n) + k_0(n)\delta(n+1) + k_0(n+1)\delta(n)\Gamma(1) \\
&\quad + \delta(n)\delta(n+1)\Xi(1)) \\
D_{MZ}(n) &= k_0(n)^2 + 2k_0(n)\delta(n) + \delta(n)^2\Xi(0) \\
\Xi(j) &= \sum_{\tau=0}^{j} \kappa(\tau) \; + \; \sum_{\tau_1=1}^{\infty}\sum_{\tau_2=0}^{\infty} \kappa(\tau_1 + j)\kappa(\tau_2)\rho^{2\max(\tau_1,\tau_2)} \\
\Gamma(j) &= \sum_{\tau=0}^{j} \kappa(\tau) + \sum_{\tau=1}^{\infty} \kappa(\tau + j)\rho^{2\tau}\,.
\end{aligned} \tag{47}$$

□

Proof of Proposition 4. The claim follows directly from Proposition 2. □

Corollary 8 (BIRs in (Smooth) Diagnostic Expectations) *Suppose that*

$$F_t[y_{t+n}] \;=\; E_t[y_{t+n}] \;+\; \theta(n)(E_t[y_{t+n}] - E_{t-J}[y_{t+n}]). \tag{48}$$

Then,

$$\begin{aligned}
\Xi(j) &= (1-\rho^{2J})^{-1}(1 - (1-\rho^{2J})^{-1}(\rho^{2(J-j)} - \rho^{2J}))\mathbf{1}_{j<J} \\
\Gamma(j) &= (1-\rho^{2J})^{-1}(1 - \rho^{2(J-j)})\mathbf{1}_{j<J}
\end{aligned} \tag{49}$$

and

$$\kappa(\tau) = (1-\rho^{2J})^{-1} \begin{cases} 1, & \textit{if } \tau = 0 \\ -1, & \textit{if } \tau = J \\ 0, & \textit{if } \tau \notin \{0, J\} \end{cases}$$

and $\delta(n) = \theta(n)(1-\rho^{2J})$.

Proof of Corollaries 3 and 8. We perform the proof directly for smooth diagnostic expectations.

We have that $\kappa(\tau)$ is given by:

$$\kappa(\tau) = (1-\rho^{2J})^{-1} \begin{cases} 1, & \text{if } \tau = 0 \\ -1, & \text{if } \tau = J \\ 0, & \text{if } \tau \notin \{0, J\} \end{cases}$$

and $\delta(n) = \theta(1-\rho^{2J})$. Then,

$$\gamma(n,\tau) \;=\; \mathbf{1}_{\tau=0} \;+\; \delta(n)\kappa(\tau)\,. \tag{50}$$

Then,

$$
\begin{aligned}
\Xi(j) &= \sum_{\tau=0}^{j} \kappa(\tau) + \sum_{\tau_1=1}^{\infty}\sum_{\tau_2=0}^{\infty} \kappa(\tau_1+j)\kappa(\tau_2)\rho^{2\max(\tau_1,\tau_2)} \\
&= (1-\rho^{2J})^{-1}\mathbf{1}_{j<J} - (1-\rho^{2J})^{-2}\sum_{\tau_1=1}^{\infty}\mathbf{1}_{\tau_1\geq 1}\mathbf{1}_{\tau_1+j=J}(\rho^{2\max(\tau_1,0)}-\rho^{2\max(\tau_1,J)}) \\
&= (1-\rho^{2J})^{-1}\mathbf{1}_{j<J} - (1-\rho^{2J})^{-2}(\rho^{2(J-j)}-\rho^{2J})\mathbf{1}_{j<J} \\
&= (1-\rho^{2J})^{-1}(1-(1-\rho^{2J})^{-1}(\rho^{2(J-j)}-\rho^{2J}))\mathbf{1}_{j<J} \\
\Gamma(j) &= \sum_{\tau=0}^{j}\kappa(\tau) + \sum_{\tau=1}^{\infty}\kappa(\tau+j)\rho^{2\tau} \\
&= (1-\rho^{2J})^{-1}(1-\rho^{2(J-j)})\mathbf{1}_{j<J}
\end{aligned}
\tag{51}
$$

□

Corollary 9 *For $J>1$, we have*

$$
B(n,j) = const\begin{cases} 1-\rho^2, & j \leq J-1 \\ 0, & j > J-1 \end{cases} \tag{52}
$$

where

$$
const = -\theta(1+\theta)((1+\rho^2)D_{MZ} - 2\rho^2(1+\theta(1-\rho^{2J})+\theta(1+\theta)(1-\rho^{2(J-1)})))^{-1} \tag{53}
$$

with

$$
D_{MZ} = 1+2\theta(1-\rho^{2J})+\theta^2(1-\rho^{2J}) \tag{54}
$$

Proof of Corollary 9. Recall that

$$B(n,j) \;=\; \frac{A(n,j)-\rho^2 A(n+1,j+1)}{D(n)}, \tag{55}$$

where

$$\begin{aligned} A(n,j) &= -\delta(n-j)\Big[\Gamma(j)+\delta(n)\Xi(j)\Big] \\ D(n) &= D_{MZ}(n)+\rho^2 D_{MZ}(n+1)-2\rho^2(1+\delta(n+1) \\ &\quad +\delta(n)\Gamma(1)+\delta(n)\delta(n+1)\Xi(1)) \\ D_{MZ}(n) &= 1+2\delta(n)+\delta(n)^2\Xi(0) \end{aligned} \tag{56}$$

Thus, for $J>1$, we have

$$\begin{aligned} A(n,j) \;&=\; -\theta(1+\theta)[1-\rho^{2(J-j)}]\mathbf{1}_{j<J} \\ D_{MZ} &= 1+2\theta(1-\rho^{2J})+\theta^2(1-\rho^{2J}) \\ D(n) \;&=\; (1+\rho^2)D_{MZ}-2\rho^2(1+\theta(1-\rho^{2J})+\theta(1+\theta)(1-\rho^{2(J-1)})) \end{aligned} \tag{57}$$

□

C Extrapolative Expectations

Recall that, in the main text (Corollary 6), we study simple extrapolative expectations in the following setting: The true data generating process has a constant mean so that $y_{t+1}=\bar{y}+\varepsilon_{t+1}$, where ε_{t+1} is i.i.d., while forecasts are assumed to evolve according to:

$$\begin{aligned} F_t[y_{t+1}] \;&=\; \mu_t \\ \mu_{t+1} \;&=\; (1-\rho-\tilde{\rho})\bar{\mu} \;+\; \rho\mu_t \;+\; \tilde{\rho}y_{t+1}, \end{aligned} \tag{58}$$

where $\tilde{\rho}$ modulates the degree of extrapolation from the most recent realization of y, ρ controls the impact of extrapolation made from the past.

In this appendix, we introduce an extension of the model of extrapolative expectations studied in the main text, allowing for persistence in the underlying process.

Corollary 10 *Suppose that* $y_{t+1} = \bar{y} + \rho_y(y_t - \bar{y}) + \varepsilon_{t+1}$ *with* $|\rho_y| < 1$ *and* ε_{t+1} *are i.i.d., whereas the agent believes*

$$\begin{aligned} F_t[y_{t+1}] &= \mu_t \\ \mu_{t+1} &= (1-\rho-\tilde{\rho})\bar{\mu} + \rho\mu_t + \tilde{\rho}y_{t+1} \\ F_t[\mu_{t+1}] &= (1-\rho-\tilde{\rho})\bar{\mu} + (\rho+\tilde{\rho})\mu_t\,, \end{aligned} \tag{59}$$

with $|\rho| < 1$. *The model is linear, autonomous, and separable, with*

$$\delta(n) = \frac{\tilde{\rho}}{\rho+\tilde{\rho}}\frac{1}{1-\rho\rho_y}((\rho+\tilde{\rho})/\rho_y)^n,\ \kappa(\tau) = (1-\rho\rho_y)(\rho/\rho_y)^\tau \tag{60}$$

We have

$$B(n,j) = \frac{A(n,j)}{D(n)}, \tag{61}$$

where

$$\begin{aligned} A(n,j) &= \mathrm{Cov}(y_{t+n} - F_{t+j}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\ &= \hat{\rho}^{n-1}\left(\rho_y^n - \hat{\rho}^{n-j-1}\tilde{\rho}\frac{\rho^{j+1}-\rho_y^{j+1}}{\rho-\rho_y}\right)\left(\tilde{\rho} - \tilde{\rho}^2\rho_y\frac{1}{1-\rho_y\rho}\right) \\ &- \hat{\rho}^{n-1}\hat{\rho}^{n-j-1}\tilde{\rho}\left(\tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y} - \rho^j\tilde{\rho}^2\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)}\right) \end{aligned} \tag{62}$$

and

$$\begin{aligned} D(n) &= \mathrm{Var}[F_t[y_{t+n}] - F_{t-1}[y_{t+n}]] \\ &= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{1-\rho\hat{\rho}-\hat{\rho}(\rho_y-\tilde{\rho}-\rho\rho_y\hat{\rho})}{(1-\rho^2)(1-\rho\rho_y)}\right). \end{aligned} \tag{63}$$

The rational expectations case corresponds to $\rho = 0, \tilde{\rho} = \rho_y$.

Proof of Corollaries 10 and 6. We have

$$F_t[y_{t+n}] \;=\; F_t[\mu_{t+n-1}] \;=\; cost \;+\; (\rho+\tilde{\rho})^{n-1}\mu_t \tag{64}$$

Without loss of generality, we normalize $\mathrm{Var}[y_t] = 1$. The AR(1) structure of y_t implies that

$$\mathrm{Cov}(y_t, y_{t+h}) \;=\; \rho_y^{|h|} \tag{65}$$

We have

$$\mu_t \;=\; \sum_{\tau=0}^{\infty}\rho^\tau\tilde{\rho}y_{t-\tau} \;+\; const\,. \tag{66}$$

and

$$E_t[y_{t+n}] \;=\; const \;+\; \rho_y^n y_t \tag{67}$$

and, hence,

$$E_{t-\tau}[y_{t+n}] \;=\; const \;+\; \rho_y^{n+\tau}y_{t-\tau} \tag{68}$$

Let

$$\hat{\rho} = (\rho + \tilde{\rho}) \tag{69}$$

Thus,

$$\begin{aligned}
\mathrm{Cov}(y_{t+n}, F_t[y_{t+n}]) \;&=\; (\rho+\tilde{\rho})^{n-1}\mathrm{Cov}(y_{t+n}, \mu_t) \\
&\underbrace{=}_{\mathrm{Cov}(y_{t+n},y_{t-\tau})=\rho_y^{n+\tau}\ by\ (65)} (\rho+\tilde{\rho})^{n-1}\mathrm{Cov}(y_{t+n}, \sum_{\tau=0}^{\infty}\rho^\tau\tilde{\rho}y_{t-\tau}) \\
&=\; (\rho+\tilde{\rho})^{n-1}\tilde{\rho}\sum_{\tau=0}^{\infty}\rho^\tau\rho_y^{\tau+n}\sigma_y^2 \\
&=\; (\rho+\tilde{\rho})^{n-1}\tilde{\rho}\rho_y^n\frac{1}{1-\rho\rho_y}\sigma_y^2,
\end{aligned} \tag{70}$$

while, by (64), with $b = \rho/\rho_y$, we get

$$\begin{aligned}
E_t[F_{t+j}[y_{t+n}]] \;&=\; E_t[(\rho+\tilde{\rho})^{n-j-1}\mu_{t+j}] = \;E_t[(\rho+\tilde{\rho})^{n-j-1}\mu_{t+j}] \\
&=\; E_t[(\rho+\tilde{\rho})^{n-j-1}\mu_{t+j}] \\
&\underbrace{=}_{(66)}\; const + E_t[(\rho+\tilde{\rho})^{n-j-1}\sum_{\tau=0}^{\infty}\rho^\tau\tilde{\rho}y_{t-\tau+j}] \\
&=\; const \;+\; (\rho+\tilde{\rho})^{n-j-1}\sum_{\tau=0}^{j}\rho^\tau\tilde{\rho}\rho_y^{j-\tau}y_t \;+\; (\rho+\tilde{\rho})^{n-j-1}\tilde{\rho}\sum_{\tau=j+1}^{\infty}\rho^\tau y_{t-\tau+j} \\
&=\; const \;+\; \hat{\rho}^{n-j-1}\tilde{\rho}\rho_y^j\frac{b^{j+1}-1}{b-1}y_t \;+\; \hat{\rho}^{n-j-1}\tilde{\rho}\sum_{\tau=j+1}^{\infty}\rho^\tau y_{t-\tau+j}\,,
\end{aligned} \tag{71}$$

and, hence,

$$\begin{aligned}
&\mathrm{Cov}(y_{t+n} - F_{t+j}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \mathrm{Cov}(E_t[y_{t+n} - F_{t+j}[y_{t+n}]], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \mathrm{Cov}\left(\rho_y^n y_t - \hat{\rho}^{n-j-1}\tilde{\rho}\Big(\rho_y^j \frac{b^{j+1}-1}{b-1} y_t \; + \; \sum_{\tau=j+1}^{\infty} \rho^\tau y_{t-\tau+j}\Big), \hat{\rho}^{n-1}\mu_t - \hat{\rho}^n \mu_{t-1}\right).
\end{aligned} \tag{72}$$

Now, by (66),

$$\begin{aligned}
\mu_t - \hat{\rho}\mu_{t-1} \; &= \; const \; + \; \sum_{\tau=0}^{\infty} \rho^\tau \tilde{\rho} y_{t-\tau} \\
&- \hat{\rho} \sum_{\tau=0}^{\infty} \rho^\tau \tilde{\rho} y_{t-1-\tau} \\
&= \; const \; + \; \tilde{\rho} y_t \; + \sum_{\tau=1}^{\infty} \rho^{\tau-1}(\rho - \hat{\rho})\tilde{\rho} y_{t-\tau} \\
&= \; const + \tilde{\rho} y_t - \sum_{\tau=1}^{\infty} \rho^{\tau-1}\tilde{\rho}^2 y_{t-\tau}
\end{aligned} \tag{73}$$

because

$$\rho - \hat{\rho} \; = \; -\tilde{\rho}. \tag{74}$$

Without loss of generality, we normalize $\text{Var}[y_t] = \sigma_y^2 = 1$. Then,

$$
\begin{aligned}
&\text{Cov}(y_{t+n} - F_{t+j}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \text{Cov}(E_t[y_{t+n} - F_{t+j}[y_{t+n}]], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \hat{\rho}^{n-1}\text{Cov}\left(\rho_y^n y_t - \hat{\rho}^{n-j-1}\tilde{\rho}\Big(\rho_y^j \frac{b^{j+1}-1}{b-1} y_t \; + \; \sum_{\tau=j+1}^{\infty} \rho^\tau y_{t-\tau+j}\Big), \mu_t - \hat{\rho}\mu_{t-1}\right) \\
&= \hat{\rho}^{n-1}\left(\rho_y^n - \hat{\rho}^{n-j-1}\tilde{\rho}\rho_y^j \frac{b^{j+1}-1}{b-1}\right)\text{Cov}\left(y_t, \tilde{\rho} y_t - \sum_{\tau=1}^{\infty} \rho^{\tau-1}\tilde{\rho}^2 y_{t-\tau}\right) \\
&- \hat{\rho}^{n-1}\hat{\rho}^{n-j-1}\tilde{\rho}\text{Cov}\left(\sum_{\tau=j+1}^{\infty} \rho^\tau y_{t-\tau+j}, \tilde{\rho} y_t - \sum_{\tau=1}^{\infty} \rho^{\tau-1}\tilde{\rho}^2 y_{t-\tau}\right).
\end{aligned}
\tag{75}
$$

Now,

$$
\begin{aligned}
&\text{Cov}\left(y_t, \tilde{\rho} y_t - \sum_{\tau=1}^{\infty} \rho^{\tau-1}\tilde{\rho}^2 y_{t-\tau}\right) \\
&= \tilde{\rho} \; - \; \sum_{\tau=1}^{\infty} \rho_y^\tau \rho^{\tau-1}\tilde{\rho}^2 \\
&= \tilde{\rho} - \tilde{\rho}^2 \rho_y \frac{1}{1-\rho_y\rho}.
\end{aligned}
\tag{76}
$$

Similarly,

$$
\begin{aligned}
&\mathrm{Cov}\left(\sum_{\tau_1=j+1}^{\infty}\rho^{\tau_1}y_{t-\tau_1+j},\tilde{\rho}y_t-\sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}^2y_{t-\tau}\right)\\
&=\sum_{\tau_1=j+1}^{\infty}\rho^{\tau_1}\mathrm{Cov}\left(y_{t-\tau_1+j},\tilde{\rho}y_t-\sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}^2y_{t-\tau}\right)\\
&=\tilde{\rho}\sum_{\tau_1=j+1}^{\infty}\rho^{\tau_1}\rho_y^{|-\tau_1+j|}\\
&-\sum_{\tau_1=j+1}^{\infty}\rho^{\tau_1}\sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}^2\rho_y^{|\tau-\tau_1+j|}\\
&=\tilde{\rho}\sum_{\tau=1}^{\infty}\rho^{\tau+j}\rho_y^{\tau}\;-\;\sum_{\tau_1=1}^{\infty}\rho^{\tau_1+j}\sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}^2\rho_y^{|\tau-\tau_1|}\\
&=\tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y}\\
&-\rho^j\tilde{\rho}^2\sum_{\tau_1=1}^{\infty}\rho^{\tau_1}\left(\sum_{\tau=1}^{\tau_1}\rho^{\tau-1}\rho_y^{\tau_1-\tau}\;+\;\sum_{\tau=\tau_1+1}^{\infty}\rho^{\tau-1}\rho_y^{\tau-\tau_1}\right) \qquad (77)\\
&=\tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y}-\rho^j\tilde{\rho}^2\sum_{\tau_1=1}^{\infty}\rho^{\tau_1}\left(\rho_y^{\tau_1}(1/\rho_y)\frac{(\rho/\rho_y)^{\tau_1}-1}{(\rho/\rho_y)-1}\;+\;\sum_{\tau=1}^{\infty}\rho^{\tau_1+\tau-1}\rho_y^{\tau}\right)\\
&=\tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y}-\rho^j\tilde{\rho}^2\sum_{\tau_1=1}^{\infty}\rho^{\tau_1}\left(\frac{\rho^{\tau_1}-\rho_y^{\tau_1}}{\rho-\rho_y}\;+\;\rho^{\tau_1}\rho_y\frac{1}{1-\rho\rho_y}\right)\\
&=\tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y}-\rho^j\tilde{\rho}^2\sum_{\tau_1=1}^{\infty}\frac{\rho^{2\tau_1}-(\rho\rho_y)^{\tau_1}}{\rho-\rho_y}\\
&-\rho^j\tilde{\rho}^2\sum_{\tau_1=1}^{\infty}\rho^{2\tau_1}\rho_y\frac{1}{1-\rho\rho_y}\\
&=\tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y}-\rho^j\tilde{\rho}^2\frac{\frac{\rho^2}{1-\rho^2}-\frac{\rho\rho_y}{1-\rho\rho_y}}{\rho-\rho_y}\\
&-\rho^j\tilde{\rho}^2\frac{\rho^2}{1-\rho^2}\rho_y\frac{1}{1-\rho\rho_y}
\end{aligned}
$$

A final transformation gives

$$\frac{\frac{\rho^2}{1-\rho^2} - \frac{\rho\rho_y}{1-\rho\rho_y}}{\rho - \rho_y} \;=\; \frac{\rho^2(1-\rho\rho_y) - \rho\rho_y(1-\rho^2)}{(1-\rho^2)(1-\rho\rho_y)(\rho-\rho_y)} \;=\; \frac{\rho}{(1-\rho^2)(1-\rho\rho_y)} \tag{78}$$

and hence,

$$\begin{aligned} &\mathrm{Cov}\left(\sum_{\tau=j+1}^{\infty}\rho^{\tau}y_{t-\tau+j}, \tilde{\rho}y_t - \sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}^2 y_{t-\tau}\right) \\ &= \tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y} - \rho^j\tilde{\rho}^2\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)}. \end{aligned} \tag{79}$$

Note also that

$$\rho_y^j\frac{b^{j+1}-1}{b-1} \;=\; \frac{\rho^{j+1}-\rho_y^{j+1}}{\rho-\rho_y} \tag{80}$$

so that, summarizing, we get

$$\begin{aligned} &\mathrm{Cov}(y_{t+n} - F_{t+j}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\ &= \mathrm{Cov}(E_t[y_{t+n} - F_{t+j}[y_{t+n}]], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\ &= \mathrm{Cov}\left(\rho_y^n y_t - \hat{\rho}^{n-j-1}\tilde{\rho}\Big(\rho_y^j\frac{b^{j+1}-1}{b-1}y_t \;+\; \sum_{\tau=j+1}^{\infty}\rho^{\tau}y_{t-\tau+j}\Big), \hat{\rho}^{n-1}\mu_t - \hat{\rho}^n\mu_{t-1}\right) \\ &= \hat{\rho}^{n-1}\left(\rho_y^n - \hat{\rho}^{n-j-1}\tilde{\rho}\frac{\rho^{j+1}-\rho_y^{j+1}}{\rho-\rho_y}\right)\left(\tilde{\rho} - \tilde{\rho}^2\rho_y\frac{1}{1-\rho_y\rho}\right) \\ &- \hat{\rho}^{n-1}\hat{\rho}^{n-j-1}\tilde{\rho}\left(\tilde{\rho}\rho^j\frac{\rho\rho_y}{1-\rho\rho_y} - \rho^j\tilde{\rho}^2\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)}\right) \end{aligned} \tag{81}$$

Similarly,

$$\begin{aligned} \mathrm{Var}[F_t[y_{t+n}] - F_{t-1}[y_{t+n}]] \;&=\; \mathrm{Cov}(F_t[y_{t+n}] - F_{t-1}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\ &= \mathrm{Cov}(F_t[y_{t+n}] - F_{t-1}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \end{aligned} \tag{82}$$

Above, we have computed

$$\begin{aligned}
&\mathrm{Cov}(F_t[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \mathrm{Cov}\left(\hat{\rho}^{n-1}\tilde{\rho}\Big(y_t + \sum_{\tau=1}^{\infty}\rho^{\tau}y_{t-\tau}\Big), \hat{\rho}^{n-1}\mu_t - \hat{\rho}^{n}\mu_{t-1}\right) \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\mathrm{Cov}\left(\Big(y_t + \sum_{\tau=1}^{\infty}\rho^{\tau}y_{t-\tau}\Big), y_t - \sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}y_{t-\tau}\right)
\end{aligned} \tag{83}$$

From (79), we know that

$$\begin{aligned}
&\mathrm{Cov}\left(\sum_{\tau=1}^{\infty}\rho^{\tau}y_{t-\tau}, \tilde{\rho}y_t - \sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}^2y_{t-\tau}\right) \\
&= \tilde{\rho}\frac{\rho\rho_y}{1-\rho\rho_y} - \tilde{\rho}^2\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)},
\end{aligned} \tag{84}$$

while, by (76),

$$\begin{aligned}
&\mathrm{Cov}\left(y_t, \tilde{\rho}y_t - \sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}^2y_{t-\tau}\right) \\
&= \tilde{\rho} - \tilde{\rho}^2\rho_y\frac{1}{1-\rho_y\rho},
\end{aligned} \tag{85}$$

implying that

$$
\begin{aligned}
&\mathrm{Cov}(F_t[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\mathrm{Cov}\left(\left(y_t \ + \ \sum_{\tau=1}^{\infty}\rho^{\tau}y_{t-\tau}\right), y_t - \sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}y_{t-\tau}\right) \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{\rho\rho_y}{1-\rho\rho_y} - \tilde{\rho}\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)}\right. \\
&\left. + 1 \ - \ \tilde{\rho}\rho_y\frac{1}{1-\rho_y\rho}\right) \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{\rho\rho_y}{1-\rho\rho_y} - \tilde{\rho}\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)}\right. \\
&\left. + 1 \ - \ \tilde{\rho}\rho_y\frac{1}{1-\rho_y\rho}\right) \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{1}{1-\rho\rho_y} - \tilde{\rho}\frac{\rho}{(1-\rho^2)(1-\rho\rho_y)}\right) \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{1-\rho\tilde{\rho}-\rho^2}{(1-\rho^2)(1-\rho\rho_y)}\right)
\end{aligned}
\tag{86}
$$

Similarly, by (84),

$$
\begin{aligned}
&\mathrm{Cov}(F_{t-1}[y_{t+n}], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \hat{\rho}^{2n-1}\tilde{\rho}^2\mathrm{Cov}\left(\left(y_{t-1} \ + \ \sum_{\tau=1}^{\infty}\rho^{\tau}y_{t-1-\tau}\right), y_t - \sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}y_{t-\tau}\right) \\
&= \hat{\rho}^{2n-1}\tilde{\rho}^2\rho^{-1}\mathrm{Cov}\left(\left(\sum_{\tau=1}^{\infty}\rho^{\tau}y_{t-\tau}\right), y_t - \sum_{\tau=1}^{\infty}\rho^{\tau-1}\tilde{\rho}y_{t-\tau}\right) \\
&= \hat{\rho}^{2n-1}\tilde{\rho}^2\rho^{-1}\left(\frac{\rho\rho_y}{1-\rho\rho_y} - \tilde{\rho}\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)}\right) \\
&= \hat{\rho}^{2n-1}\tilde{\rho}^2\rho^{-1}\left(\frac{\rho\rho_y}{1-\rho\rho_y} - \tilde{\rho}\frac{\rho(1+\rho\rho_y)}{(1-\rho^2)(1-\rho\rho_y)}\right) \\
&= \hat{\rho}^{2n-1}\tilde{\rho}^2\frac{\rho_y - \tilde{\rho} - \rho\rho_y(\rho+\tilde{\rho})}{(1-\rho^2)(1-\rho\rho_y)}
\end{aligned}
\tag{87}
$$

Thus,

$$\begin{aligned}
&\mathrm{Var}[F_t[y_{t+n}] - F_{t-1}[y_{t+n}]] \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{1-\rho\tilde{\rho}-\rho^2}{(1-\rho^2)(1-\rho\rho_y)}\right) - \hat{\rho}^{2n-1}\tilde{\rho}^2\frac{\rho_y-\tilde{\rho}-\rho\rho_y(\rho+\tilde{\rho})}{(1-\rho^2)(1-\rho\rho_y)} \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{1-\rho\tilde{\rho}-\rho^2}{(1-\rho^2)(1-\rho\rho_y)} - \hat{\rho}\frac{\rho_y-\tilde{\rho}-\rho\rho_y(\rho+\tilde{\rho})}{(1-\rho^2)(1-\rho\rho_y)}\right) \\
&= \hat{\rho}^{2(n-1)}\tilde{\rho}^2\left(\frac{1-\rho\hat{\rho}-\hat{\rho}(\rho_y-\tilde{\rho}-\rho\rho_y\hat{\rho})}{(1-\rho^2)(1-\rho\rho_y)}\right).
\end{aligned} \tag{88}$$

In the simpler case of $\rho_y = 0$, we have

$$F_t[y_{t+n}] \;=\; (\rho+\tilde{\rho})^{n-1}\mu_t,\;\; \mu_t \;=\; const \;+\; \sum_{\tau=0}^{\infty}\rho^{\tau}\tilde{\rho}y_{t-\tau} \tag{89}$$

while

$$E_t[\mu_{t+1}] \;=\; const \;+\; \rho\mu_t \tag{90}$$

$$\begin{aligned}
&E_t[F_{t+j}[y_{t+n}]] \;=\; const \;+\; (\rho+\tilde{\rho})^{n-1-j}\rho^j\mu_t \\
&E_t[y_{t+n} - F_{t+j}[y_{t+n}]] \;=\; \bar{\mu} \;-\; const \;-\; (\rho+\tilde{\rho})^{n-1-j}\rho^j\mu_t \\
&\mathrm{Cov}(E_t[y_{t+n} - F_{t+j}[y_{t+n}]], F_t[y_{t+n}] - F_{t-1}[y_{t+n}]) \\
&= \mathrm{Cov}(-(\rho+\tilde{\rho})^{n-1-j}\rho^j\mu_t, (\rho+\tilde{\rho})^{n-1}\mu_t - (\rho+\tilde{\rho})^n\mu_{t-1}) \\
&= -(\rho+\tilde{\rho})^{2n-2-j}\rho^j\mathrm{Var}[\mu] + (\rho+\tilde{\rho})^{2n-1-j}\rho^{j+1}\mathrm{Var}[\mu] \\
&\mathrm{Var}[F_t[y_{t+n}] - F_{t-1}[y_{t+n}]] \;=\; \mathrm{Var}[\mu]((\rho+\tilde{\rho})^{2(n-1)} + (\rho+\tilde{\rho})^{2n} - 2(\rho+\tilde{\rho})^{2n-1}\rho)
\end{aligned} \tag{91}$$

so that

$$B(n,j) \;=\; \frac{-(\rho+\tilde{\rho})^{2n-2-j}\rho^{j}+(\rho+\tilde{\rho})^{2n-1-j}\rho^{j+1}}{(\rho+\tilde{\rho})^{2(n-1)}+(\rho+\tilde{\rho})^{2n}-2(\rho+\tilde{\rho})^{2n-1}\rho} \;=\; \rho^{j}\frac{\rho(\rho+\tilde{\rho})-1}{(\rho+\tilde{\rho})^{j}(1+(\rho+\tilde{\rho})^{2}-2\rho(\rho+\tilde{\rho}))} \tag{92}$$

□

Proof of Corollary 5. Recall that

$$B(n,j) \;=\; \frac{A(n,j)-\rho^{2}A(n+1,j+1)}{D(n)}, \tag{93}$$

where

$$\begin{aligned} A(n,j) &= -\delta(n-j)\Big[\Gamma(j)+\delta(n)\Xi(j)\Big] \\ D(n) &= D_{MZ}(n)+\rho^{2}D_{MZ}(n+1)-2\rho^{2}(1+\delta(n+1) \\ &\quad +\delta(n)\Gamma(1)+\delta(n)\delta(n+1)\Xi(1)) \\ D_{MZ}(n) &= 1+2\delta(n)+\delta(n)^{2}\Xi(0) \end{aligned} \tag{94}$$

Thus, for $J>1$, we have

$$\begin{aligned} A(n,j) \;&=\; -\theta(1+\theta)[1-\rho^{2(J-j)}]\mathbf{1}_{j<J} \\ D_{MZ} &= 1+2\theta(1-\rho^{2J})+\theta^{2}(1-\rho^{2J}) \\ D(n) \;&=\; (1+\rho^{2})D_{MZ}-2\rho^{2}(1+\theta(1-\rho^{2J})+\theta(1+\theta)(1-\rho^{2(J-1)})) \end{aligned} \tag{95}$$

where

$$\begin{aligned} \Xi(j) \;&=\; \sum_{\tau=0}^{j}\kappa(\tau) \;+\; \sum_{\tau_1=1}^{\infty}\sum_{\tau_2=0}^{\infty}\kappa(\tau_1+j)\kappa(\tau_2)\rho^{2\max(\tau_1,\tau_2)} \\ \Gamma(j) \;&=\; \sum_{\tau=0}^{j}\kappa(\tau)+\sum_{\tau=1}^{\infty}\kappa(\tau+j)\rho^{2\tau} \end{aligned} \tag{96}$$

where $\kappa(\tau) = \delta_{0,\tau}$, so that

$$\begin{aligned} \Xi(j) &= 1 \\ \Gamma(j) &= 1\,. \end{aligned} \tag{97}$$

The claim follows. □