Forecasting with Uncertain Persistence

Joel P. Flynn* Yale

Maksim Meinert† Yale

Karthik A. Sastry‡ Princeton

March 26, 2026

Abstract

We characterize the properties of forecasts in the presence of uncertainty about persistence. Using ideas from large-deviations theory, we show that such forecasts always over-react at long horizons and the asymptotic persistence of forecasts always converges to the highest believable persistence. When prior beliefs about persistence have a downward bias, forecasts simultaneously under-react at short horizons and over-react at long horizons. These results provide a simple, unified explanation for several empirical properties of macroeconomic forecasts that have otherwise been taken as evidence of non-rational behavior. When embedded in a simple macroeconomic model, these general properties of forecasting with uncertain persistence generate boom-bust cycles in investment in the absence of any behavioral biases or market frictions.

*Yale University Department of Economics, 30 Hillhouse Avenue, New Haven, CT, 06511. Email: joel.flynn@yale.edu.

†Yale University Department of Economics, 30 Hillhouse Avenue, New Haven, CT, 06511. Email: maksim.meinert@yale.edu.

‡Princeton University Department of Economics, Julis Romo Rabinowitz Building, Princeton, NJ, 08544, Email: ksastry@princeton.edu.

We are grateful to Tim Christensen, Alexandre Kohlhas, Guido Menzio, and Jakub Steiner for helpful comments. First posted version: March 26, 2026.

1 Introduction

How persistent are important economic variables like inflation, unemployment, or output? No forecaster could credibly claim to know for certain. Indeed, a prominent academic debate in the 1980s and 1990s took place precisely because of disagreements on appropriate priors for the persistence of macroeconomic variables (see *e.g.,* Sims, 1988; Sims and Uhlig, 1991; Phillips, 1991). Particularly significant uncertainty remains about the appropriate model to forecast inflation (Stock and Watson, 2009), with some suggesting that the simple benchmark of a random walk actually outperforms other methods (Atkeson and Ohanian, 2001).

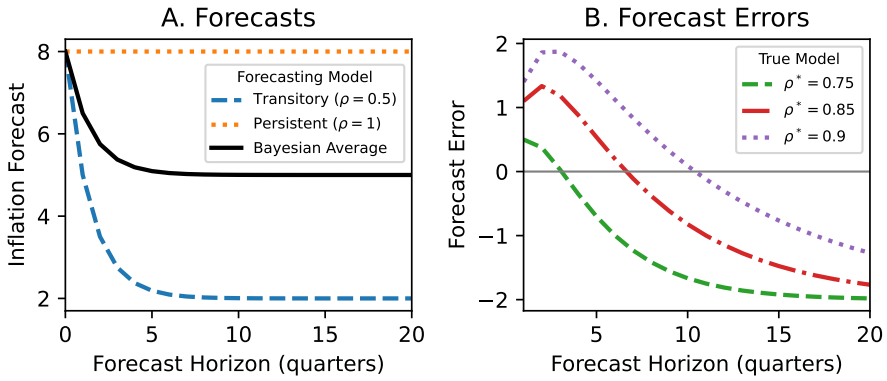
A related debate was central to the policy conversation during the global inflation surge of 2021-2023. Policymakers, academics, and market participants split their allegiances between a so-called "Team Transitory," which expected inflation to abate in a few quarters, and a so-called "Team Persistent," which thought inflation may stay elevated over the medium term (see *e.g.*, The Economist, 2024). Federal Reserve Chair Jerome Powell, in his remarks at the Jackson Hole Conference on August 27, 2021, recognized the fundamental challenge of not knowing persistence:

> Central banks have always faced the problem of distinguishing transitory inflation spikes from more troublesome developments, and it is sometimes difficult to do so with confidence in real time. (Powell, 2021)

Motivated by the centrality of model uncertainty for economic forecasting, this paper explores the implications of uncertainty about persistence for the structure of "rational" (Bayesian and correctly specified) forecasts. We find that, for general reasons that are not wedded to specific probability distributions or stochastic processes, forecasting with uncertain persistence generates a rich theory of horizon-dependent under-reaction and over-reaction. We will argue that these findings help unify many empirical findings in the literature about the properties of survey expectations regarding under- and over-reaction (see *e.g.*, Coibion and Gorodnichenko, 2015; Angeletos et al., 2021; Bordalo et al., 2020; Kohlhas and Walther, 2021; Afrouzi et al., 2023; Farmer et al., 2024) and the term structure of forecasts (Goldstein and Gorodnichenko, 2022). When applied to business cycle theory, our model of economic expectations can rationalize boom-bust cycles in forward-looking investments without any behavioral or financial frictions.

An Example. To explain our results most transparently, we begin with an example. An observer sees that inflation is currently 8% and wants to forecast it going forward. They are unsure whether to fully join "Team Transitory" or "Team Persistent," so they consider statistical models embodying each camp: an AR(1) process with a long-run mean of 2%

Figure 1: Forecasting Inflation with Unknown Persistence



Note: This figure illustrates a problem of forecasting with unknown persistence. Panel A plots forecasts under three different models: a transitory model (persistence $\rho = 0.5$), a persistent model (persistence $\rho = 1$), and a Bayesian average between the two with equal weights. Panel B shows forecast errors (realizations - forecasts) when forecasts correspond to the Bayesian average and the data-generating process is an AR(1) with a mean of 2% and the indicated persistence ρ^*.

and a quarter-to-quarter persistence of 0.5 (transitory), and a random walk (persistent). They put a 50% probability on each model being correct. As a Bayesian mean-squared-error minimizer, they forecast inflation by averaging the forecasts from these two models, weighing each by their confidence in its truth.

This average between models has a particular shape that "levels off" in the long run (Figure 1, Panel A). At short horizons, the forecast decays quickly owing to the possibility that Team Transitory is correct; but at long horizons, the forecast is persistently disanchored from 2% owing to the possibility that Team Persistent is correct. Now imagine, *ex post*, we see inflation realized with some true persistence between the two extremes. At short horizons, forecast errors may be positive, appearing as-if the forecaster "under-reacted"; whereas at long horizons, forecast errors may be negative, appearing as-if the forecaster "over-reacted" (Figure 1, Panel B). This occurs because our uncertain forecaster has, by taking an average, become "too transitory" at short horizons and "too persistent" at long horizons.

Our subsequent analysis will formalize that these two basic observations—that persistent models dominate long-run forecasts, and that this can generate simultaneous under- and over-reaction—are very general properties of forecasting under uncertainty about the stochastic process. Based on this, we will explore the broader implications of these properties for both forecasting and economic decisions.

The AR(1) Theory. We first illustrate the general principles of our analysis in the setting of forecasting an AR(1) process. We study a forecaster who observes the current value of

a stochastic process x_t, understands this process follows an AR(1) with some persistence ρ, and holds some belief about this persistence summarized by a probability distribution F.

We first describe the *term structure of forecasts*. To do this, we define an object called the *forecast persistence function* which, at each horizon $h \geq 1$, is the ratio of the h and $h-1$ horizon forecast: $P(h) = \mathbb{E}[x_{t+h}]/\mathbb{E}[x_{t+h-1}]$. We show that this function always has three key properties. First, initial persistence is governed by the prior mean: $P(1) = \mathbb{E}[\rho]$. Second, forecast persistence increases in horizon. Third, the limiting forecast persistence as the horizon increases is the *maximum believable persistence*, or the upper support point of the distribution F. All three properties follow from applying elementary concepts from large deviations theory (Dembo and Zeitouni, 2009).

Going beyond our motivating example, this result significantly generalizes the logic that forecasts based on model averages appear "more persistent" at higher horizons. This is *always* the case provided that the forecaster is averaging across different models. Long-horizon forecasts are as persistent as the most persistent model considered, be that a random walk (as in Figure 1) or even a model that implies long-run growth, a possibility considered in standard time-series econometrics (*e.g.*, Sims and Uhlig, 1991; Phillips, 1991).

We next study the implications for over- and under-reaction relative to the ground truth. We focus on the case where the forecaster is *correctly specified* in the sense that they put positive probability on the true model.[1] We show that, conditional on *any* true persistence within the support of F, forecasts at a sufficiently long horizon must be more sensitive to variation in x_t than true future realizations. That is, forecasters always asymptotically *over-react*, precisely because the maximum believable persistence determines the behavior of long-term forecasts. This coexists with *under*-reaction at short horizons if and only if the expected persistence is less than the true persistence.

We finally observe that, in a setting where multiple forecasters hold different beliefs over persistence but nonetheless agree about the *support* (*i.e.*, what degrees of persistence are even possible), then all forecasters asymptotically agree about persistence at long horizons. Crucially, this is despite the fact that they may have very different *forecasts*. In the earlier example (Figure 1), we could have a set of forecasters with different beliefs whose long-run expectations are all disanchored from 2%, but at different *levels* that depend on their differential confidence in Team Persistent.

The General Theory. We next extend our analysis to consider more general forecasting models. To do so, we study the case where a vector of economic variables X_t follows a

[1] In our simplistic example, forecasters were not correctly specified just for ease of exposition—but the predictions would be the same in a variant where the forecaster also put weight on the possibility that persistence was 0.75, 0.85, or 0.9 (Figure 1).

first-order vector autoregression (VAR) process with coefficient matrix A, and the forecaster is uncertain about this coefficient matrix. We impose some technical regularity conditions on the forecaster's prior over this matrix. We describe these later in formal detail, but we argue that the restrictions they impose are generic and thus without economic substance.

We show that our analysis of this formulation can encompass all vector regression and moving average (VARMA) models up to finite order, given appropriate definition of the state. In this way, our analysis applies to three classes of models that are commonly applied for macroeconomic forecasting: (i) univariate ARMA models beyond the AR(1) case, (ii) multivariate VARs (Sims, 1980; Del Negro and Schorfheide, 2004), (iii) and linear rational-expectations models (Blanchard and Kahn, 1980) including standard dynamic stochastic general equilibrium (DSGE) models (Smets and Wouters, 2007). Apart from being common in practice, these models are in fact *necessary* to study phenomena such as "hump-shaped" dynamics, periodic cycles, forward-looking behavior, and multi-variate relationships like Phillips curves or monetary policy rules—all of which might come into play in a practical context like our motivating example of forecasting inflation after COVID.

We find that the primary insights do indeed extend to forecasting with these models and, moreover, even richer phenomena are possible. First, if the largest eigenvalue of any possible coefficient matrix under the prior is real-valued, then the forecast persistence function for any variable in the VAR converges to this eigenvalue. In this case, this eigenvalue takes on the role of the maximum believable persistence and the persistence of forecasts always converges to this value. Forecasts continue to always asymptotically over-react in this case.

The new possibility in the more general analysis is that a complex pair of eigenvalues has the largest modulus. In this case, the forecast persistence function's magnitude is governed by the modulus of this eigenvalue pair but the function cycles with a period that is determined by the argument of the eigenvalue pair. Thus, forecasts become as asymptotically persistent as possible, but this is in terms of the modulus of a stable cycle. In this case, forecasts oscillate between over-reacting and under-reacting periodically. While this may seem pathological, we emphasize that many non-seasonally detrended variables do exhibit clear periodicity (see *e.g.,* Barsky and Miron, 1989). For such data, this cyclical asymptotic regime may be descriptive.

One interesting twist for multivariate models is that forecasts for each variable now inherit a common asymptotic persistence and—in the complex case—a common asymptotic periodicity. In this case, it is as if forecasts become maximally persistent and a dominant linear combination of the variables in the VAR system determines the structure of the forecast.

Rationalizing Macroeconomic Forecasting Puzzles. We next use our model to reconcile seemingly contradictory findings regarding macroeconomic forecasts. We do this by

primarily studying inflation (CPI) expectations from the Survey of Professional Forecasters. Our findings are summarized in five key facts that are all jointly consistent with our model.

First, at the aggregate level, long-horizon forecasts co-move considerably with short-horizon forecasts. For example, if consensus expectations for one-year-ahead inflation increase by 1%, expectations for average inflation over the next ten years increase by 0.8%. This change is more than three times what would be predicted if forecasters were using an AR(1) model calibrated to historical data. Second, as shown by Halperin and Mazlish (2025), forecasts *under-react* to information contained in forecasts and revisions at short horizons, while they *over-react* to the same information at long horizons. Halperin and Mazlish (2025) establish this by systematically studying forecasts across several variables and 89 countries, and Afrouzi et al. (2023) have a similar key finding in an experimental study. We recreate a version of this result for US inflation expectations. This finding of over-reaction at sufficiently long horizons is predicted under very general conditions in our model.

Next, we study the properties of individual forecasts. Our third fact, recreating a finding from Goldstein and Gorodnichenko (2022), is that the term structure of inflation expectations *flattens*: the relationship between forecasts at horizons h and $h-1$ gets stronger for larger h. We show that this regression model exactly estimates the forecast persistence function that we characterize theoretically and, therefore, empirically validates that the persistence function increases in horizon. Digging further into the forecaster-level data, we show as a fourth fact that this is driven by an increasing propensity to "stabilize" forecasts, or report the same value for adjacent horizons, as the horizon increases. Quite strikingly, in the *majority* of cases, stabilized forecasts four quarters out *do not* coincide with reported expectations for the very long run (*i.e.*, averages for the next ten years). This implies that forecasts are commonly "stabilized but disanchored," like the forecast in our stylized example (Figure 1). This is consistent with the model's predictions of expectations persistence and stability at long horizons that can nevertheless coexist with significant disagreements.

Our fifth fact concerns the *multivariate* structure of forecasts. Forecasts of inflation, unemployment, and interest rates are closely linked across forecasters. Moreover, the persistence of forecasters' implied *multivariate* models also increases over horizons. This is consistent with our prediction from the VAR analysis that, over longer horizons, multivariate forecasts are dominated by one persistent factor.

Investment and Boom-Bust Cycles. We conclude by describing how forecasting with unknown persistence affects economic choices in a canonical setting, capital investment. We consider a simple, stripped-down model in which a representative investor chooses, at $t = 0$, how much to invest in projects of different maturities. The realized productivity of these investments is currently positive and subsequently follows a stochastic process that is an

AR(1) with persistence unknown to the investor. Importantly, the investor can even be uncertain whether the productivity will decay to zero or grow on average.

The investor's demand for projects at different horizons is governed by the forecast persistence function characterized in our earlier analysis. As such, at longer horizons, relative demands are determined by the maximum believable persistence. This can generate a striking pattern where the investor prefers very short run projects and very long run projects, but neglects medium-run projects. Intuitively, the former have a known positive return and the latter could be extremely high under the *possibility* (even if small) of long-run growth, but the medium-horizon investments have neither desirable property.

This behavior can generate cycles of capital investment. When productivity booms, investors produce in the short term but also sow the seeds for long-horizon projects. These may come to fruition and induce an output boom if productivity remains high, or they may be unproductively wasted if productivity later goes down. These dynamics are reminiscent of boom-bust cycles and, more specifically, of the classical "Austrian" theory of the business cycle that emphasizes speculative (over-) investment (Hayek, 1931). In our setting, these properties arise without any behavioral biases or financial frictions. They arise instead from purely *rational* forecasting under model uncertainty, in a frictionless environment.

Related Literature. We start from a premise—uncertainty about the stochastic process for economic variables—that is familiar to policymakers, forecasters, and macroeconomists. In this way, our contribution is orthogonal to, but completely complementary with, studies focusing on the process of *learning* the stochastic process of macroeconomic variables (see *e.g.,* Marcet and Sargent, 1989; Eusepi and Preston, 2011; Molavi, 2022; Farmer et al., 2024; Kohlhas and Robertson, 2025). Our analysis characterizes the properties of forecasts and economic decisions *conditional* on model uncertainty. These properties are robust over a large class of learning rules and stochastic processes (true and considered) for economic variables, provided that forecasters entertain the possibility that many models are true.

Our research also relates to studies that postulate that economic agents use mis-specified dynamic models. For instance, Gabaix (2020) considers a model of *cognitive discounting* in which economic agents under-estimate persistence. García-Schmidt and Woodford (2019) and Farhi and Werning (2019) study models of level-k thinking that, in equilibrium, also imply that agents under-estimate the persistence of the economy's response to shocks. On the other hand, Angeletos et al. (2021) consider a model in which economic agents can over- or under-estimate the persistence of economic variables, and argue that the former is necessary to match certain facts in forecasting data. Molavi (2022) argues that using "over-persistent" models can arise from long-run, mis-specified learning.[2] Bordalo et al.

[2]This bears resemblance to some of our findings, but arises for different reasons. In Molavi's (2022)

(2020) and Afrouzi et al. (2023) study models of distorted information processing in which agents over-react to recent information, a phenomenon that pushes in the same direction as perceived over-persistence. Our analysis differs in two primary ways. Methodologically, we study forecasters that are Bayesian and correctly specified ("rational"), yet nonetheless uncertain about the correct model. In terms of results, our analysis implies that under- and over-reaction can *co-exist* in forecasts depending on the time horizon and structure of uncertainty, potentially helping to reconcile the contrasting perspectives of the literature.

Methodologically, our results follow from ideas in large deviations theory, in particular the Laplace principle (see *e.g.*, Dembo and Zeitouni, 2009). These ideas have also fruitfully been applied in economics by Weitzman (1998, 2001) in his analysis of the asymptotically appropriate discount rate, by Martin (2012) in his analysis of asset pricing, by Samuelson and Steiner (2025) in their analysis of investment and hedging, and by Eden et al. (2026) in their analysis of the effects of unlikely events. While our results do not overlap with those of these papers, the core mathematical idea underlying our analysis is shared.

Outline. The rest of the paper proceeds as follows. Section 2 describes the main ideas of our analysis for AR(1) models. Section 3 provides the general results. Section 4 presents five facts about professional forecasts that our framework can rationalize. Section 5 shows how uncertain persistence can rationalize boom-bust investment cycles. Section 6 concludes.

2 Forecasting an AR(1) Process with Uncertainty

We first illustrate the general principles of our analysis by considering the forecasting problem of predicting the value of an AR(1) process with uncertain persistence.

2.1 The Environment

Time is discrete $t \in \mathbb{N}$ and a forecaster believes that an economic variable $x_t \in \mathbb{R}$ (*e.g.,* the price level, the unemployment rate, the price of a stock) follows an AR(1) process:

$$x_t = \rho x_{t-1} + \varepsilon_t \tag{1}$$

where ε_t is independently and identically distributed according to some distribution G with zero mean. The forecaster is uncertain about ρ and believes that it has distribution F with

analysis, forecasters estimate a best-fitting one-factor model under model misspecification. Asymptotically, they recover the most persistent component of the true data-generating process. In our setting, forecasters are uncertain about the true model and consider a probability distribution over models. Asymptotically as the *horizon* of their forecast increases, forecasts are dominated by the most persistent model.

support contained in $[0, r]$ and has essential supremum $r > 0$. That is, they know that x_t is positively autocorrelated but do not know to what extent. They observe the realization x_t in each period t and form expectations for future values of the variable using Bayes rule, given uncertainty about the data-generating process. We denote the forecast of x_{t+h} given observation of x_t as:

$$x_{t+h|t} = \mathbb{E}_t[x_{t+h}] \tag{2}$$

We highlight that our framework can fully accommodate that the agent learns about persistence over time. This would affect how the term structure of a forecaster's expectations changes over time, but has no bearing on the term structure at any given time.

The true data-generating process of x_t is of no consequence to study only the properties of forecasts (Theorem 1), but it will be relevant for studying the properties of *ex post* forecast errors. We assume that x_t truly follows the process in Equation 1 with persistence $\rho = \rho^* \in (0, r)$. When ρ^* is in the support of F (*i.e.*, the forecaster considers the correct model), we say the forecaster is correctly specified. A correctly specified forecaster can be considered "rational" in the following sense: they do not dogmatically exclude the correct model of the world, and they incorporate new information using correct conditional probabilities.

2.2 The Persistence of Forecasts Across Horizons

We first study the *term structure* of forecasts at different horizons h, made at the same point in time t. To summarize this term structure, we define the *forecast persistence* function $P : \mathbb{N} \to \mathbb{R}$ as:

$$P(h) \equiv \frac{x_{t+h|t}}{x_{t+h-1|t}} \tag{3}$$

This function describes how persistent forecasts are between any two adjacent forecasting horizons and, combined with the base forecast $x_{t|t} = x_t$, describes the entire term structure of forecasts. Given the AR(1) structure, the forecast persistence function is:

$$P(h) = \frac{\mathbb{E}[\rho^h]}{\mathbb{E}[\rho^{h-1}]} \tag{4}$$

When F is degenerate, *i.e.,* ρ is known, one obtains that $P(h) = \rho$. That is, the persistence of forecasts across horizons is constant and equal to the persistence of the state across time. Intuitively, forecasts simply decay at a rate corresponding to the known persistence.

How does uncertainty about ρ affect the persistence of forecasts across horizons? The answer to this question is given by the following simple theorem, which leverages the Laplace principle from large-deviations theory (see *e.g.,* Dembo and Zeitouni, 2009).

Theorem 1 (Properties of Forecast Persistence)**.** *The following statements are true:*

1. *Initial forecast persistence is the mean of the persistence:*

$$P(1) = \mathbb{E}[\rho] \tag{5}$$

2. *Forecast persistence is an increasing function of the horizon, i.e., P is increasing*
3. *The asymptotic forecast persistence is the maximum believable persistence:*

$$\lim_{h\to\infty} P(h) = r \tag{6}$$

Proof. See Appendix A.1. □

To exemplify this result, consider the special case in which $\rho = rU$, where U is distributed according to a beta distribution with parameters $(\alpha, \beta) > 0$. In this case, we obtain the following analytical expression for the forecast persistence:

$$P(h) = r\frac{h+\alpha-1}{h+\alpha+\beta-1} \tag{7}$$

As claimed by the result, we can see that: (i) $P(1) = r\frac{\alpha}{\alpha+\beta}$, which is the mean persistence, (ii) $P(h)$ is an increasing function of h, and (iii) $P(h)$ limits to r. To give a visual illustration of how drastically this can differ from the known persistence case, Figure 2 plots the difference between the term structure of forecasts when ρ is known versus when it is not. We observe that unknown persistence gives rise to forecasts that are substantially more persistent than the benchmark of known persistence. We will often return to this beta-distributed example throughout the paper, as it allows for particularly simple closed-form calculations.

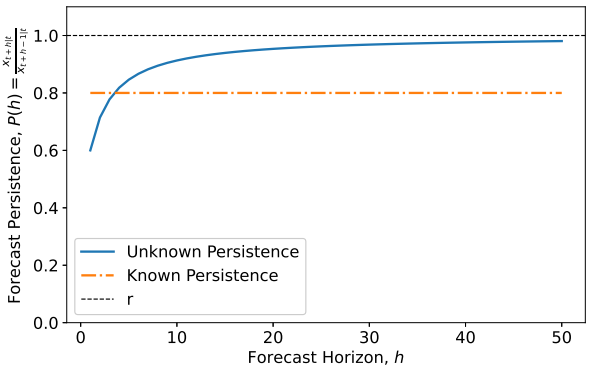
To obtain intuition for why Theorem 1 holds, it is instructive to consider an even simpler example in which the forecaster considers only two models of the world, as we described verbally in the introduction. That is, the forecaster believes that x_t is "more persistent" ($\rho = \bar{\rho}$) with probability p and "more transitory") ($\rho = \underline{\rho} < \bar{\rho}$) with probability $1-p$. By the law of iterated expectations, the horizon h forecast is given by:

$$x_{t+h|t} = p \cdot \underbrace{\mathbb{E}[x_{t+h}|x_t, \rho = \bar{\rho}]}_{\text{"Persistent" model}} + (1-p) \cdot \underbrace{\mathbb{E}[x_{t+h}|x_t, \rho = \underline{\rho}]}_{\text{"Transitory" model}} \tag{8}$$

This is an average of the forecast under each model weighted by their prior probabilities. Moreover, given the AR(1) structure, these forecasts can be written as

$$x_{t+h|t} = p \cdot \bar{\rho}^h x_t + (1-p) \cdot \underline{\rho}^h x_t \tag{9}$$

Figure 2: Illustration of the Impact of Uncertainty on Forecast Persistence



Note: This figure plots a situation in which $r = 1$ (the forecaster believes the process could be as persistent as a random walk). The blue line plots the forecast persistence function when the forecaster believes that ρ is beta-distributed between zero and one ($\alpha = 1.5, \beta = 1$), which has mean 0.6. The orange line plots the forecast persistence function when the forecaster believes that persistence is certainly 0.8.

Our first observation is that, as the horizon h increases, the forecast becomes dominated by the persistent model. This is most obvious if, for example, the persistent model is a random walk: $\bar{\rho} = 1$ and $\underline{\rho} < 1$. In this special case, as the forecasting horizon increases, the "persistent" forecast remains at x_t while the "transitory" forecast converges to 0. Therefore $x_{t+h|t} = p \cdot x_t + (1-p) \cdot 0 = p \cdot x_t$: that is, the forecast "levels off" at $p \cdot x_t$ to represent the probability, even if very small, that x_t is a random walk.

We next observe that this logic that the more persistent model dominates applies more generally, regardless of the values of $\underline{\rho} < \bar{\rho}$. To do this, we calculate the forecast persistence function as

$$P(h) = \frac{p\bar{\rho}^h + (1-p)\underline{\rho}^h}{pr^{h-1} + (1-p)\underline{\rho}^{h-1}} = \frac{p\bar{\rho} + (1-p)\underline{\rho}\left(\underline{\rho}/\bar{\rho}\right)^{h-1}}{p + (1-p)\left(\underline{\rho}/\bar{\rho}\right)^{h-1}} \tag{10}$$

where in the second equality, we simply divided both the numerator and the denominator by $\underline{\rho}^{h-1}$. From here, we can readily observe that $P(h) \in (\underline{\rho}, \bar{\rho})$ and that the function is increasing. Moreover, as h becomes large, we have that $\left(\underline{\rho}/\bar{\rho}\right)^{h-1}$ becomes small *exponentially* quickly. Thus, taking the limit of the equation above, we obtain that:

$$\lim_{h\to\infty} P(h) = \frac{p\bar{\rho} + (1-p)\underline{\rho}\lim_{h\to\infty}\left(\underline{\rho}/\bar{\rho}\right)^{h-1}}{p + (1-p)\lim_{h\to\infty}\left(\underline{\rho}/\bar{\rho}\right)^{h-1}} = \frac{p\bar{\rho} + 0}{p + 0} = \bar{\rho} \tag{11}$$

A variant of the direct calculation above would apply to any case in which the prior F has finite support. To cover the case of all distributions, the proof of Theorem 1 makes

use of a more general mathematical argument. Nonetheless, the core intuition remains that more persistent models get more weight as the forecast horizon increases. This basic idea underlies all of the results of this paper.[3]

2.3 Uncertainty and Bias: Over-Reaction and Under-Reaction

We next study how forecasts over-react and under-react relative to the ground truth. To study this, recall that the true persistence is given by $\rho^* \in (0, r)$. Denote the forecast under the true persistence by $x^*_{t+h|t} = {\rho^*}^h x_t$. To quantify the extent of over-reaction and under-reaction, we can define the *relative forecast reaction* function $R : \mathbb{N} \times (0, r) \to \mathbb{R}$:

$$R(h; \rho^*) \equiv \frac{\frac{\partial x_{t+h|t}}{\partial x_t}}{\frac{\partial x^*_{t+h|t}}{\partial x_t}} = \frac{\mathbb{E}[\rho^h]}{(\rho^*)^h} \tag{12}$$

This function describes the gap between the agent's forecast and the forecast given knowledge of the true model ρ^*. When $R(h; \rho^*) \geq 1$, forecasts over-react relative to the true model. When $R(h; \rho^*) < 1$, forecasts under-react relative to the true model. The following corollary of Theorem 1 characterizes the conditions under which under-reaction and over-reaction occur at various time horizons:

Proposition 1 (Over-Reaction and Under-Reaction)**.** *The following statements are true:*

1. *If $\mathbb{E}[\rho] \geq \rho^*$, then the forecast over-reacts at all horizons.*
2. *If $\mathbb{E}[\rho] < \rho^*$, then the forecaster initially under-reacts and then over-reacts, i.e., there exists a unique $h^* \in \mathbb{N}$ such that they under-react for $h \leq h^*$ and over-react for $h > h^*$.*

Moreover, in both cases, the asymptotic rate of over-reaction is given by:

$$\lim_{h \to \infty} \frac{1}{h} \ln R(h; \rho^*) = \ln \frac{r}{\rho^*} > 0 \tag{13}$$

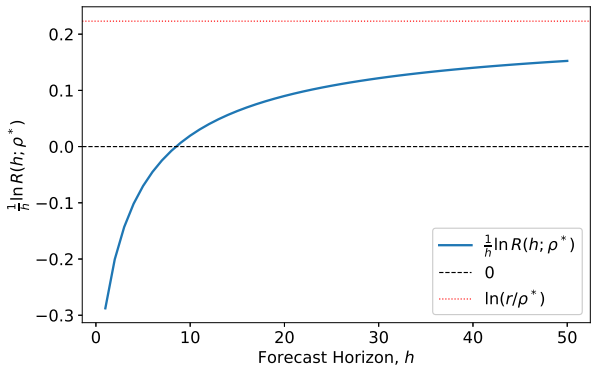
Proof. See Appendix A.2 □

This result shows that perfectly Bayesian forecasters who possess complete knowledge of the state of the process can nevertheless under-react and over-react to knowledge of that state. Moreover, their under-reaction and over-reaction can be horizon-dependent.

In fact, in the world of AR(1) processes, this result provides a tight characterization of over-reaction in terms of the *bias* in the forecaster's beliefs about persistence. Concretely, if

[3]This result can also be connected to the observation of Weitzman (2001) that, when taking present discounted values by averaging over unknown discount rates, marginal discounting increases progressively in the horizon considered.

Figure 3: Uncertainty Can Lead to Under-Reaction and then Over-Reaction of Forecasts



Note: This figure plots a situation in which $r = 1$ (the forecaster believes the process could be as persistent as a random walk). The blue line plots the normalized forecast reaction function $\frac{1}{h}\ln R(h;\rho^*)$ when the forecaster believes that ρ is beta-distributed between zero and one ($\alpha = 1.5, \beta = 1$), which has mean 0.6. We take the true value $\rho^* = 0.8$, which is halfway between the mean of the forecaster's belief and the most persistent they believe the process could be.

the forecaster has upward-biased beliefs that persistence is greater than it truly is, $\mathbb{E}[\rho] \geq \rho^*$, then the forecaster always over-reacts. This is not very surprising. What is more surprising is that if a forecaster has downward-biased beliefs that persistence is lesser than it truly is, $\mathbb{E}[\rho] < \rho^*$, then the forecaster will initially under-react but eventually *always* over-react. Moreover, this result establishes that these are the only two possible patterns.

The intuition for this result mirrors that which we gave for forecast persistence. As the horizon increases, the mere possibility—no matter how unlikely—that persistence is greater than the true persistence dominates the behavior of forecasts. This leads forecasts to asymptotically react as if the most persistent possible state is the only state that the forecaster entertains.

To give a concrete numerical example of this phenomenon, we return to the example in which $\rho = rU$ and $U \sim \text{Beta}(\alpha, \beta)$. In this case, we have that:

$$R(h;\rho^*) = \left(\frac{r}{\rho^*}\right)^h \prod_{j=0}^{h-1} \frac{\alpha + j}{\alpha + \beta + j}. \tag{14}$$

We plot the corresponding forecast reaction function in Figure 3, where we rescale by h and take the logarithm for ease of visualization. In these units, if the normalized reaction function is below zero, then the forecaster under-reacts and if the function is above zero, then the forecaster over-reacts. In this example, we have that $\mathbb{E}[\rho] < \rho^*$ and so, as guaranteed by Proposition 1, the forecaster initially under-reacts and then over-reacts.

It is further illustrative to return to the two-point prior example ($\rho = r$ with probability p and $\rho = \delta < r$ with probability $1 - p$) to provide intuition for this result. In this case:

$$R(h; \rho^*) = p \left(\frac{r}{\rho^*} \right)^h + (1 - p) \left(\frac{\delta}{\rho^*} \right)^h \tag{15}$$

At short-horizons, we therefore have under-reaction when $R(1, \rho^*) < 1$, or:

$$\mathbb{E}[\rho] = pr + (1 - p)\delta < \rho^* \tag{16}$$

That is, under-reaction at short horizons occurs if and only if expected persistence is biased below true persistence. To see why asymptotic over-reaction always happens, we can write:

$$R(h; \rho^*) = \left(\frac{r}{\rho^*} \right)^h \left[p + (1 - p) \left(\frac{\delta}{r} \right)^h \right] \implies \ln R(h; \rho^*) \sim h \ln \left(\frac{r}{\rho^*} \right) \tag{17}$$

This also allows us to see why the asymptotic rate of over-reaction is given by $\ln(r/\rho^*)$. Intuitively, no matter how unlikely the high persistence state is, forecasts are asymptotically infinitely more persistent under the high persistence state than the true state, and so the forecaster rationally over-reacts at long horizons.

2.4 Disagreement About Persistence and Reaction

Finally, we compare the persistence and over-reaction of forecasts across forecasters. To do this, suppose that there are many agents $i \in \mathcal{N} = \{0, \ldots, N\}$ with priors F_i with common support. Each agent has a persistence function P_i and a relative forecast reaction function R_i. Define the pairwise disagreement about persistence between any pair of agents i and j as $\Delta P_{i,j} = \frac{P_i}{P_j}$ and the pairwise differences in reaction as $\Delta R_{i,j} = \frac{R_i}{R_j}$.

Corollary 1 (Asymptotic Agreement About Persistence)**.** *The following statements are true:*

1. *All agents asymptotically agree about persistence, i.e., for all $i, j \in \mathcal{N}$:*

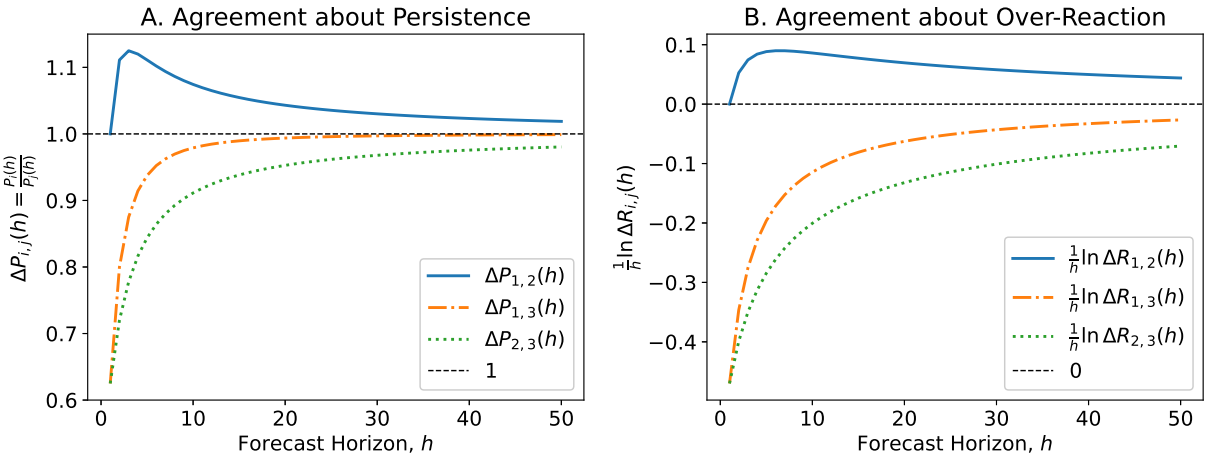
$$\lim_{h \to \infty} \Delta P_{i,j}(h) = 1 \tag{18}$$

2. *All agents asymptotically over-react to the same degree, i.e., for all $i, j \in \mathcal{N}$:*

$$\lim_{h \to \infty} \frac{1}{h} \ln \Delta R_{i,j}(h) = 0 \tag{19}$$

Proof. Immediate from Theorem 1 and Proposition 1. □

Figure 4: Asymptotic Agreement About Persistence in the Face of Uncertainty



Note: This figure plots a situation in which $r = 1$ (all forecasters believe the process could be as persistent as a random walk). There are three forecasters with priors that are all beta-distributed with parameters $(\alpha_1, \beta_1) = (1, 1)$, $(\alpha_2, \beta_2) = (2, 2)$, and $(\alpha_3, \beta_3) = (4, 1)$. In each panel, the colored lines plot the corresponding differences in the forecast persistence function (Panel A) and the normalized forecast reaction function (Panel B) across the three pairs of forecasters.

This result follows easily from the previous ones: as all that matters is the maximum possible level of persistence, exactly how likely different people perceive different levels of persistence to be is irrelevant for long-horizons. To illustrate this result, we return again to our running example in which three different forecasters have heterogeneous beta-distributed priors, with $(\alpha_1, \beta_1) = (1, 1)$, $(\alpha_2, \beta_2) = (2, 2)$, and $(\alpha_3, \beta_3) = (4, 1)$ (Figure 4). Forecasters one and two have the same expectation about the persistence of the process and thus agree about the short horizon. However, forecaster one has a more dispersed prior, and so they over-react by more at intermediate horizons. This notwithstanding, as guaranteed by Corollary 1, they eventually agree once again at long horizons and converge to believing in the same persistence and over-react to the same degree. Forecaster three has a lower mean belief, and so they under-react by more at short horizons. However, they nonetheless over-react to the same degree as forecasters one and two at long horizons.

Of course, asymptotic agreement about persistence may fail to occur if (and only if) agents disagree on the maximal possible persistence. Thus, this result isolates a one-dimensional condition that governs the presence of asymptotic agreement. This reduces the complexity of determining disagreement in general, which typically depends on infinite-dimensional heterogeneity in priors.

3 The General Structure of Forecasts with Uncertainty

In this section, we provide the main theoretical result of the paper, which establishes that—under weak regularity conditions on the forecaster's prior—the properties of forecasts that we uncovered for AR(1) processes extend to general VAR processes. In these cases, when dominant eigenvalues are real, the persistence of forecasts converges to the maximum believable persistence and forecasts *always* asymptotically over-react, as in the basic AR(1) case. When dominant eigenvalues are complex, forecasts asymptotically cycle with an amplitude governed by the maximum believable persistence. Our results also extend to VARMA processes, which encompass ARMA models for a single variable and the solutions to standard linear rational expectations models (Blanchard and Kahn, 1980) and dynamic stochastic general equilibrium models (Smets and Wouters, 2007).

3.1 The Environment

A stochastic process $X_t \in \mathbb{R}^p$ follows a VAR(1) model, if it can be represented as

$$X_t = AX_{t-1} + \Sigma\varepsilon_t \tag{20}$$

where A is the $p \times p$ coefficient matrix of the VAR process, Σ is a matrix that rotates the shocks, and the shocks $\varepsilon_t \in \mathbb{R}^p$ are drawn IID from a distribution G with mean zero.

Once again, we consider a scenario in which a forecaster is uncertain about the true process (here, within the class of VAR models) and has prior F over the coefficient matrix A. To derive our results, we make some technical assumptions on the nature of this prior. To do this, we first describe some definitions. If A is (complex) diagonalizable, then we write $A = \sum_{i=1}^{p} \lambda_i v_i w_i$, where the eigenvalues $\lambda_i \in \mathbb{C}$ are ordered by magnitude $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_p|$ and the $v_i \in \mathbb{C}^p$ and $w_i \in \mathbb{C}^p$ are the corresponding (normalized) right (column) and left (row) eigenvectors.

Definition 1 (Regular Prior)**.** *A prior F is regular if it places probabilities $\{p_i\}_{i=1}^n$ on coefficient matrices $\{A_i\}_{i=1}^n$ such that:*

1. *The coefficient matrix A_i is diagonalizable with distinct eigenvalues for all $i \in \{1, \ldots, n\}$.*
2. *Defining the spectral radius of coefficient matrix A_i as $L_i = \max_{j \in \{1,\ldots,p\}} |\lambda_{ij}|$, we have that $0 < r \equiv \max_{i \in \{1,\ldots,n\}} L_i > L_j$ for all but one $j \in \{1, \ldots, n\}$. Moreover, either a single real eigenvalue has modulus r or a single complex conjugate pair of eigenvalues has modulus r.*

Regularity encodes three properties, each of which we regard as technical rather than economically restrictive. First, it imposes that the coefficient matrix is diagonalizable with distinct eigenvalues with probability one. As diagonalizability is a generic property, we regard this assumption as entirely technical. Second, it imposes that the prior is finitely supported. As finitely supported priors can arbitrarily approximate any countably or uncountably supported prior, we also regard this property as technical in nature. Third, it imposes that the modulus of the largest eigenvalue is attained by a complex conjugate pair for a single coefficient matrix. Once again, this is a generic property.

When the prior is regular, we take A_1 as the coefficient matrix that has the largest eigenvalue (pair) and that has probability p_1. We write the corresponding largest eigenvalues in polar coordinates as $re^{\pm i\theta}$. We write the corresponding eigenvectors as (v_1, w_1) and $(\bar{v}_1, \bar{w}_1)$, where the bar denotes the complex conjugate. When the largest eigenvalue is real, we have $\theta = 0$ with unique and real eigenvectors (v_1, w_1). To compare the true model, we assume that A^* is diagonalizable with a unique largest eigenvalue pair $r^*e^{\pm i\theta^*}$ with $r > r^* > 0$ and corresponding eigenvectors (v_1^*, w_1^*) and $(\bar{v^*}_1, \bar{w^*}_1)$.

Before proceeding, we remark on how the VAR(1) analysis directly extends to VAR(q) and VARMA(q,m) models. Doing so allows us to jointly address three important classes of economic models in the analysis of this section. The first are univariate ARMA models beyond the simple AR(1) case. The second are (reduced-form or structural) vector auto-regression models, which are a standard benchmark for policy analysis and forecasting (Sims, 1980; Del Negro and Schorfheide, 2004). The third is a large class of linear rational-expectations models with "stable" solutions, for instance in the sense of Blanchard and Kahn (1980). This encompasses many benchmark dynamic stochastic general equilibrium (DSGE) models (Smets and Wouters, 2007).

Remark 1 (Representing VAR(q) Models as VAR(1) Models)**.** Our analysis below will apply to any VAR(q) model, leveraging the well-known fact that we can re-write these models as VAR(1) processes. Let $y_t \in \mathbb{R}^n$ and $\nu_t \in \mathbb{R}^n$ respectively denote the vector of variables and innovations. The variable y_t follows the process

$$y_t = \sum_{j=1}^{q} \Phi_j y_{t-j} + \tilde{\Sigma}\nu_t \tag{21}$$

where the Φ_j are $n \times n$ matrices of coefficients and $\tilde{\Sigma}$ is an $n \times n$ matrix of shock loadings. We can express these models in the form of Equation 20 using the following transformation

that "stacks" the finite dimensions of y_t (and lags) and ν_t (and lags).

$$X_t = \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-q+1} \end{pmatrix}, \qquad A = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_q \\ I_n & 0 & \cdots & 0 \\ 0 & I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_n \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \tilde{\Sigma} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{22}$$

where this matrix is commonly called the companion matrix. Importantly, for what follows, such companion matrices are generically diagonalizable despite their restricted structure. To recover the forecasts for the variables of interest, one then takes the first n elements of X_t. For univariate AR(p) models, the same reduction applies with $n = 1$. △

Remark 2 (Handling Moving Average Components)**.** As we are characterizing the asymptotic properties of forecasts, moving average (MA) terms are not relevant for our main arguments. In particular, consider a class of VARMA(q,m) stochastic processes of the form

$$y_t = \sum_{j=1}^{q} \Phi_j y_{t-j} + \sum_{j=0}^{m} \tilde{\Sigma}_j \nu_{t-j} \tag{23}$$

Continue to let X_t denote the stacked vector of y_t and lags and to construct A as in Equation 22. Then, for any $h > q$, one can show that

$$\mathbb{E}_t[X_{t+h}] = \mathbb{E}_t \left[A^h \left(C_0 X_t + \sum_{j=0}^{m-1} B_j \nu_{t-j} \right) \right] \tag{24}$$

for some matrices C_0 and $(B_j)_{j=0}^{m-1}$ that depend on the primitive VARMA coefficients. Intuitively, after horizon q, the forecaster cannot forecast any of the innovations in Equation 23. Therefore, moving average dynamics matter in the same way (summarized in the matrices B_j) for all longer-horizon forecasts, and the relevant properties of the term structure of forecasts are then determined by the properties of A. Our main result below applies almost as written for such VARMA processes (see Appendix A.4). △

3.2 General Properties of the Term Structure of Forecasts

We now generalize our concepts of the forecast persistence function and relative forecast reaction function. Let $x_t^i = (X_t)_i$ denote a specific variable among those in the VAR (*e.g.*,

inflation) We define the forecast persistence function for variable i, $P_i(h)$ as:

$$P_i(h) \equiv \frac{x^i_{t+h|t}}{x^i_{t+h-1|t}} = \frac{e_i' X_{t+h|t}}{e_i' X_{t+h-1|t}} \tag{25}$$

where e_i represents the i-th vector of the standard basis and $X_{t+h|t}$ the multivariate forecast of X_{t+h} at time t. Similarly, we can define the forecast reaction function for the ith component of the VAR(1) process as

$$R_i(h; A^*) \equiv \frac{\frac{\partial x^i_{t+h|t}}{\partial x^i_t}}{\frac{\partial x^{*i}_{t+h|t}}{\partial x^i_t}} = \frac{\mathbb{E}[A^h]_{ii}}{(A^{*h})_{ii}} \tag{26}$$

With these notions in hand, we are now equipped to generalize our results to VAR processes. This is achieved by the following result.

Theorem 2. *If the prior F is regular, then the following statements are true:*

1. *If $e_i v_1 w_1 X_t \neq 0$, then the forecast persistence function for variable i satisfies:*

$$P_i(h) = r \frac{\cos(\theta h + \vartheta_i) + o(1)}{\cos(\theta(h-1) + \vartheta_i) + o(1)} \tag{27}$$

where $\vartheta_i = \arg(e_i v_1 w_1 X_t)$. Further, when the largest eigenvalue is real, we have that:

$$\lim_{h\to\infty} P_i(h) = r \tag{28}$$

2. *Moreover, the forecast reaction function for variable i satisfies:*

$$R_i(h; A^*) = \left(\frac{r}{r^*}\right)^h p_1 \frac{|v_{1i} w_{1i}|}{|v^*_{1i} w^*_{1i}|} \frac{\cos(h\theta + \phi_i) + o(1)}{\cos(h\theta^* + \phi^*_i) + o(1)} \tag{29}$$

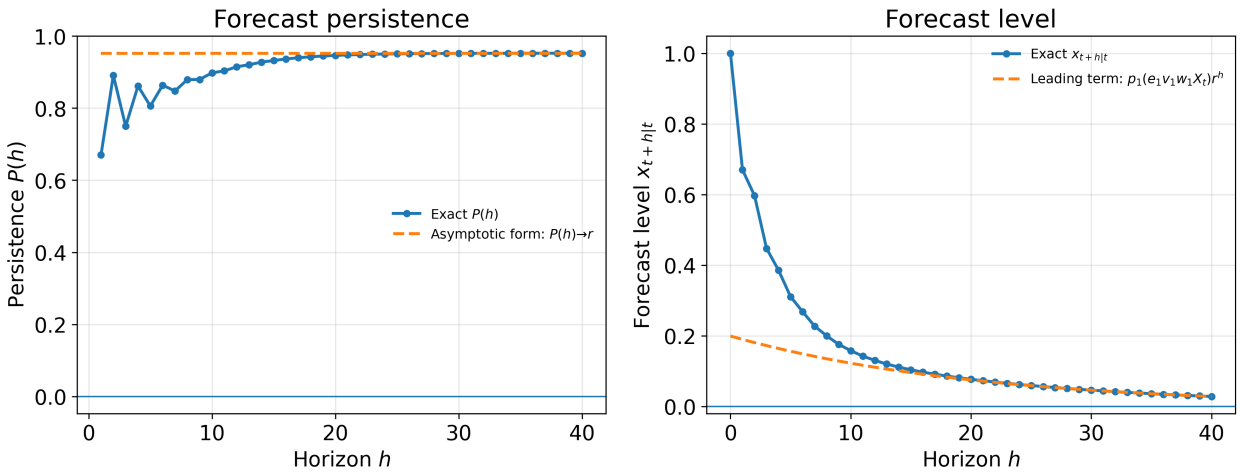
where $\phi_i = \arg(v_{1i} w_{1i})$ and $\phi^*_i = \arg(v^*_{1i} w^*_{1i})$. Moreover, when the largest eigenvalue is real, we have that:

$$\lim_{h\to\infty} \frac{1}{h} \ln R_i(h, A^*) = \ln \frac{r}{r^*} \tag{30}$$

Proof. See Appendix A.3. □

Theorem 2 shows that our results for the AR(1) case naturally transfer to general VAR processes in the case where the largest eigenvalue is real. In this case, the largest believable realization of the largest eigenvalue of the coefficient matrix A replaces that of the persistence parameter ρ in the limit of the forecast persistence function. Similarly, the true persistence parameter ρ^* is replaced by the largest eigenvalue of the true coefficient matrix A^* in the forecast reaction function.

Figure 5: An Example AR(2) Setting with a Dominant Real Eigenvalue

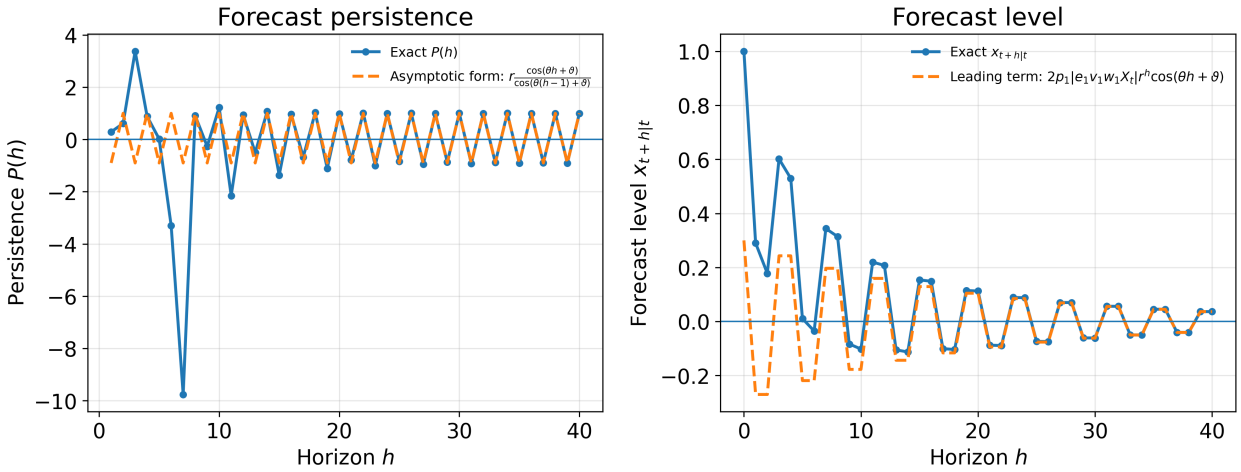


Note: This figure plots a situation in which the forecaster entertains two models: (i) an AR(2) with $\phi_1 = 0.9$ and $\phi_2 = 0.05$ with probability 0.2, and (ii) an AR(1) with $\phi_1 = 0.2$ and $\phi_2 = 0.4$ with probability 0.8. The first of these models is dominant and its largest eigenvalue is real. We initialize at $X_t = (1, 1)'$. In the left panel, we plot the forecast persistence function $P(h)$ and its asymptotic form from Theorem 2 as a function of the forecast horizon. In the right panel, we plot the forecast itself and the corresponding leading term from the eigendecomposition as a function of the forecast horizon.

The additional content of Theorem 2, relative to the AR(1) case, is that now it is possible that the largest eigenvalue comes in a complex pair. Intuitively, this case corresponds to the forecaster believing that cycles are the dominant component of the long-horizon forecast. In this case, Theorem 2 shows that the largest believable eigenvalue r governs the amplitude of cycles, while the argument of this eigenvalue intuitively governs the period of these cycles.

We illustrate each of the two cases of Theorem 2 with a simple AR(2) example. In Figure 5, we plot a situation in which the forecaster entertains two possible AR(2) models with different coefficients and the largest eigenvalue of the two corresponding companion matrices is real. Accordingly, forecasts asymptotically inherit the persistence of this largest eigenvalue, exactly as shown by the general result. This directly mimics the AR(1) case. More interestingly, in Figure 6, we plot a situation in which the forecaster entertains two AR(2) models but in which the dominant eigenvalues form a complex pair. The corresponding forecasts display seasonality and have a persistence function that oscillates asymptotically. While this pattern may seem odd, we emphasize that non-seasonally detrended data regularly display cycles (see *e.g.,* Barsky and Miron, 1989) and this forecasting regime may be relevant for such data.

Figure 6: An Example AR(2) Setting with a Dominant Complex Eigenvalue Pair



Note: This figure plots a situation in which the forecaster entertains two models: (i) an AR(2) with $\phi_1 = 0$ and $\phi_2 = -0.9$ with probability 0.3, and (ii) an AR(1) with $\phi_1 = 0.8$ and $\phi_2 = 0$ with probability 0.7. The first of these models is dominant and its eigenvalues form a complex pair. We initialize at $X_t = (1, 1)'$. In the left panel, we plot the forecast persistence function $P(h)$ and its asymptotic form from Theorem 2 as a function of the forecast horizon. In the right panel, we plot the forecast itself and the corresponding leading term from the eigendecomposition as a function of the forecast horizon.

While the mathematical content of Theorem 2 is identical for univariate and multivariate processes, there is one additional feature of interest in the multivariate context. We observe that the magnitude of persistence and over-reaction (along with the period in the complex case) is common to *all* variables. Thus, it is *as if* the forecaster believes that there is a dominant linear combination of contemporaneous variables that it uses to make all long-horizon forecasts with a persistence that is common across *all* variables. This is reminiscent of the concept of factor modeling in time-series analysis. However, rather than arising from a restriction imposed by the analyst on the model, here the presence of a dominant factor in long-run forecasting emerges endogenously from the properties of Bayesian forecasts.

The key takeaway from this analysis of more general stochastic processes is that asymptotic forecasts continue to be governed by the most persistent possible model under the forecaster's prior. This, just like in the AR(1) case, leads the forecaster to believe in high persistence at long horizons and to over-react by relatively more at long-horizons.

4 Macro Forecasting with Unknown Persistence

We now explore the implications of our findings for macroeconomic forecasts in the data. A large literature explores the empirical properties of these forecasts and uses these findings to discipline expectations formation in economic models (see, *e.g.*, the review by Angeletos et al., 2021). We show that our model of forecasting with unknown persistence is consistent with a number of facts, some known from the literature and some new, regarding the empirical term structure of expectations.

4.1 Data

We use data from the Survey of Professional Forecasts, a panel survey administered by the Federal Reserve Bank of Philadelphia. Every quarter, the survey is administered to about 40 experts from industry, government, and academia, who are asked to provide forecasts for a number of macroeconomic variables at different forward horizons. The SPF is a common reference point in academic studies of macroeconomic forecasts (e.g., Coibion and Gorodnichenko, 2015; Angeletos et al., 2021).

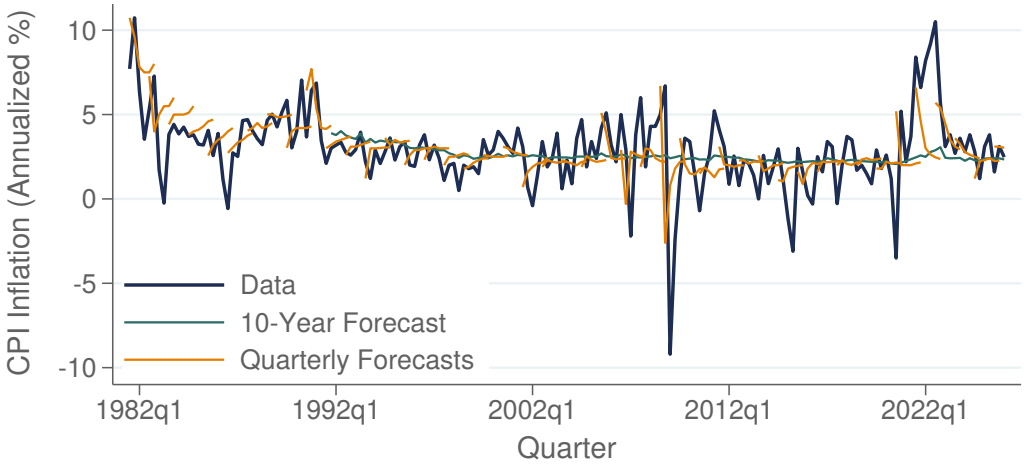
For our primary analysis, we will focus on survey forecasts of inflation measured as the Consumer Price Index. Figure 7 visualizes the basic properties of the CPI and forecasts of the CPI since 1981:Q3, when the SPF began collecting these forecasts. CPI inflation is volatile quarter-to-quarter but also experiences lower-frequency swings, such as its decline in 1981-1982 during Paul Volcker's abrupt monetary tightening and its rise and fall in 2021-22 after the Covid-19 crisis. In each SPF round, respondents are asked to forecast CPI inflation in the current quarter and the next four quarters. For a subset of time periods, we show this "term structure" of consensus (median) forecasts as orange lines in Figure 7. Finally, starting in 1991:Q4, respondents are asked to forecast annualized CPI inflation for "the next ten years." We show the consensus for this long-run forecast as the green line in Figure 7.

These raw time-series patterns give a first indication that relatively long-term expectations move considerably over the cycle. To quantify this, we observe that a simple time-series regression of the four-quarter-ahead consensus forecast on the "nowcast" reveals the relationship

$$\hat{\mathbb{E}}_t[x_{t+4}] = \underset{(0.10)}{0.50} \hat{\mathbb{E}}_t[x_t] + \varepsilon_t \tag{31}$$

where the number in parentheses is a HAC-robust standard error (bandwidth of 4 quarters). To benchmark this, we observe that CPI inflation, if modeled as an AR(1) process, has an estimated persistence of 0.75. This is the dominant assumption in the bulk of the macroeco-

Figure 7: Consumer Price Inflation and Consensus Forecasts



Note: This figure shows CPI inflation (dark blue line) and two forecasts from the Survey of Professional Forecasters: consensus quarterly forecasts (orange lines) and consensus 10-year-average forecasts (green lines). Consensus forecasts are calculated as medians among respondents. Each line for the quarterly forecasts shows their "term structure," starting with the (first-release) data available to forecasters and including forecasts at the 0, 1, 2, 3, and 4-quarter horizon. We show the forecasts only for a subset of quarterly observations, once per six quarters. Data for the 10-year forecast are available only after 1991:Q4.

nomic literature on forecasting and information frictions (e.g., Coibion and Gorodnichenko, 2015; Angeletos et al., 2021). A forecast based on an AR(1) model with this *known* persistence would generate a coefficient in Equation 31 of $(0.75)^4 = 0.32$. Perhaps more strikingly, the 10-year average forecast is even *more* responsive to the four-quarter forecast:

$$\hat{\mathbb{E}}_t[x_{t,10Y}] = \underset{(0.08)}{0.80}\ \hat{\mathbb{E}}_t[x_{t+4}] + \varepsilon_t \tag{32}$$

If inflation had a true quarter-to-quarter persistence of 0.75, then a one percent shock to inflation in four quarters' time would have an effect of $0.75^8 = 0.1\%$ in two years' time and an effect of $0.75^{36} = 0.00003\%$ after nine years. In sum, such a shock would generate an average of 0.24% of inflation per year for ten years, which is more than 3 times smaller than the estimated coefficient in Equation 32. We summarize this basic observation as follows:

Fact 1. *Long-horizon inflation forecasts vary much more over time than the benchmark model of a known, AR(1) process would predict.*

By contrast, these results are consistent with the predictions of Theorem 1, following naturally from the premise that persistence is uncertain.

4.2 Aggregate Under- and Over-reaction

We first consider the aggregate properties of forecast errors at different horizons. Theoretically, in Proposition 1, we characterized how forecasting with unknown persistence can cause forecasts to "under-react" and "over-react" at different horizons. Intuitively, at short horizons, the responses of forecasts to news are determined by the mean perceived persistence ("the average model"); at longer horizons, the responses of forecasts to news are determined by the maximum believable persistence ("the most persistent model"). Therefore, a prediction of our model is that forecasts look more "over-reactive" at longer horizons.

To test this, we adopt the regression framework of the empirical study of Halperin and Mazlish (2025), which is itself adapted from the seminal study of Coibion and Gorodnichenko (2015). Specifically, letting x_t be a macroeconomic variable (e.g., inflation) and letting $\hat{\mathbb{E}}_t$ denote consensus forecasts, these authors study the following regression model indexed by the forecast horizon h

$$\underbrace{x_{t+h} - \hat{\mathbb{E}}_t[x_{t+h}]}_{\text{Forecast error}} = \beta_1^h \cdot \underbrace{(\hat{\mathbb{E}}_t[x_{t+h}] - \hat{\mathbb{E}}_{t-1}[x_{t+h}])}_{\text{Forecast revision}} + \beta_2^h \cdot \underbrace{\hat{\mathbb{E}}_{t-1}[x_{t+h}]}_{\text{Lag forecast}} + \nu_t \tag{33}$$

In our setting, conditional on the data-generating process being an AR(1), the population estimates from this regression have a clear and simple interpretation in terms of the term structure of forecasts:

Proposition 2. *In the AR(1) setting, the Halperin and Mazlish (2025) regression (Equation 33) holds exactly with*

$$\beta_1^h = \frac{1 - R(h; \rho^*)}{R(h; \rho^*)} \tag{34}$$

Moreover, $\nu_t \perp (\mathbb{E}_t[x_{t+h}] - \mathbb{E}_{t-1}[x_{t+h}]), \mathbb{E}_{t-1}[x_{t+h}]$. Hence, a standard OLS regression estimator provides a consistent estimate of $R(h; \rho^*)$.

Proof. See Appendix A.5. □

Thus, by Proposition 1, β_1^h is initially positive and then negative if and only if $\mathbb{E}[\rho] < \rho^*$ and it is globally positive if and only if $\mathbb{E}[\rho] \geq \rho^*$. Finally, we note that while our exact derivations of this regression test are only for the AR(1) model, the basic intuition of this argument applies more broadly to other actual and perceived stochastic processes for x_t. As one useful benchmark, if forecasters have a common information set and correctly know the process for x_t—which could have complicated AR(p) dynamics and/or a latent persistent state—then $\beta_1^h = \beta_2^h = 0$ by a simple application of the law of iterated expectations.

Halperin and Mazlish (2025) estimate Equation 33 using survey data from Consensus Economics covering 89 countries and four primary outcome variables: GDP, inflation, consumption, and investment. They find consistent evidence that both β_1^h and β_2^h decline in horizons, which they vary from zero to ten years forward (see Figure 1 of Halperin and Mazlish, 2025). Moreover, both coefficients are negative for horizons greater than two years: in these cases, an increase in today's or yesterday's forecast predicts, on average, that the forecast over-estimates the economic variable. Both findings would be consistent with our model of forecasting with unknown persistence, provided that the true persistence of variables is lower than the maximum believable persistence.

We recreate these findings within our own data on inflation expectations at horizons up to three quarters, which we note are much shorter than those studied by Halperin and Mazlish (2025). Nonetheless, in our data, the evidence also points to a declining profile of β_1^h and β_2^h in horizon (Figure 15). Thus, inflation expectations react differently to shocks at different horizons. We summarize these findings below:

Fact 2 (Halperin and Mazlish (2025))**.** *At longer horizons, forecasts over-react more to shocks contained in past forecasts and forecast revisions.*

4.3 The Term Structure of Individual Forecasts

We next zero in on the term structure of *individual* forecasts using the available panel data. As a first test, we consider the raw relationship between individual forecasts of variable x (denoted by $\mathbb{E}_{it}$) at adjacent horizons h and $h-1$:

$$\mathbb{E}_{it}[x_{t+h}] = \gamma^h \cdot \mathbb{E}_{it}[x_{t+h-1}] + \varepsilon_{it} \tag{35}$$

In the AR(1) case, this exactly identifies the forecast persistence function.

Proposition 3. *In the AR(1) setting, the Goldstein and Gorodnichenko (2022) regression (Equation 35) holds exactly with:*

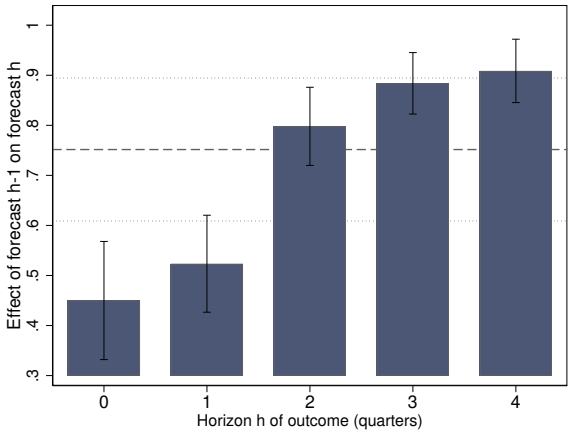
$$\gamma^h = \frac{\mathbb{E}_F[\rho^h]}{\mathbb{E}_F[\rho^{h-1}]} = P(h) \tag{36}$$

Proof. Immediate by definition of P. □

The key testable prediction follows from property 2 of Theorem 1: uncertain persistence should lead γ^h to increase with the horizon h.

Goldstein and Gorodnichenko (2022) estimate Equation 35 using data from the US SPF and find evidence that γ^h increases in horizon: that is, the term structure of individual

Figure 8: The Flattening Term Structure of CPI Forecasts



Note: This figure shows estimates of Equation 35 corresponding to forecasts of CPI inflation, replicating the analysis of Goldstein and Gorodnichenko (2022) The sample covers 1981:Q1 to 2026:Q1. Each estimate corresponds to the coefficient β from a separate regression at the indicated horizon for the outcome variable. Note that the $h-1$ "forecast" used as the regressor for $h=0$ is first-release data reported within the SPF questionnaire and therefore visible to respondents. The dashed horizontal line is the estimated persistence of CPI inflation in the same sample, and the dotted horizontal lines are the upper and lower limits of a 95% confidence interval based on HAC robust standard errors with a 4-quarter bandwidth. Error bars on the main estimates are 95% confidence intervals based on standard errors double clustered by forecaster ID and quarter.

forecasts tends to *flatten out*. We recreate their finding in Figure 8. For reference, we compare the estimates of γ^h with the estimated time-series persistence of CPI inflation (dashed line) and statistical uncertainty in this estimate (dotted lines, indicating a 95% confidence interval). This yields two additional observations. First, the data are consistent with a pattern of under-reaction at short horizons ($\gamma^h = P(h) < \rho^*$) and over-reaction at longer horizons ($\gamma^h = P(h) > \rho^*$). Second, the dramatic change in γ^h may be especially plausible given the high uncertainty for an econometrician (even *ex post*) regarding the persistence of inflation.

Other Variables. We explore the boundaries of this prediction for other variables in Figure 16. We reassuringly have very similar estimates for PCE inflation, suggesting that our findings reflect something structural about inflation dynamics rather than the specifics of one measurement of inflation. In contrast, we find no significant evidence of a flattening term structure for nominal interest rate (three-month Treasury bill rates) or unemployment expectations. But there are two fundamental differences between these cases and that of inflation: (i) these term structures are extremely persistent at *all* horizons ($\gamma^h > 0.9$) and (ii) these are more persistent time series in practice, whose persistence is moreover more

precisely estimated in the data. For both reasons, we argue that it would be natural for forecasters to face less meaningful uncertainty about the persistence of these variables, thus generating less variation in the persistence function.

Summary. We summarize these lessons about the empirical term structure below:

Fact 3 (Goldstein and Gorodnichenko (2022)). *The term structure of individual inflation forecasts flattens: at longer horizons, forecasts become more similar between adjacent horizons.*

4.4 Unpacking the Mechanism: Stability and (Dis)anchoring

So far, we have still focused on properties of the entire distribution of forecasts rather than properties of *individual* forecasts. In particular, if individuals are forecasting with unknown persistence, then their individual patterns of response should flatten as the horizon increases.

To explore this, we first calculate an individual measure of the persistence function at each horizon h by using the term structure of quarterly forecasts as well as the reported long-run forecast:

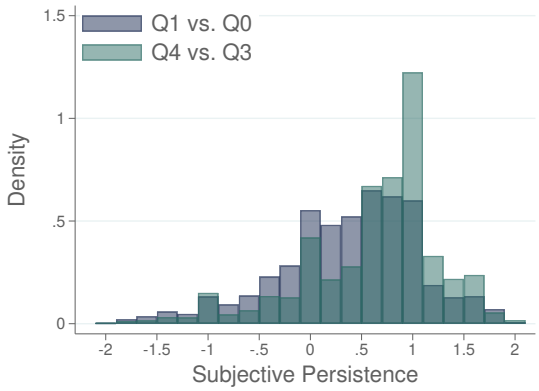
$$\hat{\rho}_{ith} = \frac{\hat{\mathbb{E}}_{it}[x_{i,t+h}] - \hat{\mathbb{E}}_{it}[x_{i,10Y}]}{\hat{\mathbb{E}}_{it}[x_{i,t+h-1}] - \hat{\mathbb{E}}_{it}[x_{i,10Y}]} \tag{37}$$

This is the ratio of forecasts in deviation from what each individual reports as the long-run mean level of inflation.[4] We compare the distribution of this statistic for $h = 1$ and $h = 4$ in Figure 9. As predicted by our model, the distribution of persistence shifts toward the right at the longer horizon. A second clearly apparent feature is a prominent spike at $\hat{\rho} = 1$ for $h = 4$. This is generated when forecasts at horizons $h = 3$ and $h = 4$ are the same as one another, but neither is equal to the 10-year forecast.

To even better understand this phenomenon, we next calculate two statistics that describe the shape of the individual term structure. We say that forecasts are "stabilized" at horizon h if they are almost the same at horizons h and $h - 1$. Given that most respondents report forecasts only up to 0.1% increments, we adopt the empirical definition of "almost the same" as being within 0.1% points in absolute value. We say that forecasts are "anchored" at horizon h if they are almost the same (by the same definition) as the 10-year forecast. Finally, with these definitions in hand, we can say that forecasts are "stabilized yet disanchored" if they satisfy the first definition but not the second. Such forecasts might be natural to report if the respondent thought inflation was affected by a highly persistent shock.

[4] We treat this as undefined if the denominator is zero, i.e., the $h - 1$ quarter forecast coincides with the long-run forecast.

Figure 9: The Individual Persistence of CPI Forecasts



Note: This figure shows the cross-sectional distribution of persistence at the one-quarter horizon, ρ_{it1} (blue distribution), and persistence at the four-quarter horizon, ρ_{it4} (green distribution), as defined in Equation 37. The sample is CPI forecasts after 1991:Q4. We truncate both distributions at ± 2, excluding 5% of each distribution in the tails.

CPI forecasts are much more likely to be stabilized and to be anchored as the horizon increases (blue and green bars in Figure 10). Strikingly, 33% of all reported forecasts are stabilized at the four-quarter horizon. In the majority of these cases, forecasts are also disanchored by our definition (red bars in Figure 10): at $h = 4$, this describes 25% of all forecasts and hence 78% of all stabilized forecasts. Moreover, these forecasts are not simply very close approximations of the long-run target: in Appendix Figure 14, we plot the gaps between four-quarter forecasts and the 10-year forecasts by category and show that stabilized forecasts are highly dispersed, to an equivalent extent as their non-stabilized counterparts.

We summarize these findings below:

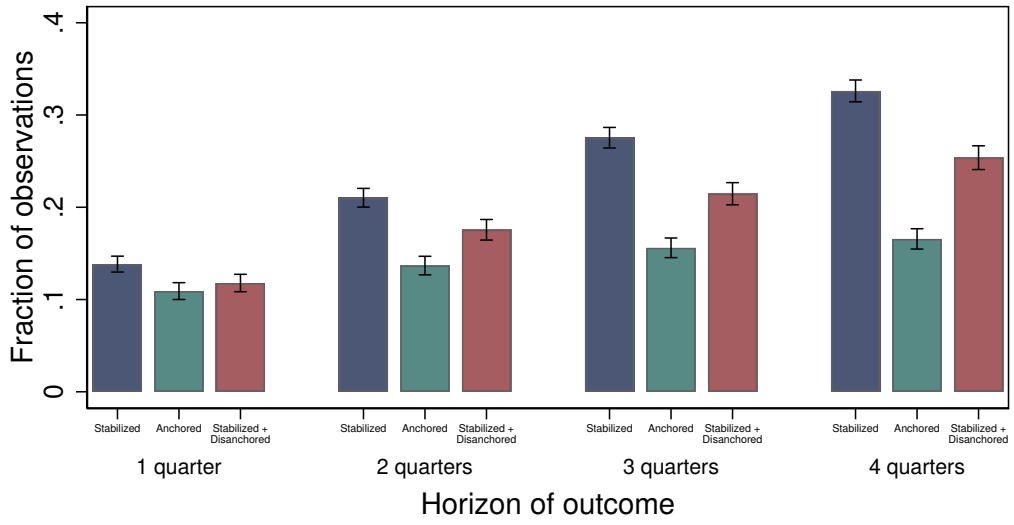
Fact 4. *At one-year horizons, individual inflation forecasts tend to stabilize on constant values. But, these stable values often differ from forecasts of long-run average inflation.*

This possibility that forecasts are "stabilized but disanchored" is natural in our theory and, in fact, was even anticipated in our motivating example in the Introduction (Figure 1). For example, this would occur any time that inflation is hit by a shock and forecasters consider model(s) in which inflation contains a unit root.

4.5 The Multivariate Term Structure of Forecasts

We finally study the multivariate patterns of forecasts. While it is especially tractable to model forecasters as having a collection of univariate models for the economy, we suspect in

Figure 10: Stability and (Dis)anchoring in CPI Forecasts

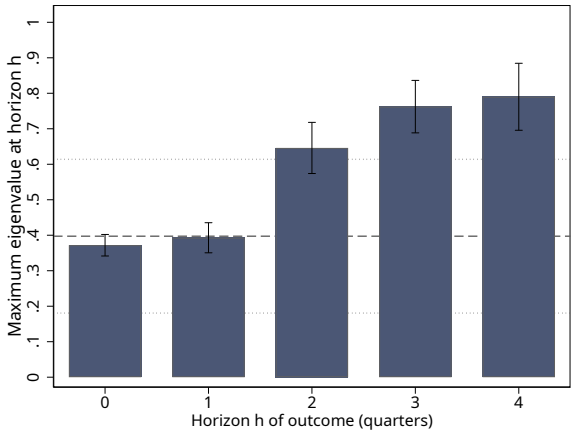


Note: This figure shows the frequency of individual CPI forecasts (i.e., forecaster-by-quarter observations) that satisfy definitions "stability" and "anchoring" at different horizons. Forecasts are stabilized at horizon h (blue bars) if forecasts for h and $h-1$ differ by at most 0.1% points. This is calculated over the full sample from 1981:Q1 to 2026:Q1. Forecasts are anchored at horizon h (green bars) if the forecast is within 0.1% points of the long-run (10-year average) forecast. This is calculated after 1991:Q1 when the 10-year question was introduced. Forecasts are stabilized but disanchored (red bars) if they satisfy the first definition but not the second. Error bars are 95% confidence intervals based on robust standard errors.

practice that most if not all macroeconomic forecasters have a multivariate view of the world. For instance, forecasters might understand that interest rates respond to unemployment and inflation due to systematic monetary policy, or that unemployment and inflation are affected by common shocks to aggregate supply and aggregate demand. In support of this view, Wu (2023) uses SPF data to show that revisions to forecasts of multiple distinct variables predict forecast errors of any given variable, including but not limited to inflation.

Our analysis extends to the case of multivariate forecasting via any finite-order VARMA (Section 3) and implies that, as the forecast horizon increases, the forecasts of an agent become more "persistent" in a specific sense. To test this, we revisit the estimating equation of Goldstein and Gorodnichenko (2022) but now consider a *multivariate* extension. Specifically, we focus on forecasts of inflation (CPI), unemployment, and interest rates (the 3-month Treasury rate), as these main variables in the canonical "three-variable" representation of the macroeconomy (the investment-savings curve, the Phillips curve, and the monetary policy

Figure 11: The Flattening Term Structure of Multivariate Forecasts



Note: This figure shows estimates of the persistence of multivariate forecasts. Each bar shows an estimate of λ_h, the largest eigenvalue of Γ_h, where Γ_h is constructed (as described in the main text) from estimates of equation (38). The sample covers 1981:Q3 to 2026:Q1. Note that the $h-1$ "forecasts" used as the regressors for $h=0$ are first-release data reported within the SPF questionnaire and therefore visible to respondents. The dashed horizontal line is the estimated value of λ from a VAR(1) estimate on the full sample, and the dotted horizontal lines are the upper and lower limits of a 95% confidence interval based on bootstrapping. Error bars on the main estimates are 95% confidence intervals based on bootstrapping estimates by forecaster ID.

rule). For each variable x in the vector $z = \{\pi, u, r\}$, we estimate the regression model

$$\mathbb{E}_{it}[x_{t+h}] = \gamma_h^{\pi,x} \cdot \mathbb{E}_{it}[\pi_{t+h-1}] + \gamma_h^{u,x} \cdot \mathbb{E}_{it}[u_{t+h-1}] + \gamma_h^{r,x} \cdot \mathbb{E}_{it}[r_{t+h-1}] + \varepsilon_{it} \tag{38}$$

and we represent the model in matrix form as $\mathbb{E}_{it}[z_{t+h}] = \Gamma_h \cdot \mathbb{E}_{it}[z_{t+h-1}] + E$. For each horizon h, we calculate the modulus of the maximum eigenvalue of Γ_h, and denote this object as λ_h. If respondents in the SPF forecasted with a dogmatic VAR(1) model of inflation, unemployment, and interest rates, then we would observe that Γ_h and, in particular, λ_h do not depend on horizon. If instead these objects depend on the horizon and, in particular, λ_h increases in horizon, this would be consistent with our prediction under model uncertainty that forecasters effectively apply higher persistence at higher horizons.

Our findings, summarized in Figure 11 are consistent with our early exploration of the univariate model: at longer horizons, forecasters use more persistent models. Moreover, our point estimates suggest that the multivariate term structure of forecasts is more persistent than the underlying economic variables at all horizons ≥ 1 quarter. That is,

Fact 5. *The term structure of multivariate forecasts flattens: at longer horizons, the vector of forecasts regarding inflation, unemployment, and interest rates becomes more persistent.*

5 Investments With Uncertain Persistence

We finally consider the economic implications of forecasting with unknown persistence. To do this, we study an intentionally stylized model of investment under uncertainty about the stochastic process for fundamentals. We characterize how uncertainty about model parameters translates to demand for investments at different horizons. We show how the model can generate a particularly striking pattern of investment: the agent desires to invest in projects with a short maturity and those with a sufficiently long maturity, but *not* those with intermediate maturity. We draw implications for boom-bust investment cycles.

5.1 Set-up

At time $t = 0$, a representative investor chooses how much to invest in each of a finite number of projects indexed by their date of maturity, $h \in \{0, 1, \ldots, H\}$. Investing x_h units in project h incurs a cost today of $p_h x_h$, where p_h is the cost of a project-specific input. This investment generates a payoff $\theta_h x_h$, delivered at time h, where θ_h is a random variable. Specifically, this variable follows an AR(1) process:

$$\theta_t = \rho\theta_{t-1} + \varepsilon_t, \qquad \forall t \geq 1 \tag{39}$$

where ε_t is drawn IID from a distribution G with mean zero and $\theta_0 > 0$ is observed at the time of investment. The investor is uncertain about the parameter ρ that governs the behavior of the shock θ_t: they believe $\rho \sim F$ where F is supported within the interval $(0, r]$, for some $r > 0$. Finally, the investor's objective is to maximize their expected discounted payoff with discount factor $\delta \in (0, 1)$. Putting this together, the investor solves:

$$\max_{(x_h)_{h=0}^T \in \mathbb{R}^H} \sum_{h=0}^{H} \left(\mathbb{E}\left[\delta^h \theta_h x_h\right] - p_h x_h \right) \tag{40}$$

To close the model, we assume that there is an outside sector that produces the project-specific inputs. This generates, for each input h, a supply curve of the form $x_h = p_h^\chi$ for some $\chi > 0$. We study an equilibrium in which the investor optimizes taken as given the price of each input, and the price of each input lies on the respective supply curve.

Interpretation. For physical production, our model can be thought of as a particular instantiation of a technology in which time is an explicit input. Such a technology is a crucial ingredient in the "Austrian" model of capital formalized classically by Böhm-Bawerk (1891) and recently revised in a modern context by Antràs and Kulesza (2026). In this particular

formulation, all types of inputs must be installed today but the investor can choose, in effect, the time to production by investing in projects with different horizons. In this context, it will be natural to consider partially elastic supply of inputs ($\chi > 0$), so variation in input demand translates to physical output.

This model also admits natural interpretations for financial investments. In equity markets, x_t might represent investments in a "dividend strip" trading strategy that buys and sells a stock in a narrow window around a stochastic dividend payout represented by θ_t. In debt markets, x_t might represent an investment in a bond with maturity t that has an uncertain *real* payoff, so θ_t is the inverse of the price level. In both of these cases, it might be natural to consider these assets in constant, inelastic supply ($\chi = 0$), so prices are determined to make the investor indifferent between taking any directional position.

5.2 The Term Structure of Investments

For each project h, the investor's expected return is $\mathbb{E}_0[\delta^h \theta_h] - p_h$. Substituting in the investor's model of θ_h, this can be written as $\theta_0 \delta^h \mathbb{E}_0[\rho^h] - p_h$. Finally, using the supply curve $p_h = x_h^{\frac{1}{\chi}}$, we arrive at the simple description of optimal investments and equilibrium prices:

$$x_h = \left(\theta_0 \delta^h \mathbb{E}[\rho^h]\right)^{\chi}, \qquad p_h = \theta_0 \delta^h \mathbb{E}[\rho^h], \qquad \forall h \in \{0, \ldots, H\} \tag{41}$$

An immediate implication is that relative prices and investment demands at any two adjacent horizons $h-1$ and h are determined by the discount factor and the persistence function:

$$\frac{x_h}{x_{h-1}} = \delta^{\chi} P(h)^{\chi}, \qquad \frac{p_h}{p_{h-1}} = \delta P(h), \qquad \text{where } P(h) = \frac{\mathbb{E}[\rho^h]}{\mathbb{E}[\rho^{h-1}]} \tag{42}$$

Thus, Theorem 1 can be applied directly to characterize the general properties of the term structure of investments and the prices of investment goods:

Corollary 2 (The Term Structure of Investments)**.** *The following statements are true:*

1. *If $\mathbb{E}[\rho] < \delta^{-1}$, then both quantities and prices of investment are initially decreasing in the investment horizon.*
2. *The ratios between adjacent investment quantities and prices in this term structure are increasing in the investment horizon.*
3. *Quantities and prices of investment can increase in the horizon if (i) it is conceivable for the project payoff to grow faster than discounting ($r > \delta^{-1}$) and (ii) sufficiently long-horizon projects are feasible (H is large enough).*

Proof. Immediate from combining Theorem 1 and Equation 42. □

We unpack these three statements in turn. First, because $P(1) = \mathbb{E}[\rho]$, both quantities and prices of investment are initially decreasing in horizon provided that $\mathbb{E}[\rho] < \delta^{-1}$. Second, because $P(h)$ is increasing, the ratio between adjacent investment quantities and prices in this term structure is increasing in horizon. Third, because $\lim_{h\to\infty} P(h) = r$, it is possible for quantities and prices of investment to *increase* in the horizon if (i) it is conceivable for the project payoff to grow faster than discounting ($r > \delta^{-1}$) and (ii) sufficiently long-horizon projects are feasible (H is large enough). Intuitively, all three properties arise because, as the horizon extends, models that feature a greater persistence (or higher growth) of θ_t become more relevant for expected payoffs.

A Simple Example. To exemplify these properties, we return to the example in which ρ follows a scaled β distribution. In this case, the expectations persistence function has a known form derived in Equation 7. As a result, we can calculate the term structure of investments and prices in closed form:

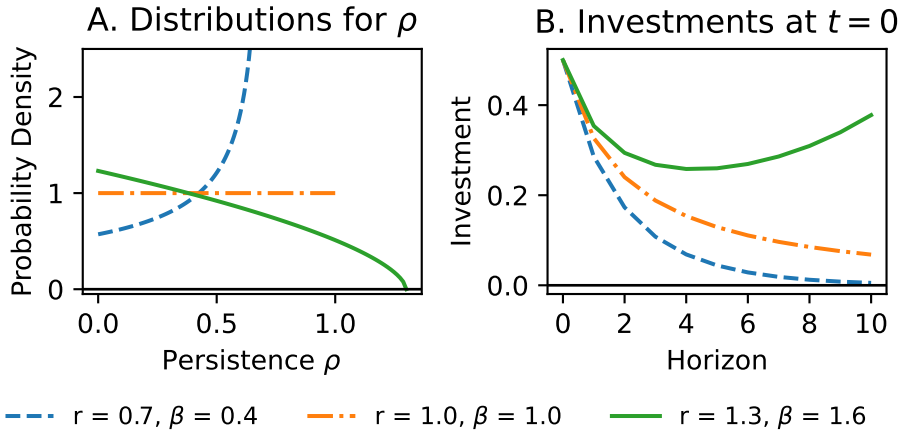
$$\frac{x_h}{x_{h-1}} = \delta^{\chi}\left(r\frac{h+\alpha-1}{h+\alpha+\beta-1}\right)^{\chi}, \qquad \frac{p_h}{p_{h-1}} = \delta r\frac{h+\alpha-1}{h+\alpha+\beta-1} \tag{43}$$

We visualize the resulting optimal investments in Figure 12, varying the shape of the β distribution across three calibrations but keeping its mean normalized at $\mathbb{E}[\rho] = \frac{1}{2}$. The three calibrations differ in the amount of weight they put on cases with higher persistence or, in one case, expected growth. For the distributions that put all mass on $\rho < 1$, the term structure of investments is downward sloping, but its decay depends on the amount of mass on more persistent models (blue dashed line and orange dotted line). For the distribution that allows for growth and, in particular, values of $\rho > \delta^{-1}$, the term structure of investments is U-shaped: the investor is particularly keen on short-horizon projects and very long-horizon projects, but not projects with an intermediate horizon.

5.3 Boom-Bust Investment Cycles

We finally finish with a simple illustration of how this model can generate endogenous *investment cycles*. To do this, we consider a dynamic extension of the model in which, in each time period $t \in \mathbb{N}$, the investor considers projects that mature at each future period $\{t, \ldots, t+H\}$. We let $x_{t,h}$ denote the h-period ahead investment made at time t. In such an economy, total output Y_t at any date t depends on all past investments that mature at

Figure 12: Investments When Payoffs Have Unknown Persistence



Note: This figure shows an example calibration of the investment model of Section 5. Panel A visualizes three possible subjective distributions for the persistence of payoffs, ρ. Each is a β distribution with $\alpha = 1$ and mean $\mathbb{E}[\rho] = 0.5$, but different values of r (scale) and β. Panel B shows the resulting term structure of investments, given a maximum horizon of $H = 10$, an initial productivity $\theta_0 = 0$, a supply elasticity of $\chi = 1$, and a discount factor $\delta = 0.98$.

this date, or the capital stock K_t:

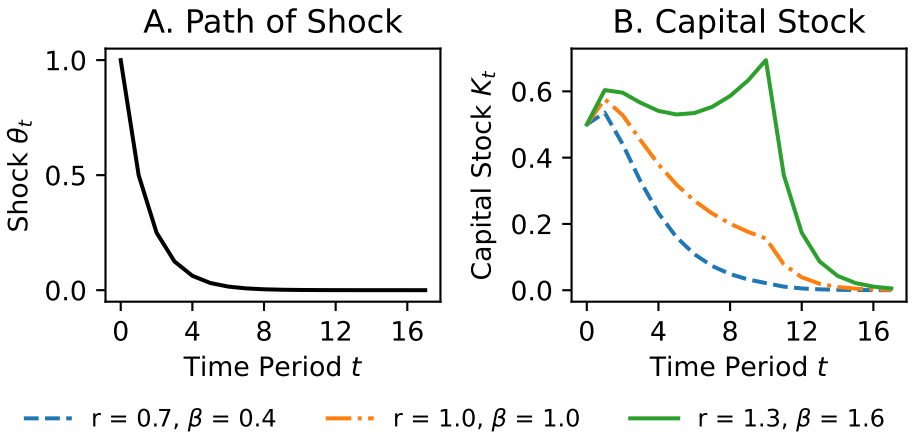
$$Y_t = \theta_t K_t, \qquad K_t = \sum_{h=0}^{H} x_{t-h,h} \tag{44}$$

We finally assume the economy starts with no projects, or $x_{t,h} = 0$ for all $t < 0$.

Figure 13 shows the output dynamics following a transitory shock to productivity θ_t. The presence of investment opportunities at multiple horizons "backloads" the effect on output in all cases. In the third model, which generated a U-shaped term structure of investments (recall Figure 12), the response of capital is *double peaked* (green solid line). When the shock hits in this economy, the investor *expects* the shock to be transitory ($\mathbb{E}[\rho]$), but considers the possibility that productivity will actually grow in the long run. In the calibration, this subjective probability that $\rho > \delta^{-1}$ is only 9%. Nonetheless, this creates strong incentives to invest in long-maturing capital. In this sample path, these capital investments turn out *ex post* to be very unproductive, as in standard Austrian theories of the business cycle (Hayek, 1931). Critically, this occurs despite the fact that investors are not over-optimistic in the standard sense: they over-invest in *ex post* worthless capital despite the fact that their expectations of persistence are unbiased.

We finally note that an entirely symmetric logic would apply to a transitory negative

Figure 13: Investments When Payoffs Have Unknown Persistence



Note: This figure shows investment dynamics after a shock in the investment model of 5. Panel A shows the time path of the valuation or productivity shock θ_t, which is mean-reverting with a persistence of $\frac{1}{2}$. Panel B shows the time path of maturing investments calculated as in Equation 44. In each panel, we vary the investor's subjective probability distribution regarding ρ. Each is a β distribution with $\alpha = 1$ and mean $\mathbb{E}[\rho] = 0.5$, but different values of r (scale) and β (see Panel A of Figure 12) The model is calibrated with a maximum horizon of $H = 10$, an initial productivity $\theta_0 = 0$, a supply elasticity of $\chi = 1$, and a discount factor $\delta = 0.98$.

shock. If our analysis is interpreted relative to an initial steady-state level of investment, then it implies that investors may disinvest from both current and future projects. This, in turn, contributes toward both an immediate bust and a "double dip" in the future.

In this way, our simple model can generate endogenous "boom-bust" cycles in investment. The key ingredient for this result is investors' uncertainty about the persistence or growth of fundamentals, a potentially very plausible ingredient for investments at both the micro and macro scales. Most interestingly, this intentionally extremely simple model generates these rich dynamics despite the absence of both behavioral biases and financial frictions.

6 Conclusion

We study the implications of uncertainty about persistence for the structure of forecasts. For AR(1) processes, we find that forecasts asymptotically are as persistent as is believable. For this reason, forecasts always over-react at long horizons. At short horizons, forecasts under-react if and only if beliefs about persistence have a downward bias. We generalize these findings to univariate AR(p) processes and multi-variate VAR processes and uncover analogous results when dominating eigenvalues are real. When dominating eigenvalues are complex,

asymptotic forecasts become maximally persistent and exhibit cycles. In the data, we show that our theory is consistent with the estimated term structure of inflation forecasts and can explain seemingly contradictory findings in the literature regarding simultaneous under-reaction and over-reaction at different horizons. In an application to investment dynamics, we show how our mechanism can generate rich patterns in investments across horizons and even generate boom-bust cycles within a purely "rational," Neoclassical framework.

Appendices

A Omitted Derivations and Proofs

A.1 Proof of Theorem 1

Proof. The first statement follows immediately from the fact that $x_{t+h|t} = \mathbb{E}[\rho]x_t = \mathbb{E}[\rho]x_{t|t}$. The second statement follows from by taking logarithms of the forecast persistence:

$$\ln P(h) = \ln M_{\ln \rho}(h) - \ln M_{\ln \rho}(h-1) = K_{\ln \rho}(h) - K_{\ln \rho}(h-1) \tag{45}$$

where $K_{\ln \rho} \equiv \ln M_{\ln \rho}$ is the cumulant generating function of the natural logarithm of the persistence. As any cumulant generating function is convex (see *e.g.,* Lemma 2.2.5 in Dembo and Zeitouni, 2009), it follows immediately that $\ln P(h)$ (and therefore $P(h)$) is increasing in h.

The final statement can be derived by using the tilting method from large deviations theory and the Laplace principle (see *e.g.,* Dembo and Zeitouni, 2009). As the ideas are elementary, we give a self-contained proof. In particular, we have that for all $\epsilon \in (0, r)$:

$$\begin{aligned} P(h) &= \frac{\mathbb{E}[\rho^h]}{\mathbb{E}[\rho^{h-1}]} \geq \frac{\mathbb{E}[\rho^h \mathbb{I}[\rho > r-\epsilon]]}{\mathbb{E}[\rho^{h-1}]} \geq (r-\epsilon)\frac{\mathbb{E}[\rho^{h-1}\mathbb{I}[\rho > r-\epsilon]]}{\mathbb{E}[\rho^{h-1}]} \\ &= (r-\epsilon)\left(1 - \frac{\mathbb{E}[\rho^{h-1}\mathbb{I}[\rho \leq r-\epsilon]]}{\mathbb{E}[\rho^{h-1}]}\right) \geq (r-\epsilon)\left(1 - \frac{(r-\epsilon)^{h-1}}{\mathbb{E}[\rho^{h-1}]}\right) \\ &= (r-\epsilon)\left(1 - \frac{(r-\epsilon)^{h-1}}{\mathbb{E}[\rho^{h-1}]}\right) \end{aligned} \tag{46}$$

Define $p(\epsilon) = \mathbb{P}[\rho > r - \epsilon/2] > 0$, where the strictly inequality follows as r is the essential

supremum of ρ. We have that:

$$\begin{aligned}\mathbb{E}[\rho^{h-1}] &= \mathbb{E}[\rho^{h-1}\mathbb{I}[\rho \leq r-\epsilon/2]] + \mathbb{E}[\rho^{h-1}\mathbb{I}[\rho > r-\epsilon/2]]] \geq p(\epsilon)\mathbb{E}[\rho^{h-1}|\rho > r-\epsilon/2] \\ &\geq p(\epsilon)(r-\epsilon/2)^{h-1}\end{aligned} \tag{47}$$

Combining the previous two equations, we have that:

$$P(h) \geq (r-\epsilon)\left(1 - \frac{(r-\epsilon)^{h-1}}{p(\epsilon)(r-\epsilon/2)^{h-1}}\right) = (r-\epsilon)\left(1 - \frac{1}{p(\epsilon)}\left(\frac{(r-\epsilon)}{(r-\epsilon/2)}\right)^{h-1}\right) \tag{48}$$

Taking the limit of both sides as $h \to \infty$, we then have shown that for all $\epsilon \in (0, r)$:

$$\lim_{h\to\infty} P(h) \geq r - \epsilon \tag{49}$$

It is moreover immediate that $P(h) = \frac{\mathbb{E}[\rho\rho^{h-1}]}{\mathbb{E}[\rho^{h-1}]} \leq r\frac{\mathbb{E}[\rho^{h-1}]}{\mathbb{E}[\rho^{h-1}]} = r$. Thus, $\lim_{h\to\infty} P(h) = r$. □

A.2 Proof of Proposition 1

Proof. To prove the first statement, observe that we can write:

$$\ln R(h;\rho^*) = K_{\ln\rho}(h) - h\ln\rho^* = h\left(\frac{K_{\ln\rho}(h)}{h} - \ln\rho^*\right) \tag{50}$$

As $K_{\ln\rho}$ is a cumulant generating function, it is convex. Moreover, if we extend the domain of $K_{\ln\rho}$ to $\mathbb{R}_+$, we can see that $K_{\ln\rho}(0;\rho^*) = 0$. Thus, we have that $K_{\ln\rho}(h)$ is non-decreasing. To see this, observe that we can write for any $h, h' : h < h'$:

$$K_{\ln\rho}(h) \leq \frac{h}{h'}K_{\ln\rho}(h') + \left(1 - \frac{h}{h'}\right)K_{\ln\rho}(0) = \frac{h}{h'}K_{\ln\rho}(h') \tag{51}$$

which implies that $K_{\ln\rho}(h)/h \leq K_{\ln\rho}(h')/h'$ for all $h < h'$. It follows that the sign of $\ln R$ can change at most once. Moreover, if $K_{\ln\rho}(1) = \ln\mathbb{E}[\rho] \geq \ln\rho^*$, then the sign never changes. In the converse case that $K_{\ln\rho}(1) = \ln\mathbb{E}[\rho] < \ln\rho^*$, we have that the sign must change at least once. To see this, we observe by the Laplace principle that (see *e.g.,* Theorem 4.3.1 in Dembo and Zeitouni, 2009) that $\lim_{h\to\infty}\frac{1}{h}\ln K_{\ln\rho}(h) = \ln r > \ln\rho^*$. This observation also yields the final claim. □

A.3 Proof of Theorem 2

Proof. Without loss of generality, we write all results for variable $i = 1$. Recall that the forecast persistence function case can be written as

$$P_1(h) = \frac{e_1' \mathbb{E}[A^h] X_t}{e_1' \mathbb{E}[A^{h-1}] X_t} \tag{52}$$

Given *regularity* (as stated in Definition 1), the forecaster places probabilities $\{p_i\}_{i=1}^n$ on coefficient matrices $\{A_i\}_{i=1}^n$. Each of these matrices is diagonalizable with distinct eigenvalues λ_{ij}, $j \in \{1, ..., p\}$. We order the eigenvalues such that $|\lambda_{i1}| \geq |\lambda_{i2}| \geq ... \geq |\lambda_{ip}|$. This allows us to write the forecast persistence function as

$$P_1(h) = \frac{\sum_{i=1}^n \sum_{j=1}^p p_i \cdot \lambda_{ij}^h e_1' v_{ij} w_{ij}' X_t}{\sum_{i=1}^n \sum_{j=1}^p p_i \cdot \lambda_{ij}^{h-1} e_1' v_{ij} w_{ij}' X_t} \tag{53}$$

The second part of the *regularity* condition (see Definition 1) specifies that there is exactly one matrix A_i with the largest spectral value r. Hence, the eigenvalues of all but that single matrix i are smaller in absolute value than r. Consequently,

$$\lambda_{kj}^h e_1 v_{kj} w_{kj}' X_t = o(r^h) \tag{54}$$

for all $k \in \{1, ..., n\}$ with $k \neq i$. The forecast persistence function is reduced to

$$P_1(h) = \frac{p_i \sum_{j=1}^p \lambda_{ij}^h e_1' v_{ij} w_{ij}' X_t + o(r^h)}{p_i \sum_{j=1}^p \lambda_{ij}^{h-1} e_1' v_{ij} w_{ij}' X_t + o(r^{h-1})} \tag{55}$$

The probability p_i cancels. By *regularity* there are two possible cases. First, the eigenvalue λ_{i1} has a larger absolute value than λ_{i2}, λ_{i1} is real, and $r = \lambda_{i1}$. In this case,

$$\lambda_{ij}^h e_1' v_{ij} w_{ij}' X_t = o(r^h) \tag{56}$$

for all $j > 1$, which implies that the forecast persistence function can be written as

$$P_1(h) = \frac{r^h e_1' v_{i1} w_{i1}' X_t + o(r^h)}{r^{h-1} e_1' v_{i1} w_{i1}' X_t + o(r^{h-1})} = r \frac{e_1' v_{i1} w_{i1}' X_t + o(1)}{e_1' v_{i1} w_{i1}' X_t + o(1)} \tag{57}$$

and $\lim_{h \to \infty} P(h) = r$. In the second case, λ_{i1} is not the only eigenvalue that attains the absolute value of r. In this case, λ_{i1} is complex, and λ_{i2} is λ_{i1}'s complex conjugate. Clearly,

$$\lambda_{ij}^h e_1' v_{ij} w_{ij}' X_t = o(r^h) \tag{58}$$

for all $j > 2$. The forecast persistence function can be written as

$$P_1(h) = \frac{\lambda_{i1}^h e_1' v_{i1} w_{i2}' X_t + \lambda_{i2}^h e_1' v_{i2} w_{i2}' X_t + o(r^h)}{\lambda_{i1}^{h-1} e_1' v_{i1} w_{i2}' X_t + \lambda_{i2}^{h-1} e_1' v_{i2} w_{i2}' X_t + o(r^{h-1})} \tag{59}$$

We write the eigenvalues in their polar coordinates $\lambda_{i1} = re^{i\theta}$ and $\lambda_{i2} = re^{-i\theta}$ and define $\vartheta = \arg(e_1' v_{i1} w_{i1} X_t)$. Define $m = |e' v_{i1} w_{i1}' X_t|$ and since $v_{i2} w_{i2}' = \bar{v_{i1}} \bar{w_{i2}}$, we have $e' v_{i1} w_{i1} X_t = me^{i\vartheta}$ and $e' v_{i2} w_{i2}' X_t = me^{-i\vartheta}$. This allows us to write the forecast persistence function in the following way

$$\begin{aligned} P_1(h) &= \frac{r^h e^{i\theta h} m e^{i\vartheta} + r^h e^{-i\theta h} m e^{-i\vartheta} + o(r^h)}{r^{h-1} e^{i\theta(h-1)} m e^{i\vartheta} + r^{h-1} e^{-i\theta(h-1)} m e^{-i\vartheta} + o(r^{h-1})} \\ &= r \frac{e^{i(\theta h + \vartheta)} + e^{-i(\theta h + \vartheta)} + o(1)}{e^{i(\theta(h-1)+\vartheta)} + e^{-i(\theta(h-1)+\vartheta)} + o(1)} \\ &= r \frac{\cos(\theta h + \vartheta) + o(1)}{\cos(\theta(h-1) + \vartheta) + o(1)} \end{aligned} \tag{60}$$

which concludes the proof of the first implication. The proof of the second implication proceeds in a very similar fashion. Consider the forecast reaction function

$$R_1(h, A^*) = \frac{\mathbb{E}[A^h]_{11}}{(A^*)_{11}} \tag{61}$$

Using *regularity*, we expand the expectation and diagonalize the matrices A_i:

$$R_1(h, A^*) = \frac{\sum_{i=1}^n p_i \sum_{j=1}^p \lambda_{ij}^h (v_{ij} w_{ij}')_{11}}{\sum_{j=1}^p (\lambda_h^*)^h (v_j^* w_j^{*\prime})_{11}} \tag{62}$$

Similar to before, *regularity implies that*

$$\lambda_{kj}^h (v_{kj} w_{kj}')_{11} = o(r^h) \tag{63}$$

for all $k \neq i$ with i being the matrix whose spectral radius attains r. Then

$$R_1(h, A^*) = \frac{p_i \sum_{j=1}^p \lambda_{ij}^h (v_{ij} w_{ij}')_{11} + o(r^h)}{\sum_{j=1}^p (\lambda_j^*)^h (v_j^* w_j^{*\prime})_{11}} \tag{64}$$

Once again, we consider two cases. If r is attained by a real eigenvalue, $\lambda_{i1} = r > |\lambda_{i2}|$ and consequently,

$$\lambda_{ij}^h (v_{ij} w_{ij}')_{11} = o(r^h) \tag{65}$$

for all $j > 1$. If, moreover, the largest eigenvalue of A^* is real, too. Since it uniquely attains the spectral radius of A^*, we have

$$(\lambda_j^*)^h (v_j^* w_j^*)_{11} = o((\lambda_1^*)^h) \tag{66}$$

such that

$$R_1(h, A^*) = \frac{p_i r^h (v_{i1} w_{i1}')_{11} + o(r^h)}{(\lambda_1^*)^h (v_1^* w_1^{*\prime})_{11} + o((\lambda_1^*)^h)} = \frac{r^h}{(\lambda_1^*)^h} \frac{p_i (v_{i1} w_{i1}')_{11} + o(1)}{(v_1^* w_1^{*\prime})_{11} + o(1)} = \frac{r^h}{(\lambda_1^*)^h} O(1) \tag{67}$$

It follows immediately that $\lim_{h\to\infty} \frac{1}{h} \ln R(h, A^*) = \ln \frac{r}{\lambda_{1^*}}$. If r is attained by two complex conjugates with $\lambda_{i1} = re^{i\theta}$ and $\lambda_{i2} = re^{-i\theta}$, we also write for A^*: $\lambda_1^* = r^* e^{i\theta^*}$ and $\lambda_2^* = r^* e^{-i\theta^*}$. Thus,

$$\lambda_{ij}^h (v_{ij} w_{ij}')_{11} = o(r^h) \tag{68}$$

for $j > 2$ and

$$(\lambda_j^*)^h (v_j^* w_j^{*\prime})_{11} = o((r^*)^h) \tag{69}$$

With this in mind, we can define $m = |(v_{i1} w_{i1}')_{11}|$ and $\phi = \arg\{(v_{i1} w_{i1}')_{11}\}$ and the respective analogs for the true model A^* so that

$$\begin{aligned} R_1(h, A^*) &= p_i \frac{r^h e^{i\theta h} (v_{i1} w_{i1}')_{11} + r^h e^{-i\theta h} (\overline{v}_{i1} \overline{w}_{i1}')_{11} + o(r^h)}{(r^*)^h e^{i\theta^* h} (v_1^* w_1^{*\prime})_{11} + (r^*)^h e^{-i\theta^* h} (\overline{v}_{i1}^* \overline{w}_{i1}^{*\prime})_{11} + o((r^*)^h)} \\ &= p_i \frac{r^h}{(r^*)^h} \frac{e^{i(\theta h + \phi)} m + e^{-i(\theta h + \phi)} m + o(1)}{e^{i(\theta^* h + \phi^*)} m^* + e^{-i(\theta^* h + \phi^*)} m^* + o(1)} \\ &= p_i \frac{r^h}{(r^*)^h} \frac{m}{m^*} \frac{\cos(\theta h + \phi) + o(1)}{\cos(\theta^* h + \phi^*) + o(1)} \end{aligned} \tag{70}$$

which concludes the proof of the second statement and therefore, the proof of Theorem 2. □

A.4 Proof of Claims in Remark 2

First, we will briefly explain how to rewrite the VARMA(q, m) conveniently. Consider a VARMA(q, m) process y_t represented as

$$y_t = \sum_{j=1}^{q} \Phi_j y_{t-j} + \sum_{j=0}^{m} \tilde{\Sigma}_j \nu_{t-j} \tag{71}$$

where y_t and ν_{t-j} are n-dimensional vectors and Φ_j and $\tilde{\Sigma}_j$ $n \times n$ matrices. Stacking these objects allows us to define

$$X_t = \begin{pmatrix} y_t \\ y_{t-1} \\ ... \\ y_{t-q+1} \end{pmatrix} \quad A = \begin{pmatrix} \Phi_1 & \Phi_2 & ... & \Phi_q \\ I_n & 0 & ... & 0 \\ 0 & I_n & ... & 0 \\ 0 & ... & & 0 \end{pmatrix} \quad B = \begin{pmatrix} \tilde{\Sigma}_0 & \tilde{\Sigma}_1 & ... & \tilde{\Sigma}_m \\ 0 & 0 & ... & 0 \\ & ... & & \\ 0 & 0 & ... & 0 \end{pmatrix} \tag{72}$$

where X_t is an $(q \cdot n) \times 1$ dimensional vector, A a $(n \cdot q) \times (n \cdot q)$ dimensional matrix, and B a matrix. Moreover, it is helpful to define

$$\tilde{\nu}_t = \begin{pmatrix} \nu_t \\ \nu_{t-1} \\ ... \\ \nu_{t-m} \end{pmatrix} \quad F = \begin{pmatrix} 0 & 0 & ... & 0 \\ I & 0 & ... & 0 \\ 0 & I & ... & 0 \\ ... & & & \end{pmatrix} \tag{73}$$

The VARMA(q, m) process can then be represented as

$$X_t = AX_{t-1} + B\tilde{\nu}_t \tag{74}$$

with the vector of stacked errors, $\tilde{\nu}_t$ following

$$\tilde{\nu}_t = F\tilde{\nu}_{t-1} + \begin{pmatrix} \nu_t \\ 0 \\ ... \\ 0 \end{pmatrix} \tag{75}$$

This representation allows us to write forecasts $X_{t+h|t}$ in the following way

$$X_{t+h|t} = A^h X_t + \left(\sum_{k=1}^{h} A^{h-k} B F^k \right) \tilde{\nu}_t = A^h X_t + \left(\sum_{k=1}^{m} A^{h-k} B F^k \right) \tilde{\nu}_t \tag{76}$$

where the second equality follows from the fact that $F^{m+1} = 0$. This implies the following expression for the forecast persistence functions $P_n(h)$:

$$P_n(h) = \frac{e_n' \mathbb{E}_F[A^h] X_t + \left(\sum_{k=1}^{m} \mathbb{E}_F[e_n' A^{h-k} B F^k] \right) \tilde{\nu}_t}{e_n' \mathbb{E}_F[A^{h-1}] X_t + \left(\sum_{k=1}^{m} \mathbb{E}_F[e_n' A^{h-k-1} B F^k] \right) \tilde{\nu}_t} \tag{77}$$

Following the same arguments as in Theorem 1, we can express the forecast persistence functions $P_n(h)$ as

$$P(h) = r\frac{e^{i\theta h}me^{i\vartheta} + e^{-i\theta h}me^{-i\vartheta} + \sum_{k=1}^{m}\left[e^{i\theta(h-k)}m_k e^{i\vartheta_k} + e^{-i\theta(h-k-1)}m_k e^{-i\vartheta_k}\right] + o(1)}{e^{i\theta(h-1)}me^{i\vartheta} + e^{-i\theta(h-1)}me^{-i\vartheta} + \sum_{k=1}^{m}\left[e^{i\theta(h-k-1)}m_k e^{i\vartheta_k} + e^{-i\theta(h-k-1)}m_k e^{-i\vartheta_k}\right] + o(1)} \tag{78}$$

where

$$\begin{aligned} m &= |e_n' v_{i1} w_{i1}' X_t| & \vartheta &= \arg(e_n' v_{i1} w_{i1}' X_t) \\ m_k &= |e_n' v_{i1} w_{i1}' \mathbb{E}[B|A_i] F^k \tilde{\nu}_t| & \vartheta_k &= \arg(e_n' v_{i1} w_{i1}' \mathbb{E}[B|A_i] F^k \tilde{\nu}_t) \end{aligned} \tag{79}$$

Note that A_i denotes the matrix whose largest eigenvalue attains r. This further simplifies to

$$P_n(h) = r\frac{m\cos(\theta h + \vartheta) + \sum_{k=1}^{m} m_k \cos(\theta(h-k) + \vartheta_k) + o(1)}{m\cos(\theta(h-1) + \vartheta) + \sum_{k=1}^{m} m_k \cos(\theta(h-k-1) + \vartheta_k) + o(1)} \tag{80}$$

Clearly, if r is attained by a real eigenvalue, $\theta = 0$, and $P(h) = r$. A close to identical derivation can be applied to yield the formula for the forecast reaction function.

A.5 Proof of Proposition 2

Proof. We guess and verify the result. We have by definition that:

$$x_{t+h} - \mathbb{E}_t[x_{t+h}] = (\rho^*)^h x_t + \sum_{j=0}^{h-1}(\rho^*)^j \epsilon_{t+h-j} - \mathbb{E}_F[\rho^h]x_t \tag{81}$$

Moreover, by definition we have that:

$$\beta_1^h \cdot (\mathbb{E}_t[x_{t+h}] - \mathbb{E}_{t-1}[x_{t+h}]) + \beta_2^h \cdot \mathbb{E}_{t-1}[x_{t+h}] + \nu_t = \beta_1^h \mathbb{E}_F[\rho^h]x_t + (\beta_2^h - \beta_1^h)\mathbb{E}_F[\rho^{h+1}]x_{t-1} + \nu_t \tag{82}$$

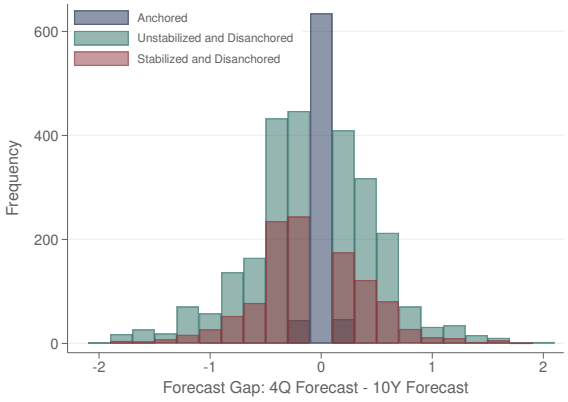
Thus, by matching coefficients, in order for Equation 33 to hold, we require that:

$$\beta_1^h \mathbb{E}_F[\rho^h] = (\rho^*)^h - \mathbb{E}_F[\rho^h] \implies \beta_1^h = \frac{(\rho^*)^h}{\mathbb{E}_F[\rho^h]} - 1 = \frac{1 - R(h;\rho^*)}{R(h;\rho^*)} \tag{83}$$

along with $\beta_1^h = \beta_2^h$ and $\nu_t = \sum_{j=0}^{h-1}(\rho^*)^j \epsilon_{t+h-j}$. Which verifies the guess and proves the claimed result for β_1^h. Moreover, as $\sum_{j=0}^{h-1}(\rho^*)^j \epsilon_{t+h-j}$ is orthogonal to x_t and x_{t-1}, the claimed orthogonality conditions hold. The claims for β_1^h then follow from Proposition 1. □

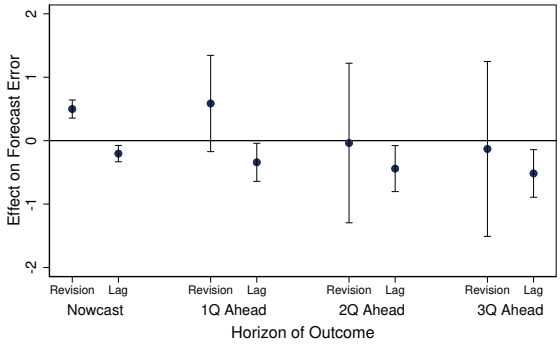
B Additional Figures

Figure 14: The Distribution of Forecast Gaps



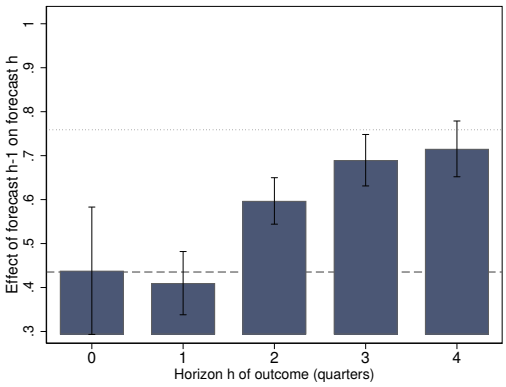
Note: This figure shows a frequency histogram of forecast gaps, defined as the individual four-quarter-ahead forecast of CPI minus the individual 10-year average forecast. We split the distribution into three exhaustive categories: forecasts that are anchored (within 0.1% points of the 10-year average; blue), forecasts that disanchored and unstabilized (not within 0.1% points of the 10-year average and not within 0.1% points of the three-quarter forecast), and those that are disanchored but stabilized (not within 0.1% points of the 10-year average but within 0.1% points of the three-quarter forecast).

Figure 15: Under- and Over-reaction in CPI Forecasts

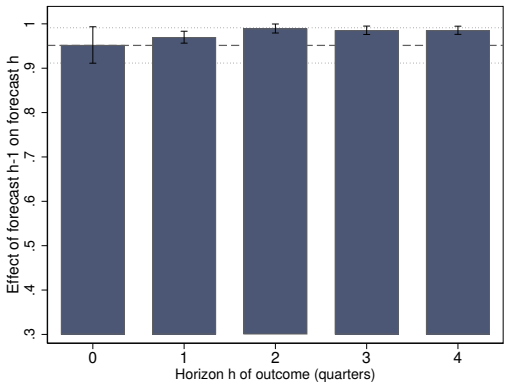


Note: This figure shows estimates of Equation 33 corresponding to forecasts of CPI inflation, replicating the general anaylsis of Halperin and Mazlish (2025) for one variable. The sample covers 1981:Q1 to 2026:Q1. Each pair of estimates corresponds to a separate regression model where the outcome is the aggregate CPI forecast error at that horizon, where consensus forecasts are the median and observed data are final-release numbers. The estimates correspond respectively to β_1 ("Revision") and β_2 ("Lag") in Equation 33. Error bars are 95% confidence intervals based on HAC-corrected standard errors with a bandwidth of 4 quarters.

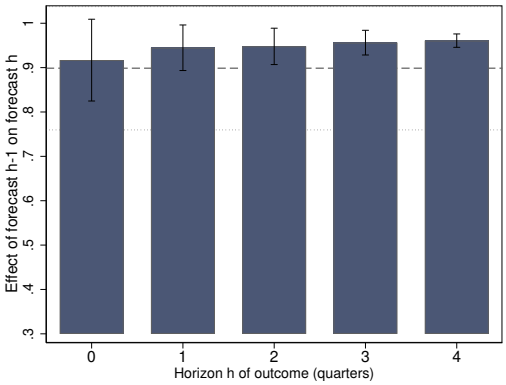
Figure 16: The Term Structure of Forecasts for Other Variables



(a) PCE Inflation



(b) 3-Month Treasury Rates



(c) Unemployment

Note: This figure shows estimates of Equation 35 corresponding to forecasts of PCE inflation, three-month Teasury rates, and the unemployment rate. Each estimate corresponds to the coefficient β from a separate regression at the indicated horizon for the outcome variable. Note that the $h-1$ "forecast" used as the regressor for $h=0$ is first-release data reported within the SPF questionnaire and therefore visible to respondents. The dashed horizontal line is the estimated persistence of each variable in the same sample, and the dotted horizontal lines are the upper and lower limits of a 95% confidence interval based on HAC robust standard errors with a 4-quarter bandwidth. Error bars on the main estimates are 95% confidence intervals based on standard errors double clustered by forecaster ID and quarter.

References

Afrouzi, H., Kwon, S. Y., Landier, A., Ma, Y., and Thesmar, D. (2023). Overreaction in expectations: Evidence and theory. *The Quarterly Journal of Economics*, 138(3):1713–1764.

Angeletos, G.-M., Huo, Z., and Sastry, K. A. (2021). Imperfect macroeconomic expectations: Evidence and theory. *NBER Macroeconomics Annual*, 35(1):1–86.

Antràs, P. and Kulesza, A. (2026). An 'Austrian' Model of International Specialization. Working Paper 34983, National Bureau of Economic Research.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, 25(1):2–11.

Barsky, R. B. and Miron, J. A. (1989). The seasonal cycle and the business cycle. *Journal of Political Economy*, 97(3):503–534.

Blanchard, O. J. and Kahn, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica*, pages 1305–1311.

Böhm-Bawerk, E. (1891). *The Positive Theory of Capital*. Macmillan and Co., London.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–2782.

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

Del Negro, M. and Schorfheide, F. (2004). Priors from general equilibrium models for vars. *International Economic Review*, 45(2):643–673.

Dembo, A. and Zeitouni, O. (2009). *Large deviations techniques and applications*, volume 38. Springer Science & Business Media.

Eden, M., Samuelson, L., and Steiner, J. (2026). The importance of unlikely events. *Working Paper*.

Eusepi, S. and Preston, B. (2011). Expectations, learning, and business cycle fluctuations. *American Economic Review*, 101(6):2844–72.

Farhi, E. and Werning, I. (2019). Monetary policy, bounded rationality, and incomplete markets. *American Economic Review*, 109(11):3887–3928.

Farmer, L. E., Nakamura, E., and Steinsson, J. (2024). Learning about the long run. *Journal of Political Economy*, 132(10):3334–3377.

Gabaix, X. (2020). A behavioral new keynesian model. *American Economic Review*, 110(8):2271–2327.

García-Schmidt, M. and Woodford, M. (2019). Are low interest rates deflationary? a paradox of perfect-foresight analysis. *American Economic Review*, 109(1):86–120.

Goldstein, N. and Gorodnichenko, Y. (2022). Expectations formation and forward information. Technical report, National Bureau of Economic Research.

Halperin, B. and Mazlish, J. Z. (2025). Overreaction and forecast horizon: longer-term expectations overreact more, shorter-term expectations drive fluctuations.

Hayek, F. A. (1931). *Prices and Production*. George Routledge & Sons, Ltd., London.

Kohlhas, A. N. and Robertson, D. (2025). Cautious expectations. *Journal of Monetary Economics*, page 103759.

Kohlhas, A. N. and Walther, A. (2021). Asymmetric attention. *American Economic Review*, 111(9):2879–2925.

Marcet, A. and Sargent, T. J. (1989). Convergence of least-squares learning in environments with hidden state variables and private information. *Journal of Political Economy*, 97(6):1306–1322.

Martin, I. (2012). On the valuation of long-dated assets. *Journal of Political Economy*, 120(2):346–358.

Molavi, P. (2022). Simple models and biased forecasts. *arXiv preprint arXiv:2202.06921*.

Phillips, P. C. (1991). To criticize the critics: An objective bayesian analysis of stochastic trends. *Journal of Applied Econometrics*, 6(4):333–364.

Powell, J. H. (2021). Monetary Policy in the Time of COVID. Speech at the Federal Reserve Bank of Kansas City Economic Policy Symposium, Jackson Hole, Wyoming. Accessed: 2026-03-25.

Samuelson, L. and Steiner, J. (2025). Growth and redistribution: The hedging perspective. *American Economic Review: Insights*, 7(2):250–267.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, pages 1–48.

Sims, C. A. (1988). Bayesian skepticism on unit root econometrics. *Journal of Economic dynamics and Control*, 12(2-3):463–474.

Sims, C. A. and Uhlig, H. (1991). Understanding unit rooters: A helicopter tour. *Econometrica: Journal of the Econometric Society*, pages 1591–1599.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(3):586–606.

Stock, J. H. and Watson, M. W. (2009). Phillips curve inflation forecasts. *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*, page 101.

The Economist (2024). Has "Team Transitory" really won America's inflation debate? Published: 2024-01-10. Accessed: 2026-03-26.

Weitzman, M. L. (1998). Why the far-distant future should be discounted at its lowest possible rate. *Journal of Environmental Economics and Management*, 36(3):201–208.

Weitzman, M. L. (2001). Gamma discounting. *American Economic Review*, 91(1):260–271.

Wu, L. (2023). Mental macro-finance models: Evidence and theory. Working Paper 4332258, SSRN.