Forecasting with Uncertain Persistence

Joel P. Flynn* Maksim Meinert† Karthik A. Sastry‡
Yale Yale Princeton

June 26, 2026

Abstract

With uncertainty about persistence, we show that forecasts *necessarily* become more persistent and over-react at long horizons. For these reasons, correctly specified and Bayesian forecasts may under-react at short horizons and over-react at long horizons. These results provide a unified explanation for several asset pricing and forecasting puzzles, including: (i) the excess responsiveness of long-horizon rates to short rates, (ii) the dominance of apparent term premia for long-term rates, (iii) the *ex post* predictability of bond yields, (iv) the excess volatility of long-horizon forward prices, (v) the excess persistence of long-horizon forecasts, and (vi) the over-reaction of long-horizon forecasts.

*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 Hassan Afrouzi, Andrew Atkeson, Jaroslav Borovĉka, Ryan Chahrour, Tim Christensen, Stefano Eusepi, Basil Halperin, Christian Hellwig, Alexandre Kohlhas, Guido Menzio, Pooya Molavi, Kris Nimark, Larry Samuelson, Jakub Steiner, and seminar participants at the Columbia Measured Beliefs Conference, the Oxford-Paris Conference on the Macroeconomics of Expectations, the Northwestern Kellogg Learning and Bounded Rationality Conference, the T2M Conference, and the Macro Theory Group for helpful comments. First posted version: June 26, 2026.

1 Introduction

How persistent are key economic variables like inflation, unemployment, or interest rates? No one 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). Significant uncertainty remains about the best 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). For interest rates, despite the importance of these forecasts for evaluating long-run investments (Weitzman, 1998; Adrian et al., 2015), there is still relatively little consensus on the relative merits of various approaches (Piazzesi, 2010; Adrian et al., 2016).

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 himself highlighted this 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, Jackson Hole Conference, August 2021

Motivated by the centrality of uncertain persistence for economic forecasting, this paper explores the implications of uncertainty about persistence for the structure of Bayesian and correctly specified forecasts. We find that, for general reasons that are not wedded to specific probability distributions or the presence (or absence) of learning, uncertain persistence *necessarily* generates: (i) forecasts that are more persistent at longer horizons and (ii) forecasts that over-react at long horizons. We show that these findings provide a unified explanation for seemingly disparate findings in finance and macroeconomics, including: (i) the excess responsiveness of long-run interest rates to short rates (Gürkaynak et al., 2005; Hanson and Stein, 2015), (ii) the dominance of apparent term premia at the long-end of the yield curve (Adrian et al., 2015), (iii) *ex post* predictable excess returns in bond yields (Fama and Bliss, 1987; Cochrane and Piazzesi, 2005), (iv) the excess volatility of long-horizon forward prices (Shiller, 1979; Giglio and Kelly, 2018), (v) horizon-dependent persistence in the term structure of survey forecasts (Goldstein and Gorodnichenko, 2022), and (vi) simultaneous under- and over-reaction in these forecasts (Coibion and Gorodnichenko, 2015; Angeletos et al., 2021; Bordalo et al., 2020; Kohlhas and Walther, 2021; Halperin and Mazlish, 2025).

The AR(1) Theory. We first illustrate the general principles of our analysis in the simplest possible setting: forecasting an AR(1) process. A forecaster observes the current value of a stochastic process x_t and believes that its persistence ρ is drawn from a non-degenerate probability distribution F.

We first describe the *term structure of forecasts*: the forecasts made at date t about future periods $t+h$. To do this, we define the *forecast persistence function* which, at each horizon $h \geq 1$, is the ratio of the h and $h-1$ horizon forecasts: $P(h) = \mathbb{E}_t[x_{t+h}]/\mathbb{E}_t[x_{t+h-1}]$. Our first main result is that forecast persistence strictly increases in horizon and that the limiting forecast persistence as the horizon increases is the *maximum believable persistence*, *i.e.,* the upper support point of the distribution F. We show that this result is a simple consequence of concepts from large deviations theory (see *e.g.,* Dembo and Zeitouni, 2009).

We next study the implications for over- and under-reaction relative to the ground truth. 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 are on average. That is, forecasters always asymptotically *over-react*, precisely because the maximum believable persistence determines the behavior of long-term forecasts. Moreover, long-horizon over-reaction can coexist with short-horizon *under*-reaction.

Economically, these results imply that uncertain persistence *must* generate forecasts that are more persistent at longer horizons and over-react at sufficiently long horizons. These results are robust in the sense that they apply for arbitrary beliefs about persistence and thus hold regardless of the underlying learning process by which agents form these beliefs. Moreover, they obtain for forecasters who follow Bayes' rule and are correctly specified in the sense that they do not dogmatically exclude the correct model of the world.

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 first-order vector autoregression (VAR) process with coefficient matrix A, and the forecaster is uncertain about this coefficient matrix. This allows for both uncertainty about persistence and cross-variable dependency. 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 these restrictions are generic and thus without economic substance.

We show that our analysis of this formulation can encompass all vector autoregression and moving average (VARMA) models up to finite order, as well as models with unobserved states. In this way, our analysis applies to five classes of models that are commonly used to make forecasts in practice and/or describe how economic agents make forecasts within models: (i) univariate ARMA models, (ii) multivariate VARs (Sims, 1980; Del Negro and Schorfheide, 2004), (iii) linear rational-expectations models (Blanchard and Kahn, 1980) including standard dynamic stochastic general equilibrium (DSGE) models (*e.g.,* Smets and Wouters, 2007), (iv) models in which the forecaster sees signals of the state (Coibion and Gorodnichenko, 2015) or must infer a hidden state (Hamilton, 1989; Crump et al., 2023), and (v) models with an uncertain long-run mean (Bauer and Rudebusch, 2020; Afrouzi et al., 2023; Farmer et al., 2024). Thus, our analysis applies to forecasters who have incomplete information about the state of the world and/or understand that economic variables have properties that cannot be described by an AR(1), such as "hump-shaped" dynamics, periodic cycles, or multi-variate relationships.

We show that the primary insights from the AR(1) analysis extend to forecasting with this larger class of models, and even richer phenomena are possible. 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 common eigenvalue, which takes on the role of the maximum believable persistence from the AR(1) analysis. Moreover, the single most-persistent factor of any possible model determines the cross-variable behavior of long-horizon forecasts, and forecasts for all variables asymptotically over-react.

If instead a complex pair of eigenvalues has the largest modulus, then the forecast persistence'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 believable, but this is in terms of the modulus of a stable cycle. In this case, forecasts oscillate between over- 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) and, for such variables, this cyclical asymptotic regime may be descriptive.

Rationalizing Asset Pricing and Forecasting Puzzles. We next use our model to provide a unified explanation for several puzzles in asset pricing and macroeconomics. While each of these puzzles has a large number of competing explanations, we show that the simple premise that persistence is uncertain necessarily implies these phenomena.

We first study a set of puzzles regarding the term structure of asset prices. First, long-horizon forward interest rates closely co-move with short-horizon interest rates (Hanson and Stein, 2015). Second, the vast majority of variation in these long-horizon rates is typically attributed to term premia rather than expected future short rates (Adrian et al., 2015). Third, there are predictable excess returns in bonds (Fama and Bliss, 1987; Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005). And fourth, for both bonds and other assets, there is excessive volatility in the prices of long-horizon claims that is inconsistent with standard models (Giglio and Kelly, 2018).

We show that all four of these findings are *necessary* outcomes of an arbitrage-free linear pricing model in which the persistence of the short interest rate (or another relevant factor) is stochastic. In such a model, forward rates correspond to the term structure of forecasts with uncertain persistence under the risk-neutral measure. Given this mapping, the four puzzling phenomena arise as corollaries of our main theoretical findings that, at long horizons, forecasts become more persistent and over-react. The former explains why long-horizon forward rates co-vary almost perfectly with one another and have excess volatility. The latter explains why term premia dominate the long-end of the yield curve and bond returns are *ex post* predictable. We emphasize that all of these results are derived within an arbitrage-free pricing model in which we have imposed an *exact* subjective expectations hypothesis: while an econometrician judges that there are forecastable excess returns *ex post*, agents in the model would not think that there are any expected excess returns.

Moreover, modest and empirically plausible levels of uncertainty about persistence suffice to explain these phenomena quantitatively. A two-parameter model with uncertain persistence, one more than under known persistence, can almost perfectly rationalize the effects of short rates on forward rates of different horizons in the data. Further, this model implies that term premia account for 45% and 84% of variation in 15-year and 30-year forward rates, respectively. In addition, the same model rationalizes why the variance ratio statistics from Giglio and Kelly (2018) exceed one and increase in horizon. All of this is despite the fact that, under our model, market participants are certain that short rates are stationary, in contrast to explanations based on permanent shocks (*e.g.,* Bauer and Rudebusch, 2020). The implied uncertainty about persistence necessary to match the data is also empirically plausible: it is similar to the uncertainty implied by an *ex post* statistical estimation of the persistence of short interest rates over the whole sample, which itself should *under*-estimate the uncertainty forecasters face in real time.

Finally, we use our model to understand two puzzles regarding the term structure of macroeconomic forecasts in surveys. First, as documented by Goldstein and Gorodnichenko (2022), forecasts become more persistent as the forecasting horizon increases. Second, as documented by Halperin and Mazlish (2025), forecasts over-react at long horizons while also under-reacting at short horizons. We show formally how these results are necessary implications of the general theory. Beyond the intrinsic interest of macroeconomic forecasts (see, *e.g.,* Angeletos et al., 2021), this analysis allows us to validate the economic mechanism underlying our rationalizations of the asset pricing facts: forecasts do, in fact, have the properties implied by our model with uncertain persistence.

Related Literature. We start from a premise—uncertainty about the stochastic process for economic variables—that is familiar to policymakers, forecasters, and economists. 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 automatically robust over a large class of learning rules and stochastic processes for economic variables, provided that forecasters entertain the possibility that many models could be true.

Our research also relates to studies that postulate that economic agents use mis-specified dynamic models. Some focus on under-reaction: Gabaix (2020) considers a model of *cognitive discounting* in which agents under-estimate persistence, and García-Schmidt and Woodford (2019) and Farhi and Werning (2019) study models of iterative reasoning that imply that agents under-estimate the persistence of the economy's response to shocks. Another set of approaches focuses on over-reaction: Angeletos et al. (2021) show that a model in which agents over-estimate the persistence of macroeconomic variables can match certain patterns in forecast data; Molavi (2022) argues that using "over-persistent" models can arise from long-run, mis-specified learning; Bordalo et al. (2020) and Afrouzi et al. (2023) study models of distorted information processing in which agents over-react to recent information; and Fuster et al. (2010) study agents who irrationally extrapolate because they place weight on an intuitive model with excessive 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 these contrasting perspectives.

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 been fruitfully 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 economic 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 shows how these results rationalize important asset pricing puzzles. Section 5 shows how our results rationalize important forecasting puzzles. Section 6 concludes.

2 Forecasting an AR(1) Process with Uncertainty

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

2.1 The Environment

Time is discrete $t \in \mathbb{N}$. A forecaster believes that a variable $x_t \in \mathbb{R}$ (*e.g.,* inflation, the unemployment rate, or the nominal interest rate) follows an AR(1) process:

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

where ε_t is a mean-zero shock that is mean-independent across time.[1] The forecaster is uncertain about the persistence parameter ρ, believing that it is drawn from some non-degenerate distribution F with a support that is contained in $(0, r]$, where $r > 0$ is the essential supremum of F. Intuitively, r is the largest believable persistence that the forecaster entertains. Our qualitative conclusions do not depend on the precise value of r, including whether it is less than or greater than one. Moreover, one should interpret F and r as describing beliefs at time t; these beliefs could of course change if the agent were learning. We are concerned with the properties of forecasts made at a given time t, so the exact process by which learning takes place is immaterial for our conclusions.

The forecaster observes the realization x_t in each period t and forms conditional expectations for future values. We denote the forecast of x_{t+h} given observation of x_t as:

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

where the expectations operator takes into account uncertainty about both the unknown persistence ρ and the future shocks. Our object of interest will be the *term structure of forecasts*, or the sequence of h-period ahead forecasts from date t: $(x_{t+h|t})_{h=0}^{\infty}$. From a purely statistical perspective, these forecasts minimize the agent's expected mean squared error (MSE). In economic applications, the term structure of forecasts emerges as a relevant object when studying the forward-looking decisions of (subjective) expected utility agents or the prices of financial assets that are consistent with no-arbitrage (see, *e.g.*, Section 4).

The true data-generating process of x_t is of no consequence to study only the properties of forecasts, 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 $\rho = \rho^* \in (0, r)$.

[1] Since our theory concerns mean forecasts, the exact distribution of ε_t is immaterial. In particular, this distribution could vary with calendar time or the state.

When ρ^* is in the support of F, the forecaster considers the correct model, and therefore 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

Our first main result describes how forecasts at longer horizons relate to forecasts at shorter horizons. Toward this, we define the *forecast persistence* function $P : \mathbb{N} \to \mathbb{R}$ as the following ratio of adjacent forecasts:

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

Combined with the base forecast $x_{t|t} = x_t$, this function describes the entire term structure $(x_{t+h|t})_{h=0}^{\infty}$. Given the AR(1) structure, the forecast persistence function is:

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

In the standard case where F is a point mass at some known ρ, we have that $P(h) = \rho$: the persistence of forecasts across horizons is constant and equals 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:*

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

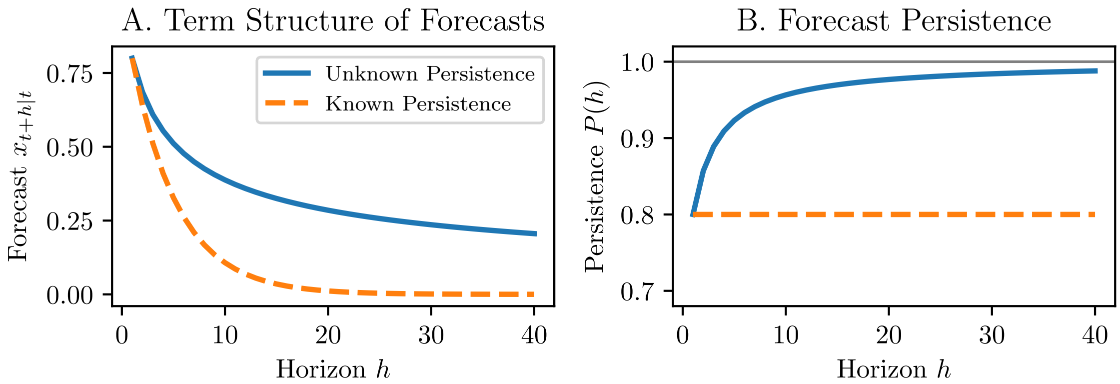
2. *Forecast persistence, P, is a strictly increasing function of the horizon*
3. *The asymptotic forecast persistence is the maximum believable persistence:*

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

Proof. See Appendix A.1. □

Figure 1 visualizes this result through a specific example that contrasts two cases: one in which ρ follows a Beta distribution with $\alpha = 2$ and $\beta = 0.5$, supported on $[0, 1]$, and one in which ρ is known and equal to the mean from that distribution, 0.8. Panel A shows

Figure 1: The Impact of Uncertainty on Forecast Persistence



Note: This figure visualizes the effects of uncertain persistence on the persistence of forecasts across horizons. In both panels, the blue solid line is a case where $\rho \sim \text{Beta}(2, 0.5)$ and the orange dashed line is a case where ρ is known to be 0.8, the mean of the Beta distribution. Panel A shows $x_{t+h|t} = \mathbb{E}_t[x_{t+h}|x_t]$ for $h \geq 1$, with $x_t = 1$. Panel B shows the forecast persistence function defined in Equation 4, $P(h) = x_{t+h|t}/x_{t+h-1|t}$. The gray solid line at $P = r = 1$ is the limiting value of $P(h)$ under the Beta distribution.

$x_{t+h|t}$, conditional on $x_t = 1$. Unknown persistence causes the term structure of forecasts to flatten out. Panel B shows the forecast persistence function. With unknown persistence, this function starts at the mean persistence, increases, and asymptotes to one. All of these properties can also be verified directly via a simple, closed-form calculation of the persistence function under the Beta distribution:

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

Because of this analytical tractability, we will return to the Beta distribution later in the analysis (see Section 4).

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. That is, the forecaster believes that x_t is persistent ($\rho = \bar{\rho}$) with probability p and transitory ($\rho = \underline{\rho} < \bar{\rho}$) with probability $1 - p$. By the law of iterated expectations, the horizon h forecast follows:

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

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

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

As the horizon h increases, the forecast becomes dominated by the persistent model in the following sense: the ratio of the persistent forecast to the transitory forecast tends to infinity. This drives the key properties of the persistence function derived in Theorem 1. To see this, we calculate

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

where in the second equality, we divide both the numerator and the denominator by $\bar{\rho}^{h-1}$. 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.

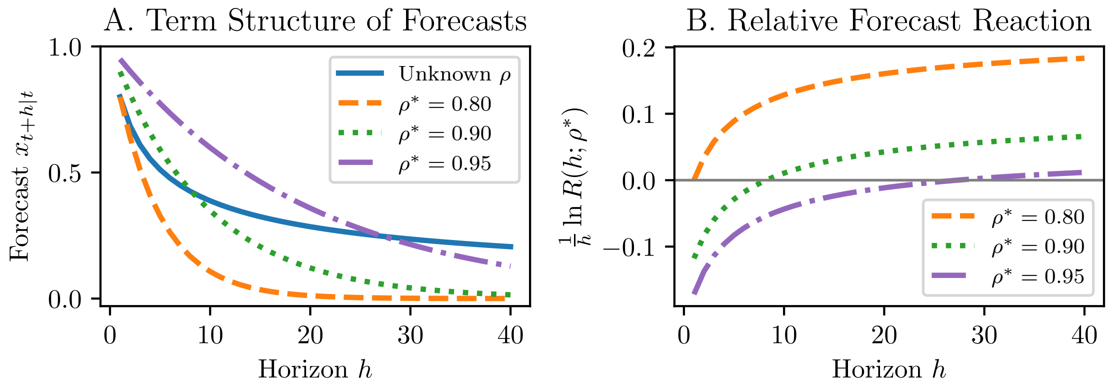
2.3 Over-Reaction and Under-Reaction

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 $R : \mathbb{N} \times (0, r) \to \mathbb{R}$:

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

This function describes the gap between the forecast by the uncertain agent 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 $R(h;\rho^*) < 1$, they under-react. The following result characterizes the conditions under which under-reaction and over-reaction occur at various time horizons:

Figure 2: Uncertainty Can Lead to Under-Reaction and then Over-Reaction of Forecasts



Note: This figure illustrates under- and over-reaction when persistence is unknown. Panel A shows the forecast when $\rho \sim \text{Beta}(2, 0.5)$ (solid blue line) and average outcomes under various possible ρ^* (other lines). Panel B shows the (transformed) relative reaction function $\frac{1}{h} \ln R(h; \rho^*)$ under the indicated values of ρ^*.

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 $h < h^*$ and over-react for $h \geq h^*$.*

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 □

To visualize this result, we return again to the case where $\rho \sim \text{Beta}(\alpha, \beta)$. Figure 2 shows the agent's forecast versus various possible ground truths (Panel A) and a transformation of the reaction function (Panel B). Even for very large ρ^*, the ground truth forecast eventually crosses the forecast with unknown persistence. For this reason, the (log) reaction function is eventually positive, corresponding to over-reaction at long enough horizons. For values of ρ^* that are higher than the mean of the unknown distribution, the ground truth crosses the forecast from above and the log reaction function crosses zero from below. That is, short-horizon forecasts under-react and long-horizon forecasts over-react.

Intuitively, as the horizon increases, the mere possibility—no matter how unlikely—that persistence is greater than the true persistence dominates the behavior of forecasts. As before, this is easiest to see in the case where the forecaster considers two models: $\rho = \bar{\rho}$ with probability p and $\rho = \underline{\rho}$ with probability $1-p$. In this case,

$$R(h;\rho^*) = p\left(\frac{\bar{\rho}}{\rho^*}\right)^h + (1-p)\left(\frac{\underline{\rho}}{\rho^*}\right)^h \tag{14}$$

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

$$p\frac{\bar{\rho}}{\rho^*} + (1-p)\frac{\underline{\rho}}{\rho^*} < 1 \implies \mathbb{E}[\rho] = p\bar{\rho} + (1-p)\underline{\rho} < \rho^* \tag{15}$$

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:

$$R(h;\rho^*) = \left(\frac{\bar{\rho}}{\rho^*}\right)^h \left[p + (1-p)\left(\frac{\underline{\rho}}{\bar{\rho}}\right)^h\right] \implies \ln R(h;\rho^*) \sim h \ln\left(\frac{\bar{\rho}}{\rho^*}\right) \tag{16}$$

As h increases, the relative reaction is determined only by the persistent model.

The proof for Proposition 1 observes that dominance of the most persistent model is a more general feature under uncertainty, just as in the proof of Theorem 1. Thus, perhaps surprisingly, the over-reaction of sufficiently long-horizon forecasts is a general property for any true persistence less than the essential supremum. Intuitively, even if it is unlikely that a stochastic process is more persistent than some value $\rho^* < r$, it is possible. This possibility dominates a Bayesian forecaster's long-horizon forecasts, and so the forecaster rationally over-reacts at these horizons. Moreover, there is no contradiction with under-reaction at shorter horizons, since this behavior is governed by average beliefs.

The upshot is that forecasters who are Bayesian, correctly specified, and in possession of complete knowledge of the state can nevertheless under-react and over-react to knowledge of that state. These properties distinguish our explanation of under- and over-reaction from others that presume that agents are mis-specified in the sense that they consider a single, incorrect model (*e.g.*, Afrouzi et al., 2023; Molavi, 2022; Fuster et al., 2010), lack complete knowledge of the state (*e.g.*, Coibion and Gorodnichenko, 2015; Angeletos et al., 2021), and/or perform inference in a way inconsistent with Bayes' rule (*e.g.*, Bordalo et al., 2020; d'Arienzo, 2020).

3 The General Structure of Forecasts with Uncertainty

In this section, we provide the main theoretical result of the paper: under weak regularity conditions on the forecaster's prior, the properties of forecasts that we uncovered for AR(1) processes extend to general finite-order vector autoregressive moving average (VARMA) processes. In these cases, when dominant eigenvalues are real, the persistence of forecasts converges to the maximum believable persistence and forecasts always asymptotically overreact. When dominant eigenvalues are complex, forecasts asymptotically cycle with an amplitude governed by the maximum believable persistence.

3.1 The Environment

We focus in this section on first-order vector autoregression (VAR) models because, as we will show momentarily, analysis of these processes suffices to establish the properties of interest in a much larger class of models. A stochastic process $X_t \in \mathbb{R}^p$ follows a VAR(1) model, if it can be represented as

$$X_t = AX_{t-1} + \varepsilon_t \tag{17}$$

where A is the $p \times p$ coefficient matrix of the VAR process and the shocks ε_t are mean-zero and mean-independent across time.

Once again, we consider a scenario in which a forecaster is uncertain about the true process 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 require 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 and left 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 positive real eigenvalue has modulus r or a single complex conjugate pair of eigenvalues with positive real component has modulus r.*

Regularity encodes three technical properties, none of which we regard as economically restrictive. First, the prior is finitely supported. We note that finitely supported priors can approximate any countably or uncountably supported prior arbitrarily well. Second, the coefficient matrix is diagonalizable with distinct eigenvalues with probability one. We note that diagonalizability is a generic property among matrices: that is, almost all matrices (viewed as elements of $\mathbb{R}^{p \times p}$) are diagonalizable. Third, the modulus of the largest eigenvalue is attained by at most a single complex conjugate pair for a single coefficient matrix. This is also a generic property.[2]

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 as $re^{\pm i\theta}$, where r is the modulus and θ is the argument. 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) models, VARMA(q,m) models, models with unobserved states, and models with uncertainty about long-run means. Doing so allows us to jointly address several classes of economic models that are commonly used to make forecasts in practice and to describe how economic agents make forecasts within models.

Remark 1 (Representing VAR(q) Models as VAR(1) Models)**.** Our analysis below will apply to any VAR(q) model. Such models are a benchmark for policy analysis and forecasting (Sims, 1980; Del Negro and Schorfheide, 2004). We leverage the well-known fact that we can rewrite 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

$$y_t = \sum_{j=1}^{q} \Phi_j y_{t-j} + \tilde{\Sigma}\nu_t \tag{18}$$

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 17 using the following transformation that "stacks" the finite dimensions of y_t (and lags) and ν_t (and lags).

$$X_t = \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-q+1} \end{pmatrix}, \qquad A = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{q-1} & \Phi_q \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \ddots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \tilde{\Sigma} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{19}$$

where A is commonly called the companion matrix. Importantly, for what follows, such

[2]The one potentially substantive restriction is that the largest eigenvalue has positive real component, *i.e.,* persistence is positive. Notwithstanding that this is likely the relevant case in almost all economic contexts, our arguments extend directly to negative persistence. We omit them in the interests of space.

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

$$y_t = \sum_{j=1}^{q} \Phi_j y_{t-j} + \sum_{j=0}^{m} \tilde{\Sigma}_j \nu_{t-j} \tag{20}$$

Under standard invertibility and stability conditions (*e.g.,* Blanchard and Kahn, 1980), standard dynamic stochastic general equilibrium frameworks have such a representation (*e.g.*, Smets and Wouters, 2007). Continue to let X_t denote the stacked vector of y_t and lags and to construct A as in Equation 19. Then, for any $h > m$, one can show that

$$\mathbb{E}_t[X_{t+h}] = \mathbb{E}_t\left[A^h\left(C_0 X_t + \sum_{j=0}^{m} B_j \nu_{t-j}\right)\right] \tag{21}$$

for some matrices C_0 and $(B_j)_{j=0}^{m}$ that depend on the primitive VARMA coefficients. Intuitively, after horizon m, the forecasts of the MA terms become zero and the initial MA terms affect the forecast only indirectly through the AR terms. 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 C.1). △

Remark 3 (Handling Models with Unknown or Hidden States)**.** Our analysis also applies essentially as written to settings in which the forecaster does not know the state of the process. This allows our analysis to apply to settings in which the forecaster observes signals about the state rather than observing the state itself (as in, *e.g.,* Coibion and Gorodnichenko, 2015) as well as settings in which hidden variables or regime shifts or unobserved transitory and persistent components drive dynamics (Hamilton, 1989; Crump et al., 2023). To see why this is the case, consider the simple AR(1) setting in which the agent does not know the state. In this case, we can write the forecast as:

$$x_{t+h|t} = \mathbb{E}_t[\rho^h x_t] = \mathbb{E}_t[\rho^h \mathbb{E}_t[x_t|\rho]] \tag{22}$$

All of our results then extend (see Appendix C.2). △

Remark 4 (Uncertainty About Long-Run Means)**.** Our main analysis considers VAR(1) models without uncertainty about long-run means. This is an expositional choice to clearly distinguish our points about persistence and cross-sectional dependency from existing points about the importance of uncertainty about the long-run or "shifting end-points" (see *e.g.* Bauer and Rudebusch, 2020; Afrouzi et al., 2023; Farmer et al., 2024). All of our results extend essentially as written to situations in which there is a long-run mean about which the forecaster is also uncertain. To be concrete, suppose that x_t follows an AR(1) process as in our baseline model but with a long-run mean that is also uncertain:

$$x_t = (1-\rho)\mu + \rho x_{t-1} + \varepsilon_t \tag{23}$$

If we define the demeaned forecast $\hat{x}_{t+h|t} = x_{t+h|t} - \mathbb{E}_t[\mu]$, then all of the properties of $x_{t+h|t}$ in our baseline model extend to $\hat{x}_{t+h|t}$. In Appendix C.3, we give an explicit demonstration of this point for the forecast persistence function. This highlights that the properties of forecasts with uncertain persistence are distinct from properties that arise from long-run mean uncertainty. △

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:

$$P_i(h) \equiv \frac{x^i_{t+h|t}}{x^i_{t+h-1|t}} = \frac{e_i' X_{t+h|t}}{e_i' X_{t+h-1|t}} \tag{24}$$

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 relative forecast reaction function for the ith component of the VAR(1) process as

$$R_i(h; A^*) \equiv \frac{\partial x^i_{t+h|t}}{\partial x^i_t} \Bigg/ \frac{\partial x^{*i}_{t+h|t}}{\partial x^i_t} = \frac{\mathbb{E}[A^h]_{ii}}{(A^{*h})_{ii}} \tag{25}$$

As before, this object depends also on the true model A^*.

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:*

$$P_i(h) = r \frac{\cos(\theta h + \vartheta_i) + o(1)}{\cos(\theta(h-1) + \vartheta_i) + o(1)} \tag{26}$$

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

$$\lim_{h\to\infty} P_i(h) = r \tag{27}$$

2. *If $|v_{1i} w_{1i}|, |v_{1i}^* w_{1i}^*| \neq 0$, then the relative forecast reaction function for variable i obeys:*

$$R_i(h; A^*) = \left(\frac{r}{r^*}\right)^h p_1 \frac{|v_{1i} w_{1i}|}{|v_{1i}^* w_{1i}^*|} \frac{\cos(h\theta + \phi_i) + o(1)}{\cos(h\theta^* + \phi_i^*) + o(1)} \tag{28}$$

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:

$$\lim_{h\to\infty} \frac{1}{h} \ln |R_i(h, A^*)| = \ln \frac{r}{r^*} \tag{29}$$

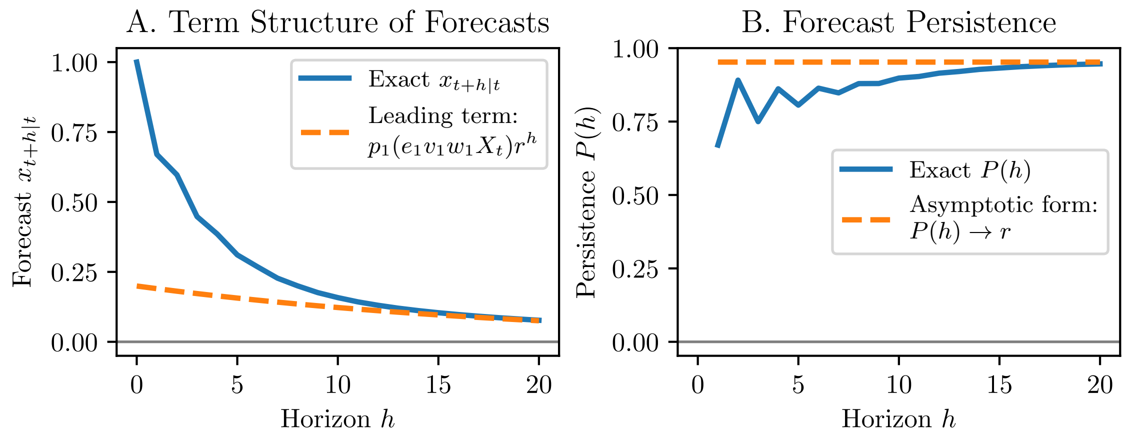
Proof. See Appendix A.3. □

When the largest eigenvalue is real, our results for the AR(1) case extend to VAR processes in a particularly simple way. The largest believable realization of the largest eigenvalue of A replaces 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.

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

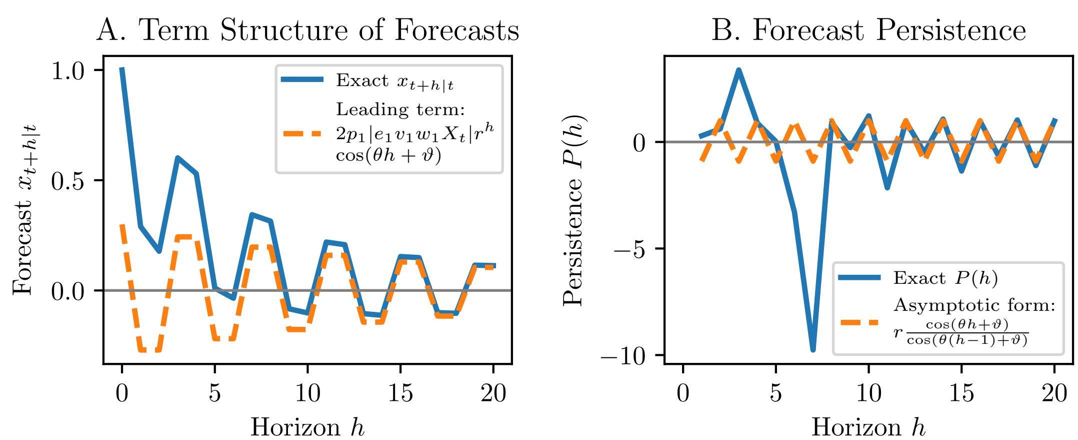
We illustrate each of the two cases of Theorem 2 with a simple AR(2) example. In Figure 3, we plot a situation in which the forecaster entertains two possible AR(2) models with different coefficients and the largest eigenvalue of the two corresponding companion matrices is real. Accordingly, forecasts 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 4, 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 have a persistence function that oscillates asymptotically. While this pattern may seem odd,

Figure 3: 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(2) 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 Panel A, we plot the forecast and the corresponding leading term from the eigendecomposition. In Panel B, we plot the forecast persistence function $P(h)$ and its asymptotic form from Theorem 2.

Figure 4: 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 Panel A, we plot the forecast and the corresponding leading term from the eigendecomposition. In Panel B, we plot the forecast persistence function $P(h)$ and its asymptotic form from Theorem 2.

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.

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. What is new here is that the dominant factor asymptotically is the one that is most persistent, and not necessarily one with high loadings at short horizons.

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.

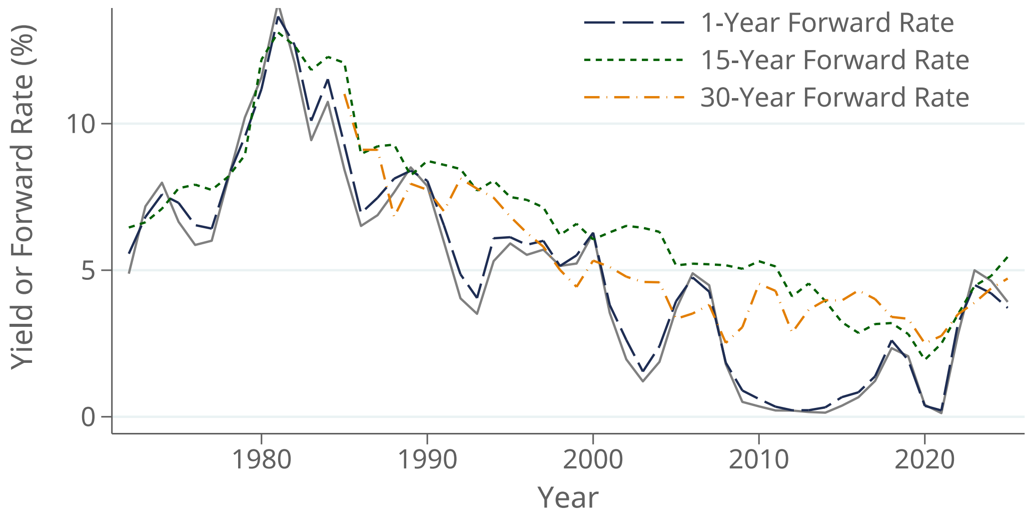
4 Explaining Term-Structure Puzzles

We now apply our model to explain a family of puzzles regarding the term structure of asset prices. The unifying theme of these puzzles is that prices of claims on short- and long-horizon payoffs (*e.g.*, bonds that mature in 1 year versus 30 years) covary "too much" given the observed stochastic properties of fundamentals. We show formally how uncertainty about the persistence of fundamentals can parsimoniously explain these findings. The reason is that agents who are uncertain about persistence behave as if they perceive higher persistence at longer horizons. We apply this idea to study the excessive responsiveness of long-horizon interest rates to news, the dominance of apparent term premia at the long end of the yield curve, and the excess volatility of long-duration asset prices.

4.1 The Excessive Responsiveness of Long Rates

We first study the excess responsiveness of the long-horizon interest rates to seemingly short-run news. The following is an example in the spirit of that which motivates Hanson and Stein (2015), but related to the monetary policy cycle of 2021-2023. On December 13, 2023, the Federal Open Market Committee (FOMC) kept the Federal Funds Rate unchanged at 5.25-5.50% and announced projections of rate cuts in 2024. In a two-day window around the announcement, one-year Treasury yields declined by 22 basis points. Perhaps more sur-

Figure 5: US Forward Rates Over Time



Note: This Figure plots interest rates at different parts of the term structure in the US since 1972 at the annual frequency. The data are based on the methods of Gürkaynak et al. (2007), as reported by the US Federal Reserve System (Board of Governors of the Federal Reserve System, 2026). The lines correspond to the one-year yield (grey solid line), the one-year forward rate (blue dashed line), the 15-year forward rate (green dotted line), and the 30-year forward rate (orange dash-dot line). Data for the 30-year forward are available only after 1982.

prisingly, 15 and 20-year nominal Treasury forwards declined by 33 and 18 basis points, respectively. That is, markets behaved as if that day's news would have significant implications for interest rates more than a decade later.

This phenomenon is not confined to specific and recent policy events. Figure 5 shows forward interest rates at different horizons in the US since 1972. Very long-run forward rates have varied significantly over time, largely in step with short-term interest rates. For example, they rise and fall with the larger cycle of tightening and loosening in the 1970s and 1980s and secularly decline in the Great Moderation of the 1990s and early 2000s.

Set-up. To understand this phenomenon through the lens of our theory, we consider an intentionally very simple model of interest rates. We suppose that there is no arbitrage and that there exists a risk-neutral measure $\mathbb{Q}$ under which the nominal short interest rate x_t follows (as in Giglio and Kelly, 2018):[3]

$$x_t = \rho_{\mathbb{Q}} x_{t-1} + \varepsilon_t \tag{30}$$

[3]Of course, if true rates $\tilde{x}_t$ have mean μ, then Equation 30 holds under the transformation $x_t = \tilde{x}_t - \mu$. This has no bearing on the subsequent analysis.

where ε_t is mean zero, mean independent across time periods.

Departing from the standard model, we treat $\rho_{\mathbb{Q}}$ as itself stochastic with distribution $F_{\mathbb{Q}}$ supported on $[0, 1]$. The date t price of an h-period forward claim on the asset is given by:[4]

$$f_t(h) = \mathbb{E}_t^{\mathbb{Q}}[x_{t+h-1}] = \mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}]x_t \tag{31}$$

where $\mathbb{E}_t^{\mathbb{Q}}[\cdot]$ is the risk-adjusted expectation operator of market participants. In our convention, $f_t(h)$ corresponds to the h-year ahead forward interest rate at time t. Thus, the term structure of (risk-adjusted) forecasts determines forward interest rates and, by implication, the yield curve. Thus, ours is a particular example of an affine one-factor model of interest rates (Piazzesi, 2010). Under suitable assumptions for the distribution of the shock ε_t, which are themselves immaterial for our analysis, the model can correspond to a discrete-time analogue of the canonical models of Vasicek (1977) or Cox et al. (1985) (see Backus et al., 2000).

The economic interpretation of a stochastic $\rho_{\mathbb{Q}}$ is that the persistence of the factor is unknown to market participants. We regard this as an uncontroversial premise for two reasons. First, even if there is a fixed and true persistence, markets may struggle to learn this persistence using only time-series data. Second, persistence itself may change over time because of, for example, changing policymakers, policy frameworks, and macroeconomic factors. Thus, we argue that persistence is likely to remain uncertain even in the long-run.

We adopt a one-factor model to most transparently demonstrate that all of the puzzles we discuss can be resolved, at least qualitatively, in the simplest possible setting with uncertain persistence. As our general analysis from Section 3 demonstrates, however, all of this analysis can be immediately extended to models with multiple pricing factors that follow richer, more quantitatively realistic dynamics, provided that agents face some uncertainty about the parameters governing those dynamics. Later, we will consider one example of such a multi-factor as an extension.

We finally note that, since these dynamics are under the risk-neutral measure $\mathbb{Q}$, uncertainty about the forward evolution of x_t could reflect uncertainty about *both* the physical

[4] In writing this, we are following standard practice in linearizing the standard bond-pricing equation. This can be justified more precisely by defining the forward yield as $f_t(h) = \ln P_t(h-1) - \ln P_t(h)$, where $P_t(h)$ is the price of the bond maturing at horizon h. Under no arbitrage, we have that:

$$\ln P_t(h) = \ln \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{j=0}^{h-1} x_{t+j}\right\}\right] \approx -\sum_{j=0}^{h-1} \mathbb{E}_t^{\mathbb{Q}}[x_{t+j}]$$

and hence that $f_t(h) \approx \mathbb{E}_t^{\mathbb{Q}}[x_{t+h-1}]$. In Figure A.2, we numerically compute the approximation error that arises from this. We find that this error is modest over the horizons we consider and is actually *smaller* in the model with uncertain persistence than without uncertain persistence.

behavior of interest rates and the behavior of risk and risk prices. Nonetheless, as we will see below, all of the arguments would hold *even if* agents were risk neutral and there were no role for risk premia and time variation thereof.

The Co-movement of Short and Long Rates. An implication of Equation 31 is that the co-movement of forward yields with short-term rates is governed by the moments of perceived persistence. In particular, if short-term interest rates change by one basis point, then forward rates at each horizon h change by $\mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}]$ basis points. Our analysis of forecast persistence in Theorem 1 implies that the ratios of these responses for different horizons h increase and asymptote to the maximum believable persistence of the interest rate factor. Formally, defining the relative response of horizon $h+1$ and horizon h rates as

$$\frac{\frac{\partial f_t(h+1)}{\partial x_t}}{\frac{\partial f_t(h)}{\partial x_t}} = \frac{\mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h}]}{\mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}]} = P(h) \tag{32}$$

we can show the following:

Corollary 1 (Long-Horizon Rates Move Together One-For-One)**.** *The relative response of horizon $h+1$ and h forward rates to the short interest rate is:*

1. *Strictly increasing in the horizon, h*
2. *Convergent to one as $h \to \infty$*

Proof. Immediate from Equation 32 and Theorem 1. □

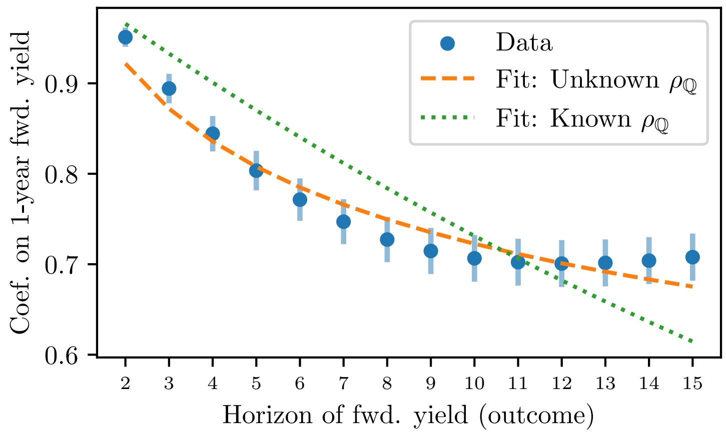
That is, forecast updates far down the yield curve "flatten out," even for rational market participants who are *sure* that the underlying factor is stationary.

Rationalizing Co-Movement via Unknown Persistence. Corollary 1 suggests that measurements of the (excess) co-movements between short- and long-horizon interest rates, in the style of Hanson and Stein (2015), can be used to identify the persistence function P. We implement such a strategy using the nominal forward yield curve dataset of Gürkaynak et al. (2007). Specifically, for different horizons h, we run regressions of the form

$$f_t(h) = \alpha(h) + \beta(h)x_t + \nu_t \tag{33}$$

where $f_t(h)$ is the h-year forward rate and x_t is the one-year forward rate. In the theory, we have that $\beta(h) = \mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}]$ and the OLS estimator $\hat{\beta}(h)$ is consistent. For our main estimates, we use the largest possible sample of forward rates since 1971, restricting us to horizons between $h = 2$ and $h = 15$.

Figure 6: Fitting the Excessive Responsiveness of Long Rates



Note: The blue dots are regression coefficients of daily forward yields of the indicated horizon on the one-year forward yield (Equation 33), where the error bars are 95% confidence intervals based on HAC-robust SE (bandwidth of 20 trading days). The orange dashed line plots predicted coefficients from a model with unknown persistence that follows a Beta distribution, with parameters $\hat{\alpha} = 2.08$ and $\hat{\beta} = 0.18$. The green dotted line plots predicted coefficients with a known but potentially mis-specified persistence, $\bar{\rho}_{\mathbb{Q}} = 0.97$.

We find that, consistent with previous observations in the literature, far-future forwards are highly responsive to short-run rates (Figure 6). For example, our estimates imply that a 100 basis point change in the short rate moves 15-year forward rates by 71 basis points. Moreover, the pattern of coefficients flattens out at longer horizons: forwards 10 years and beyond have essentially the same responsiveness to short-run forwards. These basic patterns also emerge if, following the exact approach of Hanson and Stein (2015), we estimate the model using day-to-day changes in rates, restrict the sample to FOMC meetings, or look at the real forward curve based on TIPS (Figure A.1).

This pattern would present a puzzle in any model where interest rates follow a stationary process whose parameters are known to market participants. For instance, if investors used the model in Equation 30 with some particular persistence $\bar{\rho}_{\mathbb{Q}}$, then the coefficients plotted in Figure 6 follow a geometric sequence: $f_t(h) = \bar{\rho}_{\mathbb{Q}}^{h-1} f_t(1)$. We estimate the $\bar{\rho}$ that best fits the estimated regression coefficients in terms of minimizing mean squared error, and we plot this model's implications for the co-movement of short and long rates as the green dotted line in Figure 6. The geometric decay fits poorly, predicting too much responsiveness for shorter rates (2-10 years) and not enough for longer rates (11-15 years). The difference is even starker further out on the yield curve. If we estimate the responsiveness of 30-year forwards to short rates (Equation 33), using a shortened sample available only from 1985 onward, we estimate

Table 1: *Ex Ante* Versus *Ex Post* Uncertainty About Persistence

Model	$\mathbb{P}[\rho_{\mathbb{Q}} \geq x]$, where $x =$ 0.85	0.90	0.95	1.00
Unknown $\rho_{\mathbb{Q}} \sim \text{Beta}(\hat{\alpha}, \hat{\beta})$	0.83	0.78	0.69	0.00
Non-parametric Lower Bound	0.65	0.63	0.58	0.00
Non-parametric Upper Bound	0.94	0.84	0.79	0.00
Known $\rho_{\mathbb{Q}} = \bar{\rho}_{\mathbb{Q}}$	1.00	1.00	1.00	0.00
Ex Post: "Uninformed" Prior	0.93	0.74	0.40	0.12
Ex Post: "Informed" Prior	0.98	0.86	0.52	0.16

Note: This table compares *ex ante* uncertainty about persistence $\rho_{\mathbb{Q}}$, estimated to match the co-movement of short and long interest rates, with *ex post* uncertainty, based on statistical estimation over the full sample. Each cell reports the probability that $\rho_{\mathbb{Q}}$ exceeds the indicated value. The first row corresponds to a model in which $\rho_{\mathbb{Q}} \sim \text{Beta}(2.08, 0.18)$. The second and third rows are nonparametric bounds on the CDF computed from moment inequalities (see Appendix D). The fourth row corresponds to a model in which $\rho_{\mathbb{Q}} = 0.97$. The fifth and sixth rows correspond to Bayesian posterior distributions from estimating an annual AR(1) process for short interest rates from 1972 to 2026, $x_t = c + \rho x_{t-1} + \varepsilon_t$, with priors $\rho \sim N(1, \sigma_\rho^2)$, $c \sim N(0, \sigma_c^2)$, and $\varepsilon_t \sim N(0, \sigma^2)$ where $\sigma^2 \sim \text{InvGamma}(a, b)$. Under the "uninformed" prior, $\sigma_\rho = \sigma_c = 10^4$ and $a = b = 0.01$. Under the "informed" prior, $\sigma_\rho = 0.1$, $\sigma_c = 1.0$, $a = 2$, and $b = 1$.

$\hat{\beta}(30) = 0.56$ (SE: 0.02). The model with known persistence significantly under-predicts this non-targeted moment: $\beta(h) = \bar{\rho}_{\mathbb{Q}}^{29} = 0.36$.

What if, instead, investors were uncertain about persistence? As a simple way to model this, we return to our running example in which persistence follows a Beta distribution, supported on $[0, 1]$, with two shape parameters. This model has only one more parameter than the previous one considered. We calibrate the two shape parameters to best fit the estimated regression coefficients in terms of minimal mean squared error, as before, and we plot the resulting model estimates as the orange dashed line in Figure 6. Unlike the model with known persistence, the model with unknown persistence fits the data for both short and long horizons. In particular, unknown persistence allows us to match the significant co-movement of short rates with long-horizon rates ($h \geq 10$) without over-predicting co-movement at the short end. For the untargeted responsiveness of 30-year forwards, the model's prediction of $\beta(30) = 0.60$ is very close to the empirical estimate of $\hat{\beta}(30) = 0.56$.

The Implied Beliefs. In Table 1, we describe the beliefs that rationalize the co-movement of short and long interest rates (Figure A.3 shows the full CDFs). Under the estimated model, market participants are confident that persistence is high (*e.g.*, $\mathbb{P}[\rho_{\mathbb{Q}} \geq 0.90] = 0.78$), but unsure exactly how high. On the other hand, they are sure that short interest rates are stationary under $\mathbb{Q}$ measure, a property that is potentially desirable.[5] To stress-test how much the first conclusion depends on the exact functional form of beliefs, we also calculate non-parametric pointwise bounds on CDFs supported on $[0, 1]$ that satisfy moment inequalities implied by our coefficient estimates (see Appendix D). These tell largely the same story. Finally, we contrast these estimates to the benchmark in which market participants are dogmatic but potentially mis-specified: in this case, everyone is sure that $\bar{\rho}_{\mathbb{Q}} = 0.97$.

A natural follow-up question is which, if any, of these beliefs is "reasonable" by some external metric. The gold standard would be *real-time* measurement of uncertainty *under the risk neutral measure*, which is infeasible on both counts. Instead, as a simple sanity check, we compare with the *ex post* statistical uncertainty that an econometrician might face if they took annual data on short interest from 1972 to 2025 and estimated an AR(1) model. We consider two versions of this exercise: one under an "uninformed" prior that essentially replicates frequentist inference, and one under an "informed" prior that is centered around a random walk, as is common in Bayesian time-series analysis (see, *e.g.*, Litterman, 1986).[6] In both cases, we get distributions that are comparably confident as our Beta forecaster that persistence exceeds 0.9 (probability 0.74 when "uninformed," or 0.87 when "informed"). Both beliefs are unrankably uncertain versus the Beta distribution, in the sense that the CDFs cross twice. Thus, via this simple test, we view the beliefs that rationalize asset prices as broadly reasonable.

Uncertainty in a Multi-factor Model. As mentioned earlier, we focus on a one-factor model to isolate the core economics of our argument in a very simple setting with only two free parameters controlling model uncertainty. But multi-factor models, which are common in term structure analysis (see, *e.g.*, Piazzesi, 2010), can of course also be easily analyzed in our framework. To understand the implications of our model in such a setting, we consider the case in which investors believe that p factors, stacked in the $p \times 1$ vector X_t, determine interest rates. In particular, they believe that the (demeaned) short interest rate x_t is affine in the factors and the factors follow a VAR(1) under $\mathbb{Q}$ measure:

$$x_t = \Gamma' X_t, \qquad X_t = A_{\mathbb{Q}} X_{t-1} + \varepsilon_t \tag{34}$$

[5] In standard macroeconomic models, non-stationary interest rates are not a natural outcome unless, for example, the discount rates in household preferences are non-stationary. Moreover, under non-stationary models, the second-order terms that are ignored in the approximations to derive the subjective expectations hypothesis (Equation 31) explode as the horizon increases: intuitively, if interest rates are non-stationary, holding long-horizon assets entails exposure to unboundedly high risks.

[6] There is of course a robust debate in econometrics about what is the right benchmark for "uninformed" in this context (see, *e.g.*, Sims, 1988; Sims and Uhlig, 1991; Phillips, 1991). For ease of exposition, we simply use a diffuse (but integrable) Normal-Inverse-Gamma prior.

where $\varepsilon_t \in \mathbb{R}^p$ is a mean-zero, and mean-independent random variable (under $\mathbb{Q}$ measure), and $\Gamma \in \mathbb{R}^p$ and $A_{\mathbb{Q}} \in \mathbb{R}^{p\times p}$ are parameters. To focus on the role of unknown dynamics, we assume that $A_{\mathbb{Q}}$ is the key parameter unknown to market participants, who believe that it follows distribution $F_{\mathbb{Q}}$, but otherwise know Γ and observe the current state X_t. Under no arbitrage, forward rates are

$$f_t(h) = \mathbb{E}_t^{\mathbb{Q}}[x_{t+h-1}] = \mathbb{E}_t^{\mathbb{Q}}[\Gamma' X_{t+h-1}] = \Gamma' \mathbb{E}_{F_{\mathbb{Q}}}[A_{\mathbb{Q}}^{h-1}] X_t \tag{35}$$

To bring this model to the data, we take $p = 2$ and, following standard practice, estimate the factors via principal components analysis (PCA) on our daily data on forwards between 1 and 15 years. In the cross-section of maturities, the estimated factors have the familiar interpretation of a "level" factor that loads equally on all horizons and a "slope" factor that loads more on shorter than longer horizons. Then, using the same data, we estimate regressions of each forward of horizon h on both the first and second factor:

$$f_t(h) = \alpha(h) + \beta_1(h) X_{1t} + \beta_2(h) X_{2t} + \nu_t \tag{36}$$

We note that $\hat{B}(h) := (\hat{\beta}_1(h), \hat{\beta}_2(h))'$ is a consistent estimate of $\Gamma' \mathbb{E}_{F_{\mathbb{Q}}}[A_{\mathbb{Q}}^{h-1}]$. We directly plug in $\Gamma = \hat{B}(1)$ and then back out implied beliefs about $A_{\mathbb{Q}}$ to minimize the sum of squared deviations from the elements of $(\hat{B}(h))_{h=2}^{15}$, taking each of the four elements of $A_{\mathbb{Q}} = [A_{11}, A_{12}; A_{21}, A_{22}]$ as an independent Gaussian variable with mean μ_{ij} and variance σ_{ij}^2. Finally, as in the previous analysis, we contrast this model to an alternative in which investors hold some (potentially mis-specified) point beliefs about $A_{\mathbb{Q}}$, which we also fit to minimize mean squared error with the same moments.

Our estimates, shown in Figure A.4, show that uncertain dynamics help match the responsiveness of long forward rates to the level factor. We illustrate the predictions of the two-factor model for the term structure of forecasts and the forecast persistence function in Figure A.5. With *known* but mis-specified dynamics, a two-factor model continues to predict too much reaction at short horizons and too little at long horizons. By contrast, both models can almost equally well fit responses to the slope factor. Thus, through the lens of a multi-factor model, the key mechanism for our explanation of interest rate comovement is uncertainty about the dynamics of the level factor.

Unknown Persistence Meets Other Mechanisms. Of course, there are other explanations for the excess sensitivity of forward rates to short rates. For example, Hanson et al. (2021) propose an explanation based on rate-amplifying shocks, whereby higher short interest rates temporarily induce a net supply of long-term bonds, and slow arbitrage. A potential feature of our explanation is its consistency with the absence of arbitrage. We see the two explanations as complementary in the sense that as uncertainty about persistence can offer an alternative microfoundation for the correlated demand at different parts of the yield curve that is proposed by Hanson et al. (2021) and more broadly by the literature studying "preferred habitat" models of the term structure (Vayanos and Vila, 2021).[7]

4.2 Term Premia and Excess Return Predictability

A long literature has documented predictable excess returns in US bond yields and used this as *prima facie* evidence against the expectations hypothesis of the term structure, whereby long-term forward rates coincide with expected future short rates (*e.g.*, Fama and Bliss, 1987; Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005; Hanson and Stein, 2015). A key fact in this literature is that long-horizon rates are primarily driven by factors other than the physical expectation of future short rates. That is, movements in apparent term premia determine long-horizon interest rates (see *e.g.,* Adrian et al., 2015).

In our model, there *must be* expected returns from trading strategies at the long end of the yield curve from the perspective of an econometrician looking at the data *ex post*. This is despite the fact that, for the agents who live in our model and price assets, a subjective expectations hypothesis holds and there are no predictable excess returns.

The Dominance of Term Premia at Long Horizons. Suppose that the true physical measure is $\mathbb{P}^*$ and the interest-rate factor has true persistence $\rho^*_{\mathbb{P}}$. Following standard practice in the asset-pricing literature, we can decompose asset prices into a component explained by forecasts under the *true* physical measure and a residual which we refer to as an *apparent term premium*:

$$f_t(h) = \mathbb{E}_t^{\mathbb{P}^*}[x_{t+h-1}] + \tau_t(h) \tag{37}$$

In practice, the forecast $\mathbb{E}_t^{\mathbb{P}^*}[x_{t+h-1}]$ might be constructed *ex post* using an econometric model estimated on the full sample, and it may not correspond to agents' (subjective) physical expectations at time t. In our model with unknown persistence, the apparent term premium follows:

$$\tau_t(h) = (\mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}] - (\rho^*_{\mathbb{P}})^{h-1})x_t \tag{38}$$

This reflects both (i) the difference between risk-adjusted and physical probabilities and (ii) the difference between subjective expectations at time t and those under the true model. We

[7]Moreover, changing uncertainty about persistence over time could micro-found time variation in these demand effects, thereby helping to explain the time variation in excess long-rate sensitivity uncovered by Hanson et al. (2021).

emphasize that even if risk premia are zero, the presence of uncertain persistence necessitates that apparent term premia completely drive long-horizon rates.[8]

To characterize the importance of term premia for forward rates, we define the ratio of the apparent term premium and the forward rate as

$$I(h) = \frac{\tau_t(h)}{f_t(h)} = \frac{\mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}] - (\rho_{\mathbb{P}}^*)^{h-1}}{\mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}]} = \frac{R(h-1;\rho_{\mathbb{P}}^*) - 1}{R(h-1;\rho_{\mathbb{P}}^*)} \tag{39}$$

The key properties of this ratio are determined by the relative forecast reaction function studied in Proposition 1, since it fundamentally depends on whether current (risk-neutral) forecasts over- or under-react relative to the true evolution of short rates. Therefore, building on the earlier finding that sufficiently long-horizon forecasts *must* over-react:

Corollary 2 (Term Premia Account for Long-Horizon Forward Rates)**.** *If physical rates are stationary ($\rho_{\mathbb{P}}^* < 1$), then the response of long-horizon rates is entirely driven by the response of apparent term premia and not by the physical expectation of rates, i.e.,* $\lim_{h\to\infty} I(h) = 1$.

Proof. Immediate from Proposition 1. □

We can calculate $I(h)$ in our estimated model of Treasuries. We take $\rho_{\mathbb{P}}^* = 0.93$ from an *ex post* linear regression and use our fitted distribution $\rho_{\mathbb{Q}} \sim \text{Beta}(\alpha = 2.08, \beta = 0.18)$.[9] About 36% of 15-year forward rates, the farthest horizon rate that we use for estimation, are due to term premia in the model. Extrapolating this out further down the yield curve, our model implies that 77% of the predicted variation of 30-year forwards are due to term premia. Thus, our model rationalizes the dominance of term premia at the long end of the yield curve purely from a small and empirically disciplined quantity of uncertainty about persistence.

To visualize how this helps explain the behavior of US interest rates, Figure 7 shows a decomposition of 2-year, 15-year, and 30-year forward interest rates into (i) a portion explained by our simple one-factor model with uncertain persistence and (ii) a portion predicted by the pure expectations hypothesis. For short horizons, the actual forward rate, model prediction,

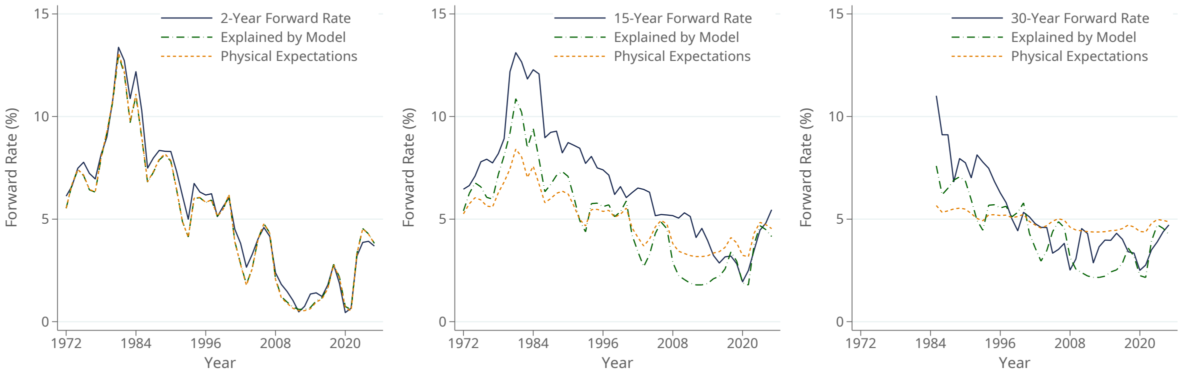
[8]To see this point, observe that we may further decompose apparent term premia as:

$$\tau_t(h) = \left[\underbrace{\left(\mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{h-1}] - \mathbb{E}_{F_{\mathbb{P}}}[\rho_{\mathbb{P}}^{h-1}]\right)}_{\text{Risk Premium}} + \underbrace{\left(\mathbb{E}_{F_{\mathbb{P}}}[\rho_{\mathbb{P}}^{h-1}] - (\rho_{\mathbb{P}}^*)^{h-1}\right)}_{\text{Uncertain Persistence}} \right] x_t$$

where $\mathbb{P}$ is the subjective probability measure of investors. Thus, even when $\mathbb{P} = \mathbb{Q}$ and there is no risk-adjustment, our subsequent results that apparent term premia dominate the long-end hold as stated.

[9]As above, the regression to obtain $\rho_{\mathbb{P}}^*$ is run on annual data from 1972 to 2025. This OLS estimate is essentially identical to the posterior mean from the "uninformed" Bayesian estimation in Table 1.

Figure 7: Explaining Forward Interest Rates



(a) 2-Year **(b)** 15-Year **(c)** 30-Year

Note: This figure decomposes US forward interest rates over time. Each panel corresponds to a different horizon of the nominal US Treasury forward curve, as reported by Board of Governors of the Federal Reserve System (2026). Data on the 2- and 15-year forwards are available since 1972, and data on the 30-year forwards are available since 1985. In each panel, the solid blue line is the forward rate from the data; the orange dashed line is the model-implied forward rate, constructed using Equation 31 and our estimated risk-neutral uncertainty about persistence, $\rho_{\mathbb{Q}} \sim \text{Beta}(\alpha = 2.08, \beta = 0.18)$ (see Figure 6); and the green dash-dot line is the component corresponding to physical expectations of short interest rates, using $\rho_{\mathbb{P}}^* = 0.93$. For both calculations, we use a mean short interest rate of 5.1%, estimated from the full sample.

and physical expectations almost exactly coincide. At long horizons (*e.g.*, 30 years), physical expectations are closely anchored to the long-run mean, and all of the model's predicted dynamics come from apparent term premia. This creates the correct low-frequency dynamics whereby long-horizon forward rates fall dramatically from 1980 to 2000, in lockstep with decreases in short interest rates (see also Figure 5). We obtain this prediction without needing to assume that market participants "just so happened" to over-extrapolate. Our estimated beliefs about persistence are close to correct on average (see Figure 6); but, because they put some probability mass on higher-than-realized values of persistence, long-horizon forwards and term premia must display the patterns in Figure 7.

Thus, our one-factor model with random persistence governed by a single additional parameter is capable of accounting for the long-horizon dominance of apparent term premia in explaining forward interest rates. This is the case even when agents put probability one on the event that rates are stationary, distinguishing our explanation from those of Bauer and Rudebusch (2020) and Farmer et al. (2024). Our explanation also differs from classical hypotheses requiring (time varying) risk compensation (see *e.g.,* Piazzesi, 2010), preferred habitats (Vayanos and Vila, 2021), or failures of the subjective expectations hypothesis

(Molavi et al., 2026). Of course, our results do not imply that these explanations are wrong. Rather, we provide a complementary explanation that long rates *must* be driven by term premia even in a setting without any additional frictions.

Predictable Excess Returns. We next show how the model necessarily generates what would appear to be *ex post* predictable returns to an econometrician, due to the aforementioned dynamics of the term premium. A long and influential literature (*e.g.*, Fama and Bliss, 1987; Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005) has documented the existence of these predictable returns. Here, we follow the language of Cochrane and Piazzesi (2005) and define excess returns as:

$$\mathrm{rx}_{t+1}(h) = p_{t+1}(h-1) - p_t(h) - x_t \tag{40}$$

where $p_t(h)$ is the (log) price at t of the bond that pays out a single unit at time $t+h$. This excess return can be interpreted as a measure of the overperformance of longer bonds over shorter bonds: that is, the return on a trading strategy of holding (buying today, selling tomorrow) an h-period bond relative to the return of holding a one-period bond to maturity.

We now study how the current interest rate x_t predicts future excess returns $\mathrm{rx}_{t+1}(h)$ for different horizons h. Formally, suppose that an analyst ran the regression:

$$\mathrm{rx}_{t+1}(h) = \xi(h) + \beta(h)x_t + \nu_{t+1} \tag{41}$$

Since our simple model has only one priced factor, any test of whether asset prices at t predict excess returns at $t+1$ boils down to this one. Under the null hypothesis of there being no forecastable excess returns, we would have that $\beta(h) = 0$. Conversely, a positive (negative) $\beta(h)$ coefficient would imply that there are larger (smaller) forecastable excess returns from holding long bonds. In our model, an econometrician would estimate such a relationship with the following properties.

Corollary 3. *The excess return forecasting regression (Equation 41) holds exactly with:*

$$\beta(h) = \sum_{j=1}^{h-1} \mathbb{E}_{F_{\mathbb{Q}}}[\rho_{\mathbb{Q}}^{j-1}]\,(P(j) - \rho_{\mathbb{P}}^*) \tag{42}$$

Moreover, $\nu_{t+1} \perp x_t$. Hence, a standard OLS regression estimator is consistent for $\beta(h)$. Furthermore, there are forecastable excess returns: $\beta(h) \neq 0$ for all but at most one $h \geq 2$.

Proof. See Appendix A.4. □

When an econometrician comes to the data *ex post*, they implicitly impose the generating process that *actually* obtained in sample, which has persistence $\rho^*_{\mathbb{P}}$. However, the agents that price assets in the model do not know the true value of risk-neutral persistence $\rho_{\mathbb{Q}}$, leading them to apparently overreact to current rates. This gives rise to *ex post* forecastable excess returns, even though assets in the model are exactly priced by a subjective expectations hypothesis. Moreover, because agents apparently over-react at long horizons, an econometrician confronting data generated by this model could find evidence that $\beta(h) > 0$ for sufficiently long time horizons: interest rates mean-revert faster under the physical law than under the risk-neutral law.

4.3 Excessive Long-Horizon Forward Volatility

Beyond excessive responses of long-horizon interest rates to short-term interest rate news, Giglio and Kelly (2018) document that long-horizon volatility is excessive across a wide range of asset classes:

> We document a form of excess volatility in prices along the term structure that is difficult to reconcile with "standard" asset-pricing models. Our central finding is that price fluctuations at different points in the term structure are internally inconsistent with each other—prices on the long end of the term structure are far more variable than justified by the behavior of short-end prices—given usual modeling assumptions. The consistency violations are highly significant statistically and economically. Perhaps most interesting, excess volatility of long-maturity prices is evident in a large number of asset classes, including claims to equity and currency volatility, sovereign and corporate credit risk, Treasury yields, commodities, and inflation. — Giglio and Kelly (2018)

They document this claim by deriving a formal test and finding strong support across asset classes. Formally, given a dataset of forward prices $f_t(h)$ at dates $t \in \{1, \ldots, T\}$ at horizons $h \in \{1, \ldots, H\}$ Giglio and Kelly (2018) propose a variance ratio (VR) statistic:

$$\text{VR}(h) = \frac{\mathbb{V}[b_h f_t(1)]}{\mathbb{V}[b_2^{h-1} f_t(1)]} \tag{43}$$

where b_h is the regression coefficient of $f_t(h)$ on $f_t(1)$. Under the null hypothesis of a deterministic affine term structure model, Giglio and Kelly (2018) show that $\text{VR}(h) = 1$ for all $h \in \{1, \ldots, H\}$. For all of the asset classes mentioned, Giglio and Kelly (2018) document two key facts: (i) the VR statistic is invariably greater than one and (ii) the VR statistic is invariably strictly increasing in the time horizon.

Uncertainty about persistence can rationalize both of these facts in a linear, one-factor model. In such a model with *deterministic* persistence, as introduced by Giglio and Kelly (2018) to represent "standard models" of asset pricing, $\text{VR}(h) \equiv 1$. But when persistence is stochastic, this no longer holds. Intuitively, the prices of long-horizon claims, in the limit, perfectly covary with one another, and therefore covary more with the prices of short-horizon claims than would be predicted by the extrapolation in the denominator of Equation 43. Therefore, the VR statistic inherits properties of the forecast persistence function:

Corollary 4 (VR Statistics Are Super-Unitary and Increasing)**.** *The VR statistic can be written as:*

$$\text{VR}(h) = \prod_{i=2}^{h-1} \left(\frac{P(i)}{P(1)} \right)^2 \tag{44}$$

Moreover, if persistence is uncertain (F is non-degenerate), then the VR statistic is:

1. *Strictly Super-unitary:* $\text{VR}(h) > 1$ *for all* $h \geq 3$*.*
2. *Strictly Increasing:* VR *is a strictly increasing function*

Proof. See Appendix A.5. □

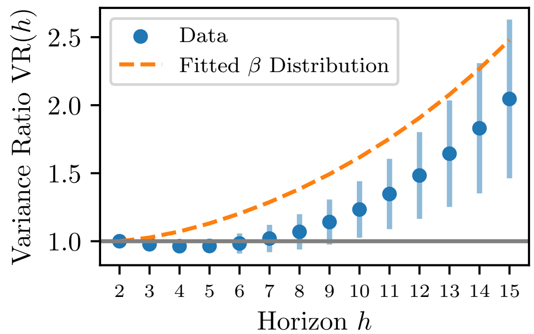
Thus, uncertainty about the persistence of interest rates can directly explain why, in an arbitrage-free model: (i) variance ratios exceed one and (ii) variance ratios increase with horizon. We emphasize that this result is entirely independent of whether investors over- or under-extrapolate, a candidate explanation suggested by Giglio and Kelly (2018). Indeed, these properties hold *whenever* $\rho_{\mathbb{Q}}$ is stochastic under the risk-neutral measure.

Returning to our setting of the Treasury forward curve, we can calculate the VR statistic under our estimated models for investors' risk-neutral uncertainty (Figure 8). Our baseline Beta-distribution model predicts a VR ratio of 2.28 when applied 15 years ahead. This is close, and even somewhat larger than, the estimate given by Giglio and Kelly (2018) for Treasury bills. Moreover, as emphasized above, we obtain this prediction in a fully rational model with no over- or under-extrapolation and no mis-pricing conditional on beliefs.

5 Explaining Macroeconomic Forecasting Puzzles

We now explore the implications of our findings for macroeconomic forecasts in the data. Beyond the potential intrinsic interest in macroeconomic forecasts (see, *e.g.*, the review by Angeletos et al., 2021), examining the properties of forecasts will allow us to evaluate the economic mechanism underlying our rationalization of the forward return puzzles. We find that our model of forecasting with unknown persistence is consistent with two key facts

Figure 8: The Variance Ratio Statistic for US Treasuries



Note: This figure plots estimates of the Variance Ratio statistic of Giglio and Kelly (2018), as defined in Equation 43. The blue dots correspond to empirical estimates using daily data on the US Treasury forward curve, defined as $\hat{\text{VR}}(h) = \hat{\beta}(h)^2/\hat{\beta}(2)^{2(h-1)}$ where $\hat{\beta}(h)$ and $\hat{\beta}(2)$ are estimated from Equation 33. Error bars are 95% confidence intervals based on block bootstrapping at the monthly level. The orange dashed line plots estimates from the estimated model of uncertain persistence, fitted as a Beta distribution with parameters $\hat{\alpha} = 2.08$ and $\hat{\beta} = 0.18$.

regarding the term structure of expectations: as the horizon increases, forecasts appear to become more persistent (Goldstein and Gorodnichenko, 2022); and forecasts appear to overreact at long horizons (Halperin and Mazlish, 2025). Both of these mechanisms were the key forces underlying our rationalization of the asset pricing puzzles, thus providing support for the economic mechanism we propose.

5.1 The Term Structure of Individual Forecasts

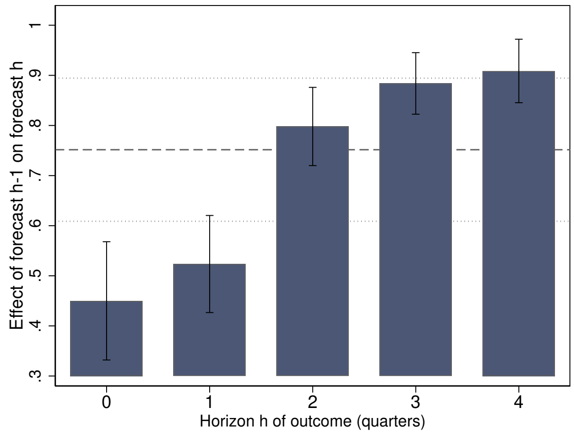
We first study the persistence of the term structure of macroeconomic forecasts, mirroring our discussion in Section 4. Using panel data on professional forecasters' inflation expectations, Goldstein and Gorodnichenko (2022) estimate the implied persistence of forecasts between horizon $h-1$ and h for forecasters indexed by i by estimating the regression model:

$$\mathbb{E}_{it}[x_{t+h}] = \xi(h) + \beta(h)\mathbb{E}_{it}[x_{t+h-1}] + \nu_t \tag{45}$$

where x_t is a macroeconomic variable of interest and $\mathbb{E}_{it}[\cdot]$ denotes individual survey expectations. This regression has a straightforward interpretation in the case in which forecasters use an AR(1) model with unknown persistence ρ, as we studied in Section 2. In this case, the sequence of coefficients $\beta(h)$ identifies the forecast persistence function.

Corollary 5. *In the AR(1) setting, the Goldstein and Gorodnichenko (2022) regression*

Figure 9: The Flattening Term Structure of CPI Forecasts



Note: This figure shows estimates of Equation 45 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. 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.

(Equation 45) holds exactly with:

$$\beta(h) = P(h) \tag{46}$$

Moreover, when F is non-degenerate, $\beta(h)$ is strictly increasing in the horizon h.

Proof. Immediate by the definition of P and from Theorem 1. □

Thus, our theory implies that the regression coefficients estimated by Goldstein and Gorodnichenko (2022) *must* increase in horizon when forecasters are uncertain about persistence. Indeed, Goldstein and Gorodnichenko (2022) find exactly this pattern. In a theoretical analysis, Goldstein and Gorodnichenko (2022) show that news about the future can explain this pattern. Our explanation is complementary and shows that uncertainty about persistence necessarily implies this pattern.

To demonstrate this, we recreate their finding in Figure 9, using individual Consumer Price Inflation (CPI) forecasts at the quarterly frequency from 1981 onward in the US Survey of Professional Forecasters (SPF). For reference, we compare the estimates of $\beta(h)$ with the estimated time-series persistence of CPI inflation (dashed line) and statistical uncertainty in this estimate (dotted lines, indicating a 95% confidence interval). The behavior of $\beta(h)$

across horizons may be especially plausible given the high uncertainty for an econometrician (even *ex post*) regarding the persistence of inflation.

Beyond the survey data used by Goldstein and Gorodnichenko (2022), there is strong evidence of greater forecast persistence at long horizons in laboratory experiments. In particular, Afrouzi et al. (2023) document that—given a wide range of experimentally varied true levels of persistence—the gap between implied persistence in subjects' forecasts grows as the forecasting horizon extends (see Figure III in Afrouzi et al., 2023).

In contrast to our previous study of asset prices, our study of survey forecasts also allows us to look at *individual* beliefs to better understand the mechanism. This allows us to assess if individual forecasters behave consistently with having uncertain persistence, rather than merely appearing to behave as such from aggregation. To do this, we calculate an individual measure of the persistence function at each horizon h by using the term structure of quarterly forecasts as well as individuals' reported forecast for average inflation over the next ten years ($\hat{\mathbb{E}}_{it}[x_{i,10Y}]$), which we take as a proxy for their estimate of the long-run mean of inflation:[10]

$$\hat{\rho}_{ith} = \frac{\hat{\mathbb{E}}_{it}[x_{i,t+h}] - \hat{\mathbb{E}}_{it}[x_{i,10Y}]}{\hat{\mathbb{E}}_{it}[x_{i,t+h-1}] - \hat{\mathbb{E}}_{it}[x_{i,10Y}]} \tag{47}$$

As described in Remark 4, demeaning by long-run forecasts allows us to recover the individual-level forecast persistence function even when forecasters have heterogeneous and/or time-varying beliefs about long-run means (as suggested by Farmer et al., 2024). We compare the distribution of this statistic for $h = 1$ and $h = 4$ in Figure 10. As predicted by our model, the distribution of persistence shifts toward the right at the longer horizon.

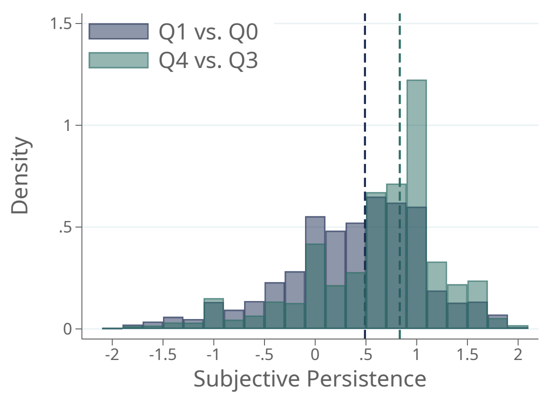
Taken together, these findings support the key implication of our general theory: forecasts appear to become more persistent as the forecasting horizon increases. This provides direct evidence for the mechanism by which we explained the excessive responsiveness of long rates and the excessive volatility of long-horizon forward prices.

5.2 Aggregate Under- and Over-reaction

We finally study the horizon-dependent under- and over-reaction of survey forecasts. To do this, we adopt the regression framework of the empirical study of Halperin and Mazlish (2025), which is itself adapted from Coibion and Gorodnichenko (2015). Specifically, letting x_t be a macroeconomic variable (e.g., inflation) and letting $\hat{\mathbb{E}}_t$ denote consensus forecasts,

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

Figure 10: 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 47. The vertical dashed lines denote the medians of each distribution. The sample is CPI forecasts after 1991:Q4. In the figure, we truncate both distributions at ± 2, excluding 5% of each distribution in the tails.

these authors study the following regression model indexed by the forecast horizon h:

$$\underbrace{x_{t+h} - \hat{\mathbb{E}}_t[x_{t+h}]}_{\text{Forecast error}} = \beta(h) \underbrace{\left(\hat{\mathbb{E}}_t[x_{t+h}] - \hat{\mathbb{E}}_{t-1}[x_{t+h}]\right)}_{\text{Forecast revision}} + \gamma(h) \underbrace{\hat{\mathbb{E}}_{t-1}[x_{t+h}]}_{\text{Lag forecast}} + \nu_t \tag{48}$$

The standard logic behind this regression is that—if forecasters know the true DGP and are Bayesian—then forecast revisions (which are known to forecasters) should not predict future forecast errors (see *e.g.,* Coibion and Gorodnichenko, 2015). Under this view, $\beta(h)$ should be zero at all horizons. If, instead, $\beta(h)$ were positive, it would imply that upward forecast revisions predict that outcomes exceed forecasts. This corresponds to an informal notion that forecasts *under-react* to recent news. Conversely, if $\beta(h)$ were negative, it would correspond to an informal notion that forecasts *over-react* to recent news.

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

Corollary 6. *In the AR(1) setting, the Halperin and Mazlish (2025) regression (Equation 48) holds exactly with*

$$\beta(h) = \frac{1 - R(h;\rho^*)}{R(h;\rho^*)} \tag{49}$$

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* $R(h;\rho^*)$*. Moreover,* $\beta(h)$ *is initially positive and then negative if and only if* $\mathbb{E}[\rho] < \rho^*$*, and it is globally negative if and only if* $\mathbb{E}[\rho] \geq \rho^*$*.*

Proof. See Appendix A.6. □

Halperin and Mazlish (2025) estimate Equation 48 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 $\beta(h)$ declines with horizon and is negative for all horizons from the two-year to the longest (10 years and beyond) horizons (see Figure 1 of Halperin and Mazlish, 2025).

Evidence from Other Domains. d'Arienzo (2020) studies horizon-dependent over-reaction of interest rate expectations measured in two ways, directly via surveys and indirectly via asset prices. His core finding is that forecast revisions and forecast errors are positively correlated at short horizons but negatively correlated at longer horizons (see, *e.g.*, Figure 3). This is consistent with the empirical findings of Halperin and Mazlish (2025) and with our summary (and rationalization) of bond-pricing puzzles in Section 4.

Afrouzi et al. (2023) conduct an experiment to study how people forecast stochastic processes, in an environment in which they can directly control the characteristics of those processes. They find that forecasts over-react to recent observations and, moreover, do so to a greater extent for longer-horizon forecasts (see, *e.g.*, Figure III).

Summary. All of these findings are necessary consequences of forecasting with unknown persistence, so long as the true persistence of variables is lower than the maximum believable persistence. This complements existing explanations based on combining two features: noisy expectations, which generate short-horizon under-reaction, and mis-specified over-extrapolation, which generates long-horizon over-reaction (d'Arienzo, 2020; Angeletos et al., 2021; Afrouzi et al., 2023; Halperin and Mazlish, 2025). Moreover, the fact that economic forecasts themselves display the key properties predicted by the theory provides validation for the mechanism underlying our rationalization of the bond pricing puzzles.

6 Conclusion

Motivated by the pervasiveness of uncertainty about the persistence of important economic variables, we study the implications of such uncertainty for forecasts. We find that long-horizon forecasts are necessarily as persistent as is believable and over-react. We demonstrate how these properties rationalize six important asset pricing and forecasting puzzles: (i) the excess responsiveness of long-run interest rates to short rates (Hanson and Stein, 2015), (ii) the dominance of apparent term premia at the long-end of the yield curve (Adrian et al., 2015), (iii) *ex post* predictable excess returns in bond yields (Fama and Bliss, 1987; Cochrane and Piazzesi, 2005), (iv) the excess volatility of long-horizon forward prices (Giglio and Kelly, 2018), (v) horizon-dependent persistence in the term structure of survey forecasts (Goldstein and Gorodnichenko, 2022), and (vi) simultaneous under- and over-reaction in these forecasts (Coibion and Gorodnichenko, 2015; Angeletos et al., 2021; Bordalo et al., 2020; d'Arienzo, 2020; Kohlhas and Walther, 2021; Halperin and Mazlish, 2025).

While there are of course many existing theories that rationalize subsets of these facts, we argue that uncertainty about persistence represents a compelling alternative explanation for three reasons. First, it is robust in the sense that these six puzzles are *necessary* implications of uncertain persistence. Second, it is parsimonious in the sense that it accounts for all of these puzzles from a single, plausible premise. Third, it is empirically reasonable: the amount of uncertainty required to explain the key asset pricing puzzles is comparable to the *ex post* statistical uncertainty about persistence at the end of the sample period, which should *under-state* in-sample uncertainty about persistence.

While we have studied the implications of uncertain persistence for forecasts and asset prices, the approach we have taken here could also be fruitfully explored in general equilibrium, macroeconomic models. To exemplify this, in Appendix E we study an equilibrium model of capital investment with uncertainty about the persistence of productivity and investors who can choose between projects of different maturities. We show that, in equilibrium, 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.

Thus, uncertainty about persistence can generate cycles of capital investment. Intuitively, 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. In the interests of space, we leave a fuller analysis of the equilibrium implications of uncertain persistence for real investment to future study, including this exercise as a proof-of-concept that interesting work remains to be done in this area.

A Proofs of Main Results

A.1 Proof of Theorem 1

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

$$\ln P(h) = \ln M_{\ln \rho}(h) - \ln M_{\ln \rho}(h-1) = K_{\ln \rho}(h) - K_{\ln \rho}(h-1) \tag{50}$$

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 for a non-degenerate distribution is strictly 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 $\epsilon \in (0, r)$:

$$\begin{aligned} P(h) &= \frac{\mathbb{E}[\rho^h]}{\mathbb{E}[\rho^{h-1}]} \geq \frac{\mathbb{E}[\rho^h \mathbb{I}[\rho > r - \epsilon]]}{\mathbb{E}[\rho^{h-1}]} \geq (r-\epsilon)\frac{\mathbb{E}[\rho^{h-1}\mathbb{I}[\rho > r-\epsilon]]}{\mathbb{E}[\rho^{h-1}]} \\ &= (r-\epsilon)\left(1 - \frac{\mathbb{E}[\rho^{h-1}\mathbb{I}[\rho \leq r-\epsilon]]}{\mathbb{E}[\rho^{h-1}]}\right) \geq (r-\epsilon)\left(1 - \frac{(r-\epsilon)^{h-1}}{\mathbb{E}[\rho^{h-1}]}\right) \\ &= (r-\epsilon)\left(1 - \frac{(r-\epsilon)^{h-1}}{\mathbb{E}[\rho^{h-1}]}\right) \end{aligned} \tag{51}$$

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:

$$\begin{aligned} \mathbb{E}[\rho^{h-1}] &= \mathbb{E}[\rho^{h-1}\mathbb{I}[\rho \leq r-\epsilon/2]] + \mathbb{E}[\rho^{h-1}\mathbb{I}[\rho > r-\epsilon/2]]] \geq p(\epsilon)\mathbb{E}[\rho^{h-1}|\rho > r-\epsilon/2] \\ &\geq p(\epsilon)(r-\epsilon/2)^{h-1} \end{aligned} \tag{52}$$

Combining the previous two equations, we have that:

$$P(h) \geq (r-\epsilon)\left(1 - \frac{(r-\epsilon)^{h-1}}{p(\epsilon)(r-\epsilon/2)^{h-1}}\right) = (r-\epsilon)\left(1 - \frac{1}{p(\epsilon)}\left(\frac{(r-\epsilon)}{(r-\epsilon/2)}\right)^{h-1}\right) \tag{53}$$

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

$$\lim_{h \to \infty} P(h) \geq r - \epsilon \tag{54}$$

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:

$$\ln R(h; \rho^*) = K_{\ln \rho}(h) - h \ln \rho^* = h \left(\frac{K_{\ln \rho}(h)}{h} - \ln \rho^* \right) \tag{55}$$

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 $K_{\ln \rho}(h)/h$ is non-decreasing. To see this, observe that we can write for any $h, h' : h < h'$:

$$K_{\ln \rho}(h) \leq \frac{h}{h'} K_{\ln \rho}(h') + \left(1 - \frac{h}{h'}\right) K_{\ln \rho}(0) = \frac{h}{h'} K_{\ln \rho}(h') \tag{56}$$

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 $\lim_{h \to \infty} \frac{1}{h} K_{\ln \rho}(h) = \ln r > \ln \rho^*$. 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

$$P_1(h) = \frac{e_1' \mathbb{E}[A^h] X_t}{e_1' \mathbb{E}[A^{h-1}] X_t} \tag{57}$$

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

$$P_1(h) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{p} p_i \cdot \lambda_{ij}^{h} e_1' v_{ij} w_{ij}' X_t}{\sum_{i=1}^{n} \sum_{j=1}^{p} p_i \cdot \lambda_{ij}^{h-1} e_1' v_{ij} w_{ij}' X_t} \tag{58}$$

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,

$$\lambda_{kj}^{h} e_1' v_{kj} w_{kj}' X_t = o(r^h) \tag{59}$$

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

$$P_1(h) = \frac{p_i \sum_{j=1}^{p} \lambda_{ij}^{h} e_1' v_{ij} w_{ij}' X_t + o(r^h)}{p_i \sum_{j=1}^{p} \lambda_{ij}^{h-1} e_1' v_{ij} w_{ij}' X_t + o(r^{h-1})} \tag{60}$$

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,

$$\lambda_{ij}^{h} e_1' v_{ij} w_{ij}' X_t = o(r^h) \tag{61}$$

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

$$P_1(h) = \frac{r^h e_1' v_{i1} w_{i1}' X_t + o(r^h)}{r^{h-1} e_1' v_{i1} w_{i1}' X_t + o(r^{h-1})} = r \frac{e_1' v_{i1} w_{i1}' X_t + o(1)}{e_1' v_{i1} w_{i1}' X_t + o(1)} \tag{62}$$

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,

$$\lambda_{ij}^{h} e_1' v_{ij} w_{ij}' X_t = o(r^h) \tag{63}$$

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

$$P_1(h) = \frac{\lambda_{i1}^{h} e_1' v_{i1} w_{i1}' X_t + \lambda_{i2}^{h} e_1' v_{i2} w_{i2}' X_t + o(r^h)}{\lambda_{i1}^{h-1} e_1' v_{i1} w_{i1}' X_t + \lambda_{i2}^{h-1} e_1' v_{i2} w_{i2}' X_t + o(r^{h-1})} \tag{64}$$

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 $v_{i2} w_{i2}' = \bar{v_{i1}} \bar{w_{i1}}'$, 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

$$
\begin{aligned}
P_1(h) &= \frac{r^h e^{i\theta h} m e^{i\vartheta} + r^h e^{-i\theta h} m e^{-i\vartheta} + o(r^h)}{r^{h-1} e^{i\theta(h-1)} m e^{i\vartheta} + r^{h-1} e^{-i\theta(h-1)} m e^{-i\vartheta} + o(r^{h-1})} \\
&= r \frac{e^{i(\theta h+\vartheta)} + e^{-i(\theta h+\vartheta)} + o(1)}{e^{i(\theta(h-1)+\vartheta)} + e^{-i(\theta(h-1)+\vartheta)} + o(1)} \\
&= r \frac{\cos(\theta h+\vartheta) + o(1)}{\cos(\theta(h-1)+\vartheta) + o(1)}
\end{aligned} \tag{65}
$$

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

$$
R_1(h, A^*) = \frac{\mathbb{E}[A^h]_{11}}{(A^{*h})_{11}} \tag{66}
$$

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

$$
R_1(h, A^*) = \frac{\sum_{i=1}^n p_i \sum_{j=1}^p \lambda_{ij}^h (v_{ij} w_{ij}')_{11}}{\sum_{j=1}^p (\lambda_j^*)^h (v_j^* w_j^{*\prime})_{11}} \tag{67}
$$

Similar to before, *regularity implies that*

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

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

$$
R_1(h, A^*) = \frac{p_i \sum_{j=1}^p \lambda_{ij}^h (v_{ij} w_{ij}')_{11} + o(r^h)}{\sum_{j=1}^p (\lambda_j^*)^h (v_j^* w_j^{*\prime})_{11}} \tag{69}
$$

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

$$
\lambda_{ij}^h (v_{ij} w_{ij}')_{11} = o(r^h) \tag{70}
$$

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

$$
(\lambda_j^*)^h (v_j^* w_j^*)_{11} = o((\lambda_1^*)^h) \tag{71}
$$

such that

$$
R_1(h, A^*) = \frac{p_i r^h (v_{i1} w_{i1}')_{11} + o(r^h)}{(\lambda_1^*)^h (v_1^* w_1^{*\prime})_{11} + o((\lambda_1^*)^h)} = \frac{r^h}{(\lambda_1^*)^h} \frac{p_i (v_{i1} w_{i1}')_{11} + o(1)}{(v_1^* w_1^{*\prime})_{11} + o(1)} = \frac{r^h}{(\lambda_1^*)^h} O(1) \tag{72}
$$

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 $\lambda_{i2} = re^{-i\theta}$, then

$$\lambda_{ij}^h (v_{ij} w_{ij}')_{11} = o(r^h) \tag{73}$$

for $j > 2$. Similarly,

$$(\lambda_j^*)^h (v_j^* w_j^{*\prime})_{11} = o((r^*)^h) \tag{74}$$

for $j > 2$ if we write the eigenvalues of A^* with the largest absolute values as $\lambda_1^* = r^* e^{i\theta^*}$ and $\lambda_2^* = r^* e^{-i\theta^*}$. Note that this subsumes the case in which this eigenvalue is real with $\theta^* = 0$. 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

$$\begin{aligned} R_1(h, A^*) &= p_i \frac{r^h e^{i\theta h} (v_{i1} w_{i1}')_{11} + r^h e^{-i\theta h} (\overline{v}_{i1} \overline{w}_{i1}')_{11} + o(r^h)}{(r^*)^h e^{i\theta^* h} (v_1^* w_1^{*\prime})_{11} + (r^*)^h e^{-i\theta^* h} (\overline{v}_{i1}^* \overline{w}_{i1}^{*\prime})_{11} + o((r^*)^h)} \\ &= p_i \frac{r^h}{(r^*)^h} \frac{e^{i(\theta h + \phi)} m + e^{-i(\theta h + \phi)} m + o(1)}{e^{i(\theta^* h + \phi^*)} m^* + e^{-i(\theta^* h + \phi^*)} m^* + o(1)} \\ &= p_i \frac{r^h}{(r^*)^h} \frac{m}{m^*} \frac{\cos(\theta h + \phi) + o(1)}{\cos(\theta^* h + \phi^*) + o(1)} \end{aligned} \tag{75}$$

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

A.4 Proof of Corollary 3

Proof. We have that bond prices follow $p_t(h) = -\sum_{k=1}^h f_t(k) = -\sum_{k=0}^{h-1} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] x_t$. Thus, by the definition of excess returns, we have that:

$$\begin{aligned} \text{rx}_{t+1}(h) &= -\sum_{k=0}^{h-2} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] x_{t+1} + \sum_{k=0}^{h-1} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] x_t - x_t \\ &= -\sum_{k=0}^{h-2} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] (\rho_\mathbb{P}^* x_t + \varepsilon_{t+1}) + \sum_{k=0}^{h-1} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] x_t - x_t \\ &= \left(\sum_{k=1}^{h-1} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] - \rho_\mathbb{P}^* \sum_{k=0}^{h-2} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] \right) x_t + \left(-\sum_{k=0}^{h-2} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] \right) \varepsilon_{t+1} \\ &= \left(\sum_{k=1}^{h-1} \left(\mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] - \rho_\mathbb{P}^* \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^{k-1}] \right) \right) x_t + \left(-\sum_{k=0}^{h-2} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] \right) \varepsilon_{t+1} \\ &= \left(\sum_{j=1}^{h-1} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^{j-1}] \left(P(j) - \rho_\mathbb{P}^* \right) \right) x_t + \left(-\sum_{k=0}^{h-2} \mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^k] \right) \varepsilon_{t+1} \end{aligned} \tag{76}$$

The result then follows immediately from the fact that P is a strictly increasing function by Theorem 1 (as $F_{\mathbb{Q}}$ is non-degenerate). □

A.5 Proof of Corollary 4

Proof. As $b_h = f_t(h)/f_t(1)$, we have that $b_h = \mathbb{E}[\rho^{h-1}]$. Hence:

$$\text{VR}(h) = \frac{\mathbb{V}\left[\mathbb{E}^{\mathbb{Q}}[\rho_{\mathbb{Q}}^{h-1}]f_t^1\right]}{\mathbb{V}\left[\mathbb{E}^{\mathbb{Q}}[\rho_{\mathbb{Q}}]^{h-1}f_t^1\right]} = \frac{\mathbb{E}^{\mathbb{Q}}[\rho_{\mathbb{Q}}^{h-1}]^2\mathbb{V}\left[f_t^1\right]}{\mathbb{E}^{\mathbb{Q}}[\rho_{\mathbb{Q}}]^{2(h-1)}\mathbb{V}\left[f_t^1\right]} = \frac{\mathbb{E}^{\mathbb{Q}}[\rho_{\mathbb{Q}}^{h-1}]^2}{\mathbb{E}^{\mathbb{Q}}[\rho_{\mathbb{Q}}]^{2(h-1)}} = \prod_{i=2}^{h-1}\left(\frac{P(i)}{P(1)}\right)^2 \tag{77}$$

The claims of the result are then immediate from the definition of P and Equation 77 along with the properties of P characterized in Theorem 1. □

A.6 Proof of Corollary 6

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

$$x_{t+h} - \mathbb{E}_t[x_{t+h}] = (\rho^*)^h x_t + \sum_{j=0}^{h-1}(\rho^*)^j \epsilon_{t+h-j} - \mathbb{E}_F[\rho^h]x_t \tag{78}$$

Moreover, by definition we have that:

$$\beta(h)(\mathbb{E}_t[x_{t+h}] - \mathbb{E}_{t-1}[x_{t+h}]) + \gamma(h)\mathbb{E}_{t-1}[x_{t+h}] + \nu_t = \beta(h)\mathbb{E}_F[\rho^h]x_t + (\gamma(h) - \beta(h))\mathbb{E}_F[\rho^{h+1}]x_{t-1} + \nu_t \tag{79}$$

Thus, by matching coefficients, in order for Equation 48 to hold, we require that:

$$\beta(h)\mathbb{E}_F[\rho^h] = (\rho^*)^h - \mathbb{E}_F[\rho^h] \implies \beta(h) = \frac{(\rho^*)^h}{\mathbb{E}_F[\rho^h]} - 1 = \frac{1 - R(h;\rho^*)}{R(h;\rho^*)} \tag{80}$$

This verifies the guess and proves the claimed formula for $\beta(h)$. 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 $\beta(h)$ then follow from Proposition 1. □

References

Adrian, T., Crump, R. K., Diamond, P. A., and Yu, R. (2015). Discounting the long run. Liberty Street Economics, Federal Reserve Bank of New York.

Adrian, T., Crump, R. K., Diamond, P. A., and Yu, R. (2016). Forecasting interest rates over the long run. Liberty Street Economics, Federal Reserve Bank of New York.

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.

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.

Backus, D., Foresi, S., and Telmer, C. (2000). Discrete-time models of bond pricing. In Jegadeesh, N. and Tuckman, B., editors, *Advanced Fixed Income Valuation Tools*. Wiley, New York.

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

Bauer, M. D. and Rudebusch, G. D. (2020). Interest rates under falling stars. *American Economic Review*, 110(5):1316–1354.

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

Board of Governors of the Federal Reserve System (2026). Nominal yield curve. https://www.federalreserve.gov/data/nominal-yield-curve.htm. Accessed: 2026-05-05.

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

Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514.

Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1):138–160.

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.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Crump, R. K., Eusepi, S., Moench, E., and Preston, B. (2023). The term structure of expectations. In *Handbook of economic expectations*, pages 507–540. Elsevier.

d'Arienzo, D. (2020). Maturity increasing overreaction and bond market puzzles. Working Paper 3733056, SSRN.

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.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, pages 680–692.

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.

Fuster, A., Laibson, D., and Mendel, B. (2010). Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives*, 24(4):67–84.

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.

Giglio, S. and Kelly, B. (2018). Excess volatility: Beyond discount rates. *The Quarterly Journal of Economics*, 133(1):71–127.

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

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American Economic Review*, 95(1):425–436.

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

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, pages 357–384.

Hanson, S. G., Lucca, D. O., and Wright, J. H. (2021). Rate-amplifying demand and the excess sensitivity of long-term rates. *The Quarterly Journal of Economics*, 136(3):1719–1781.

Hanson, S. G. and Stein, J. C. (2015). Monetary policy and long-term real rates. *Journal of Financial Economics*, 115(3):429–448.

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.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

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*.

Molavi, P., Tahbaz-Salehi, A., and Vedolin, A. (2026). Expectations and the term structure of interest rates. Technical report, Working paper.

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

Piazzesi, M. (2010). Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques*, pages 691–766. Elsevier.

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.

Shiller, R. J. (1979). The volatility of long-term interest rates and expectations models of the term structure. *Journal of Political Economy*, 87(6):1190–1219.

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.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Vayanos, D. and Vila, J.-L. (2021). A preferred-habitat model of the term structure of interest rates. *Econometrica*, 89(1):77–112.

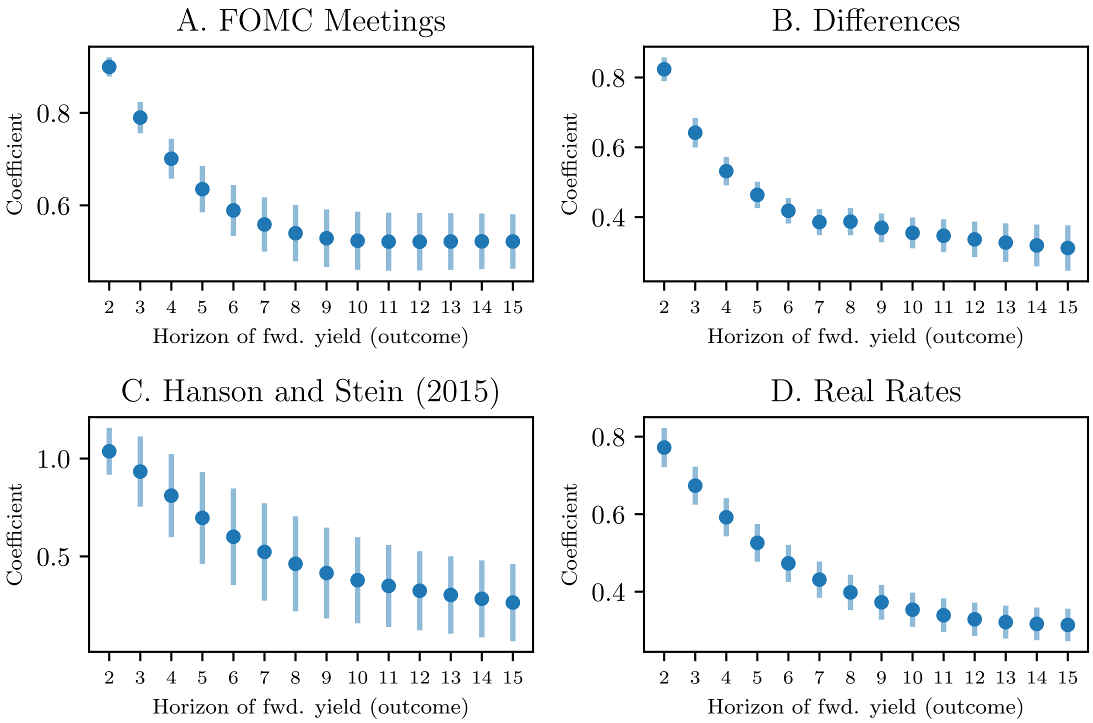
Weitzman, M. L. (1998). Why the far-distant future should be discounted at its lowest possible rate. *Journal of Environmental Economics and Management*, 36(3):201–208.

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

Supplemental Appendix to *Forecasting with Uncertain Persistence* by Flynn, Meinert, and Sastry

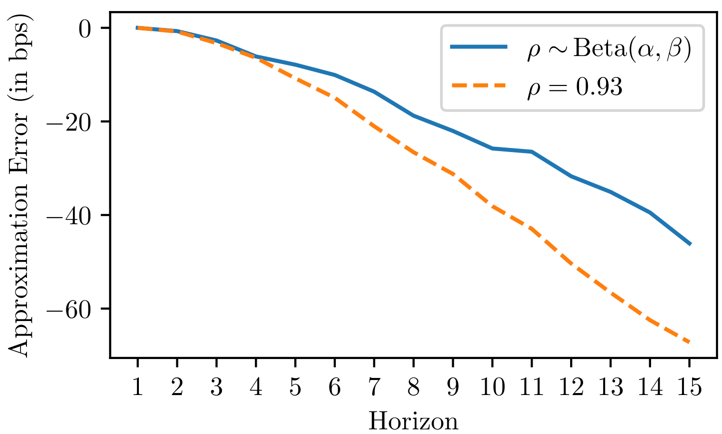
B Additional Figures

Figure A.1: Robustness of the Excessive Responsiveness of Long Rates



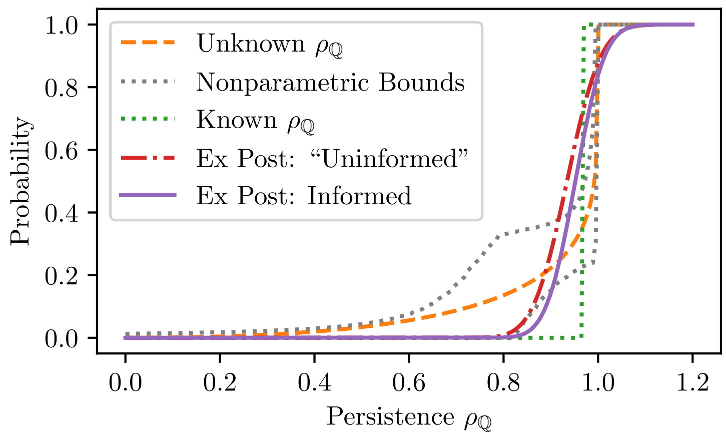
Note: This Figure shows variants of the main findings in Figure 6 regarding the excess sensitivity of long interest rates to short interest rates. In all panels, each dot is an estimated coefficient of the indicated forward yield on the one-year forward yield and the error bars are 95% confidence intervals. In Panel A, we restrict attention to the 273 FOMC meeting days in the sample from 1994 to 2026. In Panel B, we transform the outcome and regressor to two-day differences ($t + 1$ minus $t - 1$), as in Hanson and Stein (2015), on the full sample. In Panel C, we use the specification of Panel B and furthermore study the exact sample of 109 FOMC meetings from 1999 to 2012 studied by Hanson and Stein (2015). In Panel D, we plot the responsiveness of real forwards (TIPS) to the one-year nominal forward on the sample from 1999 onward (when TIPS data are available). In Panels B and C, standard errors are Eicker-Huber-White robust; in Panels A and D, they are HAC robust with a bandwidth of 20 trading days.

Figure A.2: Approximation Error for Forward Yields With and Without Uncertainty Over Persistence



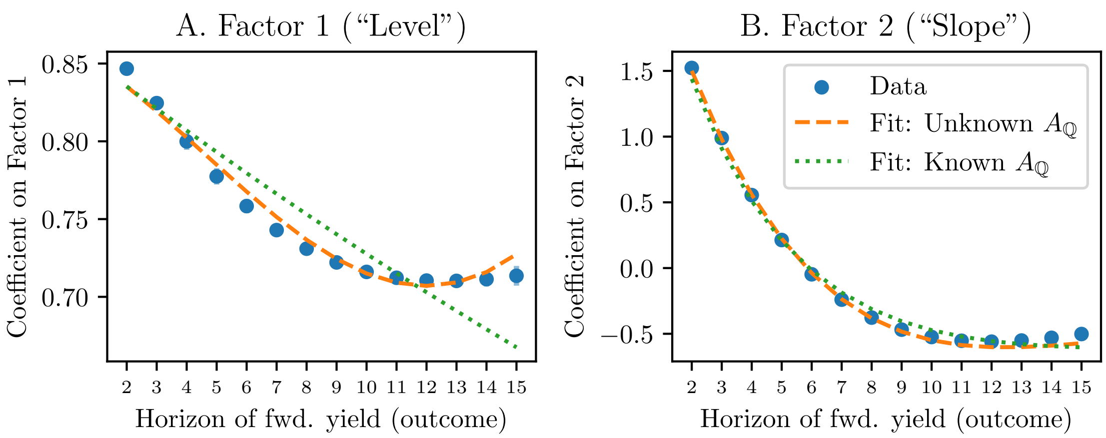
Note: This figure shows the approximation error for the forward yield defined as $\text{Err}(h) = f_t(h) - \mathbb{E}[\rho^{h-1}]x_t$ as a function of the yield horizon h. Recall that log-prices are given by $\ln P_t(h) = \ln \mathbb{E}_t^{\mathbb{Q}}[\exp\{-\sum_{j=0}^{h-1} x_{t+j}\}]$ which can be approximated to the first-order by $-\sum_{j=0}^{h-1} \mathbb{E}_t^{\mathbb{Q}}[x_{t+j}]$. The forward yield is then defined by $f_t(h) = \ln P_t(h-1) - \ln P_t(h)$ which simplifies to $f_t(h) \approx \mathbb{E}_t[\rho^{h-1}]x_t$ for the first-order approximation. We test the accuracy of this approximation under a model with uncertainty over persistence and under a model with deterministic ρ. For the former, we use a Monte Carlo simulation to compute the true forward yield under parameter values $\rho \sim \text{Beta}(\alpha, \beta)$ with $\alpha = 2.08$ and $\beta = 0.18$, $x_t = 0.03$, and $\sigma_\epsilon^2 = 0.0012$. These parameter values are taken from our regression of one-year forward yields of varying horizons on short rates (see Figure 6 for details). For the deterministic model, we use the persistence estimated from the annual time series (1972-2025), $\hat{\rho}_{\mathbb{P}} = 0.93$.

Figure A.3: Implied Beliefs About Persistence of Interest Rates



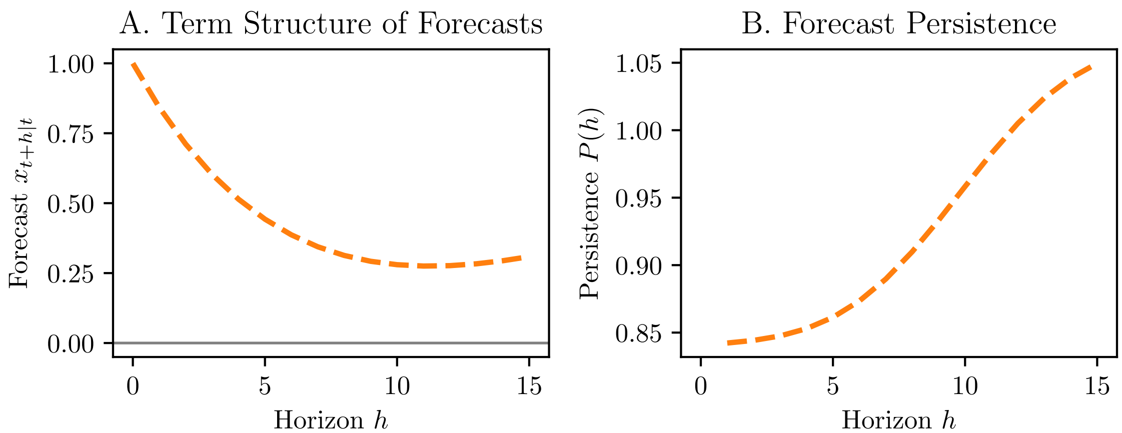
Note: This figure, which complements the summary in Table 1, reports *ex ante* and *ex post* uncertainty about the persistence of short interest rates in the form of cumulative distribution functions (CDFs). The orange dashed line corresponds to a model in which $\rho_{\mathbb{Q}} \sim \text{Beta}(2.08, 0.18)$. The grey dotted lines are nonparametric bounds on the CDF computed from moment inequalities (see Appendix D). The green dotted line corresponds to a model in which $\rho_{\mathbb{Q}} = 0.97$. The remaining lines correspond to Bayesian posterior distributions from estimating an annual AR(1) process for short interest rates from 1972 to 2026, $x_t = c + \rho x_{t-1} + \varepsilon_t$, with priors $\rho \sim N(1, \sigma_\rho^2)$, $c \sim N(0, \sigma_c^2)$, and $\varepsilon_t \sim N(0, \sigma^2)$ where $\sigma^2 \sim \text{InvGamma}(a, b)$. Under the "uninformed" prior (red dash-dot line), $\sigma_\rho = \sigma_c = 10^4$ and $a = b = 0.01$. Under the "informed" prior (purple solid line), $\sigma_\rho = 0.1$, $\sigma_c = 1.0$, $a = 2$, and $b = 1$.

Figure A.4: Fitting the Excessive Responsiveness of Long Rates in a Two-Factor Model



Note: The blue dots are regression coefficients of daily forward yields of the indicated horizon on the first factor (Panel A) and second factor (Panel B) of forward rates. The factors are estimated from principal components analysis (PCA) of daily one to fifteen year forwards, the regression model is Equation 36, and error bars are 95% confidence intervals based on HAC-robust SE (bandwidth of 20 trading days). The orange dashed line plots predicted coefficients from a model in which the two factors have unknown VAR(1) dynamics described by a coefficient matrix $A_{\mathbb{Q}}$, each element of which is an independent Gaussian. The green dotted line plots predicted coefficients with a known but potentially mis-specified model, described by some 2×2 matrix $\bar{A}_{\mathbb{Q}}$. Both sets of fitted values have parameters chosen to minimize the sum of squared residuals of model predictions versus the moments plotted in the figure.

Figure A.5: An Example VAR Setting Calibrated to the Term Structure



Note: This figure shows forecasts (Panel A) and the forecast persistence function (Panel B) in a multivariate model calibrated to match the term structure of interest rates. The model is described by Equation 34, where the forecaster believes each element of $A_{\mathbb{Q}}$ is an independent, Gaussian random variable. We set the initial state as $X_t = \frac{1}{2}[1/\Gamma_1, 1/\Gamma_2]$, so $x_t = \Gamma_1 \frac{1}{2}\frac{1}{\Gamma_1} + \Gamma_2 \frac{1}{2}\frac{1}{\Gamma_2} = 1$ and each factor has equal weight.

C Additional Proofs of Claims in Remarks

C.1 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

$$y_t = \sum_{j=1}^{q} \Phi_j y_{t-j} + \sum_{j=0}^{m} \tilde{\Sigma}_j \nu_{t-j} \tag{81}$$

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

$$X_t = \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-q+1} \end{pmatrix} \quad A = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{q-1} & \Phi_q \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \ddots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix} \quad B = \begin{pmatrix} \tilde{\Sigma}_0 & \tilde{\Sigma}_1 & \cdots & \tilde{\Sigma}_m \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{82}$$

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

$$\tilde{\nu}_t = \begin{pmatrix} \nu_t \\ \nu_{t-1} \\ \nu_{t-2} \\ \cdots \\ \nu_{t-m} \end{pmatrix} \quad F = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \ddots & 0 & 0 \\ \vdots & \vdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix} \tag{83}$$

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

$$X_t = AX_{t-1} + B\tilde{\nu}_t \tag{84}$$

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

$$\tilde{\nu}_t = F\tilde{\nu}_{t-1} + \begin{pmatrix} \nu_t \\ 0 \\ \cdots \\ 0 \end{pmatrix} \tag{85}$$

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

$$X_{t+h|t} = A^h X_t + \left(\sum_{k=1}^{h} A^{h-k} B F^k\right) \tilde{\nu}_t = A^h X_t + \left(\sum_{k=1}^{\min(m,h)} A^{h-k} B F^k\right) \tilde{\nu}_t \tag{86}$$

where the second equality follows from the fact that $F^{m+1} = 0$. Since we are interested in the asymptotic properties of forecasts, we will henceforth restrict our attention to the cases in which $h \geq m$. This implies the following expression for the forecast persistence functions $P_n(h)$:

$$P_n(h) = \frac{e_n' \mathbb{E}_F[A^h] X_t + \left(\sum_{k=1}^{m} \mathbb{E}_F[e_n' A^{h-k} B F^k]\right) \tilde{\nu}_t}{e_n' \mathbb{E}_F[A^{h-1}] X_t + \left(\sum_{k=1}^{m} \mathbb{E}_F[e_n' A^{h-k-1} B F^k]\right) \tilde{\nu}_t} \tag{87}$$

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

$$P(h) = r \frac{e^{i\theta h} \psi e^{i\vartheta} + e^{-i\theta h} \psi e^{-i\vartheta} + \sum_{k=1}^{m} \left[e^{i\theta(h-k)} \psi_k e^{i\vartheta_k} + e^{-i\theta(h-k-1)} \psi_k e^{-i\vartheta_k}\right] + o(1)}{e^{i\theta(h-1)} \psi e^{i\vartheta} + e^{-i\theta(h-1)} \psi e^{-i\vartheta} + \sum_{k=1}^{m} \left[e^{i\theta(h-k-1)} \psi_k e^{i\vartheta_k} + e^{-i\theta(h-k-1)} \psi_k e^{-i\vartheta_k}\right] + o(1)} \tag{88}$$

where

$$\begin{aligned} \psi &= |e_n' v_{i1} w_{i1}' X_t| & \vartheta &= \arg(e_n' v_{i1} w_{i1}' X_t) \\ \psi_k &= |e_n' v_{i1} w_{i1}' \mathbb{E}[B|A_i] F^k \tilde{\nu}_t| & \vartheta_k &= \arg(e_n' v_{i1} w_{i1}' \mathbb{E}[B|A_i] F^k \tilde{\nu}_t) \end{aligned} \tag{89}$$

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

$$P_n(h) = r \frac{\psi \cos(\theta h + \vartheta) + \sum_{k=1}^{m} \psi_k \cos(\theta(h-k) + \vartheta_k) + o(1)}{\psi \cos(\theta(h-1) + \vartheta) + \sum_{k=1}^{m} \psi_k \cos(\theta(h-k-1) + \vartheta_k) + o(1)} \tag{90}$$

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.

C.2 Proof of Claims in Remark 3

Proof. To save space, we write the arguments for the AR(1) case. The exact steps of Theorem 2 can be applied to generalize this to the general case, where X_t is replaced with $\mathbb{E}_t[X_t|A_i = A_1]$. First, recall that the forecast $x_{t+h|t}$ can be written as

$$x_{t+h|t} = \mathbb{E}_t[x_{t+h}] = \mathbb{E}_t[\mathbb{E}_t[x_{t+h}|\rho]] = \mathbb{E}_t[\rho^h x_t] = \mathbb{E}_t[\rho^h \mathbb{E}_t[x_t|\rho]] \tag{91}$$

If the prior F is regular, then we can write the forecast as

$$x_{t+h|t} = \sum_{i=1}^{n} p_i \cdot \rho_i^h \mathbb{E}_t[x_t|\rho_i] \tag{92}$$

This gives rise to the following expression for the forecast persistence function

$$P(h) = \frac{\sum_{i=1}^{n} p_i \cdot \rho_i^h \mathbb{E}_t[x_t|\rho_i]}{\sum_{i=1}^{n} p_i \cdot \rho_i^{h-1} \mathbb{E}_t[x_t|\rho_i]} \tag{93}$$

Define $r := \max_{j \in \{1,...n\}} \rho_j$. By assumption, $r > \rho_i$ for all, but one $i \in \{1, ..., n\}$. Then

$$P(h) = r \frac{\sum_{i=1}^{n} p_i \frac{\rho_i^h}{r^h} \mathbb{E}_t[x_t|\rho_i]}{\sum_{i=1}^{n} p_i \frac{\rho_i^{h-1}}{r^{h-1}} \mathbb{E}_t[x_t|\rho_i]} = r \frac{p_i \mathbb{E}_t[x_t|r] + o(1)}{p_i \mathbb{E}_t[x_t|r] + o(1)} \tag{94}$$

Hence, if $\mathbb{E}_t[x_t|r] \neq 0$, we have

$$\lim_{h \to \infty} P(h) = r \tag{95}$$

An analogous argument yields the forecast reaction function. □

C.3 Proof of Claims in Remark 4

Proof. Suppose x_t follows an AR(1) process with persistence ρ and long-run mean μ

$$x_t = \rho x_{t-1} + (1 - \rho)\mu + \varepsilon_t \tag{96}$$

where ε_t is independently and identically distributed according to some distribution G. If the forecaster is certain about the persistence ρ and the long-run mean μ, the forecast of x_{t+h} for some $h > 0$ is given by

$$x_{t+h|t} = \rho^h x_t + \mu(1 - \rho) \sum_{k=0}^{h-1} \rho^k = \rho^h x_t + \mu(1 - \rho^h) \tag{97}$$

Suppose the forecaster has prior F over the joint distribution of μ and ρ. We further assume that the support of ρ is finite. We define the demeaned forecast

$$\hat{x}_{t+h|t} = x_{t+h|t} - \mathbb{E}[\mu] \tag{98}$$

Then the forecast persistence function of the demeaned forecast, $\hat{P}(h)$, is given by

$$\begin{aligned}\hat{P}(h) &= \frac{\mathbb{E}[\rho^h]x_t - \mathbb{E}[\rho^h \mu]}{\mathbb{E}[\rho^{h-1}]x_t - \mathbb{E}[\rho^{h-1}\mu]} \\ &= \frac{\sum_{i=1}^n p_i \rho_i^h x_t - \sum_{i=1}^n p_i \rho_i^h \mathbb{E}[\mu|\rho_i]}{\sum_{i=1}^n p_i \rho_i^{h-1} x_t - \sum_{i=1}^n p_i \rho_i^{h-1} \mathbb{E}[\mu|\rho_i]}\end{aligned} \tag{99}$$

The essential supremum of ρ, r, is defined as $r = \max_{i\in\{1,\dots,n\}} \rho_i$. Then

$$\begin{aligned}\hat{P}(h) &= r\frac{\sum_{i=1}^n \frac{p_i\rho_i^h}{r^h} x_t - \sum_{i=1}^n \frac{p_i\rho_i^h}{r^h} \mathbb{E}[\mu|\rho_i]}{\sum_{i=1}^n \frac{p_i\rho_i^{h-1}}{r^{h-1}} x_t - \sum_{i=1}^n \frac{p_i\rho_i^{h-1}}{r^{h-1}} \mathbb{E}[\mu|\rho_i]} \\ &= r\frac{x_t - \mathbb{E}[\mu|r] + o(1)}{x_t - \mathbb{E}[\mu|r] + o(1)}\end{aligned} \tag{100}$$

which concludes our proof that

$$\lim_{h\to\infty} \hat{P}(h) = r \tag{101}$$

if $x_t \neq \mathbb{E}[\mu|r]$. Note that it is important to demean by $\mathbb{E}[\mu]$, i.e., the expected mean under the prior F, which need not coincide with the true mean of the process μ. □

D Additional Details for the Empirical Analysis

In Section 4.1, we use two methods to estimate the subjective uncertainty regarding the persistence of interest rates that rationalizes the co-movement of short- and long interest rates. We use empirical estimates from the following regression models

$$f_t(h) = \alpha(h) + \beta(h) \cdot f_t(1) + \varepsilon_t \tag{102}$$

where $f_t(h)$ denotes the date t forward nominal Treasury rate at a horizon of h years. We estimate this separately for each horizon h between 2 and 15 years using the daily-frequency Gürkaynak et al. (2007) Treasury yield curve, as reported by Board of Governors of the Federal Reserve System (2026). We observe that, in the theory, the coefficient $\beta(h)$ corresponds to $\mathbb{E}_{F_\mathbb{Q}}[\rho_\mathbb{Q}^{h-1}]$.

For the parametric method, we fit a Beta distribution by method-of-moments. That is, we choose shape parameters $\hat{a}$ and $\hat{b}$ such that

$$(\hat{a}, \hat{b}) \in \arg\min_{\mathbb{R}_+^2} \left\{ \sum_{j=1}^{14} \left(\hat{\beta}(h+1) - \frac{B(a+h, b)}{B(a,b)} \right)^2 \right\} \tag{103}$$

where the $\hat{\beta}(h)$ are our empirical estimates, B denotes the Beta function and $\frac{B(a+h,b)}{B(a,b)}$ is a closed-form expression for $\mathbb{E}[x^h]$ for $h > 0$ when $x \sim \text{Beta}(a,b)$.

For the non-parametric method, we solve linear programs based on the moment inequalities implied by our estimates. In particular, we consider candidate distributions for $\rho_{\mathbb{Q}}$ supported on a grid of N points, $X = \{x_1, \ldots, x_N\} \subset [0,1]$, such that, evaluated at these distributions, the moment conditions approximately hold. Formally, given some tolerance ε, we calculate the upper bound CDF at each point $x \in X$ as

$$\begin{aligned} \hat{F}^{\max}(x) = \max_{(p_j)_{j=1}^N} & \sum_{j:x_j \leq x} p_j \\ & \hat{\beta}(h+1) - \varepsilon \leq \sum_{j=1}^N p_j x_j^h \leq \hat{\beta}(h+1) + \varepsilon, \ \ \forall h \in \{1, \ldots, 14\} \\ & \sum_{j=1}^N p_j = 1, \ p_j \geq 0 \, \forall j \in \{1, \ldots, N\} \end{aligned} \tag{104}$$

and the lower-bound CDF via the same program with a minimum in place of the maximum,

$$\begin{aligned} \hat{F}^{\min}(x) = \min_{(p_j)_{j=1}^N} & \sum_{j:x_j \leq x} p_j \\ & \hat{\beta}(h+1) - \varepsilon \leq \sum_{j=1}^N p_j x_j^h \leq \hat{\beta}(h+1) + \varepsilon, \ \ \forall h \in \{1, \ldots, 14\} \\ & \sum_{j=1}^N p_j = 1, \ p_j \geq 0 \, \forall j \in \{1, \ldots, N\} \end{aligned} \tag{105}$$

In practice, we use a grid of 200 evenly spaced points between 0 and 1 and a tolerance of $\varepsilon = 0.02$. Up to numerical tolerance, this procedure provides pointwise upper and lower bounds on the CDFs of distributions that can be consistent with the estimated moments.

E Investments With Uncertain Persistence

In this appendix, we illustrate 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.

E.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, $h \in \{0, 1, \dots, H\}$. 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:

$$\theta_t = \rho\theta_{t-1} + \varepsilon_t, \qquad \forall t \geq 1 \tag{106}$$

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:

$$\max_{(x_h)_{h=0}^{T} \in \mathbb{R}^H} \sum_{h=0}^{H} \left(\mathbb{E}\left[\delta^h \theta_h x_h\right] - p_h x_h \right) \tag{107}$$

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 taking 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 a 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.

E.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:

$$x_h = \left(\theta_0 \delta^h \mathbb{E}[\rho^h]\right)^\chi, \qquad p_h = \theta_0 \delta^h \mathbb{E}[\rho^h], \qquad \forall h \in \{0, \dots, H\} \tag{108}$$

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:

$$\frac{x_h}{x_{h-1}} = \delta^\chi P(h)^\chi, \qquad \frac{p_h}{p_{h-1}} = \delta P(h), \qquad \text{where } P(h) = \frac{\mathbb{E}[\rho^h]}{\mathbb{E}[\rho^{h-1}]} \tag{109}$$

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 7 (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 109. □

We unpack these three statements in turn. First, because $P(1) = \mathbb{E}[\rho]$, both quantities and prices of investment are initially decreasing with the 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 Beta distribution: that is, $\rho \sim rz$ where $r > 0$ is a scaling parameter and $z \sim \text{Beta}(\alpha, \beta)$. In this case, the forecast persistence function has a known form derived in Equation 7, but now scaled by the parameter r. As a result, we can calculate the term structure of investments and prices in closed form:

$$\frac{x_h}{x_{h-1}} = \delta^\chi \left(r \frac{h + \alpha - 1}{h + \alpha + \beta - 1} \right)^\chi, \qquad \frac{p_h}{p_{h-1}} = \delta r \frac{h + \alpha - 1}{h + \alpha + \beta - 1} \tag{110}$$

We visualize the resulting optimal investments in Figure E.1, varying the shape of the Beta 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.

E.3 Boom-Bust Investment Cycles

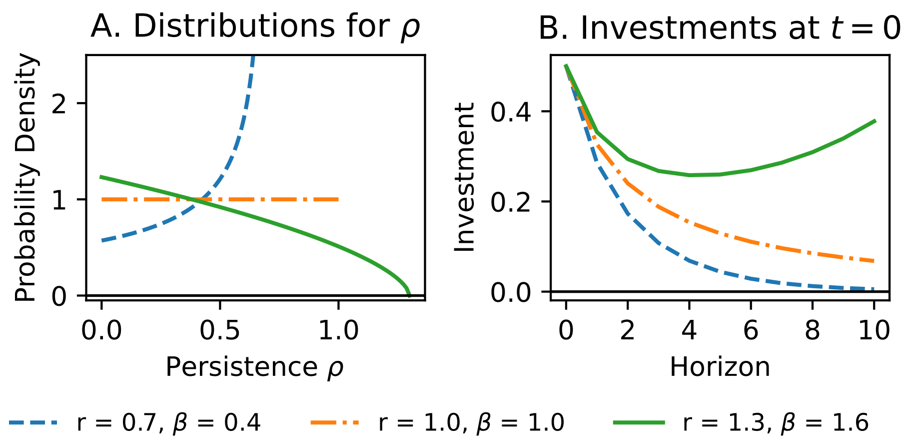
We finally finish with a simple illustration of how this model can generate endogenous *investment cycles*. To do this, we consider a dynamic extension of the model in which, in each time period $t \in \mathbb{N}$, the investor considers projects that mature at each future period $\{t, \ldots, t + H\}$. 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 this date, or the capital stock K_t:

$$Y_t = \theta_t K_t, \qquad K_t = \sum_{h=0}^{H} x_{t-h,h} \tag{111}$$

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

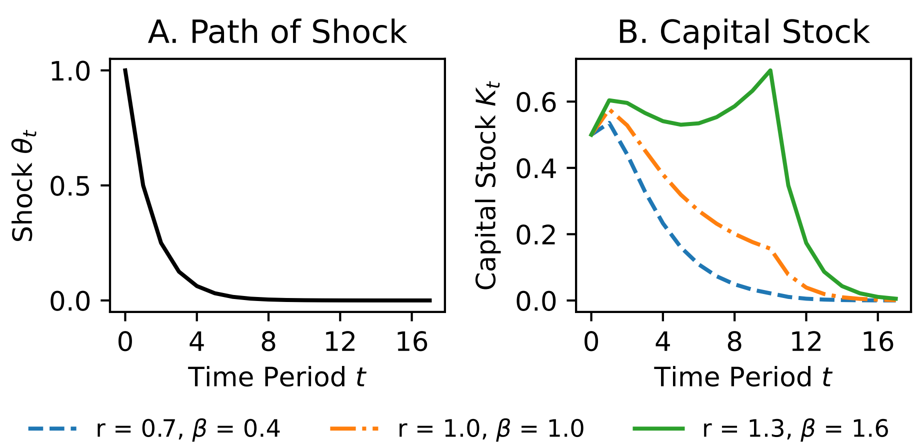
Figure E.2 shows the output dynamics following a transitory shock to productivity θ_t. The presence of investment opportunities at multiple horizons "backloads" the effect on output in all cases. In the third model, which generated a U-shaped term structure of

Figure E.1: Investments When Payoffs Have Unknown Persistence



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

Figure E.2: The Path of the Aggregate Capital Stock in Response to a Shock



Note: This figure shows investment dynamics after a shock in the investment model of E. 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 111. We vary the investor's subjective probability distribution regarding ρ. Each is a scaled Beta distribution with $\alpha = 1$ and mean $\mathbb{E}[\rho] = 0.5$, but different values of r and β (see Panel A of Figure E.1) 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 $\delta = 0.98$.

investments (recall Figure E.1), 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 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.

References

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

Board of Governors of the Federal Reserve System (2026). Nominal yield curve. `https://www.federalreserve.gov/data/nominal-yield-curve.htm`. Accessed: 2026-05-05.

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

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.

Hanson, S. G. and Stein, J. C. (2015). Monetary policy and long-term real rates. *Journal of Financial Economics*, 115(3):429–448.

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