STAFF REPORTS

NO. 1098
APRIL 2024

REVISED
NOVEMBER 2024

Is There Hope for the Expectations Hypothesis?

Richard K. Crump | Stefano Eusepi | Emanuel Moench

FEDERAL RESERVE BANK of NEW YORK

Is There Hope for the Expectations Hypothesis?
Richard K. Crump, Stefano Eusepi, and Emanuel Moench
Federal Reserve Bank of New York Staff Reports, no. 1098
April 2024; revised November 2024
https://doi.org/10.59576/sr.1098

Abstract

Most macroeconomic models impose a tight link between expected future short rates and the term structure of interest rates via the expectations hypothesis (EH). While systematically rejected in the data, existing work evaluating the EH generally assumes either full-information rational expectations or stationarity of beliefs, or both. As such, these analyses are ill-equipped to refute the EH when these assumptions fail to hold, leaving the door open for a "resurrection" of the EH. We introduce a model of expectations formation which features time-varying means and accommodates deviations from rationality. This model tightly matches the entire joint term structure of expectations for output growth, inflation, and the short-term interest rate from all surveys of professional forecasters in the U.S. We show that deviations from rationality and drifting long-run beliefs consistent with observed measures of expectations, while sizable, do not come close to bridging the gap between the term structure of expectations and the term structure of interest rates. Not only is the EH decisively rejected in the data, but model-implied short-rate expectations generally display, at best, only a weak co-movement with the forward rates of corresponding maturities.

JEL classification: D84, E43, G12
Key words: expectations formation, survey forecasts, expectations hypothesis

---

Crump: Federal Reserve Bank of New York (email: richard.crump@ny.frb.org). Eusepi: Brown University (email: stefano_eusepi@brown.edu). Moench: Frankfurt School of Finance & Management, CEPR (email: e.moench@fs.de). This paper substantially builds on and complements previous work circulated as "The Term Structure of Expectations and Bond Yields" (Crump, Eusepi, and Moench 2018). The authors thank Tobias Adrian, John Cochrane, Keshav Dogra, Marco Del Negro, Domenico Giannone, Sam Hanson, Albert Marcet, Emi Nakamura, Bruce Preston, Giorgio Primiceri, Harald Uhlig, Jonathan Wright, and participants at various seminars, workshops, and conferences for helpful comments and discussions. They also thank Kenneth Froot for sharing the Goldsmith-Nagan survey data.

1 Introduction

The expectations hypothesis of the term structure of interest rates states that yields on government bonds reflect the average short rate expected to prevail over the life of the bond. Since the 1930s, the expectations hypothesis (EH) has been the natural starting point for linking longer-term yields to short-term yields (e.g., Lutz 1940). To this day, it remains a maintained assumption in most macroeconomic models, where the monetary transmission mechanism relies on a tight relationship between expectations of future short rates and longer-term interest rates (e.g., Woodford 2003). The preeminent role of the EH stands in tension with the overwhelming empirical evidence stacked against it.

However, analyses of the EH are only as good as the expectations-formation process they are based on. Friedman (1979) and Froot (1989) first emphasized that standard tests of the EH are, in fact, *joint* tests of the EH and full information rational expectations (FIRE). Campbell and Shiller (1991) find that when the term spread is high, long-term rates tend to fall, rather than rise, as the theory would suggest. However, these patterns are perfectly consistent with the EH, under bounded rationality and learning (Sinha 2016; Farmer, Nakamura, and Steinsson 2023). Furthermore, economic agents are commonly assumed to operate in a stationary environment, where they quickly come to understand the long-run behavior of the economy. Their corresponding long-run expectations, therefore, are stable and, under the EH, would fail to match the substantial variability in longer-maturity interest rates. Conversely, models where agents constantly shift their views about the monetary regime (e.g., the central bank objectives) can produce long-term rates entirely consistent with the EH (Fuhrer 1996; Kozicki and Tinsley 2001; Gürkaynak, Sack, and Swanson 2005).[1]

In this paper, we reevaluate the empirical evidence regarding the EH by proposing a multivariate model of expectations formation that allows for deviations from FIRE and accounts for time-varying beliefs about the long-run. In the model, the agent forms forecasts about the path of the short-term interest rate based on noisy signals about the true state of the economy. Based on this information, they estimate a non-stationary trend and a stationary cyclical component and form forecasts based on their joint dynamics. Importantly, we do not impose that subjective expectations coincide with those based on the true data-generating process.

We make two contributions to the existing literature. First, we show that the class of models we propose is consistent with the whole term structure of survey-based forecasts. We estimate the process of beliefs updating using the *universe* of consensus forecasts from all U.S.

[1]Observed long-rates display "excess volatility" relative to those predicted by the EH (Shiller 1979; Campbell and Shiller 1991).

surveys of professional forecasters covering more than 600 survey-horizon pairs at a monthly frequency. This simple expectations formation model provides a tight fit to forecasts of real GDP growth, CPI inflation, and the 3-month Treasury bill for all horizons from the very-short term (current quarter) to the long-run (11 years and beyond). We show that signals about inflation and output are an important factor shaping their interest-rate expectations which is consistent with conventional views about central bank reaction functions (Andrade, Crump, Eusepi, and Moench 2016). This implication would be notably absent in univariate models; moreover, we demonstrate that a univariate version of our model fails to match the behavior of observed forecasts of the short-term interest rate.

We provide supporting evidence in favor of our expectations formation process. For example, revisions of long-horizon expectations strongly co-move with realized short-term forecast errors, which are not used in the model's estimation. This indicates that forecasts are revised in response to new information. Furthermore, these forecast revisions result in substantial variability of long-horizon forecasts even as compared to those of statistical models such as Laubach and Williams (2003) and Del Negro, Giannone, Giannoni, and Tambalotti (2018). The estimated model also exhibits meaningful deviations from full rationality. Following the approach of Coibion and Gorodnichenko (2015) we exploit our rich dataset to explore under-and over-reactions of forecasts to new information for both short- *and* long-forecast horizons. While interest rate forecasts at short-horizon under-react to new information, at longer horizons we find substantial over-reaction. Finally, as in Sinha (2016) and Farmer, Nakamura, and Steinsson (2023), treating data generated from the model as observed yields (i.e., imposing the EH), would nonetheless lead an empirical researcher to frequently reject the EH using standard statistical tests.

Our second contribution is to show that deviations from rationality and drifting long-run beliefs are sizable, and yet, they still do not come close to bridging the gap between the model-implied term structure of expectations and the term structure of interest rates. Far from resurrecting the EH, the tight connection between short-term interest rate expectations and the term structure of interest rates, assumed to hold in theory, demonstrably fails to hold in practice. Expected interest rates beyond two years have, at best, only a weak co-movement with forward rates of the corresponding maturities. In fact, the correlation between changes in longer-term forward rates and corresponding longer-horizon short rate forecasts converges towards zero as the maturity increases to ten years. It is then unsurprising that formal tests in the spirit of Froot (1989), using our model-implied expectations, result in decisive rejections of the EH. Importantly, these tests do not require any assumption about the expectations formation mechanism (e.g., do not require rational expectations).

The flip side of our results is that the wedge between observed yields and expected future short-term interest rates captures the vast majority of yield variability at medium and long maturities. In models where agents are risk averse, this wedge represents time-varying compensation for bearing risk. However, we show that this wedge is only partially explained by the underlying factors shaping beliefs about the state of the economy. This implies that any model designed to explain both the term structure of short rate expectations and the term structure of interest rates would need to involve additional drivers. In turn, these drivers represent a channel of first-order importance to better understand the effect of the term structure of interest rates on the aggregate economy that is absent in most macroeconomic models.

Related Literature This paper is related to an exploding literature using survey data to better understand economic agents' expectations formation process.[2] While the literature has not settled on a unified framework, a few lessons have emerged: survey forecasts deviate systematically from FIRE predictions, resulting in predictable forecast errors and over- and under-reaction to new information.[3] Moreover, professional forecasters' interest rate expectations, which are formed in real time, behave quite differently compared to forecasts available to an econometrician observing the full sample.[4] This indicates that agents learn about the economic environment.

Models where agents operate in a non-stationary environment, and have to constantly adapt to structural change, can generate sizable fluctuations in long-horizon expectations. For example, Fuhrer (1996) and Kozicki and Tinsley (2001) show that interest rate expectations consistent with monetary policy regime shifts during the US post-war period can generate movements in yields under the EH that well approximate the variation in observed yields. Similarly, Cogley (2005) proposes a model of the term structure of interest rates based on a VAR with drifting coefficients and shows that the EH cannot be rejected in this environment.[5] More recently, using models of bounded rationality and learning, Sinha (2016) and Farmer, Nakamura, and Steinsson (2023) provide examples where the EH would be rejected in the data even when it holds in the true data generating process. Sinha (2016) explores a New Keynesian model with long-term government bonds and bounded rationality, where agents learn about the long-run evolution of the economy. Farmer, Nakamura, and

[2]See Angeletos, Huo, and Sastry (2020) and Eusepi and Preston (2023) for recent surveys.

[3]See, for example, Eusepi and Preston (2011), Coibion and Gorodnichenko (2012), Bordalo, Gennaioli, Ma, and Shleifer (2020), and Angeletos, Huo, and Sastry (2020).

[4]See Friedman (1979), Piazzesi, Salomao, and Schneider (2015), Cieslak (2018), Singleton (2021), and Eusepi, Giannoni, and Preston (2024).

[5]Similarly, Dewachter, Iania, and Lyrio (2011) propose a structural model where agents learn about the inflation target and show that this model, based solely on macro-factors provides a better fit of observed yields than rational expectations models.

Steinsson (2023) propose a univariate trend-cycle model for the short-term interest rate. They show that their calibrated model matches existing stylized facts capturing departures from FIRE using short-term survey forecasts out to four quarters ahead.

Our results stand in stark contrast to the conclusions from this literature. Although imperfect information and learning has the potential to deliver EH-implied yields with similar features as observed yields, we show that one cannot simultaneously fit the term structure of interest rates and the term structure of short-rate expectations as measured from professional surveys. Put differently, the expected path of short-term interest rates using forward rates as the "market expectation" is very far from the expected path of short-term interest rates using survey expectations.

The argument that deviations from full-information rational expectations could invalidate tests of the EH goes back to Friedman (1979) and Froot (1989). While these earlier seminal papers use survey data to explore the validity of the EH without imposing any assumptions about the expectations formation process, limited data hindered their ability to reach clear-cut conclusions. We are the first to document the properties of survey forecasts beyond short-term forecast horizons. This is crucial as we show that the failure of the EH is most acute at forecast horizons many years out. Furthermore, our analysis expands on these early approaches by using multiple surveys for all available forecast horizons over a much longer sample.

Survey expectations of interest rates have also been used in conjunction with no-arbitrage term structure models to analyze the behavior of yields. Kim and Wright (2005) and Kim and Orphanides (2012) employ survey forecasts of the nominal short rate at a few select horizons to discipline the dynamics of the state variables under the historical measure in small samples. Wright (2011) uses expectations from an affine term structure model and survey-based expectations to study the global decline in yields beginning in the 1980s. Piazzesi, Salomao, and Schneider (2015) use survey forecasts of the short rate, inflation, and of longer-term Treasuries to disentangle subjective (i.e. survey forecasters') beliefs and objective (i.e. those of a statistician endowed with full-sample information) beliefs. In contrast to these papers, we generate the entire term structure of short-rate expectations without relying on information from yields.

In summary, we make the following distinct contributions relative to the existing literature. First, the convention in the study of expectations is to evaluate expectations formation for a single variable (e.g., Coibion and Gorodnichenko 2012, Bordalo, Gennaioli, Ma, and Shleifer 2020 and Farmer, Nakamura, and Steinsson 2023). In contrast, we present a multivariate model of expectations formation and show that modeling short-term interest rate forecasts along with those of output growth and inflation is essential to match the observed term structure of all survey forecasts—especially those for the short-term interest rate. Second, the vast majority of the aforementioned literature focuses on stationary models and the study of short-term expectations (see Crump, Eusepi, Moench, and Preston 2023 for a survey). However, observed subjective long-horizon expectations from survey data exhibit substantial time variation consistent with economic agents facing considerable uncertainty about the long-run behavior of the economy (e.g., Crump, Eusepi, Moench, and Preston 2023, 2024). Here, we propose a model in which agents learn about both the short-run *and* the long-run and show that it fits the survey data at all available forecast horizons. Such a model also produces over-reaction to new information of long-horizon forecasts, consistent with available experimental evidence (Afrouzi et al. 2023). Third, and most importantly, we show that once we discipline the expectations formation process using survey expectations, we find no evidence that imperfect information and learning can vindicate the EH.

The remainder of this paper is structured as follows. In Section 2, we introduce our model of expectations formation. In Section 3, we discuss our data and estimation approach, document the tight fit to the observed term structure of nominal short rate forecasts, and demonstrate other appealing properties of the model. In Section 4 we compare the term structure of short-rate expectations to the term structure of interest rates, formally test for the EH, and provide our main results. We end this section with a conceptual discussion comparing the average investor and the marginal investor in the bond market. Section 5 concludes. In the Appendix we present some additional details of the model and a Supplementary Appendix (hereafter, "SA1") collects additional results and robustness checks.

2 A Model of Expectations Formation

Consider an agent who observes the nominal short-term interest rate directly along with noisy signals of other key macroeconomic variables. The agent views the aggregate state of the economy to be defined by $z_t = (g_t, \pi_t, i_t)'$ where g_t is real output growth, π_t is price inflation, and i_t is the short-term nominal interest rate. The perceived law of motion for z_t is

$$z_t = \omega_t + x_t \tag{2.1}$$

where

$$\omega_t = \omega_{t-1} + \eta_t \tag{2.2}$$

is the *trend* component and

$$x_t = \Phi x_{t-1} + \nu_t \tag{2.3}$$

is a stationary *cycle* component. The innovations, η_t and ν_t, are perceived to be *i.i.d.* Gaussian innovations with joint variance-covariance matrix Σ_z. We assume that the agent perceives the trend component to be slow-moving as compared to the stationary component, implying that innovations to the former have a much smaller variance than innovations to the latter.

To form expectations, the agent requires an estimate of the unobserved components x_t and ω_t. However, the agent faces two important informational constraints. First, they can only observe noisy measures of real output growth and inflation as only the short-term interest rate is fully known at any point in time. This constraint is consistent with the fact that economic variables such as real GDP and inflation are released with a delay and feature sizable and significant subsequent revisions. In addition, observed measures of price inflation are contaminated with volatile sub-components that mask underlying inflation, π_t, which is the relevant state variable for forecasting.

Second, agents face an additional signal extraction problem. They have to infer to what extent changes in observed data are due to transitory shocks related to the business cycle, x_t, or reflect shifts in the slow-moving trend components, as captured by the innovation to ω_t. The latter component may reflect regime changes such as long-term shifts in productivity, the savings rate or fiscal policy, which directly impact the evolution of the long term real rate of interest, or shifts in the perceived long-run mean of inflation, reflecting perceived credibility of monetary policy.

We assume that the agent receives two sets of signals, collected in the $m \times 1$ vector $\mathcal{S}_t$. The first set of signals represents the noisy signals of real output growth and inflation coming from economic releases and the short-term interest rate which is perfectly observed. The agent also observes additional noisy signals on the sub-components, x_t and ω_t. This second set of signals captures information from alternative channels such as other data releases or central bank communications. For example, forward guidance about the short-term path of the policy rate (x_t) or central-bank announcements that alter perceptions about the inflation target (ω_t). The mapping between the agent's model and observed signals is then summarized by the following observation equation

$$\mathcal{S}_t = H' \begin{pmatrix} \omega_t \\ x_t \end{pmatrix} + s_t, \tag{2.4}$$

where H is an $6 \times m$ matrix with m the number of signals and s_t is a vector of measurement errors perceived to be *i.i.d.* Gaussian with variance-covariance matrix Σ_s. The agent uses the Kalman filter to estimate the latent trend and cycle components,

$$\begin{pmatrix} \omega_{t|t} \\ x_{t|t} \end{pmatrix} = F \begin{pmatrix} \omega_{t-1|t-1} \\ x_{t-1|t-1} \end{pmatrix} + f_t, \qquad F = \begin{bmatrix} I_3 & 0 \\ 0 & \Phi \end{bmatrix}, \tag{2.5}$$

where

$$f_t \equiv \mathcal{K}\left(\mathcal{S}_t - \mathcal{S}_{t|t-1}\right), \qquad \mathcal{S}_{t|t-1} = H'F \begin{pmatrix} \omega_{t-1|t-1} \\ x_{t-1|t-1} \end{pmatrix}. \tag{2.6}$$

In words, we assume the agent observes $\mathcal{S}_t$, posits a multivariate trend-cycle model, and uses the Kalman filter to construct forecasts. The current estimates of x_t and ω_t depend on their previous estimates, along with f_t, which is the "surprise" relative to the agent's prediction, $\left(\mathcal{S}_t - \mathcal{S}_{t|t-1}\right)$, scaled by the $6 \times m$ steady-state Kalman matrix $\mathcal{K}$. Given current estimates, $x_{t|t}$ and $\omega_{t|t}$, the agent then generates the *full* term structure of expectations at all forecast horizons, h, via

$$\mathbb{E}_t\left[z_{t+h}\right] \equiv z_{t+h|t} = \omega_{t|t} + \Phi^h x_{t|t}. \tag{2.7}$$

Importantly, at any point in time, the term structure of expectations is determined only by $x_{t|t}$, $\omega_{t|t}$ and Φ.

As econometricians, we observe survey forecasts but we cannot directly observe $x_{t|t}$ and $\omega_{t|t}$. In order to estimate $x_{t|t}$ and $\omega_{t|t}$ from the survey data we need to designate their joint dynamic evolution. From equation (2.5), this requires an assumption on the behavior of f_t.

2.1 Subjective Beliefs and the True Law of Motion

If rational expectations were to hold then the agent's perceived law of motion, as governed by equations (2.1)–(2.3), coincides exactly with the true data generating process of z_t. In this case, f_t is an *i.i.d.* Gaussian process with variance-covariance matrix tied directly to the model's parameters (see Appendix A for details).

However, the agent's subjective beliefs need not coincide with the true law of motion of the economy. We would like to allow for the possibility that their statistical model for $\mathcal{S}_t$, as described above, is an approximation (based on incomplete knowledge) of a complex and changing economic environment. Under these conditions, the process f_t is unlikely to satisfy the restrictions imposed by rational expectations. As econometricians, we want to avoid taking a stand on exactly how many or which signals, $\mathcal{S}_t$, the agent observes and their associated true data generating process. Instead, we accommodate possible model mis-specification and/or deviations from full rationality by allowing the forecast errors, f_t, to be autocorrelated across time. In addition, we do not impose that the variance-covariance matrix of the innovation to the forecast errors satisfies the restrictions imposed by rational expectations. Instead, we assume

$$f_t = G f_{t-1} + \varepsilon_t, \tag{2.8}$$

where ε_t is *i.i.d.* Gaussian white noise with full variance-covariance matrix Σ_ε and $G = (I_2 \otimes \phi_f)$ with $\phi_f = \text{diag}\left(\rho_{fg}, \rho_{f\pi}, \rho_{fi}\right)$ where $\otimes$ denotes the Kronecker product.[6]

By equation (2.8), we then have that the forecast errors for the true state of the economy, $z_{t|t} - z_{t|t-1}$, will be serially correlated with first-order autocorrelation coefficients of ρ_{fg}, $\rho_{f\pi}$, and ρ_{fi}, respectively. This modeling assumption aims at capturing the widely documented pattern of serial correlation in forecast errors and under-reaction to shocks in consensus survey-based forecasts (see section 3.3.3 below). This pattern is consistent with different (and possibly complementary) information frictions, such as noisy information models (e.g., Coibion and Gorodnichenko 2012; Andrade, Crump, Eusepi, and Moench 2016; Angeletos et al. 2020) and models of imperfect knowledge (e.g., Eusepi and Preston 2011). It is important to emphasize that the dynamic properties of f_t reflect (possible) model mis-specification: agents form expectations under the assumption their model is correct, i.e., that $\mathbb{E}_t\left[f_{t+1}\right] = 0$, and so forecasts satisfy equation (2.7), even when G is a nonzero matrix.

The specification in equation (2.8) is chosen for its combination of flexibility and parsimonious number of parameters. We could choose a more general time series process for f_t which could improve the fit of the model further. However, we will show that this choice already fits the data very well and so shifting to alternative specifications would not change any of the conclusions we document herein.

In sum, we make no assumption that equations (2.1)–(2.3) constitute the true data generating process for the economy. Instead, equations (2.5) and (2.8) describe the evolution of the agent's beliefs about the current state of the economy and its sub-components. Given those beliefs, equation (2.7) produces the full term structure of expectations in every time period. While we will show that our model provides a close fit to all of the survey data, the evolution of expectations is regulated by only a small set parameters: $\Theta \equiv (\Phi, \phi_f, \Sigma_\varepsilon)$.

[6] We assume the number of signals m is sufficiently large such that, given the parameters of the model, Σ_ε is nonsingular.

3 Bringing the Model to the Data

In this section, we show how to use our model of expectations formation to explain survey forecasts of professional forecasters in the U.S. We first introduce our unique data set in Section 3.1. We discuss estimation of the model in Section 3.2, and document the fit of the term structure of short-rate expectations in Section 3.3.

3.1 Data

We estimate the model's parameters using *all* available surveys of professional forecasters in the U.S. There are several advantages of using surveys of professional forecasts instead of surveys of households or firms. First, multiple surveys covering a wide range of forecast horizons spanning "nowcasts" to the very long run are available going back at least to the early 1980s. Second, professional forecasters closely watch the evolution of the economy and the conduct of monetary policy. As they often represent firms active in financial markets, their predictions are likely a good proxy to those of actual traders in the bond market. We obtain nominal and real short-rate expectations by combining all available surveys of professional forecasts of the 3-month Treasury bill rate, consumer price index (CPI) inflation, and real GDP growth covering more than 600 survey-horizon pairs at a monthly frequency.

Our data span the universe of professional forecasts for the United States. Our forecast data are obtained from nine different survey sources: (1) Blue Chip Financial Forecasts (BCFF); (2) Blue Chip Economic Indicators (BCEI); (3) Consensus Economics (CE); (4) Decision Makers' Poll (DMP); (5) Economic Forecasts: A Worldwide Survey (EF)[7]; (6) Goldsmith-Nagan (GN)[8]; (7) Livingston Survey (Liv.); (8) Survey of Primary Dealers (SPD); (9) Survey of Professional Forecasters (SPF). For each survey and each forecast horizon we use the consensus forecast, or the mean across forecasters. We focus on three sets of forecasts. For output growth we rely on forecasts of real GNP growth prior to 1992 and forecasts of real GDP growth thereafter. For inflation we use forecasts of growth in the CPI. Finally, we employ the 3-month Treasury bill (secondary market) rate as our measure of a short-term interest rate as it is by far the most frequently surveyed short-term interest rate available.[9] Combined, these surveys provide a rich portrait of professional forecasters' macroeconomic expectations. Our results are based on 627 variable-horizon pairs spanning the period 1983 to 2019. The survey data differ in frequency, forecast timing, target series,

[7]To our knowledge, the only other paper to use these survey data is Ehrbeck and Waldmann (1996).

[8]We thank Kenneth Froot for sharing the Goldsmith-Nagan survey data.

[9]For example, forecasts of the Federal Funds rate, the target rate of U.S. monetary policy are only available in two of the eight surveys we consider (BCFF and SPD).

sample availability and forecast horizons.[10] Near-term survey forecasts (target period is up to two years ahead) are available for the longest sample with CPI forecasts from the Livingston Survey beginning in the mid-1940s. Medium- and long-term forecasts (target period includes three-years ahead and longer) are available for real output growth and inflation starting in the late 1970s. However, a more comprehensive set of long-term forecasts (a target period of five or more years ahead) for all three variables is available only starting in the mid-1980s which is where we start our sample.

3.2 Estimation

To estimate the model we use only the professional survey data listed above and do not include any realized economic data. Our goal is to model the expectations formation process and reserve the realized data for our analysis of forecast errors in later sections.[11] We estimate the parameters of the model, Θ, over the sample period from January 1983 to December 2019. The linearity of the model allows us to use a standard state-space framework. The main challenge is to construct the nonlinear mapping between observed survey forecasts and the parameters implicit in model-implied expectations.

We assume that observed survey forecasts are noisy signals of the true underlying expectations. To cement ideas, suppose our dataset only included monthly forecasts of z_t for different monthly horizons. Let $\textsc{SvyExp}_{t+h|t}$ be a 3×1 vector collecting the survey forecasts of real output growth, inflation, and the nominal short rate for the forecast horizon h. Then the observation equation for each survey-horizon pair could be straightforwardly characterized as

$$\textsc{SvyExp}_{t+h|t} = \mathbb{E}_t\left[z_{t+h}\right] + o_t^{(h)} = \omega_{t|t} + \Phi^h x_{t|t} + o_t^{(h)}, \tag{3.1}$$

where $o_t^{(h)}$ is a horizon-specific measurement error which is assumed to be Gaussian white noise with a diagonal variance-covariance matrix. Even if we were able to observe $\textsc{SvyExp}_{t+h|t}$ for every forecast horizon h, equation (3.1) shows that it would still only be a noisy signal of the true expectations. In reality, we have multiple surveys for forecasts of the same variable at the same horizon but also face substantial gaps where no forecasts are available. The model allows us to combine these noisy signals in a principled way and fill gaps to obtain model-implied forecasts at all horizons.

In practice, our observation equation is more involved than that of equation (3.1). For

[10] See Crump, Eusepi, Moench, and Preston (2023) for a full discussion of the data sources.

[11] In theory, we could include the monthly realized 3-month T-bill in the estimation as the agent is assumed to observe it perfectly. In practice, forecasters are surveyed at different times in the month and may not fully observe the average short rate for that month. We verify that the model-implied monthly nowcast of the short-term interest rate is nearly identical to the realized series.

example, many of the observed survey forecasts are expressed at quarterly or yearly horizons. In addition, recorded forecasts often involve quarterly averages or annualized growth rates. We exploit the fact that all of these more general forecast targets can be well approximated by specific (weighted) linear combinations of $\text{SvyExp}_{t+h|t}$ for appropriate choices of h which maintains linearity of the observation equation. For example, a forecast formed in the last month of a quarter, for the next quarter's average of the short-term interest rate, may be well approximated by the third element of $\sum_{j=1}^{3} \zeta_j \mathbb{E}_t [z_{t+j}] + o_t$ where $\zeta_j = \frac{1}{3}$ for each j. Appendix B provides full details and additional examples.

Overall, the empirical exercise involves a sizable observation equation including 627 time series of survey forecasts for different horizons. To limit the number of parameters to be estimated, we group the variances of the measurement errors o_t by the target variable of interest and the horizon of the forecast (but not by the specific survey). In particular we group forecast horizons by: *very short term*, up to two quarters ahead, *short term*, up to two years ahead, *medium term*, from three to four years ahead, and *long term*, five or more years ahead. This allows for a parsimonious yet flexible fitting of observed forecasts.

While the state of the economy is re-estimated in every period, the model's parameters remain constant, reflecting agents' invariant priors. This simplifying assumption retains linearity. It is plausible that, over such a long sample, agents would re-evaluate their estimates of the model's parameters to adapt to structural change. We accommodate this process by allowing for a structural break in the perceived relative variances of the innovations to the trend and cycle components. We date the structural break in the first month of 1999. While the choice of the month may appear arbitrary, it reflects two considerations. First, as shown by Carvalho, Eusepi, Moench, and Preston (2023), long-term inflation expectations begin to stabilize around that period. Separately, Hanson, Lucca, and Wright (2021) show that the sensitivity of long-term interest rates to changes in short term interest rates declined substantially after 1998. This evidence suggests important revisions to agents' priors about the variability of long-term trends. This structural break allows our model to capture this revision in beliefs using the observed term structure of expectations.

Priors and Posteriors. We use a Bayesian estimation approach which requires prior distributions for the parameters of the model. In general, we use loose priors throughout. The one exception is we assume the perceived trend component, ω_t, measuring structural shifts in the economy, is slow-moving relative to the cycle component, x_t. This priors reflects the key model assumption discussed above. Consequently, we impose this restriction in our priors. To do so, we express the variance-covariance matrix of ε_t in equation (2.8) as

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma_\omega & 0 \\ 0 & \sigma_x \end{bmatrix} \times C_\varepsilon \times \begin{bmatrix} \sigma_\omega & 0 \\ 0 & \sigma_x \end{bmatrix} \tag{3.2}$$

where $\sigma_x, \sigma_\omega > 0$ are vectors collecting the standard deviations of the innovations to f_t which are linked to the evolution of $\omega_{t|t}$ and $x_{t|t}$ by equation (2.5), and C_ε is a correlation matrix. Define the three-dimensional vector measuring the element-by-element ratio $\lambda \equiv \sigma_\omega \odot \sigma_x^{-1}$ where $\odot$ denotes the Hadamard product. We let each element of this vector be distributed according to a beta distribution. As shown in Table 1, the prior mean for all elements in λ is 10% with 90% of the density between 6% and 14%. We maintain the same priors across regimes, as we expect the data to be informative about any difference in perceived beliefs about the long run across these two time periods resulting in estimates, $\lambda_{1983-1998}$ and $\lambda_{1999-2019}$. Priors on the elements of σ_x are fairly diffuse—set as an independent inverse Gamma distribution with a prior mean of 0.1 and a standard deviation of 2.

The prior on the correlation matrix, C_ε, is defined by the Lewandowski-Kurowicka-Joe (LKJ) distribution (Lewandowski, Kurowicka, and Joe 2009). In terms of moments, the prior mean is the identity matrix I_n with $n = 6$ and the scalar ψ regulates the variance of each off-diagonal entry.[12] We choose a fairly loose prior value on this parameter, setting $\psi = 2$.[13] For reference, $\psi = 1$ implies a uniform distribution across all entries, while $\psi > 1$ favors covariance matrices with stronger diagonal elements (weaker correlations). We set priors on the diagonal matrix ϕ_f so that each autocorrelation coefficient has a beta distribution with mean 0.2, reflecting a prior corresponding to a modest degree of autocorrelation in forecast errors.

Although we observe a large amount of survey data we do not observe any survey forecasts which satisfy equation (3.1) for $h = 1$. The reason is that forecast horizons in surveys of professional forecasters typically follow quarterly or annual increments while our model is estimated at the monthly frequency. This means that every element of our observation equation corresponds to Φ raised to different powers of h which makes full identification of Φ challenging. To circumvent this problem, we impose positivity on the diagonal elements by the choice of a prior which follows the gamma distribution with mean of 0.5 and standard deviation of 0.1.

Table 1 collects summary information on the prior and posterior distribution for the key

[12] Let $B(\cdot,\cdot)$ denote the Beta function. The density function is

$$p(C_\varepsilon) = 2^{\sum_{j=1}^{n-1}(2(\psi-1)+n-j)(n-j)} \times \prod_{h=1}^{n-1}(B(\psi+(n-h-1)/2, \psi+(n-h-1)/2))^{n-j} \times \det(\mathrm{C}_\varepsilon)^{\psi-1}.$$

[13] We also considered a few different choices, but $\psi = 2$ delivered the highest marginal likelihood

Table 1: **Prior and Posterior Distributions for Selected Parameters.** This table collects information about the prior and posterior distributions for the key parameters of the model. We compute the posterior parameter distribution using the Random Walk Metropolis-Hastings (RWMH) algorithm. See Section SA1-2 in SA1 for full details.

	Prior			**Posterior**		
	Dist.	Mean	[5% - 95%]	Mode	Mean	[5% - 95%]
$\lambda_{\mathbf{1983-1998}}$						
g_t	Beta	0.10	[0.06, 0.14]	0.19	0.19	[0.17, 0.22]
π_t	Beta	0.10	[0.06, 0.14]	0.44	0.44	[0.39, 0.49]
i_t	Beta	0.10	[0.06, 0.14]	0.38	0.39	[0.34, 0.44]
$\lambda_{\mathbf{1999-2019}}$						
g_t	Beta	0.10	[0.06, 0.14]	0.08	0.08	[0.07, 0.10]
π_t	Beta	0.10	[0.06, 0.14]	0.14	0.14	[0.12, 0.17]
i_t	Beta	0.10	[0.06, 0.14]	0.16	0.16	[0.13, 0.20]
Forecast Errors						
ρ_{fg}	Beta	0.20	[0.02, 0.49]	0.54	0.55	[0.49, 0.61]
$\rho_{f\pi}$	Beta	0.20	[0.02, 0.49]	0.51	0.52	[0.46, 0.58]
ρ_{fi}	Beta	0.20	[0.02, 0.49]	0.72	0.72	[0.66, 0.77]

parameters of our model. Given the wealth of survey information we employ, it is not surprising that the parameters are precisely estimated. The central tendency of the posterior distribution of λ, which is driven by the agent's belief about the relative variability of the trend component as compared to the cycle, ranges from 20 to 40 percent before 1999. In contrast, in the second part of the sample the posterior distribution of λ shifts to the left and is centered at values between 10 to 20 percent. Finally, through the lens of our model, the posterior distribution confirms the empirical regularity that survey forecasts exhibit predictable forecast errors, with the central tendency of the autocorrelation coefficients all above 0.5.

3.3 The Term Structure of Short Rate Expectations

Figure 1 shows the predicted term structure of short-rate expectations from our model, together with selected survey forecasts.[14] The model captures the evolution of consensus forecasts of nominal short rates extremely well. The close fit of the model is not a foregone conclusion, given the tight restrictions on the term structure of expectation imposed by the model, as shown in equation (2.7).[15] Given the rich survey data we employ, posterior

[14] In the second Supplemental Appendix (SA2) available at `https://www.dropbox.com/scl/fi/8x4yx91q0iclc9i50i1f2/CEM_Appendix_Fitted_2024-04-04.pdf?rlkey=trylgof2rlmw3wa2nuergboa8&dl=0` we show the fit for each of the 627 series. To construct the forecast for each horizon we sample the unobserved states $(\omega_{t|t}, x_{t|t})$ using a simulation smoother.

[15] The tight fit of our model-implied expectations to the survey data implies that the array of existing stylized facts about survey expectations (e.g., as summarized in Farmer, Nakamura, and Steinsson 2023) are

uncertainty about the model-implied short rate expectations is negligible. In fact, it is visible only for long-horizon expectations for which we have fewer observations.

The model captures the fact that professional forecasters frequently revise their forecasts at all horizons. Furthermore, interest rate expectations behave as we expect over the monetary policy cycle. Comparing expectations across horizons, the term structure of policy rate expectations typically flattens (and often inverts) at the end of economic expansions and steepens in the aftermath of recessions. For example, the term structure of short rate expectations inverts in early 1989 when expectations at short-term horizons reached their local peak, leading into the 1990-91 recession. After the short rate reached the zero lower bound in 2008, the term structure of expectations flattened at first, and then steepened again as forecasters persistently expected an imminent lift-off.[16]

With this in mind we now discuss the main properties of the expectations formation process we have proposed.

3.3.1 Role of Output Growth and Inflation Forecasts

Given that our main focus is on nominal short-rate expectations, one might ask why we employ a multivariate model of expectations formation? The short-term nominal interest rate, which is primarily governed by the choices of the central bank, is informed by assessments of current and future real activity and inflation.[17] Taking a textbook Taylor-rule type reaction function as an example and using the notation from Section 2, agents would form expectations about the short-term rate as

$$\mathbb{E}_t[i_{t+h}] = \omega^i_{t|t} + \phi_\pi \left(\mathbb{E}_t[\pi_{t+h}] - \omega^\pi_{t|t}\right) + \phi_g \left(\mathbb{E}_t[g_{t+h}] - \omega^g_{t|t}\right)$$

which clearly requires a multivariate model of expectations formation. Shifts in short- and long-term views about inflation and economic activity play a key role. In addition, while the interest rate is perfectly observed, output and inflation are not and so revisions in the corresponding elements of $\omega_{t|t}$ and $x_{t|t}$ also contribute to revisions in interest rate expectations.

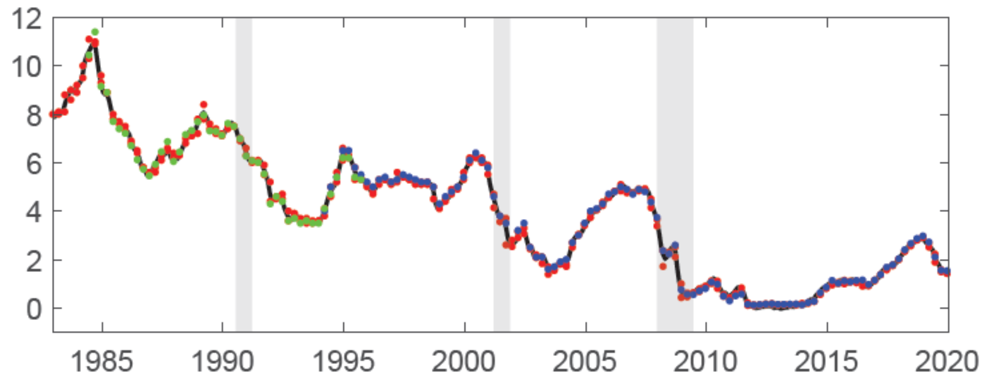
inherited by our measure of expectations.

[16] Note that these estimated measures of short rate expectations based on survey forecasts are consistent with a perceived zero lower bound (ZLB) on nominal interest rates.

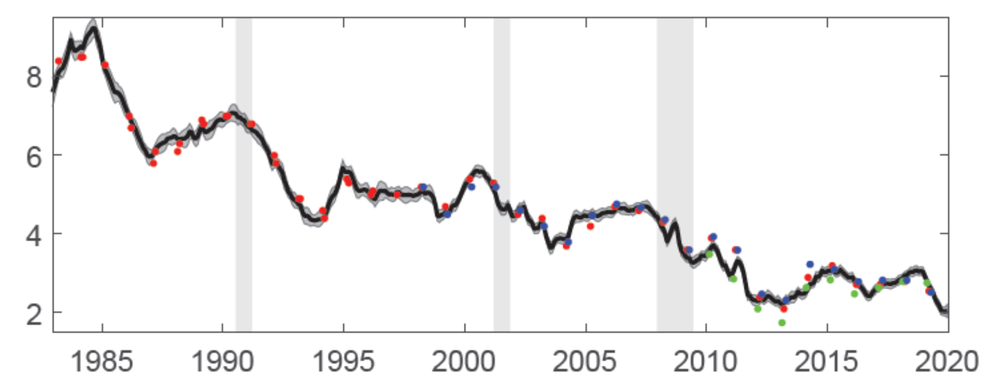
[17] See https://www.federalreserve.gov/monetarypolicy/files/fomc_longerrungoals.pdf.

Figure 1: **Term Structure of Interest Rate Expectations.** This figure compares model-implied survey forecasts to selected individual survey-horizon pairs. The top panel displays four-quarter ahead forecasts taken in the last month of the quarter ("Short-Term"); the middle panel displays forecasts for the one-year average interest rate three years ahead taken in the first half of the year ("Medium-Term"); the bottom panel features the forecast for the one-year average interest rate five years ahead taken in the first half of the year ("Long-Term"). The dots indicate forecasts from different surveys: BCEI and BCFF shown in red, CE shown in blue; EF and SPF are shown in green in the top and middle panel, respectively. The black line denotes the median prediction from the model, and the dark grey area the 99% posterior coverage interval. Light grey shaded areas denote NBER recessions. The sample period is 1983m1–2019m12.

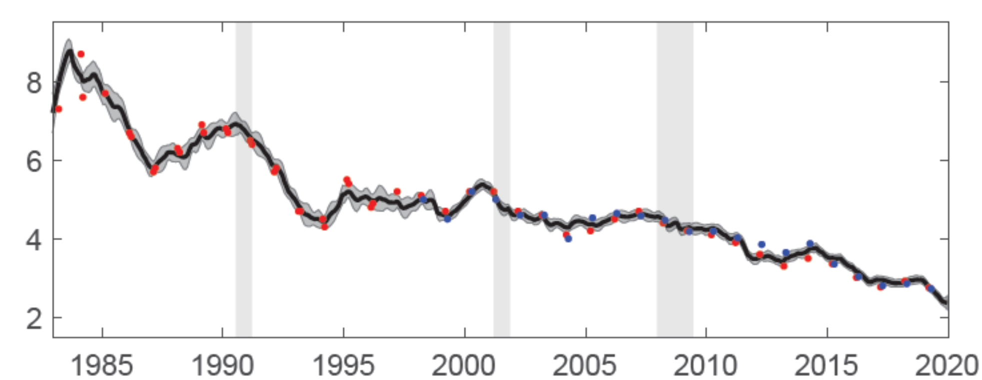
(a) *Short-Term Forecasts*



(b) *Medium-Term Forecasts*



(c) *Long-Term Forecasts*



Based on these arguments, a univariate model of the interest rate would then be misspecified. To confirm this, the top panel of Figure 2 compares the fit of our model with that of a univariate version of our model. The univariate model cannot simultaneously match short-horizon interest rate forecasts and medium or longer-horizon forecasts. The chart shows that the univariate model produces longer-horizon forecasts which are too volatile relative to the survey data.

We can also reject the univariate specification formally. Table 2 displays the marginal data density (MDD) for different model specifications. The MDD is defined as

$$p\left(\text{SvyExp} \mid \mathcal{M}_l\right) = \int p\left(\text{SvyExp} \mid \theta_l, \mathcal{M}_l\right) p\left(\theta_l \mid \mathcal{M}_l\right) d\theta_l, \tag{3.3}$$

where SvyExp includes all survey data, $\mathcal{M}_l$ denotes each of the four different model specifications we consider and θ_l is the associated vector of model's parameters.[18] In addition to our baseline model, we report results for three additional (nested) specifications. The specification labelled "bivariate" assumes nominal interest rates and inflation jointly evolve whereas output growth evolves independently. The specification labelled "univariate" assumes that all three variables evolve independently. All of these specifications allow for serial correlation in forecast errors. Finally, we also consider the baseline specification without serial correlation in f_t (labelled "$\phi_f = 0$").

Inspecting the first three columns in Table 2, it is clear that our baseline multivariate model provides, by far, the best fit of the survey-based term structure of expectations against the bivariate and univariate alternatives. The fourth column shows the fit of a restricted model with no autocorrelated forecast errors; this specification is decisively worse than our baseline, emphasizing that observed forecast display significant deviations from the FIRE benchmark. However, highlighting the importance of a multivariate structure, the multivariate model with no serial correlation in forecast errors outperforms either the bivariate or univariate models which feature serial correlation in the forecast errors.

These results show that a multivariate model is the preferred specification for fitting forecasts all three variables. However, our primary interest is on forecasts of the nominal interest rate, especially at medium to long forecast horizons. With this in mind we evaluate the following predictive likelihood[19]

$$p\left(\text{SvyExp}_L^i \mid \mathcal{I}, \mathcal{M}_l\right) = \int p\left(\text{SvyExp}_L^i \mid \theta, \mathcal{I}, \mathcal{M}_l\right) p\left(\theta \mid \mathcal{I}, \mathcal{M}_l\right) d\theta_l, \tag{3.4}$$

[18] See Appendix SA1-2 for full details.

[19] See Del Negro and Eusepi (2011) and references therein.

where $\textsc{SvyExp}^i_L$ includes short-rate forecasts at horizons longer than one year and $\mathcal{I}$ include all other forecasts, including short-term interest rate forecasts. The bottom row in Table 2 displays the above predictive likelihood, which measures the fit of interest rate expectations at horizons longer than one year, conditional on the parameter distribution delivering the best fit of short-horizon interest rate expectations and of the term structure of inflation and output growth expectations. It is important to stress that in the case of the univariate model, output and inflation forecasts provide no information in the estimation, by assumption. The predictive likelihoods confirm that observed expectations about output growth and inflation contain valuable information to explain the behavior of the term structure of interest rate expectations: a multivariate model better describes how short-term interest rate expectations are formed.

Table 2: **Model Comparison.** This table reports likelihood statistics for the baseline model and a number of alternatives. The column labelled "$\phi_f = 0$" denotes the baseline model with no serial correlation in the forecast errors (i.e., $G = 0$). The first row of the table shows the log-marginal data density (MDD) for the four alternative models (see equation (3.3)). The second row shows the log-predictive likelihood of the interest rate forecasts at horizons longer than four quarters, conditional on the remaining survey forecasts (see equation (3.4)).

MDD	**Baseline**	**Bivariate**	**Univariate**	$\boldsymbol{\phi_f = 0}$
$\ln p(\textsc{SvyExp} \mid \mathcal{M}_l)$	994.23	-274.53	-1388.5	853.09
$\ln p(\textsc{SvyExp}^i_L \mid \mathcal{I}, \mathcal{M}_l)$	-1632.8	-2008.1	-2535.5	–

These results imply that conditioning on inflation and output growth expectations is a key requirement for explaining the behavior of policy rate expectations *at all horizons*. Put differently, we find strong evidence that survey forecasts are formed jointly as predicted by our modeling framework. More importantly, it confirms the term structure of short-term interest rate expectations is broadly consistent with the type of monetary policy rules that are usually assumed in conventional monetary models (Andrade, Crump, Eusepi, and Moench 2016). This also suggests that testing theories of expectation formation using forecasts for a single variable may provide misleading results (see Crump, Eusepi, Moench, and Preston (2024) for evidence at the individual forecaster level).

3.3.2 Forecast Updating

Inspecting Figure 1 it is immediate that expectations at different forecasting horizons appear to *co-move* over time throughout the whole sample, with long-term forecasts displaying stronger co-movement before the 2000s. Through the lens of our model (equation (2.5)), revisions to forecasts at any horizon h are closely linked to current short-term forecast errors since

$$\mathbb{E}_t\left[z_{t+h}\right]-\mathbb{E}_{t-1}\left[z_{t+h}\right]=F^h f_t. \tag{3.5}$$

Under our assumptions, only the short-term interest rate is perfectly observed. Since we can observe the forecast error, it is then natural to explore the link between model-implied forecast revisions and realized forecast errors for the short rate. The bottom panel of Figure 2 shows that there is a close co-movement between interest rate forecast errors and long-term interest rate forecast revisions. The figure plots the realized forecast error for $h = 12$ (i.e., $z_t - \mathbb{E}_{t-12}\left[z_t\right]$) against the twelve-month revision of the model-implied long-term forecast (i.e., $\omega_{t|t} - \omega_{t-12|t-12}$). A positive forecast error, when realized interest rates are above expectations, leads to an upward revision to the long-term forecast. The figure also highlights the cyclicality of the revisions, with sizable forecast errors at turning points leading to changes in long-run expectations. Perhaps not surprisingly, the co-movement is stronger before the Great Recession when the interest rate reached the zero lower bound. In the prolonged aftermath to the financial crisis, other information beyond short-term interest rates such as forward guidance announcements or balance-sheet interventions likely influenced long-horizon expectations.

This evidence provides support to our proposed model of expectations formation. It is worth re-emphasizing that realized data is not included in the estimation. The fact that revisions in long-run expectations are positively correlated with interest rate surprises points to models of expectations formation where agents have to learn about the long-run properties of the economy, and update their beliefs in response to recent information.

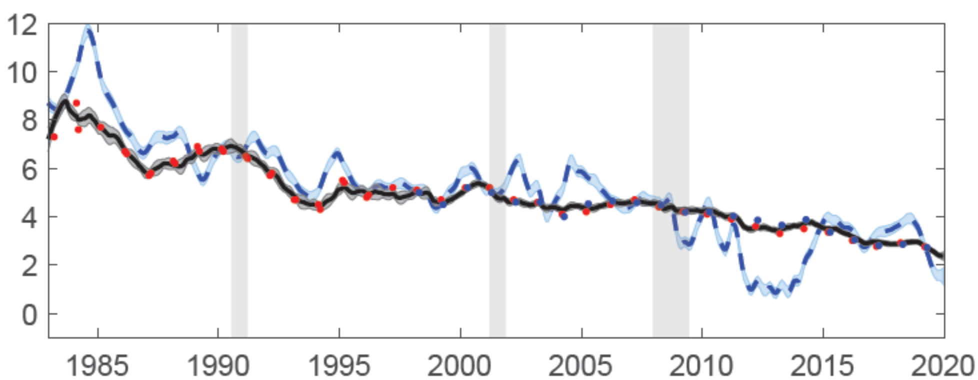
3.3.3 Deviations from FIRE

We have made no assumption about the degree of rationality embedded in observed forecasts. While the agent has imperfect information about the state of the economy, we have not constrained the beliefs' updating mechanism to be consistent with the true data generating process. Here we show that our model replicates frequently reported evidence of deviations from FIRE, and provide additional results suggesting these deviations differ by forecast horizon.

The most common statistic used to illustrate deviations from FIRE is the presence of (first-order) autocorrelation in forecast errors. Table 3 displays the sample autocorrelation of interest rate forecast errors at horizons of one-, two-, three- and four-quarters ahead, from both the raw survey data and those from the fitted model. The forecast errors are computed at a quarterly frequency, to maintain consistency with the empirical results already available in the literature. We calculate autocorrelation coefficients using non-overlapping samples when the forecast horizon is two-quarters ahead or more. For the raw survey data, the table reports results for the SPF and BCEI. While the model is fitted to all nine surveys, we focus on these surveys for two reasons. First, they are frequently used in the literature and have the longest available sample. Second, SPF and BCEI are sampled at different months in the quarter. This allows us to evaluate whether the pattern of forecast errors depends on the specific month of collection, as would be expected.

Figure 2: **Properties of the Model.** This figure demonstrates the properties of our multivariate model of expectations formation. The top chart compares the model-implied long-term forecasts from the baseline model introduced in Section 2 (black line, grey shading) with the corresponding model-implied long-term forecasts from a univariate version of the same model (blue line, blue shading). Solid lines denotes the median prediction from the model, and the shaded area denotes the 99% posterior coverage interval. The bottom chart illustrates the co-movement between revisions to long-run nominal short-rate forecasts (red line) and short-run forecast errors (black line). The sample period is 1983m1–2019m12.

(a) *Model Fit: Multivariate vs. Univariate*



(b) *Forecast Revisions and Forecast Errors*

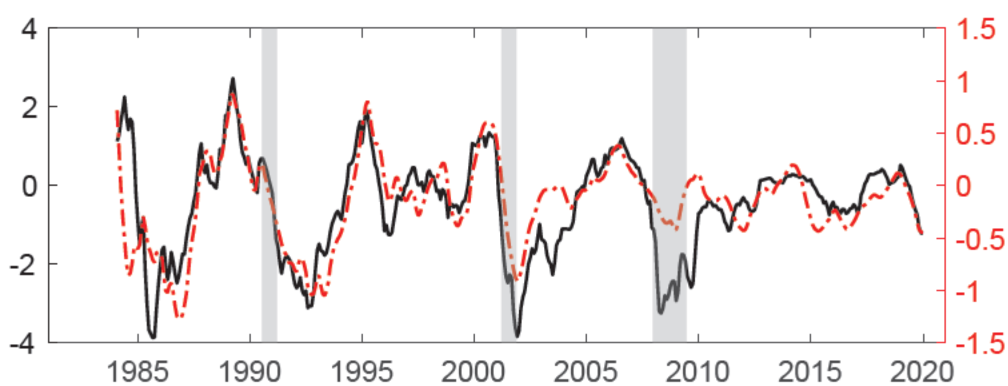


Table 3 reveals that the model can replicate three important aspects of the data. First, forecast errors are positively correlated for short-term forecasts, displaying "under-reaction" relative to the FIRE baseline. Second, the degree of autocorrelation declines with the forecast horizon. Third, the degree of autocorrelation declines with the month of the quarter in which the forecast is queried. This stands to reason as forecasts earlier in the quarter are based on less information than those later in the quarter.

A number of papers also study whether forecasts "over-" or "under-react" to new information relative to the FIRE benchmark. This is commonly measured as the estimated slope coefficient in regressions of the form (see Coibion and Gorodnichenko 2015):

$$f^Q_{t+h|t} = \alpha^{(h)}_f + \beta^{(h)}_f \left(\mathbb{E}_t[i_{t+h}] - \mathbb{E}_{t-1}[i_{t+h+1}]\right) + \xi^{(h)}_{t+h}. \tag{3.6}$$

Here, $f^Q_{t+h|t}$ is the quarterly h-period ahead forecast error from the estimated model and the regressor represents the one-period revision in the h-quarter ahead forecast also from the estimated model. We use the fitted model to estimate $\beta^{(h)}_f$ for horizons from one-quarter ahead to five-years ahead. This substantially expands upon the existing literature which has only utilized short-term survey forecasts (up to four quarters ahead).

The top panel of Figure 3 provides the estimates of $\beta^{(h)}_f$ across forecast horizons of $h = 2, \ldots, 20$ quarters. Under FIRE, this slope coefficient should be zero. In contrast we find a positive coefficient for short-term forecasts and a negative coefficient at medium and long-run forecast horizons. In the univariate framework which is common in the literature, this would be consistent with expectations adjusting slower, or "under-reacting" to new information, relative to FIRE. For these shorter forecast horizons the regression coefficient increases with the forecast horizon. This is consistent with what is commonly found using survey data directly (see, e.g., Coibion and Gorodnichenko 2015 for inflation and Farmer et al. 2023 for the short-term interest rate).

We can exploit the model and use the entire term structure of expectations which allows h to grow and include longer-horizon forecasts. The figure shows that the regression coefficient flips for forecast horizons farther than two years in the future. This is consistent with "over-reaction" of longer-run expectations to new information. The agent tends to over-extrapolate from recent information. This pattern is consistent with bounded rationality in long-term expectations (Eusepi et al. 2024) and behavioral theories of expectations formation (Bordalo et al. 2019; Bordalo et al. 2020). This is also consistent with the experimental evidence presented in Afrouzi et al. (2023). One caveat to the interpretation of these results is that our underlying model is multivariate. As such, interpreting these coefficient in terms of "over-" and "under-reaction" to new information is not clear cut. Instead we view this exercise as a set of empirical moments that any theory of expectations formation should replicate.

Table 3: **Sample First-Order Autocorrelation of Forecast Errors.** This table reports the sample estimate of the first-order autocorrelation coefficient of quarterly forecast errors of the short-term interest rate. Minimum and maximum statistics are taken across 1,000 posterior draws. The sample period is 1983m1–2019m12.

	1st Month of Quarter				2nd Month of Quarter					3rd Month of Quarter			
	Survey		Model		Survey			Model		Survey		Model	
	BCEI	Min.	Med.	Max.	BCEI	SPF	Min.	Med.	Max.	BCEI	Min.	Med.	Max.
Q1 Ahead	0.496	0.454	0.487	0.516	0.442	0.486	0.391	0.426	0.460	0.353	0.225	0.281	0.337
Q2 Ahead	0.369	0.303	0.329	0.355	0.289	0.280	0.253	0.280	0.306	0.263	0.191	0.231	0.261
Q3 Ahead	0.363	0.288	0.310	0.331	0.252	0.190	0.182	0.201	0.221	0.219	0.124	0.147	0.171
Q4 Ahead	0.200	0.145	0.167	0.190	0.150	0.094	0.095	0.115	0.140	0.118	0.082	0.104	0.123

3.3.4 Learning and the EH

An alternative measure of the degree of deviation from FIRE can be obtained by investigating regression-based tests of the EH. We consider a simulation exercise in the spirit of Sinha (2016) and Farmer, Nakamura, and Steinsson (2023) to show that our model deviates sufficiently from the FIRE benchmark so that the EH is rejected by tests which rely on rational expectations, even when the EH holds. As in Sinha (2016) and Farmer, Nakamura, and Steinsson (2023), we utilize the regression-based test of Campbell and Shiller (1991). We simulate EH-consistent yields using our fitted model parameters and show that rejections of the EH are frequent even though it holds in our simulated data by construction.

The strong form of the EH states that the yield of maturity n is equal to the average expected short rate over the life of the bond. Under the EH, longer-term interest rates are then

$$\mathsf{yld}_t^{\mathrm{EH}}(n) = \frac{1}{n}\sum_{h=0}^{n-1} \mathbb{E}_t\left[i_{t+h}\right], \tag{3.7}$$

where $\mathsf{yld}_t^{\mathrm{EH}}(1) = i_t$. We simulate interest-rate forecasts using equations (2.5), (2.8), and (3.1), with the parameter values set to their posterior mode, and the measurement error, o_t, set to zero for all forecast horizons. For each simulation, s, we obtain $i_t^{(s)}$ along with the associated expectations, $\mathbb{E}_t\left[i_{t+h}^{(s)}\right]$, for all t and $h \in 1,\dots,120$ months. We impose the EH by constructing the simulated term structure of interest rates using equation (3.7). This gives $\{(i_t^{(s)}, \mathsf{yld}_t^{(s)}(2), \dots, \mathsf{yld}_t^{(s)}(120)) : t = 1,\dots,T\}$ where T is the same length as our data sample. For each simulation, s, we can test the null hypothesis that the EH holds ($\beta = 1$)

in the Campbell and Shiller (1991) regression:

$$\mathsf{yld}_{t+1}^{(s)}(n-1) - \mathsf{yld}_t^{(s)}(n) = \alpha + \beta\left(\frac{1}{n-1}\right)\left(\mathsf{yld}_t^{(s)}(n) - i_t^{(s)}\right) + u_{t+1}^{(s)}, \tag{3.8}$$

for $n = 2, \ldots, N$.

The middle panel of Figure 3 shows the share of rejections across $1,000$ independent simulations of the model for all monthly maturities out to ten years. The nominal size of the test is 10%. For maturities beyond two years, the rejection rates are never below 40%, suggesting that an empirical researcher would frequently conclude that the EH was rejected in the data. As a robustness check, we also include the corresponding results when the data are generated without serial correlation in forecast errors (i.e., under the assumption that $G = 0$). In this case the share of rejections across simulations is essentially equivalent to the nominal size of the test. This shows that the rejections we observe are not driven by finite-sample properties of the test statistic since, under the joint null hypothesis for this test, the empirical size is very close to the nominal size. Instead this shows that it is specifically the deviation from rational expectations which generates the erroneous rejections of the EH.

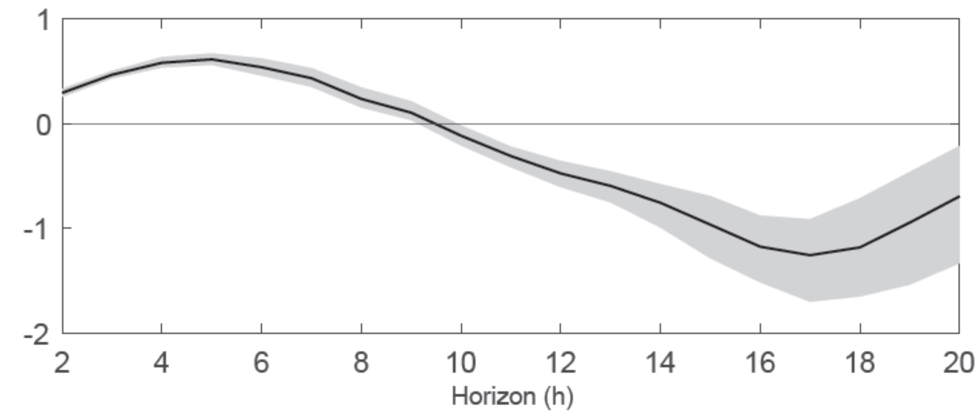
3.3.5 Time-Varying Longer-Run Expectations

Given the variability of long-maturity yields, a necessary condition for the EH to hold is that longer-horizon expectations exhibit significant time variation. We have documented in Figure 2 that professional forecasters revise their long-run expectations as a function of their short-term forecast errors. This behavior underscores that the agent has shifting views about the long term. Here, we further illustrate the degree of variability of the model-implied forecasts by comparing long-horizon forecasts of the real interest rate directly to those from two well-known statistical models. For this exercise, we use real interest rates as long-run inflation expectations are roughly constant in the second part of the sample. To compute the model-implied expected real rate we exploit the multivariate nature of our model and use forecasts of the nominal short rate and inflation.[20]

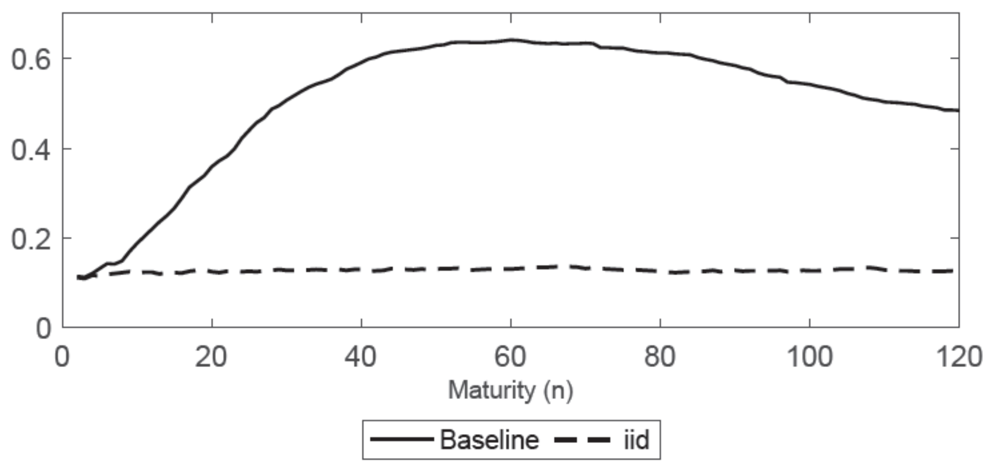
[20] Specifically, the ex-ante expected real rate for horizon h is defined as the expected nominal short rate for $h-1$, minus the expected inflation rate for horizon h.

Figure 3: **Properties of Model-Implied Expectations.** The top chart shows the estimated slope coefficient from the regression in equation (3.6) for $h \in \{2, \dots, 20\}$ quarters. The grey shaded region denotes the minimum and maximum estimated slope coefficients across $1,000$ posterior draws. The middle chart of this figure reports empirical rejection rates across $1,000$ independent simulations using our baseline model (labelled "baseline") and under the assumption that forecast errors are not serially correlated (labelled "i.i.d."). Hypothesis tests are formed using a t-statistic with the equal-weighted cosine variance estimator of Lazarus et al. (2018). The bottom chart displays the model-implied long-run forecast of the real interest rate using the results from Section 2 (labelled "Model-Implied") as compared to those from Laubach and Williams (2003) (labelled "LW") and Del Negro et al. (2018) (labelled "DGGT"). The sample period is 1983Q1–2019Q4.

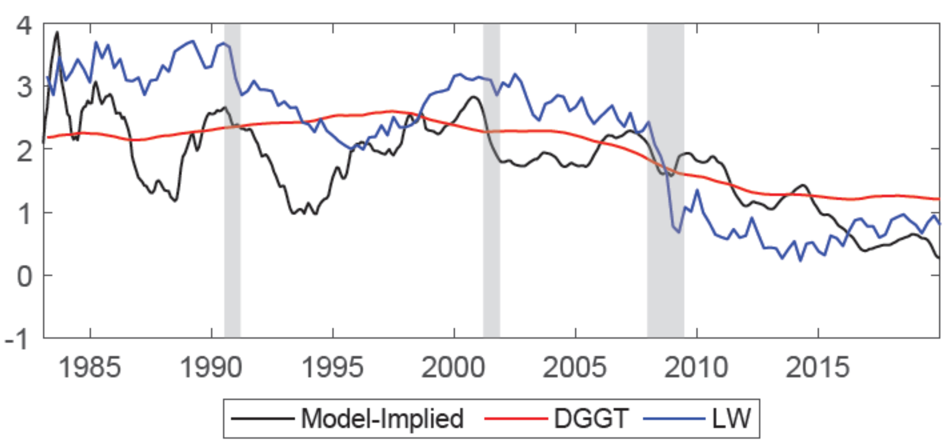
(a) *Coibion-Gorodnichenko Regressions*



(b) *Simulated EH Rejection Rates*



(c) *Long-Term Real-Rate Expectations*



As shown in the bottom panel of Figure 3, the long-run estimate of the real interest rate from our model displays a significant degree of variation when compared to popular alternative model-based measures that explicitly allow for trending interest rates. The red line shows the evolution of the 30-year forecast from the VAR model with time-varying means of Del Negro, Giannone, Giannoni, and Tambalotti (2018) which uses information from both macroeconomic and financial variables. The blue line is the long-run estimate of the short-term real rate from the unobserved components model of Laubach and Williams (2003). We use the one-sided estimate for a better comparison with our survey-based expectations model (black line). The chart shows that our measure of long-run expectations is more volatile than its counterpart in these statistical models. In particular, the standard deviation of twelve-month changes in our model, at 0.43, is higher than 0.35 for Laubach and Williams (2003) and 0.07 for Del Negro, Giannone, Giannoni, and Tambalotti (2018).

Summing up, we have documented several properties that our estimated model satisfies: (i) expectations for key macroeconomic variables are formed *jointly*; (ii) agent revise expectations at every horizon, including the very long run, in response to new information; (iii) forecast revisions are substantial, leading to volatile long-run expectations; (iv) expectations show significant deviations from FIRE; (v) short-term forecasts tend to "under-react" to available information, while longer-term forecasts tend to "over-react"; (vi) deviations from full rationality are sizable enough that standard tests of the EH based on rational expectations may fail, even though the EH holds in the data.

4 Is There Hope for the Expectations Hypothesis?

Our model-implied term structure of short-rate expectations can now be compared directly to the term structure of interest rates at all maturities. To separate longer-term from short-term expectations, we conduct our analyses in terms of *forward rates*, defined as the current yield of an n-year bond maturing in $n+m$ years:

$$\mathsf{fwd}_t(n,m) = \frac{1}{n}[(n+m)\cdot \mathsf{yld}_t(n+m) - m\cdot \mathsf{yld}_t(m)]. \tag{4.1}$$

Because our empirical model of expectations is estimated at a monthly frequency, we construct annual forward rates as the annual average of monthly forward rates. For example, a 4Y1Y forward (i.e., the one-year rate, four-years ahead) would set $n = 12$ and $m = 48$.

Expressing the EH in equation (3.7) in terms of forward rates gives

$$\mathsf{fwd}_t^{\mathrm{EH}}(n,m) = \frac{1}{n}\sum_{h=m+1}^{n+m}\mathbb{E}_t\left[i_{t+h-1}\right] = \frac{1}{n}\sum_{h=m+1}^{n+m}\mathbb{E}_t\left[r_{t+h-1}+\pi_{t+h}\right]. \tag{4.2}$$

In words, the expectations hypothesis implies that the forward rate $\mathsf{fwd}_t^{\mathrm{EH}}(n,m)$ equals the average expected nominal short-term interest rate over the n months starting m months hence. This can be further decomposed into the average expected path of real short rates plus expected inflation. Note that this is an identity: there are no implicit assumptions about the rationality or bias of expectations or the data generating process for yields, expectations, or the difference between the two.

We can contrast the term structure of interest rates with the corresponding rates implied by the EH. We define the wedge $\mathsf{w}_t(n,m)$ between the two as

$$\mathsf{w}_t(n,m) = \mathsf{fwd}_t(n,m) - \mathsf{fwd}_t^{\mathrm{EH}}(n,m), \tag{4.3}$$

so that observed forwards can be decomposed into three components: the expected average real short rate, expected inflation, and the EH wedge as

$$\begin{aligned}\mathsf{fwd}_t(n,m) &= \mathsf{fwd}_t^{\mathrm{EH}}(n,m) + \mathsf{w}_t(n,m) && (4.4)\\ &= \frac{1}{n}\sum_{h=m+1}^{n+m}\mathbb{E}_t\left[r_{t+h-1}\right] + \frac{1}{n}\sum_{h=m+1}^{n+m}\mathbb{E}_t\left[\pi_{t+h}\right] + \mathsf{w}_t(n,m). && (4.5)\end{aligned}$$

Under the strong-form of the EH, we have that $\mathsf{w}_t(n,m) = 0$ for every t and pair (m,n). The weak form of the EH, implies that $\mathsf{w}_t(n,m) = \bar{\mathsf{w}}(n,m)$ is constant for every t. This is the form of the EH that plays a key role in standard structural models in macroeconomics. Beyond the exact form of the EH, we also aim to assess the relative importance of the wedge in the evolution of forward rates.

We now turn to our formal test of the EH. We use the zero-coupon U.S. Treasury yield data from Gurkaynak et al. (2007) to obtain $\mathsf{fwd}_t(n,m)$ for all (n,m) pairs.[21] The first specification regresses EH-consistent forward spreads on actual forward spreads ("Specification 1"):

$$\mathsf{fwd}_t^{\mathrm{EH}}(n,m) - i_t = \alpha_s^{(n,m)} + \beta_s^{(n,m)}\left(\mathsf{fwd}_t(n,m) - i_t\right) + \xi_{s,t}^{(n,m)}. \tag{4.6}$$

Under the weak form of the EH we should have that $\beta_s^{(n,m)} = 1$ for all (n,m) pairs. Under the null hypothesis, we can subtract i_t from either side which ameliorates the trending behavior

[21] Data available at https://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

in yields and improves the sampling properties of the OLS estimator.

Equation (4.6) is identical to the implementation of Froot (1989) except that we can investigate a much wider range of forward maturities, including medium and long horizons, and over a much longer sample. We also consider a specification in 12-month differences ("Specification 2")

$$\Delta_{12m}\mathsf{fwd}_t^{\mathrm{EH}}(n,m) = \alpha_d^{(n,m)} + \beta_d^{(n,m)}\Delta_{12m}\mathsf{fwd}_t(n,m) + \xi_{d,t}^{(n,m)}. \tag{4.7}$$

Similarly, under the weak form of the EH $\beta_d^{(n,m)} = 1$ for all (n,m) pairs.

It is important to stress that neither test requires an assumption about the expectations formation mechanism and, in particular, we do not require rational expectations. Furthermore, we have shown that the model provides a tight fit to the term structure of expectations so that fitting error will not contaminate the results. That said, for conservativeness, we use the model-implied expectations as the dependent variable. Finally, to accommodate the uncertainty arising from the model-generated expectations, we perform all regressions using $1,000$ posterior draws and report the minimum and maximum p-values across these draws.

The regression-based tests of the EH based on Froot (1989) rely on the assumption that short-rate expectations at all future horizons are not constant. If they were constant, one would be bound to find coefficients $\hat{\beta}_s^{(n,m)}$ and $\hat{\beta}_d^{(n,m)}$ close to zero, since the actual forward rates on the RHS of equations (4.6) and (4.7) feature considerable time variation. In other words, a necessary condition for the EH to hold in the data is that short-rate expectations must vary over time at all forecast horizons. As shown in the previous section, this condition is clearly satisfied by the observed short-rate expectations which our model tightly fits across all forecast horizons.

The upper panel of Table 4 shows the test results for the first specification. The EH is overwhelmingly rejected across maturities. The estimated $\hat{\beta}_s^{(n,m)}$ are comfortably below the theoretical coefficient equal to one implied by the expectations hypothesis across horizons. The associated maximal p-values are below four percent (and mostly below one percent) across horizons showing strong statistical support for rejections of the EH along the term structure.

The bottom panel of Table 4 reports the corresponding range of estimates $\hat{\beta}_d^{(n,m)}$ and their associated p-values for the second specification, relying on twelve-month differences of model-implied and actual spreads. The coefficients are well below 0.3 across horizons with p-values that are very close to zero, providing further support for rejections of the EH across all forecast horizons including the very long-run. Note that this latter difference specification allows to eliminate trends in both the dependent and explanatory variables of the Froot

Table 4: **Regression Tests of the EH.** This table presents linear regression estimates and the associated p-values (in percent) for the null hypothesis $\beta_s^{(n,m)} = 1$ and $\beta_d^{(n,m)} = 1$. Hypothesis tests are formed using a t-statistic with the equal-weighted cosine variance estimator of Lazarus, Lewis, Stock, and Watson (2018). Minimum and maximum statistics are taken across 1,000 posterior draws. The sample period is 1983m1–2019m12.

					Specification 1					
	1Y	1Y1Y	2Y1Y	3Y1Y	4Y1Y	5Y1Y	6Y1Y	7Y1Y	8Y1Y	9Y1Y
Min. $\hat{\beta}_s^{(n,m)}$	0.37	0.44	0.55	0.64	0.69	0.71	0.72	0.71	0.71	0.71
Max. $\hat{\beta}_s^{(n,m)}$	0.38	0.46	0.58	0.67	0.72	0.74	0.74	0.74	0.74	0.74
Min. p-val (%)	0.00	0.04	0.66	1.92	2.39	1.81	1.10	0.67	0.46	0.40
Max p-val (%)	0.00	0.07	1.06	3.06	3.98	3.30	2.28	1.55	1.17	1.06
					Specification 2					
	1Y	1Y1Y	2Y1Y	3Y1Y	4Y1Y	5Y1Y	6Y1Y	7Y1Y	8Y1Y	9Y1Y
Min. $\hat{\beta}_d^{(n,m)}$	0.73	0.51	0.33	0.20	0.12	0.08	0.05	0.04	0.04	0.03
Max. $\hat{\beta}_d^{(n,m)}$	0.75	0.55	0.38	0.28	0.21	0.18	0.16	0.15	0.14	0.14
Min. p-val (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max p-val (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(1989) regressions, and as such, some of the co-movement between expectations and yields that may be induced by common trends. This is an issue we return to in the next section.

4.1 Why do the Tests Reject the EH?

The previous section shows that the EH is rejected decisively; however, a statistical test is a dichotomous outcome. Instead we can ask by how much does the EH fail? To dissect this finding, we provide a simple variance decomposition of forward rates into expected inflation, expected real rates and the wedge using 12-month changes in each variable. We have that

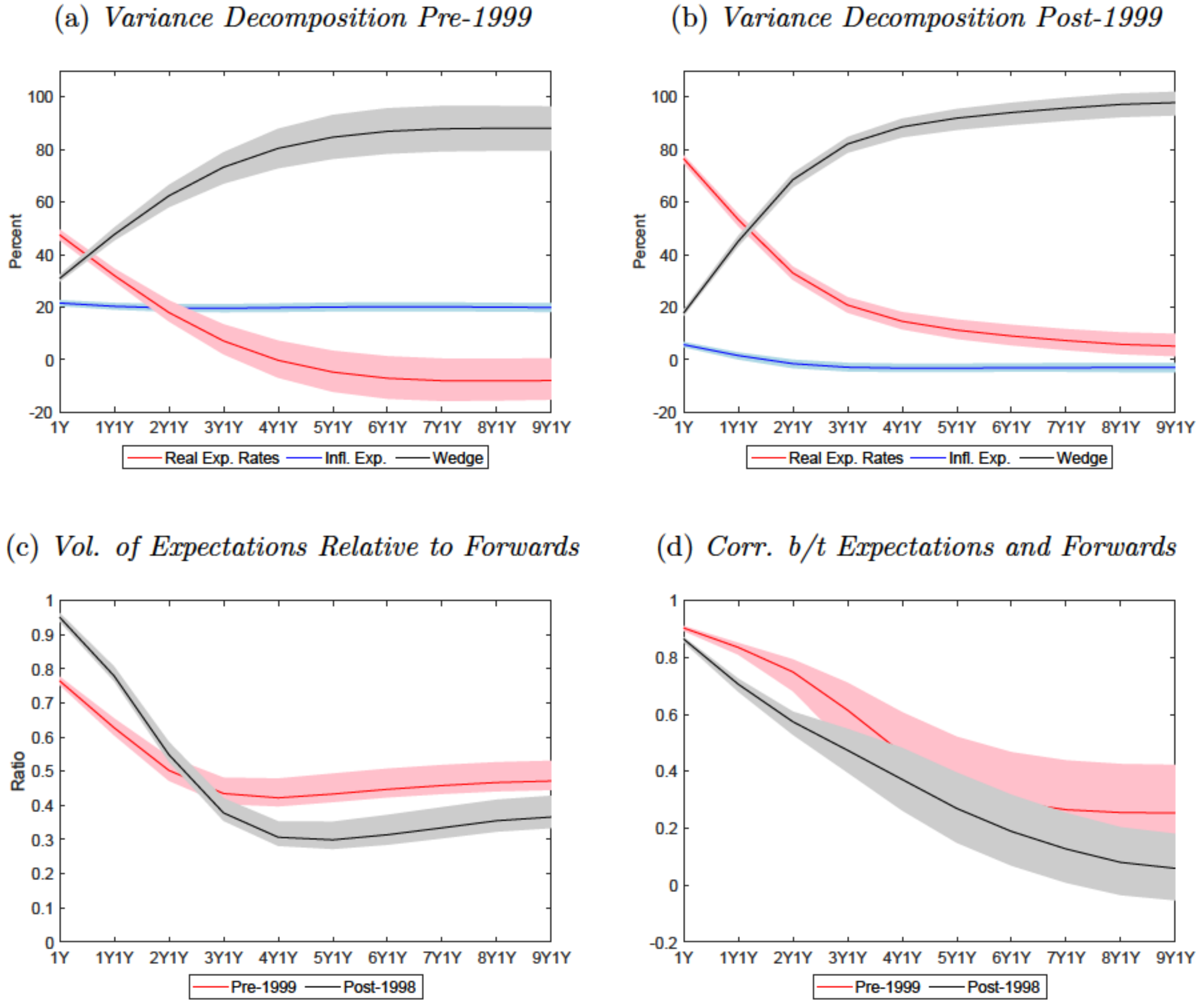
$$\hat{\mathfrak{s}}\left(n^{-1}\sum\nolimits_{h=m+1}^{n+m}\Delta_\tau\mathbb{E}_t\left[r_{t+h-1}\right]\right)+\hat{\mathfrak{s}}\left(n^{-1}\sum\nolimits_{h=m+1}^{n+m}\Delta_\tau\mathbb{E}_t\left[\pi_{t+h}\right]\right)+\hat{\mathfrak{s}}\left(\Delta_\tau\mathsf{w}_t(n,m)\right)=1. \quad (4.8)$$

Here,

$$\hat{\mathfrak{s}}(\cdot)=\frac{\widehat{\text{Cov}}\left(\Delta_\tau\mathsf{fwd}_t(n,m),\cdot\right)}{\widehat{\text{Var}}(\Delta_\tau\mathsf{fwd}_t(n,m))} \quad (4.9)$$

is the ratio between the corresponding sample covariance between forward rates and each constituent component and the sample variance of the forward rate, and $\Delta_\tau\varsigma_t = \varsigma_t - \varsigma_{t-\tau}$ for any time series ς_t.

Figure 4: **Decomposing Forward Rates.** The top panels show the variance decomposition of the 12-month changes of forward rates at different maturities in expected real rates (red line), inflation expectations (blue line) and the wedge (black line) components. The bottom left panel displays the volatility of 12-month changes in interest rate expectations relative to forwards for the pre-1999 (red line) and post-1998 (black line) sample. Solid lines denote posterior medians while shaded regions denote the range of 1,000 posterior draws. The sample period is 1983m1–2019m12.



The top panels in Figure 4 show the contributions of the three components to the variance of forwards in the sample before and after 1998.[22] The decomposition highlights the striking finding that beyond a maturity of three years (i.e., the 2Y1Y forward), over eighty percent of the variation in forward rates is driven by movements in the wedge rather than interest-rate expectations. Expected nominal short rates explain between 69% and 82% for the 1-year yield and between 52% and 55% for the 1Y1Y forward rate, and less than 50% for all other

[22]Note that individual contributions must sum to one but can be negative, depending on the sign of the covariance.

forward maturities. As the forward maturity increases, the contribution of expectations to the variability of forward rates declines sharply. This pattern is not affected by the specific sample selected, suggesting that the volatility of forward rates has also declined in the post-1998 period. The only notable difference between sub-samples is the contribution of inflation expectations. In the pre-1999 period, inflation expectations contribute about twenty-percent to the variance of forward rates across all horizons. Conversely, their contribution is close to zero at all horizons in the post-1998 sample. This is mirrored by an increased role of expected real rates at shorter maturities. Finally, each plot also provides shaded regions reflecting sampling uncertainty about the model-generated expectations across $1,000$ posterior draws. All the patterns and conclusions are unchanged when considering any of the individual posterior draws.

The bottom panels in Figure 4 offer a further insight into the failure of the EH. The sample variance share of the components, $\hat{\mathfrak{s}}(\cdot)$, can be re-expressed in terms of sample standard deviations and correlations

$$\hat{\mathfrak{s}}\left(\cdot\right)=\widehat{\mathrm{Corr}}\left(\Delta_{\tau}\mathsf{fwd}_{t}(n,m),\cdot\right)\left(\frac{\widehat{\mathrm{Var}}(\cdot)}{\widehat{\mathrm{Var}}(\Delta_{\tau}\mathsf{fwd}_{t}(n,m)}\right)^{1/2}. \tag{4.10}$$

As we have already shown, the term structure of short-rate expectations implied by our estimated model exhibits substantial volatility. The bottom left panel supports this claim: the volatility of expectations remains above forty percent of the variability of forward rates across all maturities. This would suggest a more important role for expectations even at the longest horizons. However, as indicated by the right chart, the correlation between forward rates and expectations declines steadily towards zero as the maturity increases.

Importantly, these conclusions do not depend on the frequency of interest rate changes. We choose one year changes because it eliminates high frequency fluctuations in asset prices that may be short-lived. Figure 5 replicates the structure of Figure 4 for the 9Y1Y forward only but now for 1-year, 2-year, 5-year and 10-year changes.[23] When we expand the frequency up to five-year changes the conclusions from Figure 4 are essentially unchanged. The wedge explains over 85 percent of the variation in the forward rates for either the first or second half of the sample. At the same time, the relative volatility of expectations as compared to forwards is higher than for one-year changes. For 10-year changes, which are below business-cycle frequency, the presence of the downward trend in both yields and expectations in the first half of the sample plays a larger role. In the pre-1999 sample, the relative volatility of expectations as compared to the longer-horizon forward is close to one. Despite that,

[23] In Section SA1-3 of SA1, we present the full counterpart of Figure 4 for 2-year, 5-year and 10-year changes for all forward maturities.

the 10-year change in expected short-rates only explains about 55% of the variation in the 9Y1Y forward. In the second half of our sample, when the downward trend is diminished, the variance decomposition is not significantly altered when calculated with 10-year changes as compared to the higher frequency changes. In fact, the wedge explains over 75% of the variation in longer-horizon forward rates. Taken in sum, not only is the EH decisively rejected in the data, but model-implied short-rate expectations generally display, at best, only a weak co-movement with the forward rates of corresponding maturities.

4.2 Do the Driving Forces of Expectations Explain the Wedge?

Despite the results presented in the previous section, it is possible that the same underlying forces that drive expectations also drive the time variation in the wedge, $\mathsf{w}_t(n,m)$. Recall from Equation (2.7) that expectations are linear in six underlying variables, $\omega_{t|t}$ and $x_{t|t}$, which we filter as $\hat{\omega}_{t|t}$ and $\hat{x}_{t|t}$. Thus, we can span forecasts at all horizons with only these six variables and consider regressions of the form,

$$\Delta_\tau \mathsf{fwd}_t(n,m) = c^{(n,m)} + d_x^{(n,m)\prime}\Delta_\tau \hat{x}_{t|t} + d_\omega^{(n,m)\prime}\Delta_\tau \hat{\omega}_{t|t} + \xi^{(n,m)}_{\mathsf{fwd},\tau,t}, \tag{4.11}$$

and

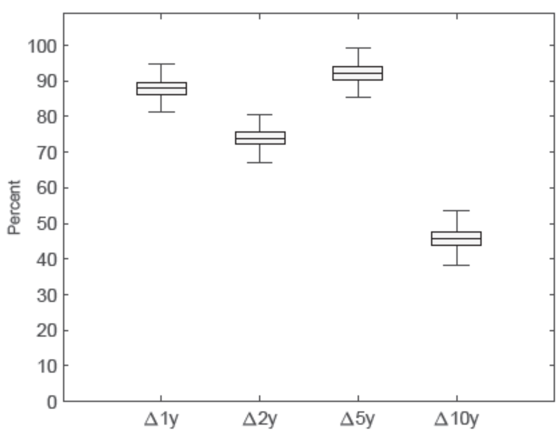
$$\Delta_\tau \mathsf{w}_t(n,m) = a^{(n,m)} + b_x^{(n,m)\prime}\Delta_\tau \hat{x}_{t|t} + b_\omega^{(n,m)\prime}\Delta_\tau \hat{\omega}_{t|t} + \xi^{(n,m)}_{\mathsf{w},\tau,t}. \tag{4.12}$$

where τ is 12, 60 or 120 months. For each of these regressions we assess the goodness of fit to examine how well forwards and the wedge are spanned by the state variables that drive beliefs about the short-term interest rate.

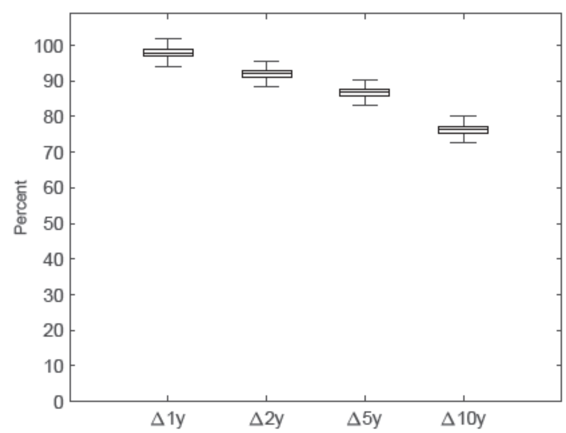
The top panel of Table 5 reports the regression R^2 for Equation (4.11). Consistent with the evidence we have already presented, the co-movement between forwards and $\hat{x}_{t|t}$ and $\hat{\omega}_{t|t}$ is much stronger at shorter maturities. At medium and longer maturities, the R^2 decline monotonically. A useful benchmark for these results is the class of affine term structure models (Duffee (2002), Ang and Piazzesi (2003), Joslin, Singleton, and Zhu (2011), Adrian, Crump, and Moench (2013)) which explicitly allow for time-variation in the price of risk. In this class of models, yields and forward rates are affine functions of the set of risk factors with coefficients obeying specific cross-maturity restrictions. As such, an affine term structure model using $\hat{\omega}_{t|t}$ and $\hat{x}_{t|t}$ as factors can have no higher R^2 than those reported in the top panel of Table 5. Furthermore, if the first three principal components of yields were instead chosen as the risk factors in an affine model—as is the case, e.g., in Joslin, Singleton, and Zhu (2011)—we would observe $R^2 \approx 1$ for all maturities and choice of τ.

Figure 5: **Decomposing Long-Horizon Forwards at Different Frequencies.** This figure replicates Figure 4 for the 9Y1Y forward rate for different choices of Δ_τ. We choose τ as 1-year, 2-year, 5-year and 10-year increments. The top row shows the variance share of forwards attributed to the wedge over the pre-1999 and post-1998 samples for different values of τ. The bottom left chart displays the relative volatility of changes in 9Y1Y short-rate expectations compared to the 9Y1Y forward rate for the pre-1999 (red) and post-1998 (black) sample. The bottom right chart shows the correlation between changes in expectations and forward rates for the pre-1999 (red) and post-1998 (black) sample. For all charts, the boxes indicate the interquartile range (with center line equal to the median) and the whiskers represent the maximum and minimum across 1,000 posterior draws. The sample period is 1983m1–2019m12.

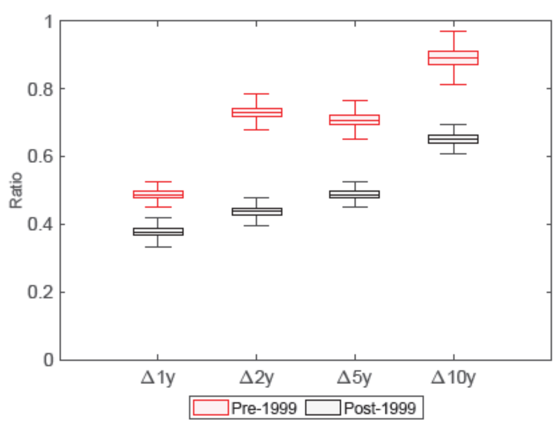
(a) *Variance Decomposition Pre-1999*



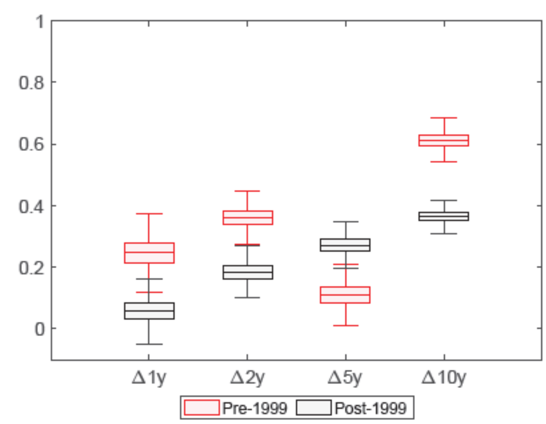
(b) *Variance Decomposition Post-1999*



(c) *Vol. of Expectations Relative to Forwards*



(d) *Corr. b/t Expectations and Forwards*



We can now investigate directly whether the state variables that drive beliefs also explain the behavior of the wedge. The bottom panel of Table 5 reports the regression R^2 for Equation (4.12). When using the wedge as the left-hand side variable, the R^2 values drop precipitously at the short end of the curve and remain low elsewhere. Across all three choices of τ (and across all 1,000 posterior draws), the R^2 is no higher than 51%. In

Table 5: **Spanning Tests.** The top panel (bottom panel) presents R^2 from regressions of changes in the forward rate (wedge) on a constant and the six state variables, $\hat{x}_{t|t}$ and $\hat{\omega}_{t|t}$. Minimum and maximum statistics are taken across 1,000 posterior draws. The sample period is 1983m1–2019m12.

	Forward Rates: $\mathsf{fwd}_t(n,m)$									
	1Y	1Y1Y	2Y1Y	3Y1Y	4Y1Y	5Y1Y	6Y1Y	7Y1Y	8Y1Y	9Y1Y
	12-month Changes									
Min. R^2 (%)	79.6	65.6	55.1	47.3	41.9	38.5	36.4	35.1	33.6	32.8
Max. R^2 (%)	81.6	71.0	63.8	58.0	53.3	49.9	47.6	46.0	44.9	44.2
	60-month Changes									
Min. R^2 (%)	92.6	84.2	72.7	60.8	51.6	45.8	42.2	39.9	38.2	37.5
Max. R^2 (%)	93.4	86.1	77.2	68.3	61.6	57.2	54.3	52.0	50.2	49.7
	120-month Changes									
Min. R^2 (%)	89.2	80.4	73.5	68.4	64.5	61.5	59.1	57.0	55.4	54.5
Max. R^2 (%)	90.2	83.3	78.1	74.3	71.4	68.9	67.3	65.9	64.5	63.8

	Wedges: $\mathsf{w}_t(n,m)$									
	1Y	1Y1Y	2Y1Y	3Y1Y	4Y1Y	5Y1Y	6Y1Y	7Y1Y	8Y1Y	9Y1Y
	12-month Changes									
Min. R^2 (%)	12.1	18.3	23.5	27.9	30.8	32.8	33.8	34.3	34.5	34.5
Max. R^2 (%)	19.3	27.2	33.5	38.1	41.1	43.1	44.1	44.5	44.5	44.5
	60-month Changes									
Min. R^2 (%)	23.2	29.5	34.5	37.6	39.2	39.6	40.1	41.0	42.1	42.8
Max. R^2 (%)	29.4	37.4	43.4	46.9	48.1	48.4	49.1	49.7	50.3	50.8
	120-month Changes									
Min. R^2 (%)	38.7	43.1	42.3	37.5	33.1	31.1	30.6	30.7	30.6	30.5
Max. R^2 (%)	44.2	50.0	50.9	47.4	43.6	41.1	40.8	40.8	40.7	40.5

Section SA1-4 of SA1, we report the results for a quadratic specification corresponding to all interactions of the six right-hand side variables (27 regressors in all). Although the R^2 values meaningfully increase in this specification they are still far from fully explaining the movements in the wedges. This evidence is strongly suggestive that other forces than those which drive expectations play a key role in explaining the wedge.

To conclude, these results suggest that allowing for time variation in risk premia in otherwise standard macroeconomic models would be insufficient to capture the behavior of the term structure of interest rates, as differences between observed yields and model-implied short rate forecasts seem to be driven by factors separate from those that explain the term structure of short rate expectations.

4.3 Consensus vs. the Marginal Investor

According to the EH, yields reflect "the market" expected path of the short-term interest rate, generally taken to be the average opinion (i.e., the consensus) among market participants. Following the existing literature, starting from Friedman (1979), we use consensus measures from professional survey forecasts as a noisy indicator of the average opinion of the market. The previous sections show that expectations consistent with this consensus measure deviate substantially from observed forward rates. This result is unlikely to stem from severe mis-measurement of the true unobserved market average opinion: professional forecasters are, after all, often drawn from market participants. Furthermore, we have already shown that, at short forecast horizons, expectations and forward rates tend to be close so that the divergence occurs at longer forecast horizons. But we also showed that revisions to long-run forecasts co-move with short-term forecast errors consistent with a trend-cycle model. In order for mis-measurement in expectations to be the explanation for our results, we would require that the true "market expectation" obey a fundamentally different process that is largely uncorrelated to our model's predictions.

An alternative criticism is that the "market expectation" does not correspond to that of the average investor but instead to the expectation of the *marginal* investor in the bond market. In this case, our wedge would be sourced to a large and time-varying gap between the beliefs of the average and the marginal investor. Such a large gap is consistent with evidence from individual forecasts in professional survey data. Surveys indicate a sizable and time-varying disagreement about the expected path of the short-term interest rate. Importantly, Andrade, Crump, Eusepi, and Moench (2016) show that the term structure of disagreement for the short-rate is upward sloping: belief dispersion increases at longer forecast horizons. This is consistent with the evidence we have presented where longer-maturity expectations and the corresponding forwards have little to no relation. Furthermore, Cao, Crump, Eusepi, and Moench (2021) show that professional forecasters' disagreement about the short-term interest rate co-moves with our wedge for long forward maturities, providing empirical evidence for the "disagreement channel."[24]

These concepts are formalized in a theoretical literature that explores equilibrium bond prices with heterogenous beliefs. In these models, dispersed beliefs can lead to speculative behavior in the bond market, driving a wedge between bond prices and the EH-implied (consensus) expectations. For example, Xiong and Yan (2009), Barillas and Nimark (2017) and Buraschi and Whelan (2022) show that in standard economies with constant risk aversion and no market frictions, equilibrium bond prices violate the EH and depend on time-varying

[24] Additionally, Giacoletti, Laursen, and Singleton (2021) show that short-term disagreement in bond yield expectations helps predict bond returns.

disagreement. In Xiong and Yan (2009) bond prices are driven by the weighted average of individual investors' expectations, where the weights are time-varying and depend on the relative wealth of investors.

In these models, a gap between the average and marginal investor naturally arises. The presence of such a gap is entirely consistent with our finding that the EH does not hold: a key conclusion from this theoretical literature is that, in a market with dispersed beliefs, there does not exist a "market expectation" in the conventional sense. Jouini and Napp (2007) and Xiong and Yan (2009) show, in a standard complete markets economy, that equilibrium asset prices under heterogenous expectations can be replicated by a "representative investor" with a specific set of beliefs about the expected path of the interest rate. However, this representative investor's beliefs do not satisfy standard probability laws, such as the law of iterated expectations. In other words, since the marginal investor changes over time, this representative investor's expectations could not follow an expectations formation process like the one described in Section 2. Thus, heterogeneity of beliefs, which would drive a gap between the average and marginal investor's views, in fact, takes us farther away (not closer) from the EH holding.

5 Conclusion

In this paper, we reevaluate the empirical evidence regarding the EH by proposing a model of expectations formation that allows for deviations from FIRE and accounts for time-varying beliefs about the long-run. This class of models has shown promise to bridge the gap between EH-implied and observed yields, fueling hopes for a "resurrection" of the EH. We estimate the model using the universe of consensus forecasts from all U.S. surveys of professional forecasters covering more than 600 survey-horizon pairs at a monthly frequency. While model-implied short-rate expectations move considerably at all horizons and suggest significant departures from rational expectations, they do not come close to matching the observed term structure of interest rates. Instead, the EH-implied short-rate expectations generally display, at best, only a weak co-movement with the forward rates of corresponding maturities. Not surprisingly, formal tests of the EH are soundly rejected.

These results suggest alternative explanations for the behavior of observed bond yields such as heterogenous beliefs, financial market frictions, nonstandard risk preferences and behavioral theories of asset pricing. Accommodating such features in models of equilibrium bond prices can have important implications for macroeconomic models, including in the transmission mechanism of monetary policy. In standard models, used by both academics and policymakers, the monetary transmission channel is based solely on the EH. The central bank can exert a tight control on longer-term interest rates by responding to changing economic conditions in a systematic manner, i.e. adhering to time-invariant policy rules, or by communicating directly about likely future policy moves through forward guidance. The sizable deviation of observed interest rates from the EH, which we document, calls in to question this conventional framework.

References

Adrian, T., Crump, R. K., Moench, E., 2013. Pricing the term structure with linear regressions. Journal of Financial Economics 110, 110–138.

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

Andrade, P., Crump, R. K., Eusepi, S., Moench, E., 2016. Fundamental disagreement. Journal of Monetary Economics 83, 106–128.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 50, 745–787.

Angeletos, G.-M., Huo, Z., Sastry, K. A., 2020. Imperfect macroeconomic expectations: Evidence and theory. In: *NBER Macroeconomics Annual 2020*, National Bureau of Economic Research, vol. 35, pp. 1–86.

Barillas, F., Nimark, K. P., 2017. Speculation and the term structure of interest rates. Review of Financial Studies 30, 4003–4037.

Bordalo, P., Gennaioli, N., La Porta, R., Shleifer, A., 2019. Diagnostic expectations and stock returns. Journal of Finance 74, 2839–2874.

Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2020. Overreaction in macroeconomic expectations. American Economic Review 110, 2748–82.

Buraschi, A., Whelan, P., 2022. Speculation, sentiment, and interest rates. Management Science 68, 2308–2329.

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

Cao, S., Crump, R. K., Eusepi, S., Moench, E., 2021. Fundamental disagreement about monetary policy and the term structure of interest rates. Staff Report 934, Federal Reserve Bank of New York.

Carvalho, C., Eusepi, S., Moench, E., Preston, B., 2023. Anchored inflation expectations. American Economic Journal: Macroeconomics 15, 1–47.

Cieslak, A., 2018. Short-rate expectations and unexpected returns in Treasury bonds. Review of Financial Studies 31, 3265–3306.

Cogley, T., 2005. Changing beliefs and the term structure of interest rates: Cross-equation restrictions with drifting parameters. Review of Economic Dynamics 8, 420–452.

Coibion, O., Gorodnichenko, Y., 2012. What can survey forecasts tell us about informational rigidities? Journal of Political Economy 120, 116–159.

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

Crump, R. K., Eusepi, S., Lucca, D., Moench, E., 2014. Which growth rate? It's a weighty subject. Liberty Street Economics Blog.

Crump, R. K., Eusepi, S., Moench, E., 2018. The term structure of expectations and bond yields. Staff Report 775, Federal Reserve Bank of New York.

Crump, R. K., Eusepi, S., Moench, E., Preston, B., 2023. The term structure of expectations. In: Bachmann, R., Topa, G., van der Klaauw, W. (eds.), *Handbook of Economic Expectations*, Elsevier, pp. 507–540.

Crump, R. K., Eusepi, S., Moench, E., Preston, B., 2024. How do we learn about the long run?, working paper.

Del Negro, M., Eusepi, S., 2011. Fitting observed inflation expectations. Journal of Economic Dynamics and Control 35, 2105–2131.

Del Negro, M., Giannone, D., Giannoni, M. P., Tambalotti, A., 2018. Safety, liquidity, and the natural rate of interest. Brookings Papers on Economic Activity 49, 235–94.

Dewachter, H., Iania, L., Lyrio, M., 2011. A new-Keynesian model of the yield curve with learning dynamics: A Bayesian evaluation, working paper.

Duffee, G., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

Ehrbeck, T., Waldmann, R., 1996. Why are professional forecasters biased? agency versus behavioral explanations. Quarterly Journal of Economics 111, 21–40.

Eusepi, S., Giannoni, M., Preston, B., 2024. The short-run policy constraints of long-run expectations. Journal of Political Economy Forthcoming.

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

Eusepi, S., Preston, B., 2023. A short history in defense of adaptive learning. In: Lenel, L., Nützenadel, A., Köhler, I., Streb, J. (eds.), *The Routledge Handbook of Economic Expectations in Historical Perspective*, Routledge, forthcoming.

Farmer, L., Nakamura, E., Steinsson, J., 2023. Learning about the long run. Journal of Political Economy 132, 3334–3377.

Friedman, B. M., 1979. Interest rate expectations versus forward rates: Evidence from an expectations survey. Journal of Finance 34, 965–973.

Froot, K. A., 1989. New hope for the expectations hypothesis of the term structure of interest rates. Journal of Finance 44, 283–305.

Fuhrer, J. C., 1996. Monetary policy shifts and long-term interest rates. Quarterly Journal of Economics 111, 1183–1209.

Giacoletti, M., Laursen, K. T., Singleton, K. J., 2021. Learning from disagreement in the u.s. treasury bond market. Journal of Finance 76, 395–441.

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

Gurkaynak, R. S., Sack, B., Wright, J. H., 2007. The U.S. Treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 2291–2304.

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press.

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

Joslin, S., Singleton, K. J., Zhu, H., 2011. A new perspective on Gaussian dynamic term structure models. Review of Financial Studies 24, 926–970.

Jouini, E., Napp, C., 2007. Consensus consumer and intertemporal asset pricing with heterogeneous beliefs. Review of Economic Studies 74, 1149–1174.

Kim, D. H., Orphanides, A., 2012. Term structure estimation with survey data on interest rate forecasts. Journal of Financial and Quantitative Analysis 47, 241–272.

Kim, D. H., Wright, J. H., 2005. An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. Finance and Economics Discussion Series 2005-33, Federal Reserve Board.

Kozicki, S., Tinsley, P. A., 2001. Shifting endpoints in the term structure of interest rates. Journal of Monetary Economics 47, 613–652.

Laubach, T., Williams, J. C., 2003. Measuring the natural rate of interest. Review of Economics and Statistics 85, 1063–1070.

Lazarus, E., Lewis, D. J., Stock, J. H., Watson, M. W., 2018. HAR inference: Recommendations for practice. Journal of Business & Economic Statistics 36, 541–559.

Lewandowski, D., Kurowicka, D., Joe, H., 2009. Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis 100, 1989–2001.

Lutz, F. A., 1940. The structure of interest rates. Quarterly Journal of Economics 55, 36–63.

Piazzesi, M., Salomao, J., Schneider, M., 2015. Trend and cycle in bond premia, working paper.

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

Singleton, K. J., 2021. Presidential address: How much "rationality" is there in bond-market risk premiums? Journal of Finance 76, 1611–1654.

Sinha, A., 2016. Learning and the yield curve. Journal of Money, Credit and Banking 48, 513–547.

Woodford, M., 2003. Interest and prices: foundations of a theory of monetary policy. Princeton University Press.

Wright, J. H., 2011. Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. American Economic Review 101, 1514–34.

Xiong, W., Yan, H., 2009. Heterogeneous expectations and bond markets. Review of Financial Studies 23, 1433–1466.

Appendix

A Model

The perceived law of motion is as described in the main text. Revisions to the estimates of the state variables evolve as:

$$\begin{pmatrix} \omega_{t|t} \\ x_{t|t} \end{pmatrix} = \begin{pmatrix} \omega_{t|t-1} \\ x_{t|t-1} \end{pmatrix} + \mathcal{K}\left(\mathcal{S}_t - \mathcal{S}_{t|t-1}\right)$$

where

$$\mathcal{K} \equiv \bar{P}H\left(H'\bar{P}H + \Sigma_s\right)^{-1}$$

and $\bar{P}$ solves[25]

$$\bar{P} = F\left[\bar{P} - \bar{P}H\left(H'\bar{P}H + \Sigma_s\right)^{-1} H'\bar{P}\right]F' + \Sigma_z.$$

Under the correct model (rational expectations) the estimates of the state variables evolve according to

$$\begin{pmatrix} \omega_{t|t} \\ x_{t|t} \end{pmatrix} = \begin{pmatrix} \omega_{t|t-1} \\ x_{t|t-1} \end{pmatrix} + \Sigma_\varepsilon^{1/2}\epsilon_t = F\begin{pmatrix} \omega_{t-1|t-1} \\ x_{t-1|t-1} \end{pmatrix} + \Sigma_\varepsilon^{1/2}\epsilon_t$$

where the shock vector ϵ_t is normally distributed with zero mean and identity variance-covariance matrix and

$$\Sigma_\varepsilon \equiv \mathcal{K}\left(H'\bar{P}H + \Sigma_s\right)\mathcal{K}' = \bar{P}H\left(H'\bar{P}H + \Sigma_s\right)^{-1} H'\bar{P}.$$

We do not make assumptions about rationality. The agent may have the "wrong" model and departures from rational beliefs can manifests itself in two ways: First, the agent might have the correct law of motion but the incorrect model parameters, so that their forecast error follows a different distribution. That is, the variance covariance of the innovations may differ from $\bar{P}H\left(H'\bar{P}H + \Sigma_s\right)^{-1} H'\bar{P}$. Second, forecast errors can display serial correlation, as discussed in the main text.

Now we can write the state space of our model as

$$\begin{pmatrix} z_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \\ z_{t-4} \\ \xi_{t|t} \\ f_t \end{pmatrix} = \begin{bmatrix} 0_{3\times3} & 0_{3\times12} & \begin{bmatrix} I_3 & I_3 \end{bmatrix} F & \begin{bmatrix} I_3 & I_3 \end{bmatrix} G \\ I_3 & 0_{3\times12} & 0_{3\times6} & 0_{3\times6} \\ 0_{3\times3} & I_3 & 0_{3\times15} & 0_{3\times6} \\ 0_{3\times6} & I_3 & 0_{3\times12} & 0_{3\times6} \\ 0_{3\times9} & I_3 & 0_{3\times9} & 0_{3\times6} \\ 0_{6\times3} & 0_{3\times12} & F & G \\ 0_{6\times3} & 0_{6\times12} & 0_{6\times6} & G \end{bmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-2} \\ z_{t-3} \\ z_{t-4} \\ z_{t-5} \\ \xi_{t-1|t-1} \\ f_{t-1} \end{pmatrix}$$

[25] See Chapter 13 in Hamilton (1994).

$$
+ \begin{bmatrix} \begin{bmatrix} I_3 & I_3 \end{bmatrix} \\ 0_{3\times 6} \\ 0_{3\times 6} \\ 0_{3\times 6} \\ 0_{3\times 6} \\ I_6 \\ I_6 \end{bmatrix} \Sigma_\varepsilon^{1/2} \varepsilon_t
$$

where

$$
\begin{aligned}
F &\equiv \begin{bmatrix} I_3 & 0 \\ 0 & \Phi \end{bmatrix} \\
G &\equiv \begin{bmatrix} \Phi_f & 0 \\ 0 & \Phi_f \end{bmatrix},
\end{aligned}
$$

and $\xi_{t|t} = (\omega_{t|t}, x_{t|t})'$. Importantly, the measurement equation of the model (from the econometrician's viewpoint) used to fit the survey forecasts is not a function of G. For example,

$$
E_t z_{t+h} = e_3 \times \begin{bmatrix} 0_{3\times 3} & 0_{3\times 12} & \begin{bmatrix} I_3 & I_3 \end{bmatrix} F & 0_{3\times 6} \\ I_3 & 0_{3\times 12} & 0_{3\times 6} & 0_{3\times 6} \\ 0_{3\times 3} & I_3 & 0_{3\times 15} & 0_{3\times 6} \\ 0_{3\times 6} & I_3 & 0_{3\times 12} & 0_{3\times 6} \\ 0_{3\times 9} & I_3 & 0_{3\times 9} & 0_{3\times 6} \\ 0_{6\times 3} & 0_{3\times 12} & F & 0_{6\times 6} \\ 0_{6\times 3} & 0_{6\times 12} & 0_{6\times 6} & 0_{6\times 6} \end{bmatrix}^h \begin{pmatrix} z_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \\ z_{t-4} \\ \xi_{t|t} \\ f_t \end{pmatrix}
$$

where e_3 selects the first three rows. Said differently, the agent is assumed to form expectations under the assumption $\mathbb{E}_t[f_{t+h}] = 0$ for all h.

B Measurement equation

In this section we provide greater detail on how we map survey forecasts to our modeling framework discussed in Section 3.2. Forecasts for the three-month Treasury bill rate are either a simple average over a period or end of period. For the latter we assign these forecasts to the last month in the period. For real output growth and inflation, survey forecasts come in three possible forms: quarter-over-quarter (QoQ) annualized growth, annual average growth and Q4/Q4 growth. The distinction between these growth rates are best illustrated through examples. In these examples we will ignore measurement error for simplicity. Let G_{2013Q1} and G_{2013Q2} be the level of real GDP in billions of chained dollars in the first and second quarter of 2013, respectively. Then, QoQ annualized growth rate is defined as $100\cdot((G_{2013Q2}/G_{2013Q1})^4 - 1)$. In our model we filter a month-over-month (annualized) real GDP growth rate series. To map the monthly series into this specific quarterly growth rate we follow Crump et al. (2014) and use

$$
100 \cdot ((G_{2013Q2}/G_{2013Q1})^4 - 1) \approx \frac{1}{9} \left(g_{2013m2} + 2 \cdot g_{2013m3} + 3 \cdot g_{2013m4} + 2 \cdot g_{2013m5} + g_{2013m6} \right),
$$

where, for example, g_{2013m2} represents month-over-month annualized real output growth in February 2013.

Annual average growth rates follow a similar pattern. For example, let G_{2012} and G_{2013} be the average level of real GDP in billions of chained dollars in the years 2012 and 2013, respectively. Then the annual average growth rate is $100 \cdot (G_{2013}/G_{2012} - 1)$ which we approximate via,

$$\begin{aligned} 100 \cdot (G_{2013}/G_{2012} - 1) \approx \frac{1}{24} & (g_{2012m2} + 2 \cdot g_{2012m3} + 3 \cdot g_{2012m4} + \cdots + 12 \cdot g_{2013m1} \\ & +11 \cdot g_{2013m2} + 10 \cdot g_{2013m3} + \cdots + 2 \cdot g_{2013m11} + g_{2013m12}) . \end{aligned}$$

Finally, Q4/Q4 growth rates are calculated, for example, by $100 \cdot (G_{2013Q4}/G_{2012Q4} - 1)$ and approximated via

$$100 \cdot (G_{2013Q4}/G_{2012Q4} - 1) \approx \frac{1}{12} (g_{2013m1} + g_{2013m2} + g_{2013m3} + \cdots + g_{2013m12}) .$$

The above shows that certain survey forecast horizons will implicitly include time periods which have already occurred. To avoid taking a stand on how forecasters treat past data (e.g., do forecasters use realized data, filtered versions or another measure?) we exclude all survey forecast horizons that include past months' values of y_t. The only exception we make is to include current quarter (Q0) and one-quarter ahead (Q1) forecasts for real output growth (which extend back, at most, four months and one month, respectively). This is why our measurement equation contains lags of z_t up to z_{t-4}. We do so to help pin down monthly real output growth since the actual series is only available at a quarterly frequency. Finally, for simplicity, forecasts which involve averages over multiple years are mapped as simple averages over the corresponding horizons.

Supplemental Appendix (SA1): "Is There Hope for the Expectations Hypothesis?"

Richard K. Crump, Stefano Eusepi & Emanuel Moench

November 5, 2024

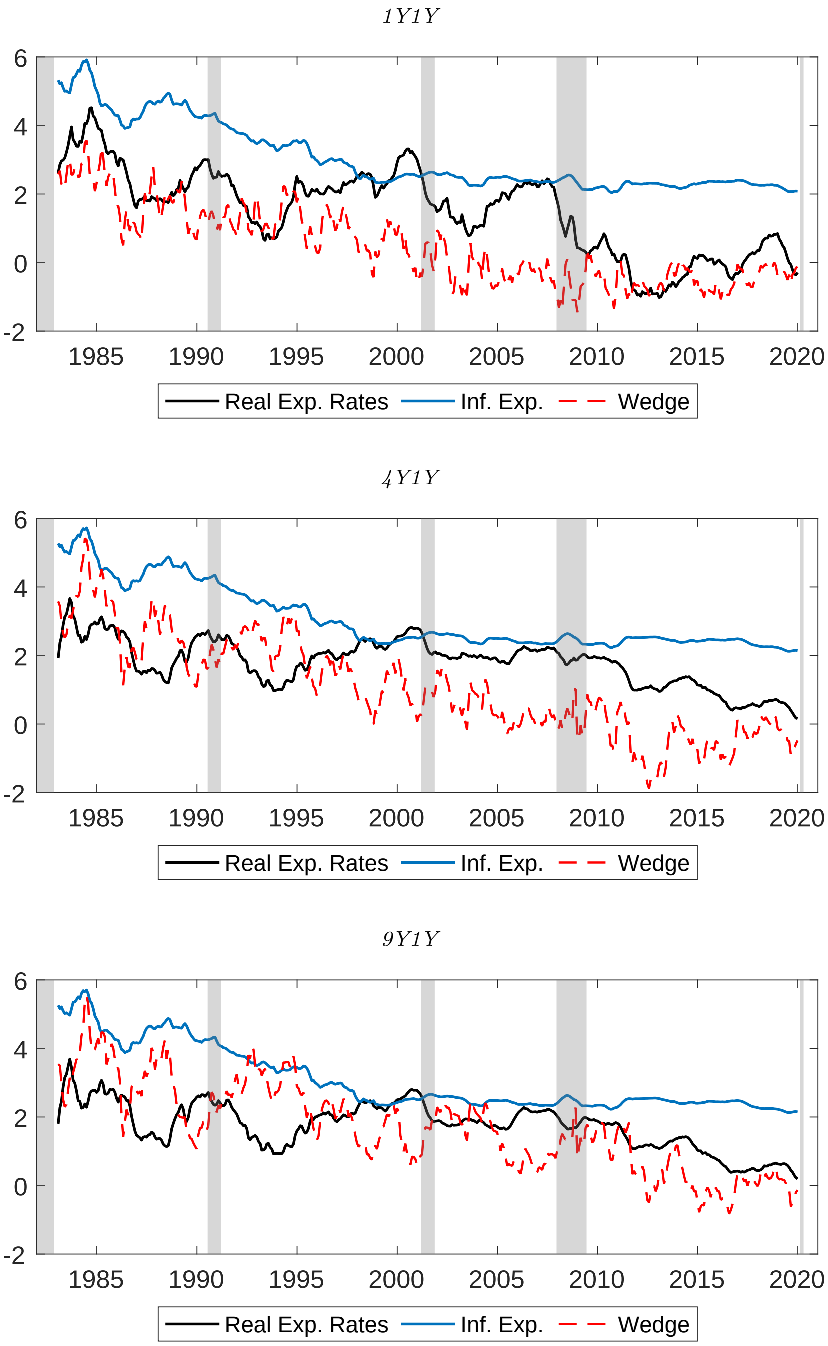
SA1-1 Decomposition of Forward Rates

Figure SA1-1 visualizes these three components of nominal Treasury forward rates for the 1Y1Y, 4Y1Y and 9Y1Y forward horizons (top, middle, and bottom panel, respectively).[1] Clearly, the wedges are sizable at all maturities and display substantial variation over time. Casual inspection of the figure suggests the expected real rate plays a dominant role at low maturities, but its relative importance decreases at longer horizons. In contrast, inflation expectations exhibit subdued cyclical fluctuations across all maturities. Both real rate and inflation expectations display lower variation after 1999, as captured by the structural shift in the second sample. All components display a downward trend in our sample. While inflation expectations stabilize, both expected real rates and wedges further decline in the aftermath of the financial crisis, with the wedges turning negative during this period.[2]

[1] We obtain zero-coupon forward rates from the data set of Gurkaynak et al. (2007) available at `https://www.federalreserve.gov/data/nominal-yield-curve.htm`.

[2] Our finding of a secular decline in the wedges is consistent with the evidence in Wright (2011) who uses an affine term structure model to show that term premiums in the U.S. and in other developed countries have experienced sizable and persistent declines between 1990 and mid-2009.

Figure SA1-1. Decomposing Forward Rates. The figure shows the individual components of forward rates at different maturities. The blue line measures expected inflation, the black line the expected short-term real rate, and the red line the wedge. From top-to-bottom the figure visualizes the components of 1Y1Y, 4Y1Y and 9Y1Y forward rate respectively: at each maturity, the sum of the three components return the forward rate.



SA1-2 Estimation and Alternative Model Specifications

For all model specifications, we compute the posterior parameter distribution using the Random Walk Metropolis-Hastings (RWMH) algorithm. We estimate the mode of the posterior distribution by maximizing the log posterior function. In order to parametrize the proposal distribution (which is assumed to be multivariate normal), we use the Hessian matrix to produce 80,000 draws from the RWMH algorithm. In a second step, we employ the variance-covariance matrix obtained from these initial draws in order to refine the proposal distribution. We then generate 5 chains of 400,000 draws: a step size of 0.3 gave a rejection rate of around 65 percent in each sample (and each model specification). We evaluate convergence using the Gelman and Rubin potential scale reduction factor (the factor is well below 1.01 for all estimated parameters, in every model specification). The posterior distribution is obtained by combining the 5 chains. The model's posterior coverage intervals are obtained using the Carter and Kohn simulation smoother. Model predictions and (independent) samples are obtained from 20,000 selected draws from the posterior distribution. The marginal data density (MDD) is obtained using Geweke's harmonic estimator.

Here we briefly discuss how we obtain the predictive likelihood following Del Negro and Eusepi (2011). We seek to evaluate the predictive likelihood[3]

$$p\left(\textsc{SvyExp}^i_L \mid \mathcal{I}, \mathcal{M}_l\right) = \int p\left(\textsc{SvyExp}^i_L \mid \theta, \mathcal{I}, \mathcal{M}_l\right) p\left(\theta \mid \mathcal{I}, \mathcal{M}_l\right) d\theta_l$$

where $\textsc{SvyExp}^i_L$ includes short-rate forecasts at horizons longer than one year and $\mathcal{I}$ include all other forecasts, including short-term interest rate forecasts. To obtain this we use

$$p\left(\textsc{SvyExp}^i_L \mid \mathcal{I}, \mathcal{M}_l\right) = \frac{p\left(\textsc{SvyExp} \mid \mathcal{M}_l\right)}{p\left(\mathcal{I} \mid \mathcal{M}_l\right)}$$

$$p\left(\textsc{SvyExp}^i_L \mid \mathcal{I}, \mathcal{M}_l\right) = \frac{p\left(\textsc{SvyExp}^i_L, \mathcal{I} \mid \mathcal{M}_l\right)}{p\left(\mathcal{I} \mid \mathcal{M}_l\right)}$$

where the last line simple clarifies that the dataset $\left(\textsc{SvyExp}^i_L, \mathcal{I}\right)$ includes all survey data. The predictive likelihood is obtained using the marginal likelihood using all data divided by the marginal likelihood obtained from estimating the model on a data set that excludes interest-rate forecasts with horizon longer than one year.

[3]See Del Negro and Eusepi (2011) and references therein.

SA1-3 Variance Decompositions: 2-year, 5-year and 10-year Changes

Figure SA1-2. Decomposing Forward Rates (2-Year Changes). The top panels show the variance decomposition of the 24-month changes of forward rates at different maturities in real expected rates (red line), inflation expectations (blue line) and the wedge (black line) components. The bottom left panel displays the volatility of 24-month changes in interest rate expectations relative to forwards for the pre-1999 (red line) and post-1998 (black line) sample. Solid lines denote posterior medians while shaded regions denote the range of 1,000 posterior draws. The sample period is 1983m1–2019m12.

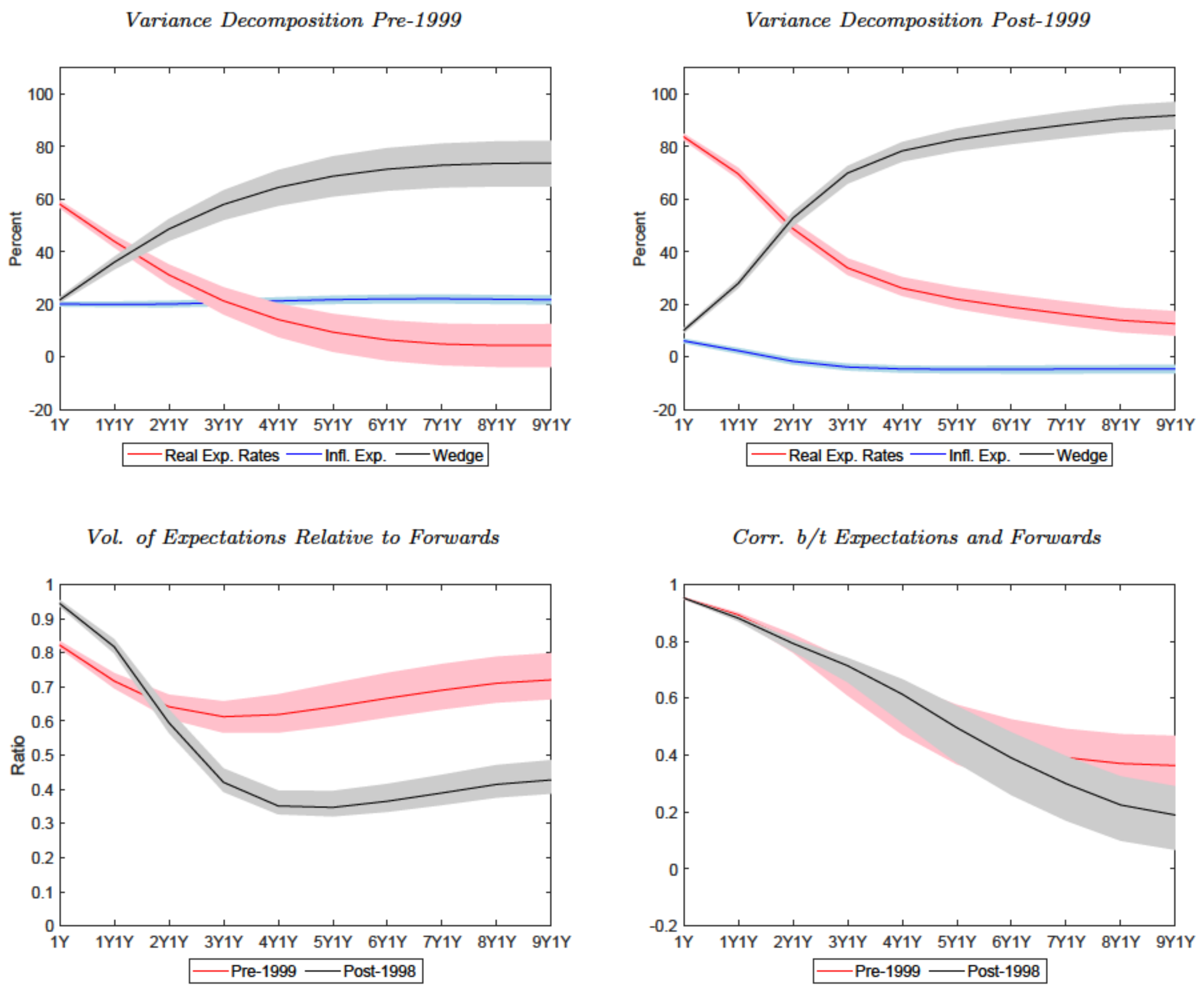


Figure SA1-3. Decomposing Forward Rates (5-Year Changes). The top panels show the variance decomposition of the 60-month changes of forward rates at different maturities in real expected rates (red line), inflation expectations (blue line) and the wedge (black line) components. The bottom left panel displays the volatility of 60-month changes in interest rate expectations relative to forwards for the pre-1999 (red line) and post-1998 (black line) sample. Solid lines denote posterior medians while shaded regions denote the range of 1,000 posterior draws. The sample period is 1983m1–2019m12.

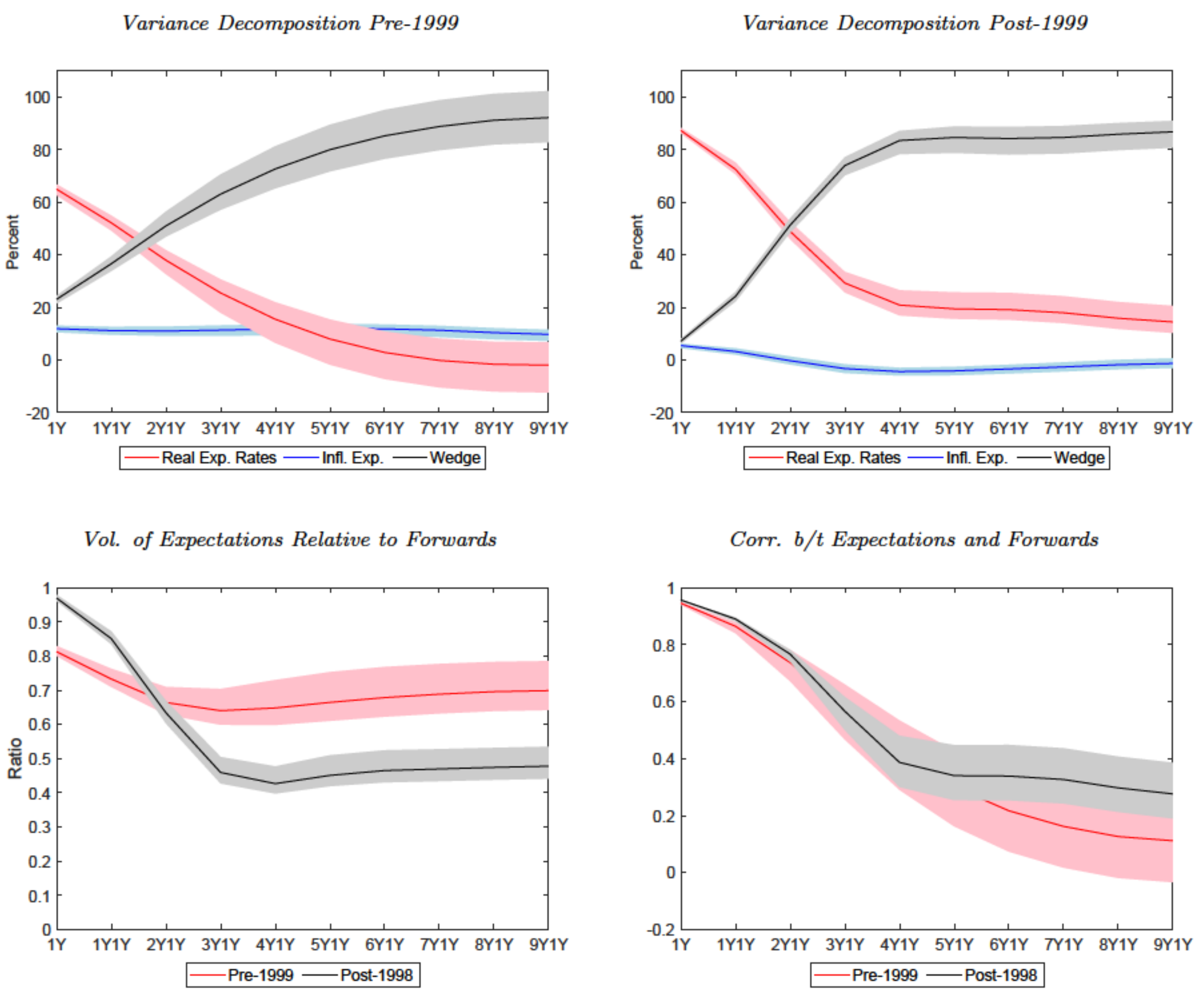
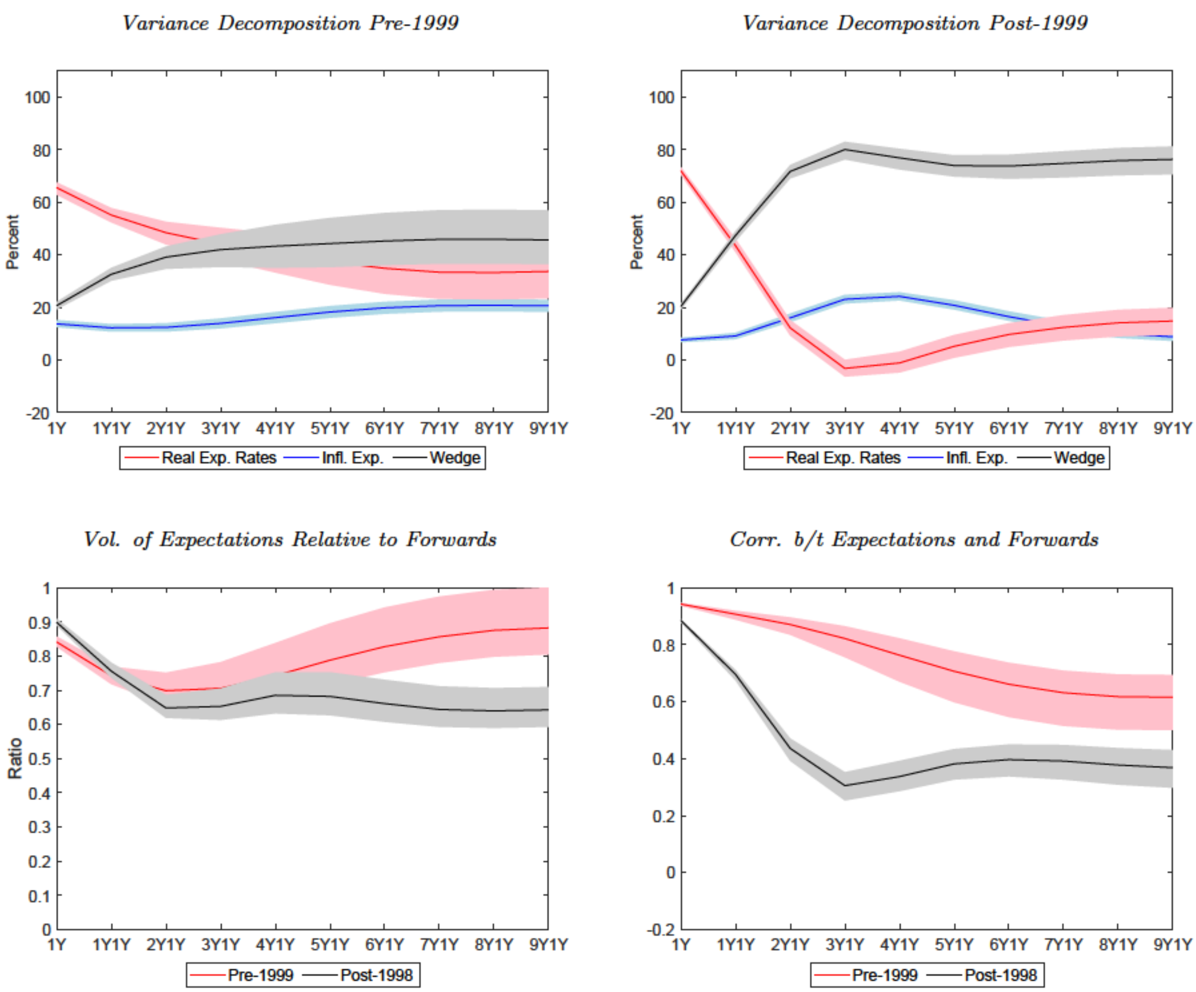


Figure SA1-4. Decomposing Forward Rates (10-Year Changes). The top panels show the variance decomposition of the 120-month changes of forward rates at different maturities in real expected rates (red line), inflation expectations (blue line) and the wedge (black line) components. The bottom left panel displays the volatility of 120-month changes in interest rate expectations relative to forwards for the pre-1999 (red line) and post-1998 (black line) sample. Solid lines denote posterior medians while shaded regions denote the range of 1,000 posterior draws. The sample period is 1983m1–2019m12.



SA1-4 Spanning Tests: Quadratic Specification

In this section we generalize the linear specification used in Section 4.2 to a quadratic specification. Specifically, along with the linear terms (as given by equations (4.11)–(4.12)) we include quadratic terms along with the interactions between all variables. This can be interpreted as a nonparametric series estimator for the unknown conditional expectation (e.g., Chen 2007). In practice, this results in 27 explanatory variables (including the constant). Table SA1.1 presents the raw R^2 values from this regression using either forward rates, $\mathsf{fwd}_t(n,m)$, or the wedges, $\mathsf{w}_t(n,m)$, as the dependent variable.

Since our attention is on the behavior of the wedge we will focus on the lower panel of Table SA1.1. The results do show that allowing for nonlinearities does meaningfully increase the R^2. However, the R^2 values are still far from fully explaining the movements in the wedges. For example, for 1-year and 10-year changes, the explained variation is, at most, between 56% and 67%. At 5-year changes, the explained variation rises, but never explains even 80% and is often substantially lower (across all $1,000$ posterior draws). Taken in sum, even by taking the most conservative approach of allowing for many regressors without regularization and using raw R^2 values, we still conclude that there are important additional drivers of the wedges which are not incorporated in the factors which drive short-term interest rate expectations.

Table SA1.1. Spanning Tests. The top panel (bottom panel) presents R^2 from regressions of changes in the forward rate (wedge) on a constant and all interactions of the six state variables, $\hat{x}_{t|t}$ and $\hat{\omega}_{t|t}$ (27 regressors in all). Minimum and maximum statistics are taken across 1,000 posterior draws. The sample period is 1983m1–2019m12.

	Forward Rates: $\mathsf{fwd}_t(n, m)$									
	1Y	1Y1Y	2Y1Y	3Y1Y	4Y1Y	5Y1Y	6Y1Y	7Y1Y	8Y1Y	9Y1Y
	12-month Changes									
Min. R^2 (%)	82.9	71.3	62.5	56.4	52.1	49.1	47.2	46.0	44.9	44.2
Max. R^2 (%)	86.9	77.1	71.0	67.0	64.7	63.7	63.2	63.0	62.8	62.6
	60-month Changes									
Min. R^2 (%)	95.6	90.3	83.3	76.7	72.2	69.8	68.7	68.1	67.7	67.6
Max. R^2 (%)	96.6	92.5	87.6	83.4	80.8	79.6	79.0	78.5	78.0	77.7
	120-month Changes									
Min. R^2 (%)	92.7	87.0	82.7	80.1	78.6	77.2	76.0	74.5	73.0	72.1
Max. R^2 (%)	94.1	89.4	86.1	84.4	83.7	83.0	82.0	80.7	79.1	78.1

	Wedges: $\mathsf{w}_t(n, m)$									
	1Y	1Y1Y	2Y1Y	3Y1Y	4Y1Y	5Y1Y	6Y1Y	7Y1Y	8Y1Y	9Y1Y
	12-month Changes									
Min. R^2 (%)	26.4	29.9	34.6	39.5	42.7	45.0	46.6	47.6	48.0	48.1
Max. R^2 (%)	42.2	44.1	48.8	53.7	57.6	60.3	62.0	63.1	63.8	64.0
	60-month Changes									
Min. R^2 (%)	54.0	57.0	60.0	63.0	65.2	66.4	67.6	68.8	69.9	70.5
Max. R^2 (%)	63.8	66.4	69.5	72.5	74.8	76.4	77.5	78.4	79.1	79.5
	120-month Changes									
Min. R^2 (%)	58.9	61.7	62.1	61.0	58.9	57.7	57.4	57.1	56.3	55.6
Max. R^2 (%)	66.3	68.9	70.0	69.1	68.6	68.5	68.3	67.9	67.3	66.7

References

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J. J., Leamer, E. E. (eds.), *Handbook of Econometrics*, Elsevier, vol. 6B, pp. 5549–5632.

Del Negro, M., Eusepi, S., 2011. Fitting observed inflation expectations. Journal of Economic Dynamics and Control 35, 2105–2131.

Gurkaynak, R. S., Sack, B., Wright, J. H., 2007. The U.S. Treasury yield curve: 1961 to the present. Journal of Monetary Economics 54, 2291–2304.

Wright, J. H., 2011. Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. American Economic Review 101, 1514–34.